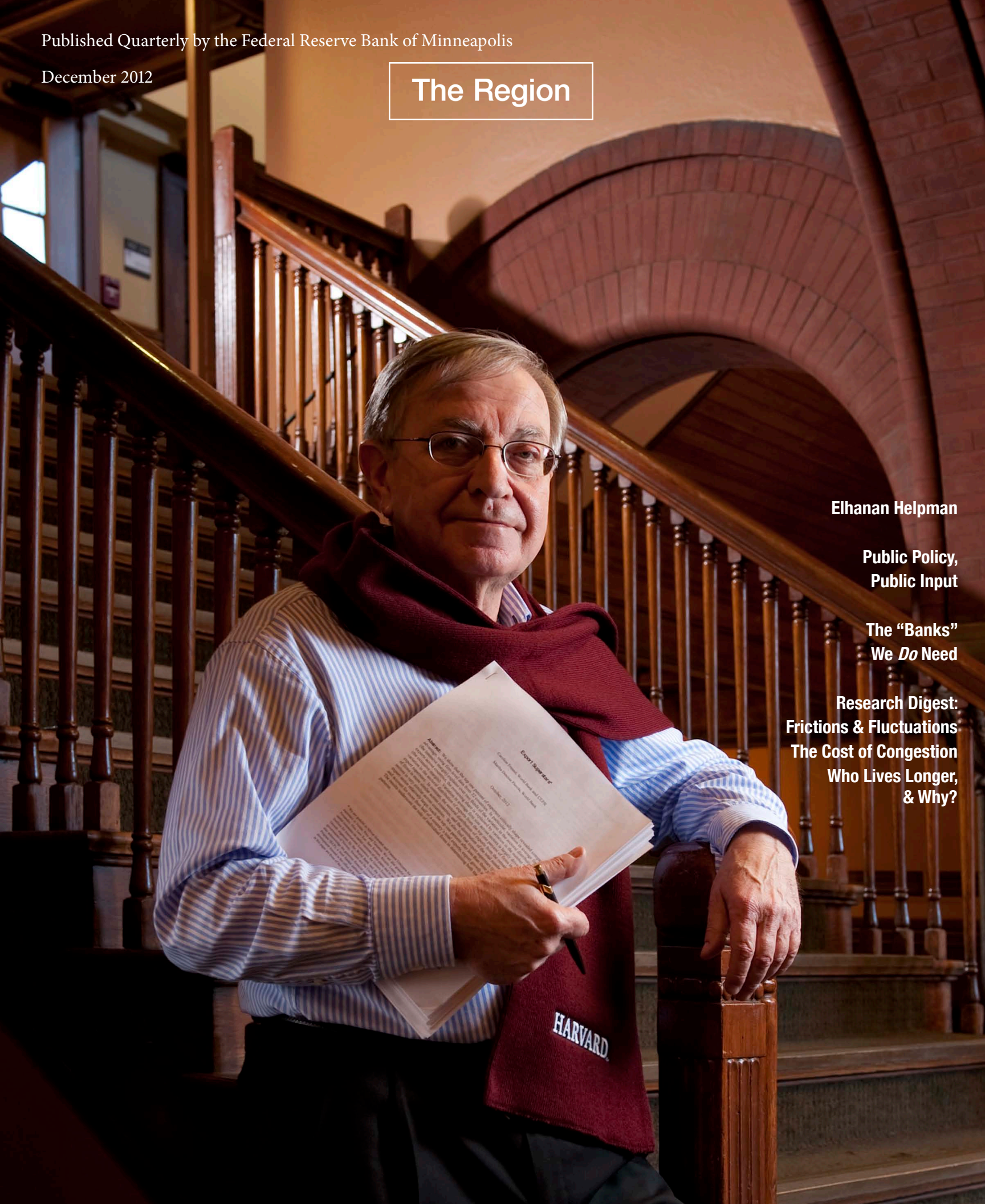
Published Quarterly by the Federal Reserve Bank of Minneapolis

December 2012

The Region



Volume 26 Number 4
December 2012
ISSN 1045-3369

The Region

The Region
Federal Reserve Bank of Minneapolis
P.O. Box 291
Minneapolis, MN 55480-0291

E-mail: letters@mpls.frb.org
Web: minneapolisfed.org

Cover Photograph by Peter Tenzer

NINTH DISTRICT NOTES

Public Policy, Public Input

Narayana Kocherlakota

President
Federal Reserve Bank of Minneapolis

Over the course of my three-year tenure as president of the Federal Reserve Bank of Minneapolis, I have given many talks across the Ninth District. Often, during the course of those talks, I have described how I obtain the economic information that helps inform my policymaking. I do this because I believe that it is important for people to know how their policymakers receive information that helps them make policy. As well, this process of information gathering reinforces the point that the Federal Reserve System is grounded in regional representation from across the country. Indeed, that's why our central bank is a "system" and not just one monolithic operation in Washington. So for this column, I will describe the information-gathering efforts that I employ and then conclude by discussing my decision to make those efforts public.

First, it would not surprise you to learn that Federal Reserve policymakers have large amounts of data at their disposal. Indeed, that is an understatement. There is no end to data on all aspects of the economy, from inflation to employment to manufacturing and trade, and hundreds of data points in between. Nonetheless, these data do not tell the whole story that I need as a policymaker. Some data can be weeks or months old at the time a decision needs to be made. In addition, as important and revealing as data are, there is still much to learn from people making decisions on the ground. For example, what are businesses' hiring plans? What are expectations for future sales? What about price pressures? What are banks experiencing in terms of loan demand?

Those questions and many others provide answers that help bring data to life and inform the choices that I have to make. But who answers those questions? How do we gather such information? The sources are many and reflect the broad and varied constituency of the Ninth District. First, there is our board of directors. In my previous column, for the September 2012 *Region*, I described the role of the directors in some detail, so I will only stress here the important role they play in relaying economic information. For every meeting, a subset of the directors is charged with answering questions about the condition of the general economy and also about their particular industry. They do this not only by relaying information pertinent to their own companies, but also based on many phone calls and conversations with colleagues in their industries.

In addition to the board of directors, I meet twice a year with three advisory councils, for a total of six meetings. These meetings are held expressly for the purpose of gathering information about the Ninth

It is these connections—this two-way communication with people in local economies—that ensure that the American people have input into monetary policy

District economy. These councils represent small business and labor, agriculture and community financial institutions. Who sits on these councils? The answer is as varied as any town's Main Street and rural highways: Ranchers, farmers, owners of retail shops, credit unions, labor representatives and small manufacturers, among many others, are represented in these meetings. By the way, information about the board of directors and these advisory councils, including the members, is available at minneapolisfed.org.

Beyond these more formal channels of communication, my travels around the Ninth District give me another way to get useful input. Whenever I give a speech outside the Twin Cities, I schedule meetings with local business owners and bankers, as well as with organizations that serve consumer groups, especially those that provide services to low- and moderate-income communities and households. I am always grateful for the time that these busy people take from their schedules to discuss their views about the economy and the choices facing businesses and individuals. I always come away with more insight than I had before.

Finally, just as I go out to the district to meet with people and to learn from them, others come to the Bank to visit with me. These people might be representatives from industries, consumer groups, labor groups, neighborhood coalitions and even representatives from other countries.

All told, I meet with a wide variety of people with insights from throughout the economy. So who are these people? Well, as I indicated earlier, you can visit our website to see who sits on our board of directors and on our advisory councils. As to those other groups I meet with throughout the Ninth District and at the Bank, you can see for yourself who they are, as they are now listed on our public website. Earlier this year I began posting my schedule online. In part, this move was a matter of transparency—that is, just as I think it is important for people to understand how I receive economic information, they should also know who gives me that information. But equally important, I publish my schedule because it reinforces the connections that I have with many segments of the Ninth District economy. As I often tell people when I give talks, it is these connections—this two-way communication with people in local economies—that ensure that the American people have input into monetary policy. R

The "Banks" We *Do* Need

Services once said to justify traditionally structured banks are now available through more efficient, less risky financial vehicles

V. V. Chari

Christopher Phelan*

University of Minnesota and
Federal Reserve Bank of Minneapolis

Introduction

The financial crisis of 2007-08 and consequent Great Recession generated substantial discussion and debate over future banking regulation. Largely absent, however, has been a careful reexamination of whether the beneficial services provided by traditional banks outweigh the inherent financial fragilities of those banks and their associated costs to society.

Three major benefits are usually said to justify traditional bank reliance on short-term debt, the source of their inherent fragility. In a previous article, we assessed—and found wanting—two of these proposed rationales: (1) the benefit of maturity transformation, or creation of long-term financial assets from shorter-term assets and (2) the benefit of efficient monitoring of bank managers, through appropriate alignment of investor incentives. (See Chari and Phelan 2012a.)

Here we discuss the third justification, that traditional banks are beneficial and necessary because they provide payments services essential to the efficient function of modern economies. We conclude that while this rationale was compelling in an earlier historical era—prior to modern advances in information and communication technology that facilitate transactions of all sorts—the necessary services can now be provided through existing financial vehicles that do not rely on traditionally structured, inherently fragile banks.

We begin by briefly reviewing the structural source of traditional bank fragility and proceed

Economic Policy Papers *are based on policy-oriented research by Minneapolis Fed staff and consultants. The papers are an occasional series for a general audience. Views expressed are those of the authors, not necessarily of others in the Federal Reserve System.*

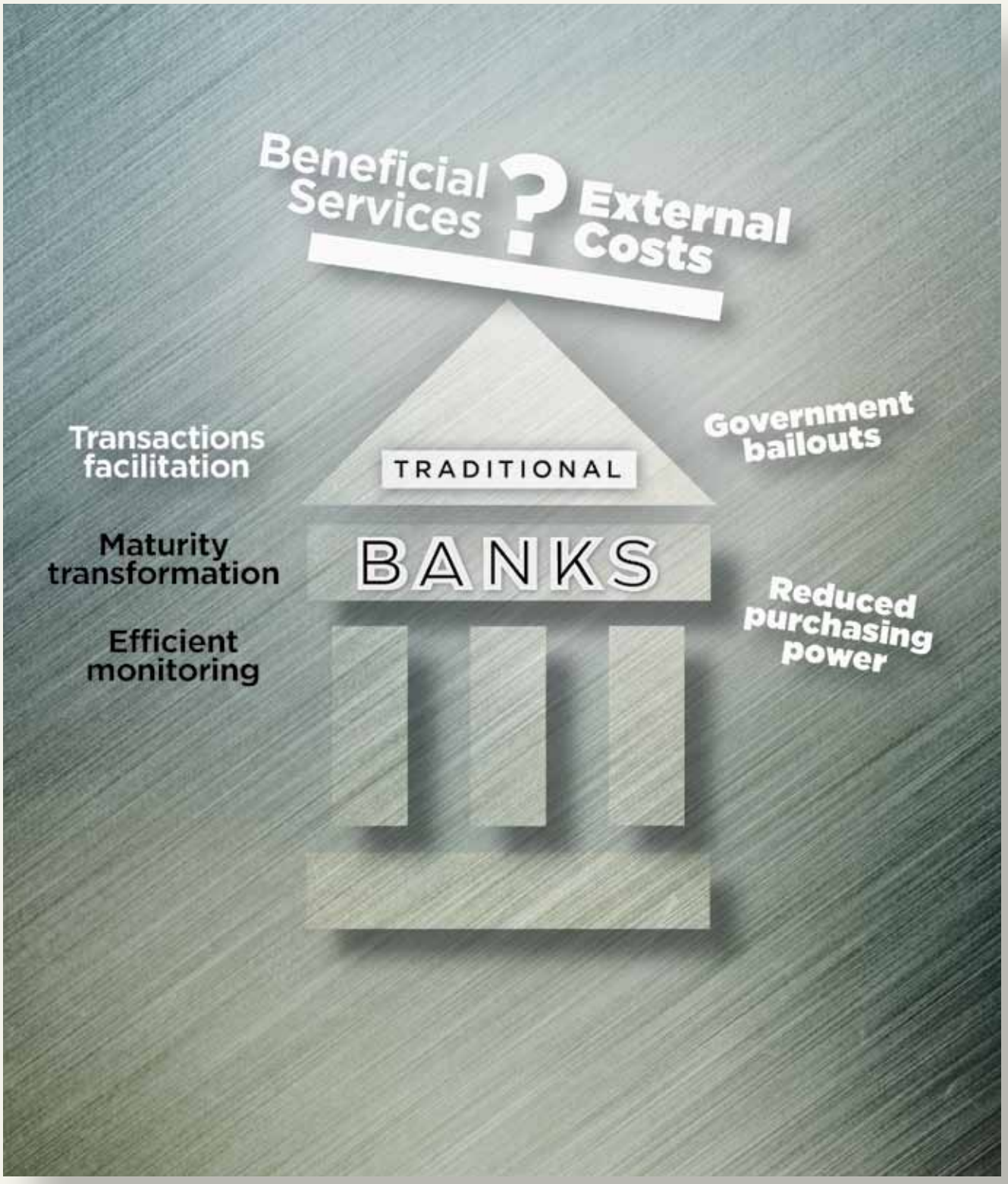
ABSTRACT

Banks are prone to panic-induced runs due to their traditional structure of short-term, unconditional liabilities and long-term, illiquid assets. To avoid systemic crises caused by such panics, governments tend to bail out failing banks. Traditional banking systems thus impose external costs.

Three major theoretical benefits are often used to justify a banking system that relies on short-term debt despite these costs: (1) *maturity transformation,* (2) *efficient monitoring* of bank managers and (3) *facilitation of financial transactions.* In a previous paper, we argued that the first two justifications, while seemingly compelling, actually suggest financial arrangements very different from our current system.

In this paper, we examine the third justification, that a banking system reliant on short-term debt is essential for the facilitation of transactions. We find, in fact, that this reliance is more costly than generally recognized and, moreover, that socially beneficial financial transactions can and should be provided at less cost and risk by both restricting and broadening the payments system. Transactions should be restricted to institutions that continuously mark to market the value of their assets and issue equity claims to owners. Such accounts should also be broadened to include financial vehicles that are readily available, thanks to advances in information and communication technologies, and possibly quite different from current banks.

