



More Money: Understanding Recent Changes in the Monetary Base

William T. Gavin

The financial crisis that began in the summer of 2007 took a turn for the worse in September 2008. Until then, Federal Reserve actions taken to improve the functioning financial markets did not affect the monetary base. The unusual lending and purchase of private debt was offset by the sale of Treasury securities so that the total size of the balance sheet of the Fed remained relatively unchanged. In September, however, the Fed stopped selling securities as it made massive purchases of private debt and issued hundreds of billions of dollars in short-term loans. The result was a doubling of the size of the monetary base in the final four months of 2008. This article discusses the details of the programs that the Fed has initiated since the crisis began, shows which programs have grown as the monetary base grew, and discusses some factors that will determine whether this rapid increase in the monetary base will lead to rapid inflation. (JEL E31, E42)

Federal Reserve Bank of St. Louis *Review*, March/April 2009, 91(2), pp. 49-59.

The monetary base is the sum of currency in circulation and bank deposits at Federal Reserve Banks. Between mid-September and December 31, 2008, the U.S. monetary base increased from approximately \$890 billion to \$1,740 billion, doubling in a little more than 3 months.¹ This is a concern because, under normal circumstances, we would associate such a rapid rise in the monetary base with a sharp acceleration of inflation. But today, more people seemed to be worried about deflation than a sudden rebound of inflation. The purpose of this article is to explore the sources of growth in the monetary base and to ask whether or not

we should expect to see high inflation following such rapid monetary growth.

Figure 1 shows that this rapid surge in the monetary base is concentrated entirely in the accumulation of bank reserves. (Throughout this article, the generic word “bank” is used instead of the official term, “depository financial institution.”) Bank deposits at the Fed include three components. Two are small and have changed little since the economic crisis began in August 2007; they are deposits used to satisfy reserve requirements and those used to satisfy required clearing balances.² The third component, “excess reserves,” accounts for the doubling of the monetary base. This rapid increase is directly related to Federal Reserve programs initiated or expanded

¹ These data are derived from the Fed’s H4.1 release (www.federalreserve.gov/releases/h41/). This measure of the monetary base is named as the series WSBASE on the Federal Reserve Bank of St. Louis’s FRED database. For technical reasons (adjustments for seasonal factors, reserve requirements, carryover, “as of,” and cash items in process of collection), the numbers here do not correspond to either the Board of Governor’s measure of the monetary base on the H3 release or the St. Louis adjusted monetary base.

² See Stevens (1993) for a description of required clearing balances. See Anderson and Rasche (2001) for a description of the sweep programs that reduced the amount of required reserves essentially to that which would normally be held as a buffer for clearing checks and meeting uncertain cash withdrawals.

William T. Gavin is a vice president and economist at the Federal Reserve Bank of St. Louis. The author thanks Dick Anderson, Bob Rasche, and Dave Kemme for helpful comments. Chris Martinek provided research assistance.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

since September that seek to improve the functioning of financial markets under stress.

Since lowering its federal funds rate target to the range of 0 to 0.25 percent, the Federal Open Market Committee (FOMC) has referred to its latest policy actions as “credit policy.” These new programs are distinguished from traditional monetary policy by the type of assets purchased by the Federal Reserve. Traditional programs involve the purchase and sale of U.S. Treasury securities, whereas the new credit-oriented policies involve the purchase of non-Treasury securities, including commercial paper and asset-backed securities. By purchasing such assets, the Fed hopes to reduce risk premiums and improve flows through the specific private markets (Bernanke, 2009). Yet, although the emphasis of these credit programs is on the types of non-Treasury securities being purchased (that is, the *composition* of the Fed’s assets), nontraditional and traditional programs share one common characteristic: The purchase of *any* asset by the Fed, unless offset by some other action, increases simultaneously *both* the Fed’s balance sheet assets *and* its liabilities.

The next section of this article shows the Fed’s balance sheet in January 2007 and again in January 2009, highlighting the balance sheet changes since the summer of 2007.³ It highlights the new programs and shows which have contributed most to the recent surge in the monetary base. The following section then discusses economic and institutional factors that will influence the Fed’s ability to maintain price stability as the economy recovers from the recession and the financial crisis.

THE FED’S BALANCE SHEET— BEFORE AND AFTER

It would be an error to believe that the Fed’s new programs to improve the functioning of credit markets began only in September 2008; they did

not. But a major change did occur in September: The Fed stopped selling Treasury securities as it increased lending to financial institutions and purchased non-Treasury assets. As Figure 1 shows, before mid-September 2008, the Fed’s practice of selling Treasuries as it purchased other financial instruments largely insulated the monetary base from these new programs—bank deposits at the Fed increased little.⁴