*The authors thank Narayana Kocherlakota, Dick Todd and Kei-Mu Yi for useful comments and Doug Clement for editorial assistance. V. V. Chari thanks the National Science Foundation for supporting the research that led to this paper.


Beneficial Services
?
External Costs
Transactions facilitation
Maturity transformation
Efficient monitoring
TRADITIONAL
BANKS
Government bailouts
Reduced purchasing power

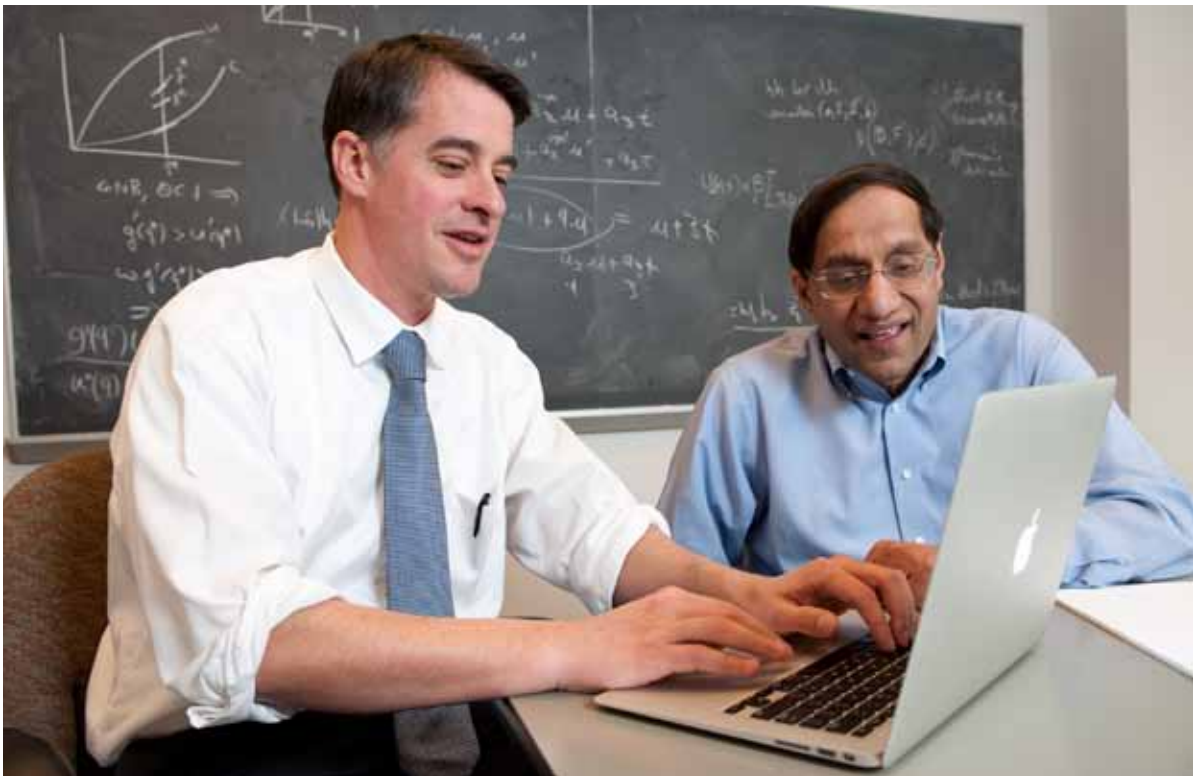
to consideration of the necessity of banks, despite this fragility. We then address the main topic of this paper: the argument that banks as currently structured are necessary because their demand deposits facilitate financial transactions. We conclude that the current structure of banks is unduly costly to society and that essential payments services can, with modern information and communication technologies, be provided with less fragile and more efficient financial institutions.

The inherent fragility of banks

In what sense are banks and similar financial institutions fragile? Our previous paper discusses this question in detail; here we provide a synopsis, referring interested readers to the earlier discussion.

The assets of financial institutions are, by and large, financial assets, and claims on them are primarily financial liabilities. Their financial *assets* consist mainly of conditional promises to deliver dollars at future dates. These assets, such as home mortgages, are often long term and illiquid. Their financial *liabilities* consist mostly of a variety of obligations to deliver dollars at particular dates, under certain circumstances. Banks in particular have liabilities that are mostly short term and unconditional, such as demand deposits and certificates of deposit.

Governments have a strong incentive to intervene to bail out debt holders of banks in order to prevent the entire financial system from failing. Paradoxically, expectations of such bailouts can increase the incidence and depth of financial crises.



PHOTOGRAPH BY STAN WALDHAUSER

Christopher Phelan V. V. Chari

This reliance on short-term debt makes banks fragile in that they are particularly vulnerable to the risks of insolvency and the possibility of confidence crises. Since bank assets are much longer term and illiquid than their liabilities and because the value of these assets fluctuates, a bank's net worth also fluctuates a great deal.

The illiquidity of banks' assets and the demandable structure of their liabilities thus expose banks to crises of confidence. Since a bank typically will not be able to meet the demands of all depositors within a short period of time should they all choose to withdraw, banks are vulnerable to self-fulfilling panics in which depositors withdraw their funds simply because they believe other depositors will do so. This panic is an entirely rational response even if the bank is solvent (though illiquid).

Governments have a strong incentive to intervene to bail out debt holders of banks in order to prevent the entire financial system from failing. Paradoxically, expectations of such bailouts can increase the incidence and depth of financial crises. Once depositors believe that their deposits will be protected in the event of systemic failure, they have less incentive to monitor bank managers, who, in turn, have increased incentive to take on risk, knowing their failures are implicitly insured by taxpayers.

In this way, expectations of bailouts can lead financial systems to rely excessively—from a social perspective—on short-term debt to fund long-term assets. Fragile banking systems thus impose external costs, and regulation may therefore be socially desirable.

Are banks necessary?

The fragility of the banking system together with the reality that such fragility may well lead to occasional massive bailouts compel us to ask why societies would choose regulatory systems that allow financial institutions to fund illiquid assets whose value can fluctuate rapidly with short-term debt and demand deposits.

One could perhaps argue that banks were necessary prior to the electronic information age because no other forms of financial intermediation were feasible. With the advent of high-speed computers and modern communications, however, alternative financial institutions can provide similar services with far less potential for crises. We discuss such alternatives later in this paper.

We now examine the possible social *benefits* of a financial system in which illiquid assets with volatile values are funded by demand deposits and short-term debt. This cost-benefit analysis facilitates the design of a better regulatory system for banks, clearly a matter of considerable importance.

The previous paper examined two of the three major theoretical justifications for the reliance of the banking system on short-term debt: (1) demand deposits allow banks to engage in socially useful *maturity transformation* and (2) demand deposits allow for *efficient monitoring* of bank managers. This paper considers the third major justification: (3) demand deposits *facilitate financial transactions.*

To anticipate our conclusion, we believe that while all three justifications are compelling, they point us to a financial system very different from the one currently in place. The first two justifications suggest that it is important to have institutions that finance long-term assets with short-term debt, but we have argued that the assets that are so funded should not have close substitutes in publicly traded markets. In this paper, we will argue two main points regarding the usefulness of banks in facilitating transactions. First, we argue that regardless of technology, the *social* benefit to using banks to facilitate transactions is lower than the *private* benefit, thus potentially explaining why the historical ubiquity of bank-facilitated transactions does not imply their efficiency. Second, we argue that the necessity of bank-facilitated transactions is much less obvious than it was a century ago, before advances in information and communication technologies allowed us to create very different institutions than we currently have to facilitate transactions.

Our analysis will suggest a framework for thinking about regulatory policy for institutions that facilitate payments. The economic case for regulating such institutions is convincing, given that the failure of the payments system imposes significant external costs. We argue that institutions that facilitate payments should primarily issue equity-like claims such as those issued by standard mutual funds. Current practice hopelessly conflates these two economic cases into a single institution

called "banks" and exposes the economy to unnecessary risks and recurrent costly bailouts.

Assessment of the transactions facilitation view

The most obvious service that banks provide now, and have offered throughout their ubiquitous existence, is payments services. Historically, banks have allowed individuals and firms to pay for goods and services through their provision of bank checks and other widely accepted claims. Therefore, those individuals and firms haven't had to resort to costly barter or specie trade.

Here, we raise the possibility that banks exist because they provide a *privately* useful function—the facilitation of transactions in a form that pays households interest—but the social usefulness is less than the private usefulness.

The starting point of our assessment involves the central bank and monetary policy. The central bank creates money, which, for simplicity, we will call "cash." Cash typically earns no interest. Our first key point is that, to the extent that monetary policy is conducted so as to keep inflation—and thus the (nominal) interest rate—inefficiently high, private agents have strong incentives to develop private payments systems to economize on the use of cash. Interest-bearing demand deposits (checking accounts) at banks are one example of such a private payments system. Because of the interest received in such accounts, households and firms will find it advantageous to switch from cash to these private deposits as their means of payment. Clearly, then, there would be private benefits to the introduction of payments systems like checking accounts.

But do these *private* benefits imply equivalent *social* benefits? If one household's use of demand deposits imposed no costs on other households, the answer would be yes. But if use of such demand deposits does indeed impose costs on other households, the net social benefit of demand deposits will be lower and can, in fact, be negative. In the appendix, we present an example economy where these net social benefits from demand deposits are indeed negative, even though each household finds it in its interest to use them (since the private benefits are positive). In Chari and Phelan (2012b), we present a more general model where the net social benefits from interest-bearing means of payments can be either positive or negative, but are nevertheless always less than the private benefits.

The reason one household's use of demand deposits imposes costs on other households is as follows: Introducing bank-provided payments leads to an expansion of the "means-of-payment" supply, now defined to include both cash and the amount of demand deposits. This higher means-of-payment supply leads to higher prices in the aggregate economy, which reduces the purchasing power of other households' deposits and cash—but individual firms or households do not take this into account when they choose to use demand deposits over cash. This pecuniary externality (that is, an external cost imposed through prices rather than real resources) can cause households to use deposits instead of cash in cases where they wouldn't, had they internalized this cost imposed on other households, and this externality implies that net social benefits of demand deposits are lower than net private benefits.

With net private benefits of banking exceeding net social benefits, it is clear that the banking system will be inefficiently large. In the model presented in the appendix (online at minneapolisfed.org), because the net social benefits are negative, not only is the banking system inefficiently large, the optimal size of banks is zero.

The model in the appendix is but a simple example, and the implications from it seem unrealistic. However, we would argue that recent developments in communication technologies and financial innovations may in fact make the model's implications more than just a hypothetical scenario.

Historically, communication costs and limited development of financial markets have led to the use of systems in which only a fraction of a household's financial wealth could be used for payments. With improvements in communication and financial markets, however, we can conceive of a world in which each individual can instantaneously access all of his or her financial wealth to make payments. We can also imagine a world in which settlement of transactions is instantaneous. In this world, cash becomes unnecessary, and precisely because cash is unnecessary, there is little or no need for payments systems that arise from the need to economize

What should "banks" look like, if not the traditional but fragile demand-deposit bank?

Alternative financial institutions can provide similar services to the transactions facilitation services that traditional banks offer with far less potential for crises.

on cash, that is, arise because monetary policy is setting the inflation rate too high.

In the 1800s, it would have been inconceivable to pay for groceries, for example, by using a debit card associated with one's mutual fund or stock portfolio (and in doing so, stocks were immediately sold, and the grocery received its settlement while the shopper was still at the counter). But today, this scenario is not far-fetched. In a world with virtually costless communication, banks as specialized providers of transactions services would simply be obsolete. These observations lead us to conclude that the actual importance of banks in the payments system is likely small today and will likely become even smaller in the near future. This is the third and final key point in our assessment of the transactions facilitation view.

What *should* "banks" look like, if not the traditional but fragile demand-deposit bank? As mentioned in the introduction, alternative financial institutions can provide similar services to the transactions facilitation services that traditional banks offer with far less potential for crises. One such example is the open-end mutual fund. These funds do not owe their shareholders a fixed dollar amount, but instead only the value of their percentage of the fund on the day the shareholder wishes to withdraw. If an unexpected surge of withdrawals occurs, the fund simply sells a sufficient quantity of the fund's assets and gives the proceeds to the withdrawing shareholders. After this, the remaining shareholders still hold exactly the same assets per share as before. No shareholder gains by being earlier in line than other shareholders. Therefore, a *belief* that a run will occur cannot cause a run for a mutual fund—the self-fulfilling nature of runs that afflicts banks with demand deposits is thus avoided.

MMMFs are *not* open-end mutual funds

One modern financial institution, the *money market mutual fund* (MMMF), which appears to resemble an open-end mutual fund as described above, is quite different in practice. MMMFs were perceived as promising one dollar for each share held as opposed to a claim to a pro rata share of the fund's assets. MMMFs in this sense resemble banks more closely than they do ordinary mutual funds.

During the financial crisis of 2007-08, there were no runs on ordinary mutual funds, including mutual funds invested in assets very similar to the assets held by MMMFs, nor were there any concerns by policymakers about runs on such ordinary mutual funds. In sharp contrast, after the fall of Lehman Brothers in September 2008, the Reserve MMMF was subject to significant withdrawals. It suspended withdrawals from the fund and eventually returned 98 cents on the dollar to shareholders. Policymakers instituted a variety of policies, including a program to insure the shareholders of *all* MMMFs.

Implications for policy

Banks have been a durable part of the economic landscape for many centuries, and economic theory does explain why it might be efficient to set up institutions that fund long-term assets with short-term debt. Theory also illustrates that it might be optimal for private agents, but undesirable for society at large, to establish such institutions. These competing lessons from economic theory also provide guidance for regulation of such institutions.

As discussed in the earlier paper, both the maturity transformation and the efficient moni-

toring views suggest that, given the costs imposed by crises and attendant bailouts, it may be desirable to allow financial institutions to issue short-term debt *only* if their assets do not have close publicly traded substitutes. Further, to minimize the incentive of governments to bail out institutions if a crisis occurs, such institutions should be separated from the payments system.

Any regulatory system must also take seriously the central role that banks have long played in the payments system. We have argued that this role may well be an artifact of a bygone era. Advances in information and communication technology make it feasible to access a wide array of assets, from stocks in public firms to portfolios of home equity loans, to undertake transactions. We have also argued that payments systems that require the use of demand deposits expose the economy to confidence crises and that it is possible to devise payments systems that do not require the use of debt-like claims, but instead use equity-like claims for transactions purposes.

These considerations suggest that the payments system should be both restricted and broadened. Transactions accounts should be restricted to institutions that mark the value of their assets to market continuously and that issue mutual-fund-like equity claims to owners. Such accounts should be broadened to institutions that are possibly very different from modern-day banks to include institutions such as stock and bond mutual fund companies.

We emphasize that the money market mutual fund as currently structured resembles a bank more than it does a mutual fund and therefore should not be allowed to issue transactions accounts. So, for example, Vanguard's money market mutual fund (as currently structured) would no longer be allowed to serve as a transactions account, but Vanguard's 500 Index Fund would.

The framework for regulatory policy implied by our analysis would lead to a banking system that is radically different from the one we currently have. Institutions that issue large amounts of short-term debt relative to their assets would be regulated and required to hold relatively little of their assets in publicly traded securities. The liabilities of such institutions would not serve as means of payment. The payments system would consist of institutions that issue equity claims.

Economic theory tells us that we do need banks. Theory also points us to constructive ways in which we can reform the financial system to make it more efficient and to ensure that crises that affect particular financial institutions do not spill over into the rest of the economy. R

References

Chari, V. V., and Christopher Phelan. 2012a. What Assets Should Banks Be Allowed to Hold? Economic Policy Paper 12-3, (May) Federal Reserve Bank of Minneapolis. Online at minneapolisfed.org.

Chari, V. V., and Christopher Phelan. 2012b. On the Social Usefulness of Fractional Reserve Banking. Working paper. University of Minnesota. Online at http://www.econ.umn.edu/~cphelan/research.html.

Elhanan Helpman

Since the days of Adam Smith, international trade and long-run growth have engrossed economists. Global trade, after all, is the exchange of goods and services—at the core of economics—writ large. And long-term growth is how countries do (or do not) permanently raise their standards of living.

Several giants in economics have made important contributions to one of these fields or the other, but very few have had an enduring and transformational influence on both. Among the latter is Harvard's Elhanan Helpman, one of the world's foremost authorities on international trade and economic growth, and a leading figure in several other areas, including political economy.

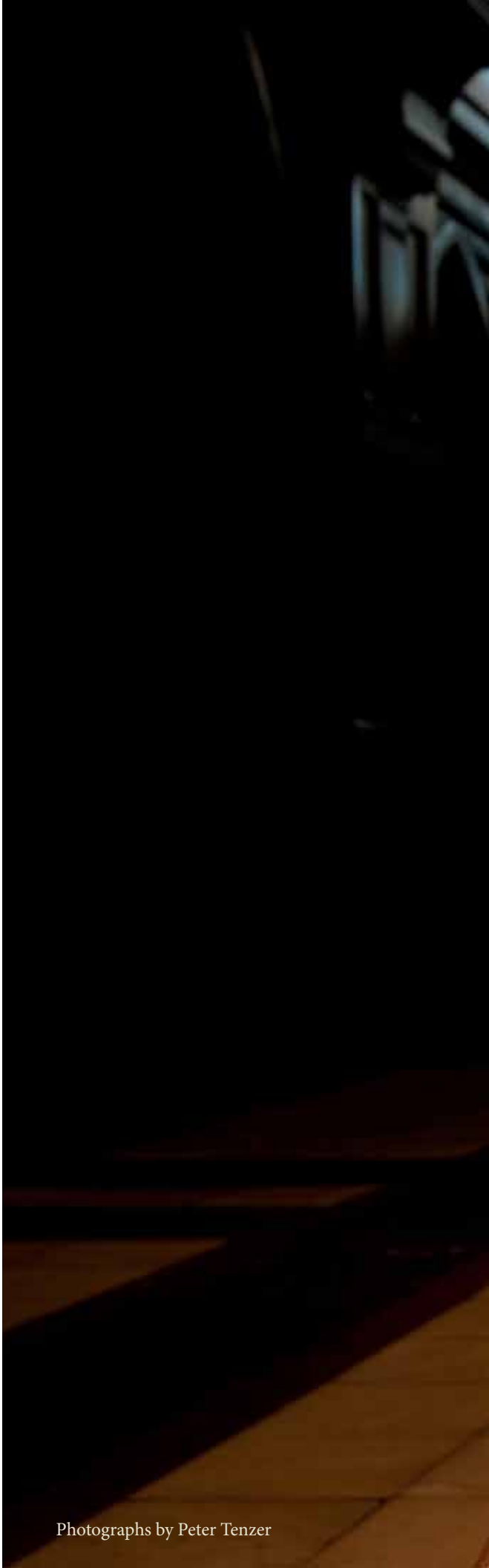
In 2010, the prestigious Nemmers Prize in Economics, awarded biennially to recognize "work of lasting significance," was given to Helpman "for fundamental contributions to the understanding of modern international economics and the effects of political institutions on trade policy and economic growth." (Five of the previous eight Nemmers Prize recipients later received the Nobel Prize.)

It was in the early 1980s that he helped develop "new trade theory," a fundamental concept that explained what traditional comparative advantage theory could not: The vast majority of international trade takes place among quite similar countries and sectors. He later developed key insight into the ways modern firms organize production not at a single factory but in multiple stages, sites and nations—leading to global trade flows never envisioned by earlier economists.

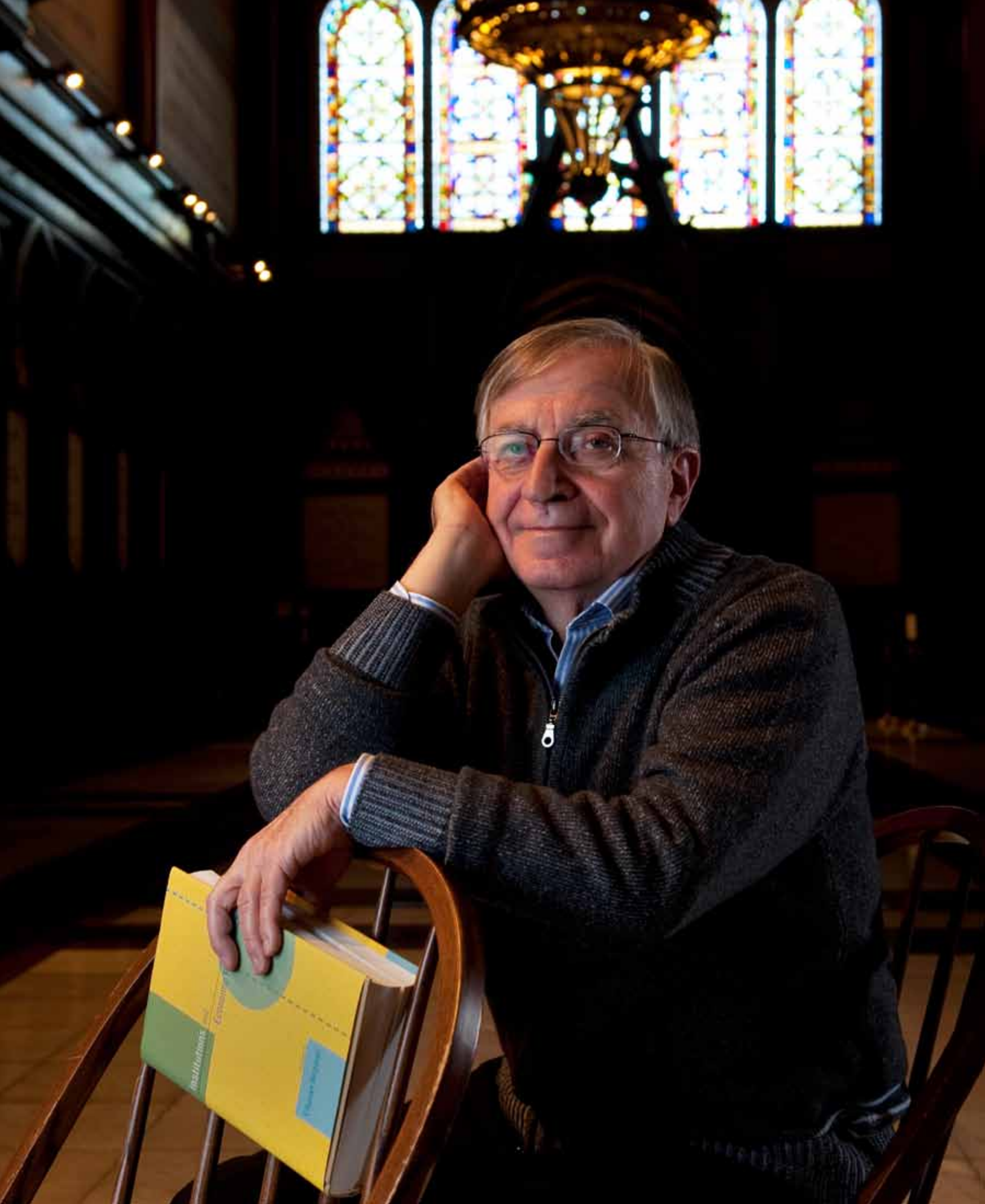
In addition, with Gene Grossman of Princeton, Helpman pioneered the extension of "new growth theory"—the idea that information, ideas and technology (not just capital and labor) are central determinants of economic growth—into settings with international trade. They explored the importance of international research and development, and spillovers thereof, to technological innovation and growth itself.

More recently, Helpman has investigated the role of institutions—legal regimes, education systems and the like—in both growth and the political systems that determine trade policy. Currently, he is studying why economic inequality often accompanies greater trade flows across borders—contrary to predictions of traditional trade theory—but then diminishes.

In the following interview, he describes the history and current frontiers of his pathbreaking research, sharing insights gained through decades of research into the riddles of economic growth and global trade.



Photographs by Peter Tenzer



Helpman: So, what would you like to talk about?

Region: In truth, I'm a bit overwhelmed by all the fields in which you've worked, but perhaps we could focus on three that I think are among your primary areas: new growth theory, new trade theory and trade (and policy) related to market structure. That's a lot to cover, but if we have time, perhaps we'll be able to visit a few other topics.

NEW GROWTH THEORY

Region: As you know, Paul Romer, Bob Lucas and others pioneered what was later termed "new," or "endogenous," growth theory, emphasizing increasing returns associated with new knowledge, ideas, technology and spillovers. You extended this new growth theory into settings with international trade, often with Gene Grossman, looking at the importance of research and development and spillovers from industrial research.

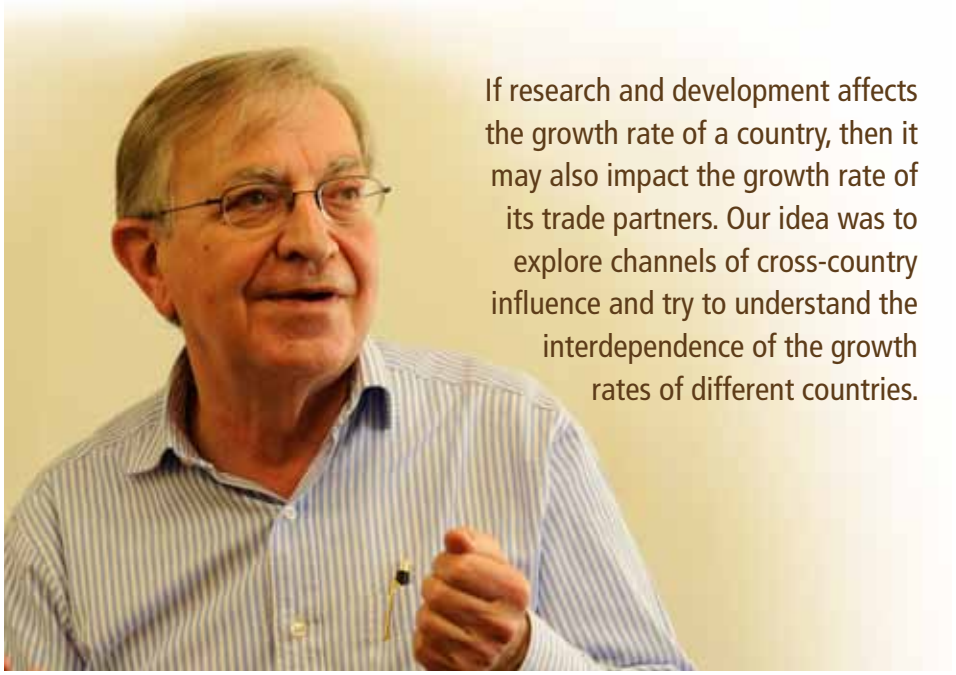
Could you tell us, why are R&D spillovers central to economic growth?

Helpman: In the previous episode of growth theory, the view was taken, following [Robert] Solow, that economic growth is driven mostly by capital accumulation. Some of Solow's students also discussed the accumulation of human capital, which Lucas essentially then extended and turned into a major view of economic growth.

However, at the same time, there were a number of people who worked on the impact of research and development. Zvi Griliches, for example, worked on it from a microeconomic perspective rather than from a macroeconomic perspective. They developed the concept of R&D capital stocks and the type of externalities that they generate. Moreover, they estimated these external effects.

Region: They developed spillover estimates that early on?

Helpman: They did, yes, and the typical estimate was that the social rate of return on R&D could be twice as high as the private return. So when Gene Grossman and I entered the field, we had this in the background. We knew that there were R&D spillovers, and we knew that the social rate of return was high. Then the question that Paul Romer and other people asked was, to what extent can you explain growth with investment in research and development, rather than assume, as Solow had done, that the rate of technical change is exogenous? Because if you could tie growth to the rate of technical change, obviously, you potentially could explain a lot of aggregate growth.



If research and development affects the growth rate of a country, then it may also impact the growth rate of its trade partners. Our idea was to explore channels of cross-country influence and try to understand the interdependence of the growth rates of different countries.

Region: Which is what made it "endogenous" growth?