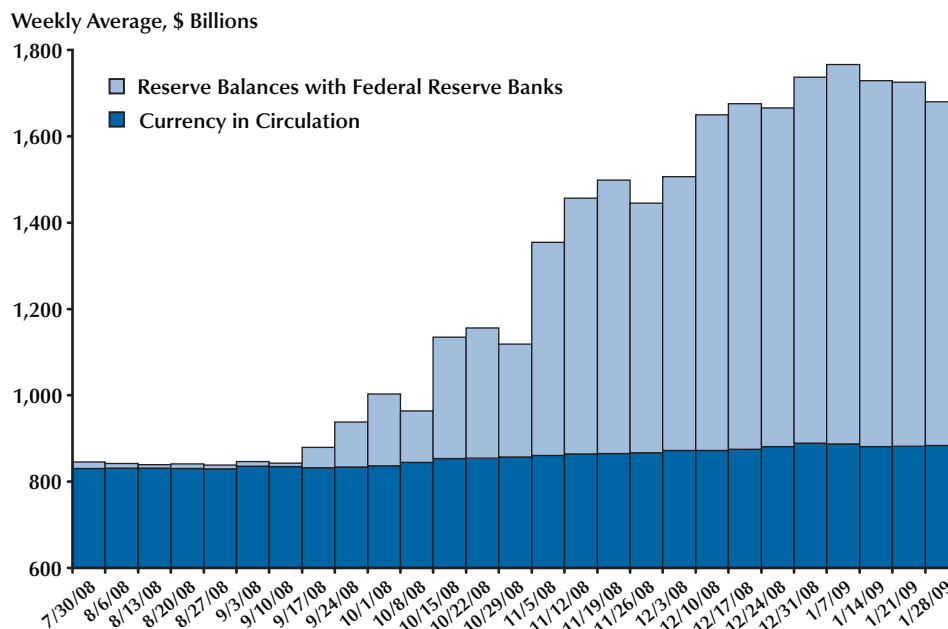
Table 1 presents a somewhat simplified view of the Fed’s balance sheet to help illustrate the sources and uses of the monetary base today.⁵ The big changes in the traditional items on the asset side of the balance sheet are in outright holdings of Treasury securities, which fell from \$778.9 billion in 2007 to \$475.2 billion in 2009. This has virtually wiped out the Fed’s holdings of Treasury bills, which fell from \$277 billion to \$18.4 billion. The Fed’s holding of notes and bonds was reduced from \$467.9 billion to \$412.9 billion. But of those the Fed still holds, \$125.1 billion has been lent through the Term Securities Lending Facility (TSLF) to securities dealers. Note that changes in the TSLF do not affect the size of the balance sheet, but they do reduce the liquidity of the Fed’s security portfolio.

Holdings of federal agency debt rose from 0 to \$26.7 billion. Note that in December 2008, the FOMC authorized the Trading Desk of the Federal Reserve Bank of New York to purchase up to \$100 billion of agency debt in the first half of 2009. Repurchase agreements (repos) decreased from \$27.5 billion to \$17.1 billion. Another big change among the traditional assets was the big increase in primary lending: from \$0.3 billion in the week ending January 17, 2007, to \$65.0 billion in the week ending January 28, 2009. A new part of this lending program was the very public campaign to eliminate the “stigma” associated with borrowing at the discount window, an obstacle to overcome if these lending facilities were to be implemented as intended.

³ See Balbach and Burger (1976) for an elementary introduction to the derivation of the monetary base from the central bank’s balance sheet. Their appendix includes an application of their method to the Fed’s balance sheet in 1976. See Anderson and Rasche (1996) for the technical details in the derivation of the St. Louis adjusted monetary base.

⁴ See Thornton (2009).

⁵ This information can be found in the H4.1 release in the table showing factors that supply and absorb reserves. These factors are found primarily on the Fed’s balance sheet, but also include monetary items from Treasury’s balance sheet.

Figure 1**Monetary Base**

From January 17, 2007, to January 28, 2009, the traditional items on the asset side of the balance sheet decreased by \$224.5 billion. The decline in outright holdings of Treasury securities, repos, and float was partially offset by increases in primary lending and federal agency debt. The other longtime items on the balance sheet are either unchanged or relatively small. These include the gold stock, special drawing rights, other assets, lending through the traditional channels, secondary and seasonal lending, and float.

Next, we turn to the new programs established after the crisis began. The Trading Desk has just begun to buy mortgage-backed securities (adding \$6.8 billion as of January 28, 2009) under instructions from the FOMC to purchase as much as \$500 billion in the first half of 2009. Also included on the balance sheet is a total of \$415.9 billion in loans to banks through the Term Auction Facility (TAF).⁶ Figure 2 shows the history of lending

under this program initiated in December 2007—the same time that the Fed began lending securities to primary dealers through the TSLF.

On January 28, 2009, there were \$32.1 billion in loans outstanding in the Primary Dealer Credit Facility (PDCF) and other broker-dealer loans, \$14.6 billion in loans through the Asset-Backed Commercial Paper Money Market Mutual Fund Liquidity Facility (AMLF), and \$38.3 billion in direct loans to the American Insurance Group (AIG). Figure 3 shows the history of lending under these three new programs, as well as the traditional discount window lending of primary credit.

Certain new programs operate as special-purpose vehicles (SPVs) wholly owned by the Federal Reserve Bank of New York. The assets and liabilities of these SPVs are included on the Federal Reserve's balance sheet. The next item in Table 1 (under "New program portfolio") is the sum of the private assets purchased under new programs; the total was \$389.9 billion the week ending January 28, 2009. Among these SPVs is the Commercial Paper Funding Facility (CPFF)

⁶ For information about the acronyms and new programs, go to www.federalreserve.gov/newsevents/recentactions.htm.

Table 1
Factors Affecting Reserve Balances (\$ billions)