Helpman: Endogenous, correct. I mean, at some level in economics, almost everything is endogenous; it only depends on how far back you want to stand. You can ask the question, OK, research and development is endogenous, but it depends on the economics environment, which is treated as exogenous. If, however, you are willing to take one step back, you realize that some elements of this economics environment, such as the patent system, are in fact endogenous too.

In any case, Gene Grossman and I thought that this is very important for links across countries. In other words, that if research and development affects the growth rate of a country, then it may also impact the growth rate of its trade partners. Our idea was to explore channels of cross-country influence and try to understand the interdependence of the growth rates of different countries.

Region: This was your "quality ladders" paper?

Helpman: Well, the quality ladder was just a model. We actually wrote an earlier paper, less known than the others, I think it was in '89, which was very specific in many details. So when we wrote the quality ladder paper, we already had a better view of the world, and we could write something more appealing.

And, of course, there was the work of Paul Romer that we could build on. Our aim was to integrate this view of the growth process into a worldwide system in order to explore these interdependencies across countries. So when we wrote,

later, the book, we showed that the same mechanisms work, whether you explain growth by quality ladders or by extending product variety.

From a macroeconomic perspective, these two alternative views are pretty good substitutes. For some issues, they are not. For example, if you want to measure the deviation of what the market generates from what is the best for society, these two views will give you somewhat different answers. But from a perspective on how the growth process behaves, they provide very similar results.

Region: The book you refer to is *Innovation and Growth in the Global Economy*?

Helpman: Yes, *Innovation and Growth*.

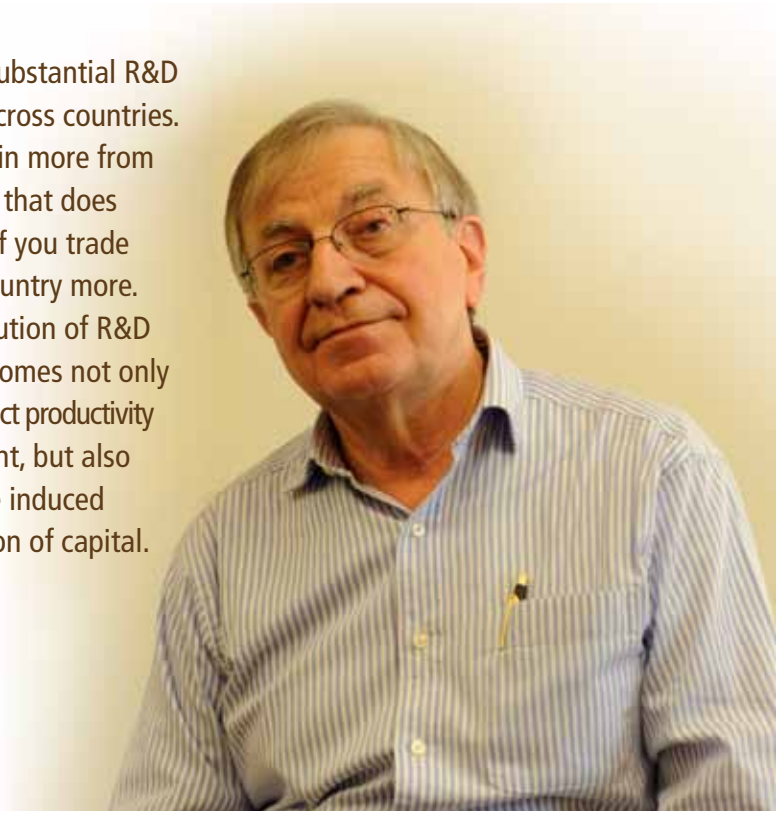
Region: Still considered the main reference on trade and endogenous growth, over 20 years after its publication.

EMPIRICAL IMPACT OF R&D, HUMAN CAPITAL AND INSTITUTIONS

Region: Earlier you mentioned Griliches' empirical estimates. Though you're mainly a theorist, you too have studied the empirical impact of R&D—both domestic and international—on productivity and therefore growth. I think your first empirical work on this was in 1995 with David Coe. And you later extended that with Coe and Alexander Hoffmaister to look at human capital and institutions.

Would you summarize your findings from that work? Which variables—R&D, human capital, institutions—have the most significant impact on productivity? And what does that suggest for policy?

Helpman: Well, I'm a little shy about policy recommendations. But I can talk about the findings; it's easier. Yes, so, our first empirical paper on R&D spillovers, international R&D spillovers I should say, is the paper we wrote in '95. We wrote a couple of other papers as well. In '97, Coe, Hoffmaister and I had a paper; then [Tamim] Bayoumi, Coe and I wrote another later on.



We found substantial R&D spillovers across countries. And you gain more from the country that does more R&D if you trade with this country more. The contribution of R&D to growth comes not only from the direct productivity improvement, but also through the induced accumulation of capital.

In the 1995 paper, essentially, we asked the following question—we know how Griliches has estimated R&D spillovers across firms; there also existed estimates of spillovers across industries—so we asked the question, can we estimate spillovers across countries?

We computed productivity growth in a variety of OECD [Organisation for Economic Co-operation and Development] countries in this particular paper. We constructed R&D capital stocks for countries, rather than for industries, which is what Griliches had done. Then we estimated the impact of the R&D capital stocks of various countries on their trade partners' productivity levels.

And we found substantial spillovers across countries. Importantly, in those data, these spillovers were related to the trade relations between the countries. And we showed that you gain more from the country that does more R&D if you trade with this country more. This produced a direct link between R&D investment in different countries and how trading partners benefit from it.

In the '97 paper with Coe and Hoffmaister, we looked at developing countries because the '95 paper was about industrialized countries. The developing countries don't do much R&D. The overwhelming majority of R&D is done in industrialized countries, and this was certainly true in the data set we used at the time.

So we asked the following question: If you look at developing countries, they trade with industrialized countries. Do they gain from R&D spillovers in the industrialized countries, and how does that gain depend on their trade structure with these industrialized countries? We showed empirically that the less-developed countries also benefited from R&D spillovers. And the more they trade

with industrialized countries that engage heavily in R&D, the more they gain.

The exercise Bayoumi, Coe and I then did is also quite interesting. The International Monetary Fund had an econometric model for its midterm projections. We integrated the equations that Coe and I had estimated previously into this IMF econometric model. Then we could simulate it using our specification of the relationship between R&D levels and productivity levels across countries. In this way, we could essentially decompose the growth process. How much of it is driven by capital accumulation? How much is driven by productivity growth due to R&D?

One of the important findings—which analytically is almost obvious, but many people miss it—is that, if you have a process that raises productivity, such as R&D investment, then this also induces capital accumulation. So then, the contribution of R&D to growth comes not only from the direct productivity improvement, but *also* through the induced accumulation of capital. When you simulate the full-fledged model with these features, you get a very clear decomposition. You can see how much is attributable to each one.

With this, we could handle a relatively large number of countries in all different regions of the world, and [run some] interesting simulations. We could ask, for example, if all the industrialized countries raise their investment in R&D by an additional half percent of gross domestic product, who is going to benefit from it? Well, you find that the industrialized countries benefit from it a lot, but the less-developed countries benefit from it *also* a lot.

It was still the case that the industrialized countries would benefit more, so in some way it broadened the gap between the industrialized and the less-developed countries. Nevertheless, all of them moved up significantly.

This was quite fascinating—both the research itself and the implications we found in these simulations.

Of course, I've also been involved in an attempt to understand how institutions affect growth, and in the more recent paper, we looked at the role of institutions in enhancing the contribution of R&D to growth. And we've found that they're quite important. For example, patent protection is an important tool. Countries that have better patent protection systems benefit more from R&D investment and also from R&D invested in other countries.

R&D, INSTITUTIONS AND INTERNATIONAL SPILLOVERS

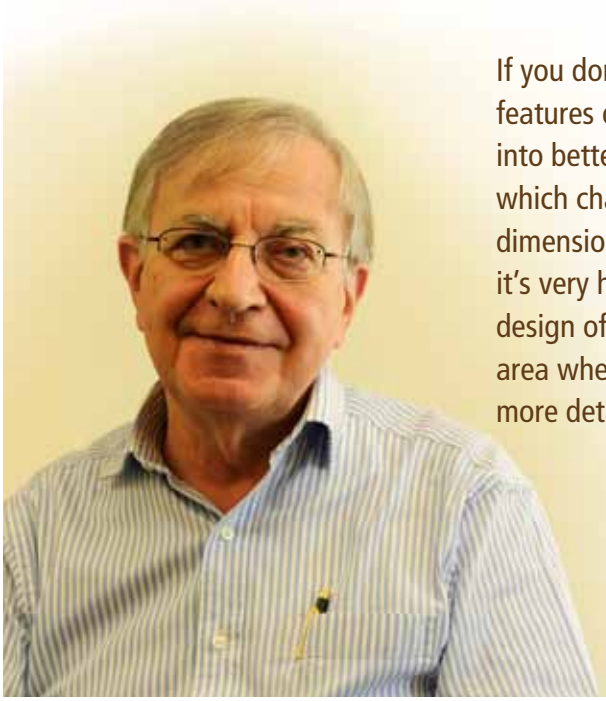
Region: What's your general sense then, from this entire body of both theoretical work and empirical research, of the importance of international R&D spillovers and institutions in contributing to economic growth?

Helpman: My sense is that institutions are very important. Of course, there has been a lot of work by other scholars on the subject, and my contribution is at the margin, to some extent. But you know, institutions impact growth through a variety of channels; R&D investment is just one of them. It is relatively less-researched than some of the others.

Let me make some general remarks on this subject. If you look at the empirical work on institutions and growth, or institutions on economic performance more broadly, then I think we, many of us, have become convinced that there exists a robust relationship between the quality of institutions and economic performance.

However, most of the empirical work is based on a broad-brush sweep. And it's hard to identify from that precise mechanisms through which institutions affect performance. There are, of course, exceptions to the rule. Generally speaking, we have these robust correlations, which in fact some people dispute, too, but let's agree that these are robust correlations.

The more important understanding that will have clear policy implications requires studies of specific mechanisms and how they work through the system in order to translate features of institutions into features of economic performance. For example, think about correlations that tell you that different legal systems have a different effect on income per cap-



If you don't understand what features of [a legal system] feed into better performance, through which channels and in what dimensions of performance, it's very hard to think about the design of policies. So, this is an area where we need a much more detailed understanding.

ita. And suppose that you're convinced that one system is better than another.

Region: That perhaps the British legal system is better for economic growth than, say, the French system.

Helpman: Right. But if you don't understand what features of the British system feed into better performance, through which channels and in what dimensions of performance, it's very hard to think about the design of policies. So, this is an area where we need a much more detailed understanding in order to be able to actually translate these broad correlations into concrete policy recommendations.

CURRENT DEVELOPMENTS IN GROWTH THEORY

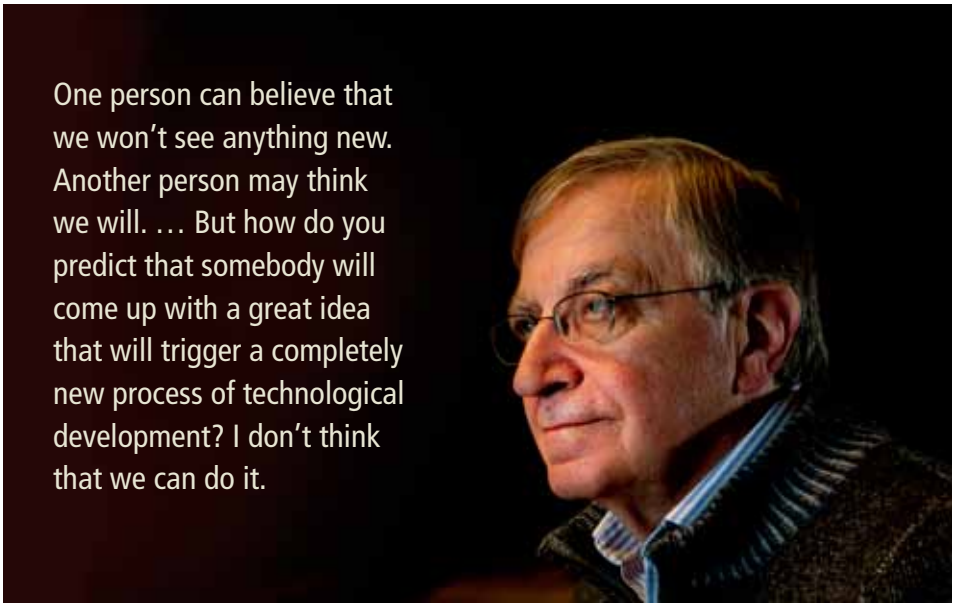
Region: Perhaps that leads to a question about current research developments. The world economy and theory itself have changed a great deal since the research of the early 1990s, so it's probably a misnomer to still call it new growth theory.

Helpman: Oh, right, it's not "new" anymore.

Region: Indeed. So, what current directions in growth theory, then, do you consider most promising? Which avenues should be pursued?

Helpman: There really hasn't been that much work on economic growth lately. A lot of work, for example, has tried to identify distortions in resource allocation, mostly empirical research. And there has been work like that of Daron Acemoglu on induced technical change.

But altogether, there hasn't been a big change in the view of the profession on economic growth. Frankly speaking, despite the fact that many papers have been published dealing with various aspects of this subject, there has been no major change in the view of the growth process.



AN END TO GROWTH?

Region: Curiously, that lack of change in the view of the growth process brings to mind a recent paper by Robert Gordon on stagnation in economic growth itself. He argues that a number of factors suggest that the rates of economic growth seen in the United States specifically over the past 250 years are not likely to be seen again. Does that seem plausible to you?

Helpman: No. I mean this is his own personal judgment, right, and that's fine. Essentially, he talks there about technologies that I would term "general-purpose technologies," which is a subject on which people worked in the past. Again, there hasn't been much work recently, but in the '90s, there was quite a bit of work on this.

So, what's a general-purpose technology? It is a type of technology on which other technological developments build. And it usually induces more specific technical change and the development of inputs that build on this technology for further production.

Region: His examples are steam engines and locomotives, I believe, electricity and …

Helpman: Yes, the steam engine was a general-purpose technology; electricity was a general-purpose technology. The microprocessor was a general-purpose technology. So there are technologies like this, which appear from time to time. And sometimes at the beginning they cause some havoc …

Region: An end to the buggy whip industry, say.

Helpman: Right. But then eventually, they trigger a process of development and growth that can be very fast and can last very long. Therefore, it is true that a number of these general-purpose technologies were big contributors to growth. But there was at least one more recently, the microprocessor.

Moreover, I don't see how we can predict how many of these technologies will emerge in the future. So, one person can believe that we won't see anything new in the near future. Another person may think that we will. I don't think we have the capability actually to predict these developments. It's easier to predict what will happen once the general-purpose technology emerges. That's not entirely easy either, but at least you have some-

thing to build on in terms of predictive power.

But how do you predict that somebody will come up with a great idea that will trigger a completely new process of technological development? I don't think that we can do it.

NEW TRADE THEORY

Region: Let me ask about new trade theory. Of course, new growth theory relates to your work on new trade theory. In the 1980s, new trade theory expanded upon neoclassical trade theory, comparative advantage based on factor proportions, labor productivity. You and Paul Krugman were the foremost leaders in developing this new work, bringing [Edward] Chamberlin's theory into the mix.

What inadequacies in traditional theory required better answers? And how did new trade theory address those weaknesses?

Helpman: When I was a student, the type of trade theory that was taught in colleges was essentially based on Ricardo's 1817 insight, Heckscher's 1919 insights and then Ohlin's work, especially as formulated by [Paul] Samuelson later on.

This view of trade emphasized sectoral trade flows. So, one country exports electronics and imports food, and another country exports chemicals and imports cars. This was the view of trade.

The whole research program was focused on how to identify features of economies that would allow you to predict sectoral trade flows. In those years, there was actually relatively little emphasis on Ricardian forces, which deal with relative productivity differences across sectors, across countries, and there was much more emphasis on differences across countries in factor composition.

In parallel, some work tried to deal with industrial organization issues in trade. One of my teachers, Richard Caves, had done at one time quite a bit of work on it, but the theory of industrial organization and trade was very slim. More generally, there was little integration of that theory with the empirical work in trade.

Two interesting developments in the 1970s triggered the new trade theory. One was the book by Herb Grubel and Peter Lloyd in which they collected a lot of detailed data and documented that a lot of trade is not *across* sectors, but rather within sectors. Moreover, that in many countries, this is the great majority of trade.

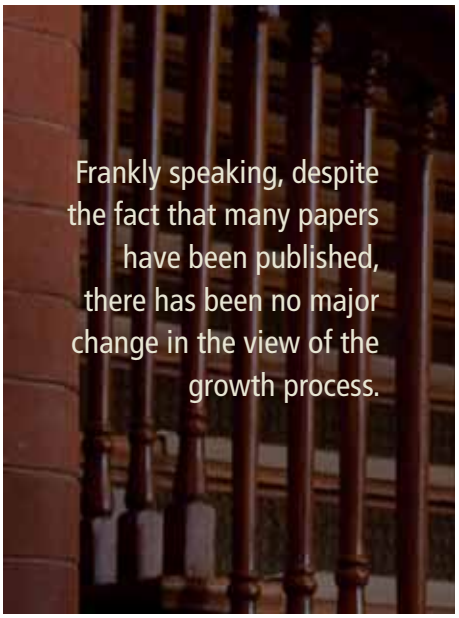
So, if you take the trade flows and decompose them into, say, the fraction that is exchanging [within sectors] cars for cars, or electronics for electronics, versus [across sectors] electronics for cars, then you find that in many countries, 70 percent—sometimes more and sometimes less—would have been what we call intra-industry trade, rather than across industries.

Region: So, for instance, looking at trade flows between the United States, Japan and Germany in, say, cars.

Helpman: Yes. You export cars, you import cars; you export electronics, you import electronics. So, Grubel and Lloyd did a great service by devising an index, which allowed a decomposition that showed the relative magnitudes of these trade flows.

The other observation that also started to surface at the time was that when you looked at trade flows across countries, the majority of trade was across the *industrialized* countries. And these are countries with similar factor compositions. There were obviously differences, but they were much smaller than the differences in factor composition between the industrialized and the less-developed countries. Nevertheless, the amount of trade between developed and developing countries was much smaller than among the developed countries.

This raised an obvious question. If you take a view of the world that trade is driven by [factor composition] differences across countries, why then do we



have so much trade across countries that look pretty similar?

Some other empirical studies raised various issues, like the work of Béla Balassa on the formation of the European Common Market. But this would take too much time to explain.

These were the empirical developments. Then, on the theoretical front, monopolistic competition was introduced forcefully by both Michael Spence in his work, which was primarily about industrial organization, and [Avinash] Dixit and [Joseph] Stiglitz in their famous 1977 paper. These studies pointed out a way to think about monopolistic competition in general equilibrium. And trade is all—or, at least then, was all—about general equilibrium.

So combining these new analytical tools with the empirical observations enabled scholars to approach these empirical puzzles with new tools. And this is how the new trade theory developed. At some level, you know, the answers are pretty simple …

Region: Well, simple in retrospect, perhaps.

Helpman: [Laughs.] Well, yes, yes. There were people like Béla Balassa who actually had the right insight. I mean, he

didn't write down the model, but when he looked at the data and he saw this, he told a story that is not that different from what the models told.

Region: I read your conversation with Daniel Trefler, in which you describe the process of writing your 1981 *Journal of International Economics* paper—a landmark paper, as he said. Arriving at those "simple" answers sounded very difficult.

Helpman: Yes, indeed. It wasn't easy at all. It was very hard actually. It's very difficult to write down a detailed economic model that describes new phenomena in a convincing way. It's very difficult.

Nevertheless, we have to do it, because this not only imposes a discipline on how we think about the problem, but there are typically unintended consequences of model building. You build a model to explain a phenomenon, but the model then has other types of predictions, and you ask yourself, are these other predictions consistent with the evidence? If they are not consistent with the evidence, then maybe there is something wrong with this model.

Generally speaking, I think this is one of the nicer things that have happened in economics in the last few decades: this interplay between theory and empirical findings. There used to be—in trade, this was definitely the case—a pretty sharp division between empirical work and theoretical work. And these new questions, and the construction of models to handle them, brought theory and empirical work much closer together.

These new models looked at product differentiation within industries. And they looked mainly at manufacturing. Today there is substantial trading in services, but at that time, it was negligible. Manufactured products are what countries used to trade. And in manufacturing industries, product differentiation is everywhere.

Region: So you look at trade flows of, say, Chevrolets and BMWs, for instance. Both cars …

Helpman: Right, both cars.

Region: But very different.

Helpman: Yes, they are different cars. And countries produce different cars. And, you know, countries produce different electronic equipment, and they produce different chemicals. And they trade them.

The first obvious conclusion you reach is that if one country produces different brands of a product from its trade partner, then they're going to exchange these brands and then you'll get intra-industry trade. This may beef up the trade volume across quite similar countries to an extent that you wouldn't be able to predict if you wanted to use differences in factor proportions across countries as drivers of trade.

So, these models provided some nice predictions that could be brought to the data. They provided indices you could look at, and they started a new research program, which has been sustained to this very day with the more recent revolution in trade research.

NEW TRADE THEORY AND MULTINATIONALS

Region: Let's move to that. It seemed to me that your 1984 paper was one of the first to develop a theory of trade and multinationals. And then firm-level data became available toward the end of the '80s, early '90s, that pointed out the importance of understanding how firms differ in their levels of trade involvement. Within the same industry, some firms trade a lot internationally while others don't.

In addition, your paper "Trade, FDI and the Organization of Firms" points out that new research is looking at the *structure* of industries and providing what you call "new explanations for trade structure and patterns of FDI and new sources of comparative advantage."

What are those theoretical refinements—the *new* new trade theory, if you will? And what are the new explanations they offer?

Helpman: In the 1990s, a lot of effort went into the integration of trade and growth. In parallel, a lot of excellent empirical work was being done. Part of it actually focused on the more traditional explanations based on differences across countries in factor proportions. This started

with the work of [Edward] Leamer in the mid '80s and some of his co-authors and continued with Trefler in a famous 1995 paper and a variety of papers that followed. This was one line of research.

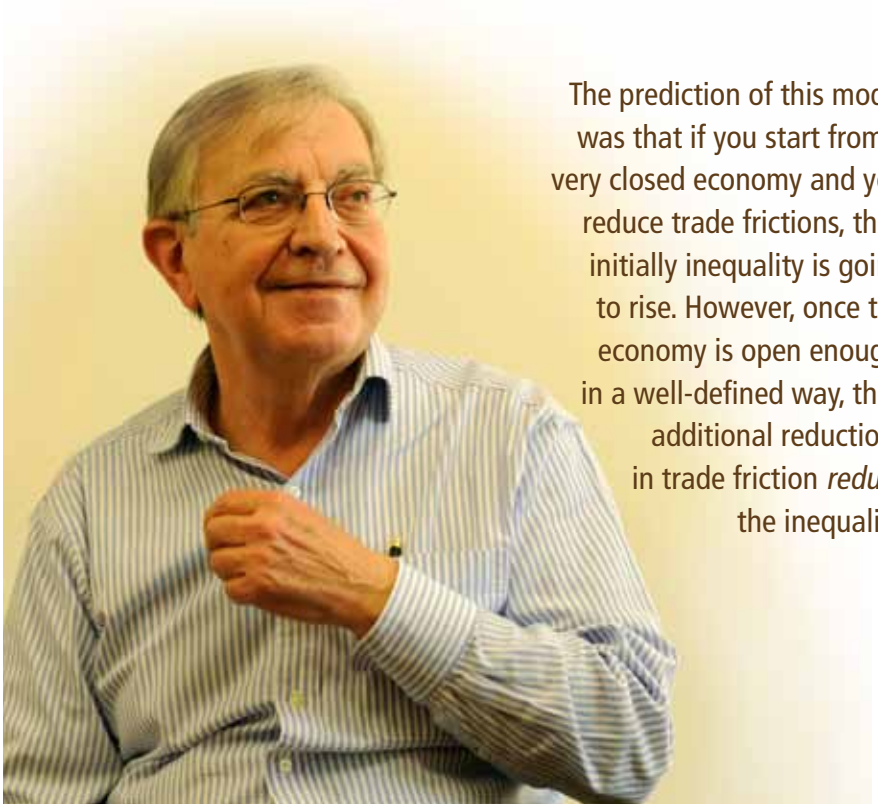
There was another line of research that evolved. Andy Bernard, for example, from Dartmouth, was a big contributor to this one. This work started to look at firm-level data sets.

In the older new trade theory—that's a funny term, no?—in the older new trade theory, there were firms, obviously, but we didn't pay much attention to the differences across firms within an industry, basically. It's not that we didn't know there was a size distribution of firms in every industry, but the questions that we asked didn't seem to require this added complexity in order to answer them. Therefore, we assumed all these firms were, basically, symmetric.

Now, the important thing about the empirical work in the 1990s that used firm-level data sets is that they identified systematic relationships between firm characteristics and their involvement in foreign trade. The key observation was that if you look across these data sets, then you find that only a fraction of firms in every industry exports, and it's not a large fraction, actually.

But this is not a random sample of firms in the industry. This is a skewed sample. In particular, the bigger and more productive firms engage in foreign trade, and the others don't. Moreover, those that export still serve the domestic market with a large share of their output.

Thus, we accumulated some insights into what you might call stylized facts about the relationship between trade and firm characteristics. And this is what triggered a reevaluation of the old new trade theory, which was then developed further, by Marc Melitz primarily but by other people as well, into the new new trade theory. The interesting thing here was that Melitz' paper—which essentially provided a theoretical explanation of these stylized facts—triggered a huge literature. And it triggered a huge literature in more than one way, one of them related to the multinational issue.



The prediction of this model was that if you start from a very closed economy and you reduce trade frictions, then initially inequality is going to rise. However, once the economy is open enough, in a well-defined way, then additional reductions in trade friction *reduce* the inequality.

Parallel to this, there was an independent development that allowed a new approach to multinationals, namely, the one based on contractual frictions. This is an interesting story because the work by [Sanford] Grossman and [Oliver] Hart on contractual frictions is from 1986. Then there was a paper by Hart and [John] Moore in 1990.

Evidently, Hart's work had been around for a while. However, it had not been integrated into international trade. And parallel to Marc Melitz' contribution, research was being built on these contractual frictions, particularly by Pol Antràs. Melitz' paper and Antràs' paper actually were published in the same year, but they dealt with very different issues.

Eventually, these two branches were integrated. As a result, we have a very comprehensive and detailed model of international trade where you can think simultaneously about the choices of firms to export, to engage in foreign direct investment, how this is related to the degree of heterogeneity of productivity within industries, how it is related to the severity of contractual frictions.

So it opens new windows, which are quite fascinating. And this research program that continues to this very day led to much better empirical work, more-refined theory; it has been a fantastic period for people working in this area.

Region: You chose your field well.

Helpman: One gets lucky from time to time. [Laughs.]

TRADE AND INEQUALITY

Region: I'd like to ask you about trade and inequality. Conventional trade theory, since Heckscher-Ohlin at least, has argued that trade should result in greater income and wage equality among nations and workers. But empirical studies generally haven't borne that out. They find an inverse relationship.

Recently, with Oleg Itskhoki and Stephen Redding, you've done a great deal of

work on the impact of trade on inequality and come up with an interesting explanation for this seeming anomaly, showing that trade seems to increase inequality initially, in contradiction to traditional theory, but eventually decreases it.

Essentially you argue for an inverted U-shaped curve between wage inequality and openness to trade, reminiscent of a Kuznets curve but for trade liberalization. Would you summarize that work and perhaps refer to your work on wage inequality in Brazil?

Helpman: Let me step back a little bit. Most of the work on trade and inequality in the neoclassical tradition was focused on inequality across different inputs. So, for example, skilled workers versus unskilled workers, or capital versus labor, and the like. There was a lot of interest in this issue with the rise in the college wage premium in the United States, which people then found happened also in other countries, including less-developed countries.

Region: So, the idea of skilled-biased technological change.