	Week Ending	
	January 17, 2007	January 28, 2009
Assets supplying reserves		
Gold stock	11.0	11.0
Special drawing rights	2.2	2.2
Treasury securities*	778.9	475.2
Bills	277	18.4
Notes and bonds (nominal)	467.9	412.9
Notes and bonds (inflation-indexed)	30.2	39.4
Inflation compensation	3.8	4.5
Federal agency debt	0	26.7
Mortgage-backed securities	NA (01/05/09) [†]	6.8
Repurchase agreements	27.5	17.1
Term Auction Facility (TAF) [‡]	NA (12/17/07)	415.9
Loans and discounts including float	0.1	148.0
Primary	0.3	65.0
Secondary	0.0	0.0
Seasonal	0.0	0.0
Float	-0.2	-2.0
Primary Dealer Credit Facility (PDCF) and other broker-dealer loans	NA (03/17/08)	32.1
Asset-Backed Commercial Paper (ABCP)	NA (09/19/08)	14.6
Money Market Mutual Fund (MMMF) Liquidity Facility (AMLF)		
Loans to American Insurance Group (AIG)	NA (09/16/08)	38.3
New program portfolio	—	389.9
Commercial Paper Funding Facility (CPFF)	NA (10/27/08)	316.2
Money Market Investor Funding Facility (MMIFF)	NA (11/24/08)	0.0
Maiden Lane	NA (06/26/08)	27.0
Maiden Lane II	NA (11/10/08)	19.7
Maiden Lane III	NA (11/25/08)	27.0
Central bank liquidity swaps	NA (12/12/07)	465.9
Other assets	39.6	44.4
Memo item		
Treasury coin outstanding (TCO)	38.3	38.8
Total factors supplying reserves = Total assets + TCO	897.5	2,041.9

Table 1, cont'd**Factors Affecting Reserve Balances (\$ billions)**

	Week Ending	
	January 17, 2007	January 28, 2009
Liabilities absorbing reserves		
Federal Reserve notes held by Treasury	0.2	0.3
Reverse repos, international accounts, and dealers (RRP)	30.5	73.1
Treasury deposits	4.7	230.4
General account	4.7	55.5
Supplemental financing account	NA (09/17/08)	174.8
Foreign official	0.1	0.2
Service-related and other demand deposits	7.2	7.9
Other liabilities and capital	36.7	50.4
Total factors other than the monetary base absorbing reserves	79.4	362.3
Monetary base	818.1	1,679.6
Currency in the hands of the public (in circulation)	757.6	830.6
Vault cash in depository institutions	50.3	53.5
Counted as required reserves	32.3	41.2
Counted as surplus cash	18	12.3
Depository institution reserve balances	10.2	795.5
Held as required reserves	6.1	26.3
Held as excess reserves	4.1	769.2
Total liabilities plus Treasury coin outstanding	897.5	2,041.9

NOTE: Components may not sum to totals because of rounding. Data are from the H4.1 release (Federal Reserve Statistical Release: Factors Affecting Reserve Balances); see www.federalreserve.gov/releases/h41/.

*Includes \$125.1 billion lent to primary dealers through the Term Securities Lending Facility.

†NA refers to an item that was created after the financial crisis was under way. The date refers to the day that the Fed announced the program or began to purchase this asset.

‡For information about the new programs created to solve the financial crisis, go to: www.federalreserve.gov/newsevents/recentactions.htm.

created to support activity in the market for the highest-rated (A1/P1) commercial paper. This program, created in the aftermath of the troubles at AIG, Lehman Brothers, and Merrill Lynch, has grown rapidly to \$316.2 billion, with the Fed providing most of the new lending in this market. Also included are three new structured investment vehicles created to buy and hold certain troubled assets of specific insolvent institutions. Through Maiden Lane, the Fed owns \$27.0 billion of the poorer-quality assets from the troubles at Bear Stearns. Through Maiden Lane II and Maiden Lane III, it owns \$19.7 billion and \$27.0 billion, respectively, of troubled assets purchased in sup-

port of the insurance firm AIG. Figure 4 shows the history of assets purchased under these new programs.

On December 12, 2007, the Federal Reserve established temporary swap lines with foreign central banks. Under these swap arrangements, the Federal Reserve provides U.S. dollar deposits at the Federal Reserve in exchange for an amount of foreign currency deposits at the foreign central bank. The amount is determined by the prevailing exchange rate. The currency is swapped back at a future date at the swap exchange rate used in the original transaction. All exchange rate risk is the burden of the borrowing central bank. Figure 5