Helpman: Yes, the conclusion was that skilled-biased technological change drove wage inequality. Because if you wanted to use a trade explanation, then you should have seen opposite movements in inequality between skilled and unskilled workers in countries at different levels of development. This was one line of inquiry and debate in the literature on the impact of trade on inequality.

The other interesting thing that happened was that labor economists who worked on these issues also identified another source of inequality. They called it "residual" wage inequality, which is to say, if you look at wage structures and clean up wage differences across people for differences in their observed characteristics, such as education and experience, there is a residual wage difference, and wages are still quite unequal across people. In fact, it's a big component of wage inequality.

Our aim in this research project, which has lasted now for a number of years, was to try to see the extent to which one can explain this inequality in residual wages by trade. It wasn't an easy task, obviously, but the key theoretical insight came from the observation that once you have heterogeneity in firm productivities within industries, you might be able to translate this also into inequality in wages that different firms pay.

You know, it's not obvious that bigger and more productive firms have to pay higher wages, although empirically this is true. You can write down a model in which this doesn't happen. Now, I was interested in the question of how do different countries respond to trade when they have different labor market frictions? This is partly related to some readings of what happened in Europe in terms of labor market policies.

Region: Sure. Greater rigidity in European labor markets has been considered a source of economic underperformance in many respects.

Helpman: We tried to combine these insights, labor market frictions on the one hand and trade and firm heterogeneity on the other, and the question is, can we generate a link between trade and unequal wages paid by different firms when there are labor market frictions?

We managed eventually, after significant effort, to build a model that has this feature but also maintains all the features that have been observed in the data sets previously. It was really interesting that the prediction of this model was that if you start from a very closed economy and you reduce trade frictions, then initially inequality is going to rise. However, once the economy is open enough, in a well-defined way, then additional reductions in trade friction reduce the inequality. Now, it is not clear that this is a general phenomenon, but our analytical model generated it.

Region: So, it's an inverted U curve.

Helpman: Yes, it's an inverted U shape, and the driving force there is the following. If, within an industry, you have firms with different productivity levels, they make different strategic decisions about how to organize their production and how to integrate into foreign markets.

What happens is the bigger, more productive firms export, as we observe in the data. But the key point here is the following: Look at two firms with very close productivity levels. And say the one with lower productivity chooses not to export because this is what maximizes its profits, and the one with a somewhat higher productivity level (even if just marginally higher) chooses to export.

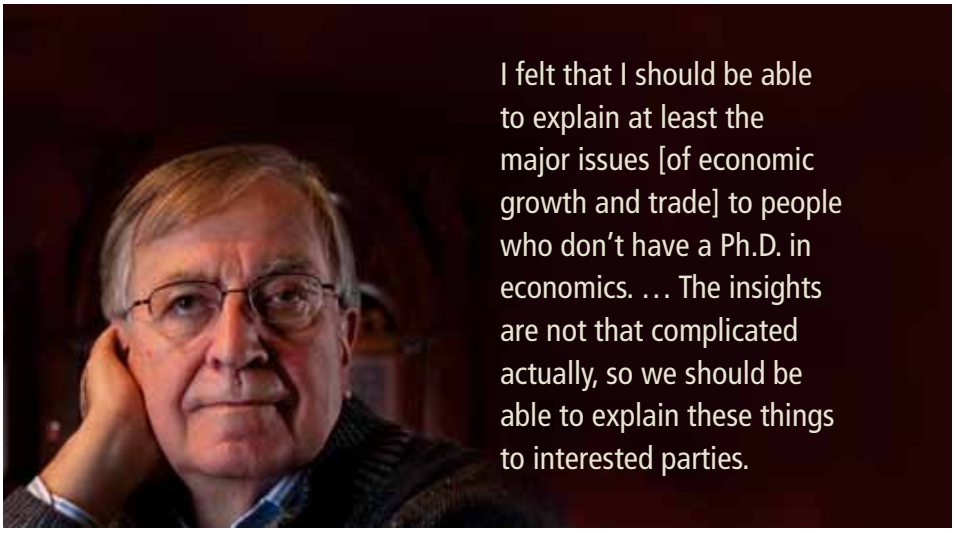
The exporter is going to respond in a discontinuous way; it will perform a big jump. Why is this? Because to export, it has to cover the fixed cost of penetrating a foreign market. Therefore, it will be significantly larger than the firm with the slightly lower productivity level. Now, if you have a mechanism—as we do in our model—in which firms screen workers and then bargain over wages, which results in a positive correlation between wages and firm productivity, then you're going to have a big jump in wages when the firm goes from nonexporting to exporting.

This generates inequality, but now, it depends where this jump takes place. If the jump takes place very close to autarky, so just a tiny number of firms in the country export, then when you remove the barriers so that more firms export, this is going to raise inequality.

But if it's a nation where almost all the firms export, yes, then the inequality is not so large because the firms with the significantly lower wages employ very few people. So now, when you liberalize trade again, and you expand the range of firms that export, you actually reduce inequality.

Region: And, empirically, you found that Brazil's trade liberalization experience was consistent with the model's prediction.

Helpman: Right. In this paper on Brazil, we wanted to see to what extent this type of model fits the data. To assess a model like this, you need very detailed data, what we call matched employer/employee data. These are data where you know in which firm every person works. In addition, you need to know the wages of every worker, their education, their experience. You need to know if the firm exports, doesn't export. Very detailed data. So we have this huge data set from Brazil on which we estimated the model, and then when we simulate the model we get the inverted U shape.



REACHING A LAY AUDIENCE

Region: Let me ask just one more question. In 2004, you wrote a wonderful book for a lay audience, *The Mystery of Economic Growth*. Then in 2011, *Understanding Global Trade*. Both books provide concise, lucid descriptions, in nontechnical language, of the historical and current research in each area, growth theory and trade theory.

I'm curious to know why, after years of deeply technical research—work that has expanded frontiers of both areas—you've chosen recently to write for a lay audience. Why did you decide, twice, to devote a substantial amount of time to each book? The opportunity cost of working on those two books was enormous, given the research time you sacrificed. Why did you consider it important to reach the lay audience as opposed to continuing to work with your colleagues to further expand the frontiers of research in either growth or trade theory?

Helpman: You know, maybe I made the wrong cost/benefit calculation. [Laughs.]

Region: I certainly don't think so.

Helpman: I just had the urge to do it, frankly speaking. It came from the fact that very few people outside the profession—or even in the profession who were not close to this research line—really understood or knew about the importance of the research that has been done.

So it started with growth theory, yes? I was engaged in work on growth for a long time, and at the Canadian Institute for Economic Research, we have had a group that worked on this subject. But each time I talked to people from other fields in economics and certainly to people outside economics, they knew relatively little about the subject.

And by that time, I felt that we had a good enough understanding of this research that we could explain it actually in nontechnical terms. It is not always possible and it usually takes a long time, for whatever reason. I don't know exactly why. It's just something about how our brains work that over time we gain a better understanding of things, even if we are not working on them, necessarily. But obviously, you keep thinking about these issues time and again. You try to explain them to other people.

I felt that the topic was obviously very important—economic growth—and I felt that I should be able to explain at least the major issues to people who don't have a Ph.D. in economics, basically, or who have an economics Ph.D. but work in labor economics markets or economic development or whatever. So I don't know, I developed this urge to do it, and I sat down and did it.

Region: And did it again, seven years later.

Helpman: Yes, it was the same. However, I would not have done it on the trade book if we had not had these wonderful 10 years of research where we had an explosion of new insights. Again, I felt the work had become more and more technical, on both the theoretical side and the empirical side. Nevertheless, the insights are not that complicated actually, so we should be able to explain these things to interested parties. And, well, I decided to do it.

Region: I'm very glad you did.

Helpman: Well, actually, I'm glad I did it too. It took a lot of time obviously and, you know, we don't have too much time. [Laughs.]

Region: Very true, and I've used more than my share of yours. Thank you so much.

Helpman: You're very welcome. Good to meet you.

—Douglas Clement
Oct. 25, 2012

More About Elhanan Helpman

Current Positions

Galen L. Stone Professor of International Trade, Harvard University; on faculty since 1997

Emeritus Professor, Tel Aviv University; on faculty 1974-2004

Director, Program on Institutions, Organizations and Growth, Canadian Institute for Advanced Research, since 2004

International Research Fellow, Kiel Institute of World Economics, since 2002

Research Fellow, CESifo, since 2002

Fellow, Canadian Institute for Advanced Research, since 1992

Research Fellow, Center for Economic Policy Research, since 1992

Research Associate, National Bureau of Economic Research, since 1986

Previous Affiliations

Editor, *Quarterly Journal of Economics*, since 2008

Board of Editors, *Journal of Economic Integration*, since 2003

Member, National Council for Research and Development, Government of Israel, 1995-96

Member, Board of Directors, Bank Hapoalim, 1993-96

Member, Council for National Economic Planning, Ministry of Economics and Planning, Government of Israel, 1992-96

Member, Advisory Board and Advisory Committee, Bank of Israel, 1988-89

Honors

Corresponding Fellow, British Academy, since 2012

Onassis Prize in International Trade, London, 2012

Distinguished Fellow, American Economic Association, 2010

Honorary Doctorate, Catholic University of Louvain, Belgium, 2010

Erwin Plein Nemmers Prize in Economics, Northwestern University, 2010

Fellow, European Economic Association, since 2004

Member, European Academy of Sciences and Arts, since 2004

EMET Prize, A.M.N. Foundation for the Advancement of Science, Art and Culture, 2002

Rothschild Prize, Yad Hanadiv Foundation, 2002

President, Econometric Society, 2000; Fellow, since 1986

Bernhard Harms Prize, Kiel Institute for World Economics, 1998

Foreign Honorary Member, American Academy of Arts and Sciences, since 1993

Honorary Member, American Economic Association, since 1991

Israel Prize, 1991

Mahalanobis Memorial Medal, Indian Econometric Association, 1990

President, Israeli Economic Association, 1989-91

Member, Israeli Academy of Sciences and Humanities, since 1988

Publications

Author or co-author of seven books on international trade, economic growth and political economy, including, most recently, *Understanding Global Trade* (Belknap Press of Harvard University, 2011) and *The Mystery of Economic Growth* (Belknap Press, 2004); editor or co-editor of additional books on those and other subjects; author of numerous journal articles about balance of payments, exchange rate regimes, stabilization programs and foreign debt, among other topics

Education

Harvard University, Ph.D., economics, 1974

Tel Aviv University, M.A. (*summa cum laude*), economics, 1971

Tel Aviv University, B.A. (*cum laude*), economics, statistics, 1969

Research Digest

In this issue of *Research Digest*, we summarize work by Cristina Arellano, Yan Bai and Patrick Kehoe on their model of the U.S. macroeconomy during the Great Recession, Sam Schulhofer-Wohl and Taryn Dinkelman on the cost of congestion, and Victor Ríos-Rull and Josep Pijoan-Mas on factors most significant in determining longevity.

Due to an editing oversight, some editions of the September 2012 *Region* included extraneous text in the Digest of "Financial Frictions and Fluctuations in Volatility," Minneapolis Fed Staff Report 466. "A promising parable," printed here, is a corrected version. That issue also included a misspelling of Juan Pablo Nicolini's name in the photo caption accompanying the Digest of "Unconventional Fiscal Policy at the Zero Bound," Minneapolis Fed Working Paper 698. The editors deeply regret both errors.

"A promising parable" *Cristina Arellano, Patrick Kehoe and Yan Bai develop a model that convincingly generates macro patterns of the Great Recession*

PHOTOGRAPH AT LEFT: STEVE NIEDORF; AT RIGHT: STAN WALDHAUSER

Cristina Arellano (seated) and Yan Bai; Patrick Kehoe at right

Research Digest

In "Financial Frictions and Fluctuations in Volatility," a Minneapolis Fed staff report published in July (SR466 online at minneapolisfed.org), economists Cristina Arellano and Patrick Kehoe of the Minneapolis Fed and Yan Bai of the University of Rochester develop a model that can convincingly generate several central macroeconomic patterns seen in U.S. data during the Great Recession. In particular, the economists explore the financial and microeconomic underpinnings of sharp declines in employment and economic output between 2007 and 2009, accompanied by relatively stable labor productivity. In almost all recessions, productivity and output *both* decline, but in the most recent downturn, *productivity was nearly unchanged*. What economic mechanisms account for this anomaly?

One clue that informs their investigation is the severe credit contraction during the recent U.S. financial crisis. Another clue, at the micro level, is the large increase in dispersion of growth rates among firms—that is to say, growth at some companies suffered very little during the crisis, while other firms contracted dramatically. Even during normal times, companies grow at different rates, of course, but during the 2007-09 recession, the range between the highest and lowest growth rates nearly doubled.

These observations are building blocks for a quantitative model with heterogeneous firms (for which growth rates can differ) and financial frictions (meaning that credit markets don't function smoothly). The economists' goal is to create a model in which increasing volatility at the firm level leads to higher dispersion in firms' growth rates along with declines in both aggregate labor and economic output, but stable labor productivity. Their aim, in short, is to better understand the U.S. economy during the recent recession by building a model that can replicate its behavior between 2007 and 2009.

Central to the model: Risk, and firms hedging against it by trimming financial obligations wherever feasible—specifically, by hiring fewer inputs. "They key idea in the model," write the economists, "is that hiring inputs to produce output is a risky endeavor."

Firms receive revenue from selling their output only *after* they have already paid for inputs, such as employees, necessary to produce that output. Hiring labor (or buying materials or purchasing machinery) therefore entails risk, since demand for a firm's output may fall after the input expenditure is incurred. If financial markets were "complete," as economists say, firms could protect themselves against that event by borrowing against future profits;

The economists' goal is to create a model in which increasing volatility at the firm level leads to higher dispersion in firms' growth rates along with declines in both aggregate labor and economic output, but stable labor productivity.

but in this model, financial market frictions mean that firms must bear the risk themselves.

"This risk has real consequences if, when firms cannot meet their financial obligations, they must experience a costly default," observe the economists. "In such an environment, an increase in uncertainty arising from an increase in the volatility of idiosyncratic shocks leads firms to pull back on their hiring of inputs." (Though the word "hiring" suggests employees only, here it applies to other inputs as well: raw materials, capital equipment and the like.)

If we build it, will it work?

The economists proceed in stages. First, they build a "benchmark" model. Then they calibrate and quantify it to gauge how well it matches real U.S. data. They create two alternatives to their benchmark model to pinpoint whether the results are driven by both factors

Research Digest

(imperfect financial markets and volatility shocks) or just one. Lastly, they extend their model with refinements that bring it closer to how economists believe economies truly work.

The model has three key pieces:

(1) Firms hire inputs before knowing how much demand they'll experience for their output.

(2) Financial markets don't necessarily provide firms with credit, and they're especially averse when the economy is volatile; as a result, firms default if they're unable to pay their debts.

(3) Since firms pay a fixed cost to start their operations, they make positive profits in the future to cover those fixed costs; the cost of default is the loss of future expected profits.

These three essential parts mean that firms trade off expected risk and return whenever they choose their inputs. Hiring more inputs enables them to make more profit as long as they don't default. But because more hiring raises their financial obligations, it also increases the chance of defaulting. It's a tough choice, and becomes more so when the broader economy is looking uncertain—or, in the idiom of economics, "when the variance of idiosyncratic shocks increases."

The model includes identical households, heterogeneous firms and financial intermediaries. Households buy goods produced by firms, but the demand for each good is subject to idiosyncratic demand shocks. The volatility of these demand shocks varies over time, and this is the source of aggregate fluctuations in the model.

Firms are the guinea pigs in this model. They differ from one another, and they face not only volatile demand for their products, but imperfect or incomplete financial markets that don't allow them to insure against fluctuations in that demand. Thus, they may sink or swim based in large part on those fluctuations, as well as their hiring decisions. If they default on their debts, they fail: They "exit the market."

Benchmark and beyond

The benchmark model is calibrated to the U.S. economy with standard values for such variables as interest rates, annual sales growth for firms and the like. The economists test the model with these parameters by checking whether it can match U.S. data accurately; it does—with, for example, the fraction of labor employed by new firms at 1.8 in both data and model, and the liability-to-sales ratio at 5.5 in the data versus 5.6 from the model. A near-perfect fit.

Then they see how it responds to "impulses"—that is, how the model's mechanism reacts to a sudden increase in demand volatility. In this test, just as in the actual U.S. economy during the recent crisis, the model's output and labor (that is, employment) drop strongly when volatility increases, but labor productivity (defined as the ratio of gross domestic product to aggregate employment) increases slightly at first and then stabilizes. "The overall response," the economists write, referring to labor productivity, "is fairly flat compared to the responses of output and labor."

In addition, wages fall about 1.4 percent after the volatility shock and then continue a slow decline, and the interest rate drops just a bit initially and remains slightly depressed. The benchmark, in short, works well as a representation of the U.S. economy during the financial crisis, at least for one-time shocks in demand volatility.

They then build two alternate versions of the benchmark to investigate whether this success is due primarily to its inclusion of incomplete financial markets or to its volatility shocks. This investigation finds that "*both* financial frictions and the source of the shocks—volatility instead of productivity—are critical to our benchmark model's results" (emphasis added). In other words, neither financial frictions by them-

Research Digest

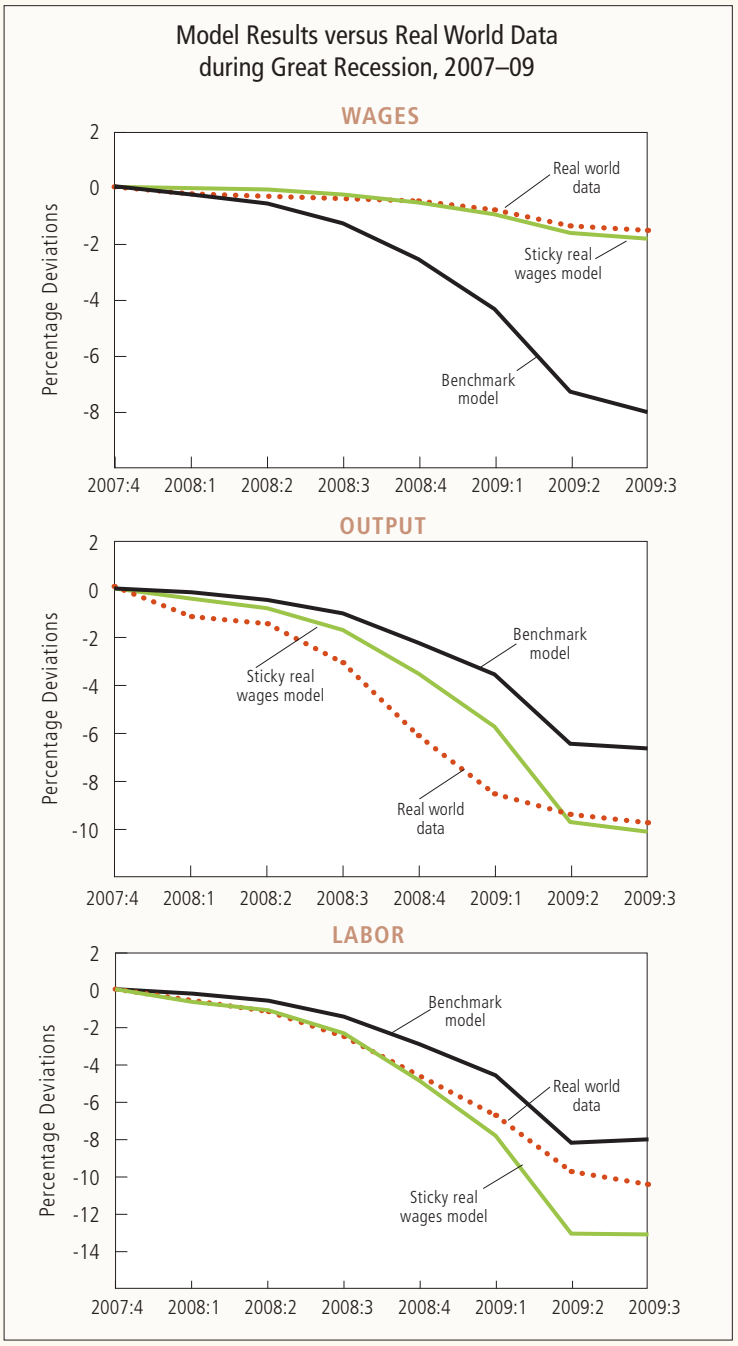
selves, nor just volatility shocks, are able to generate economic responses that resemble the real world during the Great Recession.

Real world testing

But the fundamental question is, how well can this model account not for a theoretical *one-time* volatility shock, but for a *series* of shocks like those experienced in the real economy during the Great Recession? The answer: very well. "We show that our model can account for much," the economists write.

To reach that conclusion, they first find the volatility shock sequence that generates dispersion among firms' sales growth rates similar to that actually measured in U.S. data between late 2007 and the third quarter of 2009. The data reveal nearly a doubling in this range of growth rates, from 17 percent to 31 percent. The economists feed that shock sequence into their model and see what happens to macroeconomic output, labor and productivity.

Given how crude the model is—in the sense of leaving out countless aspects of an actual national economy—it does a remarkable job of generating results similar to real world figures. "The model generates a decline in output of 6.5 percent, whereas in the data output declines 9.7 percent," they find. And it "produces about an 8 percent decline



Research Digest

The fundamental question is, how well can this model account not for a theoretical one-time volatility shock, but for a series of shocks like those experienced in the real economy during the Great Recession? The answer: very well. "We show that our model can account for much."

in labor, whereas in the data labor declines about 10 percent."

While not dead on, the model's results are quite close, suggesting that the mechanisms at its heart are what drive the actual economy, through good times and bad. When the economists summarize the overall results, they conclude that the model "can explain 67 percent of the overall contraction of output and 73 percent of the contraction in labor during the Great Recession." The model produces a fairly flat productivity profile for the recession, while in real data, productivity first falls and then rises modestly. But "both in the model and in the data, productivity at the end of this event is essentially unchanged … even though output has fallen 10 percent."

Refinement

The economists explore several dimensions of, and refinements to, their model. One is to alter the model by introducing "sticky wages," the idea that in the real world, most prices don't change instantly. A gallon of gasoline may rise or fall in price several times a day or week, but wages, automobiles and even items on a restaurant menu take a while to adjust to trends in the economy—to a broad recession or to a rise in the cost of health care, steel or eggs. This factors into the model, since in the benchmark version of the model, wages fall when volatility increases, and such response dampens the labor adjustment firms make.

And indeed, by making the model's input prices less responsive to volatility, the economists find that sticky prices "diminish offsetting equilibrium effects." The charts on page 35 show their results. They compare real wage trends in the data, the benchmark model and the sticky real wage model for the entire span of the Great Recession and show that while they drop by about 2 percent in the data and over 8 percent in the benchmark model, "in the sticky real wage economy, real wages drop about the same as in the data." Sticky real wages also amplify the output and employment effects of increased volatility.

Thus, Arellano, Bai and Kehoe's model, with key features and additional enhancements, does a striking job of duplicating patterns seen in the U.S. economy in recent years. "Hence," they conclude, "we think of the model as a promising parable for the Great Recession of 2007-2009."

—Douglas Clement

Research Digest

PHOTOGRAPH BY STEVE NIEDORF

Sam Schulhofer-Wohl

Tragedy of the commons

Measuring the unintended consequences of infrastructure projects in developing countries

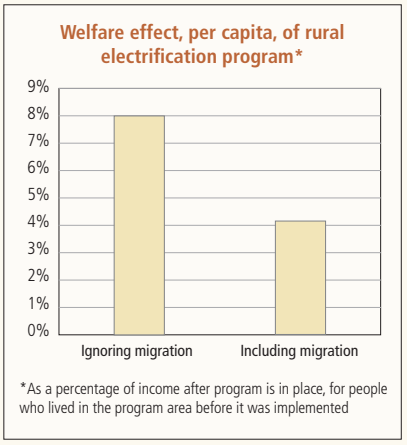
The benefits of government infrastructure projects in developing countries are obvious: Irrigation systems increase crop yields; schools produce educated, productive citizens; health clinics and sewage treatment plants enhance wellbeing. What isn't so evident—in large part because it's difficult to measure—is what happens when people flock to an area to take advantage of these benefits. New infrastructure may raise incomes and improve quality of life, but it may also put pressure on other community resources such as housing or transportation.

To measure the "congestion" effects from migration, economists typically use land prices as a proxy; new arrivals invariably drive up rents. But reliable price data simply aren't available in many parts of the world. Recent research by Taryn Dinkelman, an economist at Dartmouth College, and Sam Schulhofer-Wohl, a senior research economist with the Minneapolis Fed, demonstrates an alternative method for gauging often overlooked migration effects in less-developed countries.

In "Migration, Congestion Externalities, and the Evaluation of Spatial Investments"

Research Digest

Unlike standard analytical methods, the model doesn't rely on land prices to estimate the impact of migration on welfare. Instead, the model looks at income and population data to determine the net effect of infrastructure improvements.



(Minneapolis Fed Working Paper 700, online at minneapolisfed.org), the economists use population growth as a yardstick for congestion and find that the impact of migration can be considerable, especially in areas where land is not priced. Rural South Africa is a case in point; in studying the consequences of an electrification project in that country, Dinkelman and Schulhofer-Wohl estimate that congestion effects, including crowded settlements and schools, cut the project's per capita benefits in half.

If you build it …

The notion that congestion can diminish the benefits of location-specific infrastructure programs is well established, although migration effects have received less attention than direct, positive outcomes of those programs, such as higher incomes and improved health. A new public amenity like a hospital or a water treatment plant will continue to draw people until crowding of other shared public resources becomes so severe that in-migration ceases. "That intuition is pretty well understood in economics," Schulhofer-Wohl said in an interview. "The challenge is how to measure that effect."

To investigate the welfare impact of migration, the economists develop a model in which infrastructure upgrades in a rural area induce people to move there from the city. The government-funded facility raises local incomes (by allowing women to work outside the home, for example), but also increases the population—and demand for other public goods such as subsidized housing, schools and bus service.

Unlike standard analytical methods, the model doesn't rely on land prices to estimate the impact of migration on welfare. Instead, the model looks at income and population data to determine the net effect of infrastructure improvements. "The existing methods work if there is a land market and you can observe the prices," Schulhofer-Wohl said. "What we contribute is how to analyze these programs if either there's a land market but you can't observe the prices, or there isn't a land market."

That is the case in many rural areas of developing countries, including Dinkelman's native country of South Africa. Large expanses of that country are state owned or communally held and are allocated based on tribal or family ties.

To put their model to the test—and illustrate the importance of migration in assessing the worth of infrastructure projects—the researchers analyze a government electrification project in KwaZulu-Natal (KZN), a South African province with high unemployment and no land market. The project extended electrical service to about 200,000 households in the late 1990s, and the primary impacts on labor market outcomes were documented by Dinkelman in a 2011 paper. [1]

Research Digest

Welfare drain

Electrification in KZN increased employment, raising average incomes; but it also led to dramatic population increases in comparison with communities that didn't receive electrical hookups. One outcome of population gain was crowded schools; student-teacher ratios rose by two-thirds relative to villages that remained off the grid.

To calculate the net per capita welfare impact—the degree to which congestion effects offset income gains—the economists feed into their model summary income and Census data gleaned from over 1,800 rural KZN communities. The output of the model is the monetary value of the project to residents, measured as a fraction of monthly income. It turns out that when congestion effects are accounted for, roughly half of that value—the per capita welfare gain from the electrification project—disappears (see chart). Thus, the study "provides the first empirical evidence from a developing-country context that congestion effects exist and can be quantitatively large," the researchers write.

The model also shows that migration undercuts the benefits of infrastructure projects the most in places like rural South Africa that lack a functioning land market. Without rising land prices to signal increasing congestion, people keep moving into the rural area, consuming more communal land and other public goods and reducing welfare gains for all—an outcome that the researchers view as a version of a tragedy of the commons. Migration exacts a lower toll in areas with land markets because congestion is less severe, and landowners benefit from higher rents.

Dinkelman and Schulhofer-Wohl see "broad relevance" for their model in gauging the costs and benefits of infrastructure projects in developing countries, where people are becoming increasingly mobile.