Figure 2

Term Auction Facility Assets

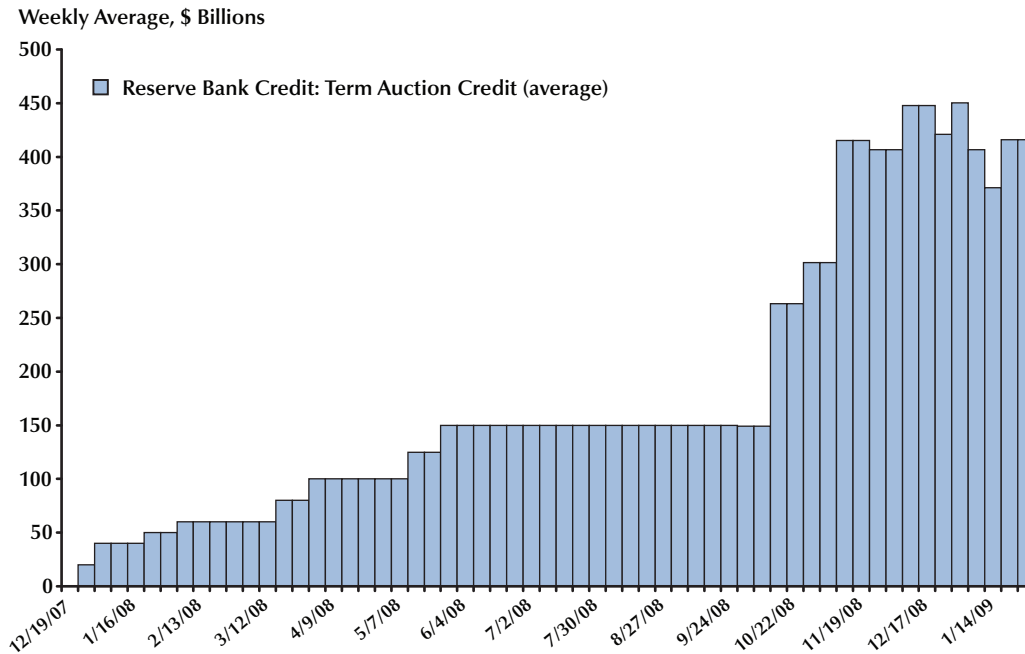


Figure 3

Total Loans and Discounts

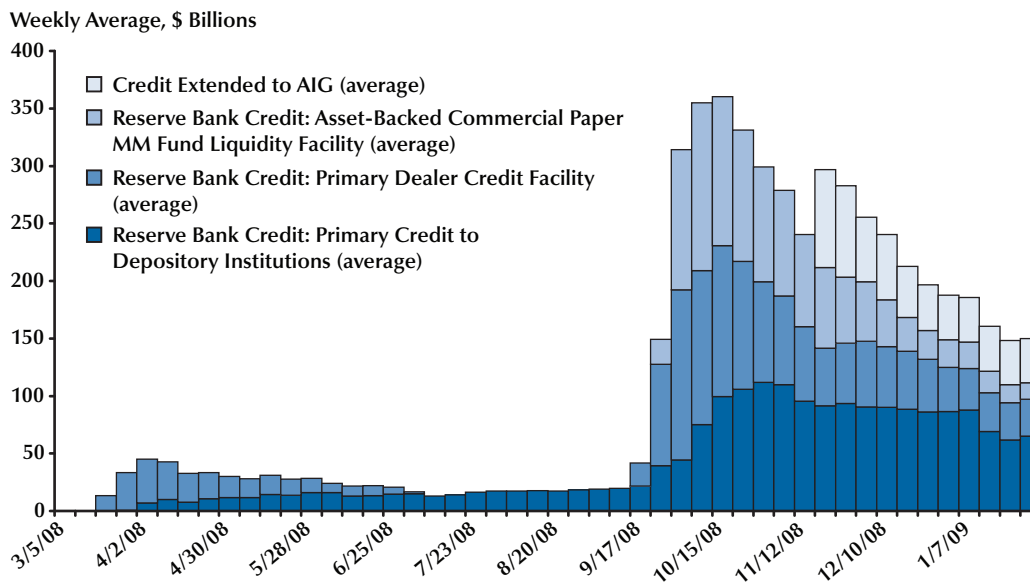


Figure 4

New Program Portfolio

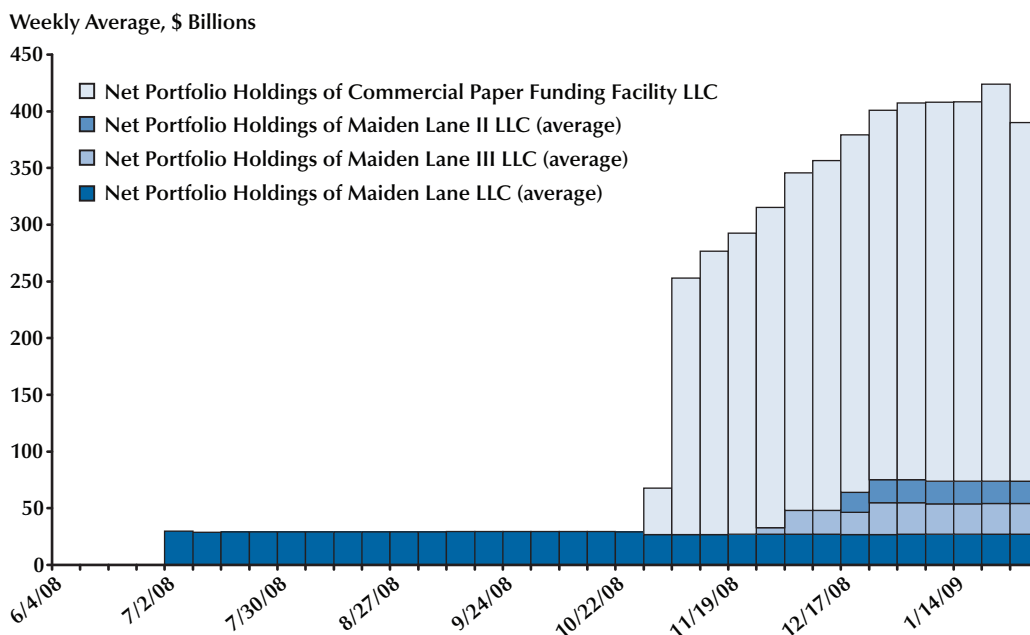


Figure 5

Foreign Exchange Swaps

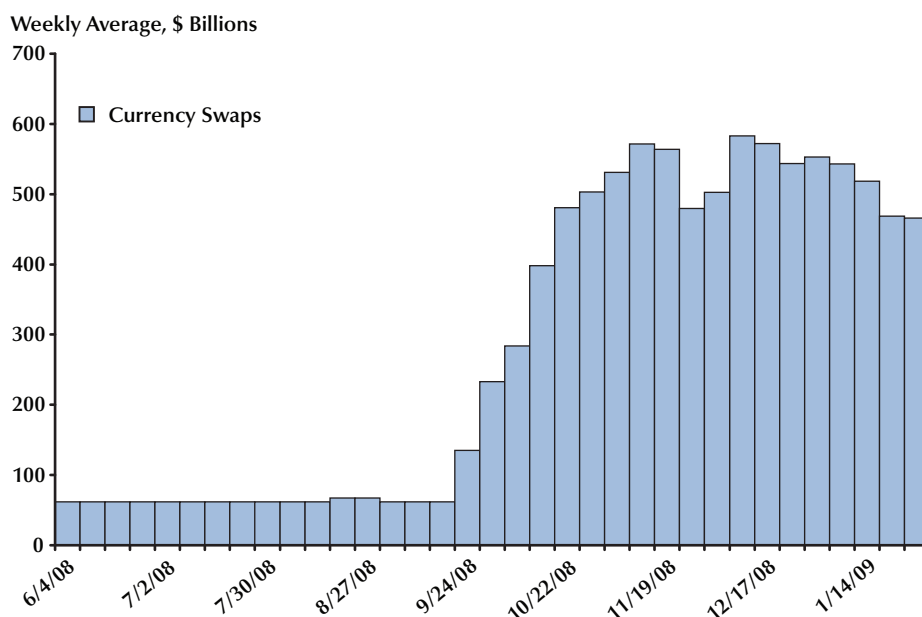
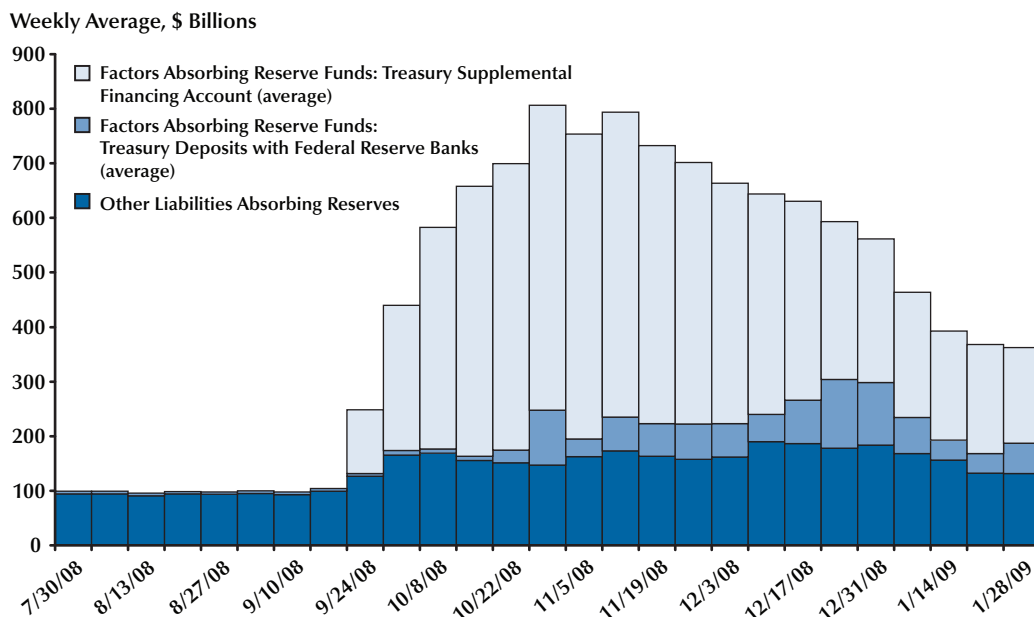


Figure 6**Total Factors Other than Monetary Base Absorbing Reserves**

shows the dramatic rise in foreign exchange assets from currency swap arrangements; these swaps totaled \$465.9 billion as of the week ending January 28, 2009.

Overall, total assets added by the Federal Reserve's nontraditional credit programs were \$1,363.5 billion the week ending January 28, 2009. Total factors supplying monetary base skyrocketed that week to a little more than \$2 trillion, with almost all the increase coming after mid-September 2008.

Next, we turn to the liability side of the Fed's balance sheet. Not all liabilities of the Fed are included in the monetary base. Most items on the liability side of the Fed's balance sheet that are not included in the monetary base are small or relatively unchanged. There has been a relatively large increase—from \$30.5 billion to \$73.1 billion—in reverse repos with dealers, foreign official, and other international accounts. The other—and the most important item absorbing reserves—is the Treasury deposit account. The

Treasury general account rose from \$4.7 billion to \$55.5 billion; for more than 20 years, the Fed has maintained its general account (used for tax collection and government disbursements) near a \$5 billion balance. In September 2008, the Treasury created a new "supplemental financing account," which held \$174.8 billion on January 28, 2009. At inception, this account held \$500 billion obtained by the Treasury as