Accounting for mobility

Dinkelman and Schulhofer-Wohl see "broad relevance" for their model in gauging the costs and benefits of infrastructure projects in developing countries, where people are becoming increasingly mobile. In some cases, the net benefits of such programs may be less than supposed, because of resulting strains on public resources that are slow to respond to population inflows.

In areas without land markets or reliable price data, a means of quantifying congestion effects could help planners mitigate welfare losses—by spreading out projects, for example, or simultaneously expanding other public services such as schools or health clinics.

"Our hope is that people will use our work as a building block to be able to account for migration in evaluating these programs," Schulhofer-Wohl said.

—Phil Davies

[1] Dinkelman, Taryn. 2011. "The effects of rural electrification on employment: New evidence from South Africa." *American Economic Review* 101 (7): 3078-3108.

Research Digest

PHOTOGRAPH BY STEVE NIEDORF

Victor Ríos-Rull

Who lives longer?

Education appears to be a far more powerful factor than wealth or marital status

Reprinted with permission from voxeu.org

Economists have long been worried about income inequality and its effects on welfare. For instance, workers with a college degree earn on average much more than those who did not complete high school. This disparity translates into large differences in consumption levels and hence welfare (see, for instance, Heathcote, Storesletten and Violante 2010). We argue, however, that these welfare differences are dwarfed by the differences in longevity between individuals in different socioeconomic groups, and mainly by differences in longevity between individuals of different educational levels.

In recent research (Pijoan-Mas and Ríos-Rull 2012), we use the Health and Retirement Study (HRS) to document the expected longevities at age 50 of different population subgroups of white men and white women. In particular, we look at the different expected longevities by educational groups, wealth quintiles, labor market status and marital status.

Research Digest

Main results

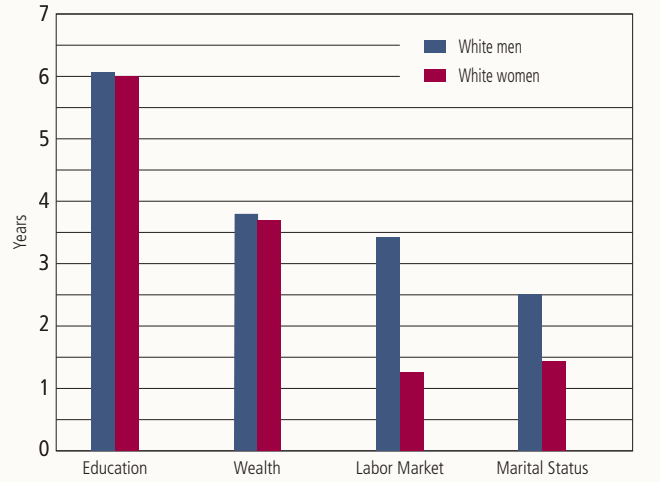
Figure 1 shows that the most important differences are linked to education, which turns out to be much more important than wealth. At age 50, a college-educated white man can be expected to live 6.1 more years than a high school dropout; in contrast, a white man in the top quintile of the wealth distribution is expected to live 3.8 more years than a white man in the bottom quintile. Very similar differentials hold for women.

In addition, we find that a white man fully attached to the labor market (as a full-time worker or as an unemployed worker actively looking for a job) is expected to live 3.4 more years than an inactive individual; and a married white man can be expected to live 2.5 more years than an unmarried one. The differentials for women are substantially smaller, but still large.

To obtain these differentials, we did not compute life expectancies. Instead, we estimated a hazard model for survival, with the socioeconomic characteristic of interest and (self-assessed) health status as stochastic endogenous covariates. Then we used these estimates to compute expected life durations at age 50 for each group. Our methodology allows us to bypass the two problems associated with the use of life expectancy. The first problem is that people's socioeconomic characteristics evolve over the life cycle (except for education) and hence so do the relevant mortality rates. For instance, one-third of white women who are married at age 50 become divorced or widowed before age 70. The second problem is that mortality rates tend to decline over time, and this may happen at different rates for people in different socioeconomic groups.

Figure 1

Differentials in expected longevities at age 50 between population subgroups



Decomposition

When we look at these longevity differences in more detail, we learn that they must be due to factors that evolve slowly with age. In particular, we use our estimates to decompose the differentials in expected longevity into three components:

(a) differences in health among socioeconomic types already present at age 50,

(b) different evolution of health conditional on socioeconomic status, and

(c) different mortality rates by individuals with identical health but different socioeconomic status.

As Figure 2 (men) and Figure 3 (women) show, the differences in

Research Digest

longevity are mainly due to the health-protecting nature of good socioeconomic conditions over the years, which is found both in the health differences at age 50 and in the different evolution of health afterward. In contrast, differences in mortality matter very little. For instance, the difference in the initial distribution of health between college graduates and high school dropouts generates 1.7 years of life expectancy difference for men and 1.1 years for women. Then, the fact that health deteriorates less for highly educated people generates a life expectancy gap of 4.7 years for men and 4.9 years for women. Finally, the effect of education-specific mortality is very small: 0.0 years for men and 0.3 years for women.

Time trends

We obtained our results with the pooled HRS data, which range from 1992 to 2010. The large temporal span of the HRS can be used to obtain some information about how these differentials in expected longevity have evolved over time. Previous estimates document large increases in life expectancy differences between education groups (see, for instance, Preston and Elo 1995; Meara, Richards and Cutler 2008; and Olshansky et al. 2012). Consistently, we find that the differentials for education have increased, between 1992 and 2008, by 1.8 years

Figure 2

Decomposition of longevity differentials, white men

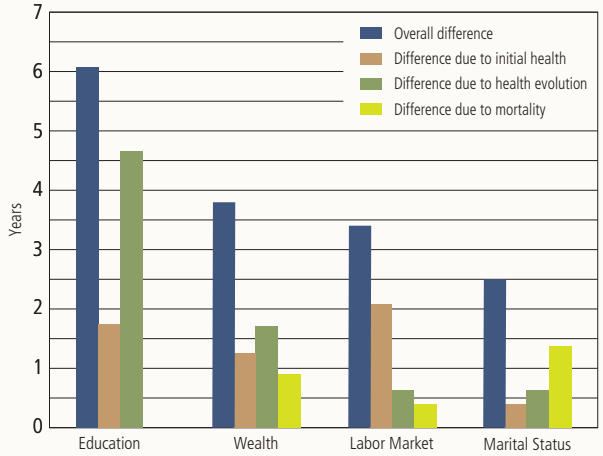
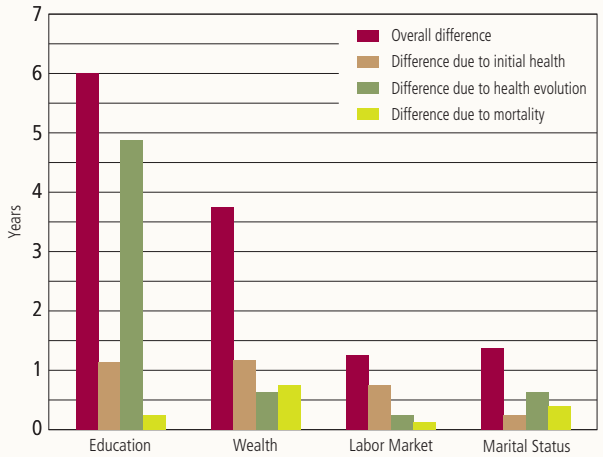


Figure 3

Decomposition of longevity differentials, white women



Research Digest

for men and 1.7 years for women. In addition, we also document important increases for wealth (1.4 years for men, 0.7 years for women), for labor market attachment (0.7 years and 0.6 years) and for marital status (1.0 years and 1.5 years).

These large increases happened during a time period when there was a sizable increase in income and wealth inequality. Although we do not want to make any causal statement, it is hard to avoid thinking that the increase in income inequality lurks behind the increase in the socioeconomic gradient of longevity. If so, we should conclude that the upsurge of income inequality in recent decades has had welfare implications much stronger than previously thought. Our results also show, however, that education seems to matter more than wealth. Therefore, it might very well be that the increase in the socioeconomic gradient of longevity is also tightly related to selection: Over the years, the pool of less-educated or unmarried people has become worse off in terms of their ability to survive.

—Josep Pijoan-Mas and Victor Ríos-Rull

Although we do not want to make any causal statement, it is hard to avoid thinking that the increase in income inequality lurks behind the increase in the socioeconomic gradient of longevity. If so, we should conclude that the upsurge of income inequality in recent decades has had welfare implications much stronger than previously thought.

References

Heathcote, Jonathan, Kjetil Storesletten and Giovanni Violante. 2010. "The macroeconomic implications of rising wage inequality in the United States." *Journal of Political Economy* 118 (4): 681-722.

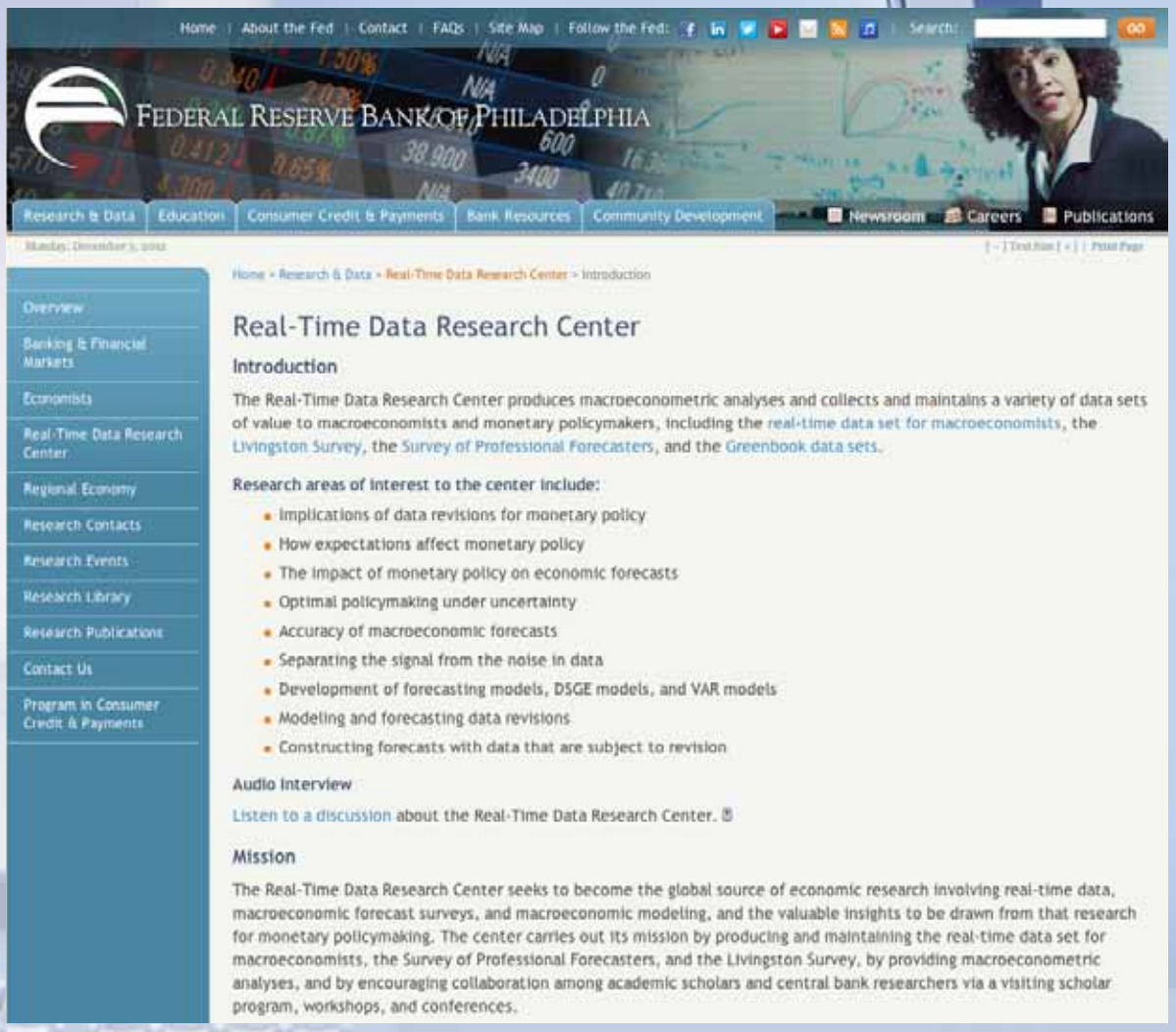
Meara, Ellen R., Seth Richards and David M. Cutler. 2008. "The gap gets bigger: Changes in mortality and life expectancy, by education, 1981-2000." *Health Affairs* 27 (2): 350-60.

Olshansky, S. Jay, Toni Antonucci, Lisa Berkman, Robert H. Binstock, Axel Boersch-Supan, John T. Cacioppo, Bruce A. Carnes, Laura L. Carstensen, Linda P. Fried, Dana P. Goldman, James Jackson, Martin Kohli, John Rother, Yuhui Zheng and John Rowe. 2012. "Differences in life expectancy due to race and educational differences are widening, and many may not catch up." *Health Affairs* 31 (8): 1803-13.

Pijoan-Mas, Josep, and Victor Ríos-Rull. 2012. "Heterogeneity in expected longevities." CEPR Discussion Paper 9123.

Preston, Samuel H., and Irma T. Elo. 1995. "Are educational differentials in adult mortality increasing in the United States?" *Journal of Aging and Health* 7 (4): 476-96.

Virtual Fed



Right on time

It has been said that trying to gauge the health of the economy is like driving with only the rearview mirror as a guide. The data that are available to economists come out with a lag—sometimes a week, sometimes several months after the fact. While looking at the recent past is informative, it doesn't quite tell you where you are right now, or where you're truly headed.

To get the most current read on the economy, the Philadelphia Fed established the Real-Time Data Research Center. Along with forecasts and economic research, the center collects and makes available the most up-to-date economic data. While its primary audience is economists and policymakers, the site is useful for anyone curious about the state of the economy. Of particular interest is the center's real-time business conditions index, which is updated every few days as new data are released.

Don't lag behind: philadelphiafed.org/research-and-data/real-time-center/

—Joe Mahon