proceeds from selling a special issue of Treasury bills to the public. The mechanism for this sale was quite simple: Each purchaser of a Treasury bill paid with a bank check or debit. When these transactions cleared, the Federal Reserve transferred the amounts from bank deposits (reserve accounts) to the Treasury account (which absorbs reserves, but is not part of the monetary base). In the second half of January, the Treasury allowed the special issue of Treasury bills to mature, to avoid hitting statutory debt limits. As the Treasury repaid the owners of these bills, deposits were transferred from the Treasury's deposits at the Fed to the

deposits of the banks whose customers owned the maturing bills. Doing so increases the monetary base, dollar for dollar.

Overall, other factors absorbing reserves rose from \$79.4 billion in January 2007 to \$362.3 billion in January 2009. Figure 6 shows the total factors other than the monetary base that absorb reserves, with detail shown for Treasury deposits (both the general account and the supplemental financing account).

The bottom section in Table 1 lists the components of the monetary base. The “monetary base” is defined as currency in circulation and bank demand deposits at the Fed. Currency in circulation includes vault cash (coin and Federal Reserve notes held by the depository institutions) and cash held by the general public. Bank reserve deposits are sometimes referred to as federal funds. Thus, the federal funds rate is the interest rate that banks pay to borrow Federal Reserve deposits from other banks. The monetary base includes coin and Federal Reserve notes held overseas because we have no measure of the amount held overseas, only the total amount outstanding.⁷ And, regardless of where they are held, Federal Reserve notes are a liability of the Fed.

Between January 17, 2007, and January 28, 2009, currency in the hands of the public grew from \$757.6 billion to \$830.6 billion. Vault cash grew from \$50.3 billion to \$53.5 billion. In January 2007, \$32.3 billion of the vault cash was used to meet reserve requirements; in January 2009, \$41.2 billion of vault cash was used. In the early period, \$18.0 billion of vault cash was counted as surplus vault cash; in January 2009, \$12.3 billion was counted as surplus vault cash.

The interesting change in the monetary base was in reserve balances held in demand deposits at the Federal Reserve. These deposits grew from \$10.2 billion the week ending January 17, 2007, to \$795.5 billion the week ending January 28, 2009. Of this large amount, \$769.2 billion was held as excess reserves at the Fed.

WILL RAPID GROWTH IN THE MONETARY BASE CAUSE RAPID INFLATION?

The enormous accumulation of excess reserves began at the time of the Lehman bankruptcy and rescue of AIG in mid-September. Whether this large increase in the monetary base is a harbinger of rapid inflation in the future depends on how the Federal Reserve and the U.S. government act when financial markets return to more-normal behavior and the recession ends.

The difficulty of maintaining price stability will depend on the size the balance sheet reaches before the crisis ends, the quality of the assets in the portfolio, and the policy followed to manage the interest rate paid on reserves. Any attempt to predict whether inflation will occur must rely on predictions about the Fed’s response to events and its exit from these new programs (that is, reducing the size of the balance sheet) as the economy recovers from recession and financial crisis.

Analysis of future monetary policy must consider the October 1, 2008, Congressional authorization for the Fed to pay interest to banks on both required reserve and excess reserve balances. By increasing this rate relative to the federal funds rate target, the Fed provides an incentive for banks to hold more deposits at the Fed. By reducing this rate the Fed encourages banks to expand their lending—and the money supply. When the FOMC set the federal funds rate target to the range 0 to ¼ percent on December 16, 2008, it also set the interest paid on both required and excess reserves equal to ¼ percent.

A logical question might be why depository institutions would choose to hold \$800 billion in excess reserves that are earning so little. Two answers are important, one at the level of the individual bank and one at an aggregate level. First, for the individual bank, the risk-free rate of ¼ percent must be the bank’s perception of its best investment opportunity. Note that on January 28, 2009, the interest rate on the 3-month Treasury bill was less than ¼ percent. The other is that, perhaps because of market conditions—the dramatic decline in the price of bank stocks and the fall in the market value of assets—the bank finds

⁷ See Porter and Judson (1996) for estimates of the amount of currency held abroad.

Table 2**New Program Use and Authorization (\$ billions)**

Assets supplying reserves	Announced authorization	Week ending January 28, 2009
Mortgage-backed securities	500	6.8
Term Auction Facility (TAF)	600	415.8
Primary Dealer Credit Facility (PDCF)	No announced limit	32.1
Asset-Backed Commercial Paper (ABCP) Money Market Mutual Fund (MMMF) Liquidity Facility (AMLF)	No announced limit	14.6
Loans to American Insurance Group (AIG)	60	38.3
Commercial Paper Funding Facility (CPFF)	No announced limit	316.2
Money Market Investor Funding Facility (MMIFF)	540	0
Maiden Lane	29	27.0
Maiden Lane II	22.5	19.7
Maiden Lane III	30	27.0
Swaps	No announced limit	465.9

itself undercapitalized. In such conditions, the bank is likely to hold relatively more safe assets while it builds capital by cutting costs, raising fee income, and hoping for a recovery in both the economy and its stock price.

Second, the banking system as a whole cannot create or destroy bank deposits at the Fed. Only the Fed (and technically, the Treasury) can create or destroy bank reserves. If one bank makes a loan and the funds are deposited in another bank, then the ownership of the deposits at the Fed would change, but the total bank deposits *at* the Fed would remain the same. In theory, the banking system *reduces* excess reserves—but only by expanding loans and the money supply in a way that *increases* required reserves by an equivalent amount. The key is that the Fed will have to drain reserves when the economy begins to recover if it is to prevent a rapid acceleration of inflation. That necessity drives the current discussion of exit strategies.⁸

The ease with which the Fed can reduce the size of its balance sheet in the future depends on many factors, including the term of its loan portfolio, the quality of assets that it holds outright, and the market's appetite for repurchasing these

financial instruments. The authorization for the ultimate size of new programs varies and has grown since the beginning of the crisis. Table 2 lists each new program and the upper limit authorized as of January 28, 2009. Of course, in some cases the limits will be determined by the available assets and/or by the demand for the program. (Note that there are zero assets in the MMIFF, which has an authorization of \$540 billion.)

When the time comes to shrink the monetary base, the Fed could allow the lending programs to expire as loans mature and sell the assets that it holds outright. If the crisis is over, the assets should be priced in the market and the Fed should expect to recover most of its investment in such assets.

Inflation does not appear to be a risk in the current environment: The economy is in recession. Inflation is falling and is not expected to return before the recession ends. If inflation resumes but the economy does not recover, policy-makers will face a difficult choice. Monitoring the size and composition of the monetary base as the economy recovers will help us understand what actions are needed (and should be taken) by the Fed and the Congress to prevent a return to a high-inflation economy.

⁸ See Bernanke (2009).

REFERENCES

- Anderson, Richard G. and Rasche, Robert H. "Retail Sweep Programs and Bank Reserves, 1994-1999." Federal Reserve Bank of St. Louis *Review*, January/February 2001, 83(1), pp. 51-72; <http://research.stlouisfed.org/publications/review/01/0101ra.pdf>.
- Anderson, Richard G. and Rasche, Robert H. "A Revised Measure of the St. Louis Adjusted Monetary Base." Federal Reserve Bank of St. Louis *Review*, March/April 1996, 78(2), pp. 3-14; <http://research.stlouisfed.org/publications/review/96/03/9603ra.pdf>.
- Balbach, Anatol and Burger, Albert E. "Derivation of the Monetary Base." Federal Reserve Bank of St. Louis *Review*, November 1976, pp. 2-8; http://research.stlouisfed.org/publications/review/76/11/Derivation_Nov1976.pdf.
- Bernanke, Ben S. "The Crisis and the Policy Response." The Stamp Lecture, London School of Economics, London, England, January 13, 2009; www.federalreserve.gov/newsevents/speech/bernanke20090113a.htm.
- Porter, Richard D. and Judson, Ruth A. "The Location of U.S. Currency: How Much Is Abroad?" *Federal Reserve Bulletin*, October 1996, 82, pp. 883-903; www.federalreserve.gov/paymentsystems/coin/1096lead.pdf.
- Stevens, Edward J. "Required Clearing Balances." Federal Reserve Bank of Cleveland *Economic Review*, Quarter IV 1993, 29(4), pp. 2-14; <http://clevelandfed.org/Research/Review/1993/93-q4-stevens.pdf>.
- Thornton, Daniel L. "The Fed, Liquidity, and Credit Allocation." Federal Reserve Bank of St. Louis *Review*, January/February 2009, 91(1), pp. 13-21; <http://research.stlouisfed.org/publications/review/09/01/Thornton.pdf>.



Foreign Direct Investment, Productivity, and Country Growth: An Overview

Silvio Contessi and Ariel Weinberger

The authors review the empirical literature that studies the relationship between foreign direct investment, productivity, and growth using aggregate data and focus on two questions: Is there evidence of a positive relationship between foreign direct investment and national growth? And does the output of the “multinational sectors” exhibit higher labor productivity? The authors also briefly discuss how the microeconomic evidence and a number of aggregation and composition problems might help explain the ambiguous results in this literature. (JEL E32, F21, F32, F36)

Federal Reserve Bank of St. Louis *Review*, March/April 2009, 91(2), pp. 61-78.

“Today’s policy literature is filled with extravagant claims about positive spillovers from FDI but the evidence is sobering.”¹

The notable growth of foreign direct investment (FDI) in the past 30 years continues to trigger conflicting reactions, in both industrial and emerging countries (Coughlin, 1992). In short, FDI is an investor’s acquisition of “long-term influence” in the management of a firm in another country. (See the next section for a more complete definition.) In the developed world, countries that export capital and countries that import capital both raise concerns about FDI: The former are concerned that capital leaving their countries might be detrimental to domestic investment; the latter’s politicians and workers fear foreign ownership of domestic firms. Emerging, transition, and developing countries (and at times local governments) usually welcome FDI, assuming that investment through this multinational activity will bring additional capital, managerial expertise, and technology.

In economics, multinational activity (essentially foreign firms with U.S. production units and U.S. firms with foreign production units, described in more detail later) is also viewed as a positive contribution to the technological progress of the host economies. An established literature that dates back to Findlay (1978) develops models in which multinational firms own and transfer technology—which may not be available in the host country—that allows them to be more productive and profitable than firms that are not multinational in nature. Because such a transfer is assumed to contribute to the technical progress of the host economies, it is also assumed to contribute ultimately to their growth.

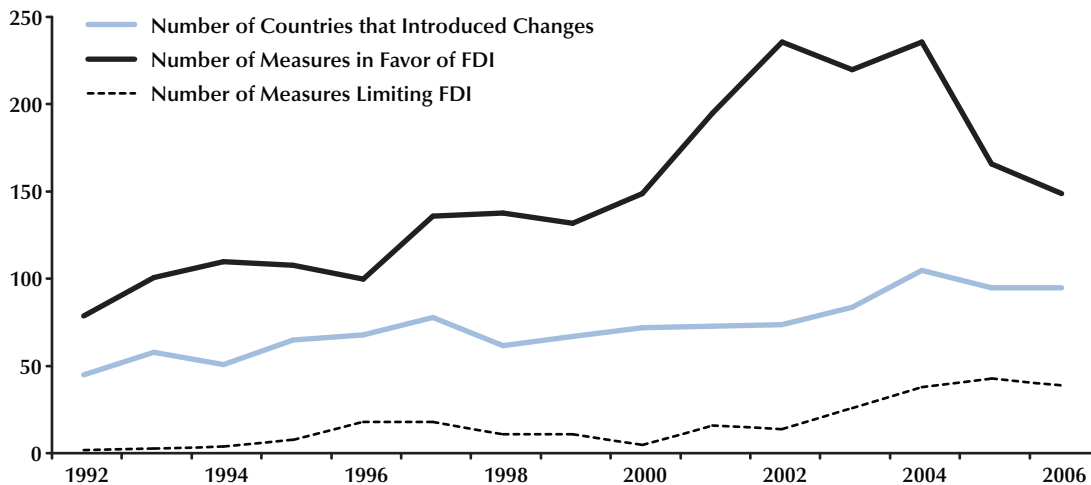
Rivera-Batiz and Rivera-Batiz (1991) develop a formal model that allows for increasing returns due to specialization as a result of FDI. Borensztein, De Gregorio, and Lee (1998) stress the interaction between FDI and investment in human capital. Helpman, Melitz, and Yeaple (2004) and Yeaple (2008) show that only the most productive firms in a country become multinationals, whereas progressively less productive firms enter progressively more attractive countries.

¹ From Rodrik (1999).

Silvio Contessi is an economist and Ariel Weinberger is a research analyst at the Federal Reserve Bank of St. Louis.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Figure 1
National Regulatory Changes, 1992-2006



SOURCE: UNCTAD Database on National Laws and Regulations.

Some other studies highlight reasons why FDI may *not* accelerate growth: Aitken and Harrison (1999) argue that increased local competition caused by multinationals may crowd out domestic firms; Boyd and Smith (1992) show that FDI distorts resource allocation and slows growth when other distortions are present in the financial sector, prices, or trade. This would imply that FDI does not necessarily contribute to growth, and countries could be harming their economies with provisions that favor FDI.

As mentioned, overall FDI has increased in many countries. In Figure 1, we plot an index of the time series of the number of national regulatory changes between 1992 and 2006, which we obtained from various annual surveys on national laws and regulations.² If we consider these series as proxies for the amount of intervention aimed at expanding and restricting FDI activities, the graph illustrates clearly the existence of a grow-

ing trend over the past 15 years of introduction of policies aimed at promoting FDI. Since 1992 at least 80 percent of regulatory changes have been favorable to FDI, particularly those in the 1990s.³ Furthermore, the absolute number of favorable changes has steadily increased since 1992, with some countries introducing more provisions and others that previously had no favorable provisions now passing legislation to encourage foreign investment.

This paper attempts to lay out the empirical evidence on each side. We review a number of macroeconomic studies that mostly fail to show convincing evidence that FDI contributes to growth. In particular, we organize our discussion around two questions: Is there evidence of a positive relationship between FDI and growth in macroeconomic data? And does the output of the

² Specifically, we used various issues of *The World Investment Report*, an annual publication from the United Nations Commission on Trade and Development (UNCTAD) that focuses on FDI trends based partially on public data and partially on proprietary datasets and surveys and covers a different special topic every year: www.unctad.org.

³ Specific examples of positive changes include the creation of new special economic zones in India, many of which offer tax holidays or other incentives, and variations in corporate taxes such as the change in Egypt's corporate tax from a base rate of 40 percent (32 percent for industrial and export activities) to a standard rate of 20 percent. Examples of negative changes include the restriction to a ceiling of 49 percent participation in Algerian state-owned oil and gas enterprises or the restriction to foreign participation in the Russian strategic sector, such as defense-related activities, aviation, and natural resources.

multinational sector exhibit higher labor productivity? After explaining how to define and identify FDI in the next section, we discuss the evidence based on aggregate data and then the evidence regarding labor productivity. The final section offers a possible interpretation of this evidence.

DEFINITIONS, MEASUREMENT, AND RELEVANCE

The Definition of FDI

The most widely accepted definition of FDI is known as “the IMF/OECD benchmark definition” because it was provided by a joint workforce of these two international organizations with the objective of providing standards to national statistical offices for compiling FDI statistics.⁴ The gist of the definition is that FDI is an international venture in which an investor residing in the home economy acquires a long-term “influence” in the management of an affiliate firm in the host economy. According to the definition, the existence of such long-term influence should be assumed when voting shares or rights controlled by the multinational firm amount to at least 10 percent of total voting shares of rights of the foreign firm.⁵

Aggregate FDI flows are the sum of equity capital, reinvested earnings, and other direct investment capital; hence, aggregate FDI flows and stocks include all financial transfers aimed at financing of new investments, plus retained earnings of affiliates, internal loans, and financing of cross-border mergers and acquisitions. FDI flows can be observed from the perspective of the host economy, which records them as *inward* FDI along with other liabilities in the balance of payments, or from the perspective of the home economy,

which records them as *outward* FDI, a category of assets.⁶

The sum of all direct capital owned by non-residents in a given country j in a certain time period t constitutes the existing stock of FDI at that time. We will refer to the stock of foreign direct capital as K_{jt}^I . Hence, in each period t , FDI_{jt} is the per-period increase in the stock of foreign direct capital, $FDI_{jt} = K_{jt}^I - K_{jt-1}^I$, in country j .

These measures can be sufficiently accurate in the short run. However, the value of the capital stock changes over longer periods, causing problems with the adjustment of its valuation. Over 20 years, the value of the stock of FDI at current prices may become three times as large as its historical value. For example, Ihrig and Marquez (2006) show that if one simply adds up net direct investment flows from 1982 to 2004, then the United States has net claims on foreigners of approximately \$250 billion; whereas, if one adjusts the values of assets and liabilities for inflation and changes in exchange rates (current cost), then net claims on foreigners in 2004 soar to almost \$600 billion. The difference between these two measures of the net direct investment position results from valuation adjustments over this time period. Another way to adjust the value is to calculate the net position at market value, a procedure that brings the net direct investment position to \$500 billion in 2004.

Table 1 describes the composition of U.S. outward FDI, using stocks and flows in 2006. The data we use for this table do not capture the ultimate destination of flows that most statistical agencies try to report, but only the initial destination. Initial and final destinations might differ because multinational firms try to minimize the tax burden from multinational activity by exploiting the differences of fiscal regimes across countries. For example, the main FDI partners of the United States in 2006 include the Netherlands,

⁴ The fourth revision of the OECD Benchmark Definition of Foreign Direct Investment (OECD, 2008) can be found here: www.oecd.org/dataoecd/26/50/40193734.pdf. This collection of operational guidelines sets the world standard for FDI statistics, indicating to national statistical agencies the best practice to measure FDI activity. The benchmark definition was first published in 1983 based on a report by the OECD Group of Financial Statisticians, with revisions in 1990, 1992, and 2008.

⁵ Voting shares give the stockholder the right to vote on matters of corporate policymaking and on the appointment of the board of directors. If the equity share of control is 50 percent or more, the controlled firm is often defined as a *subsidiary*.

⁶ Three main types of private capital flows appear in the balance of payment accounts of a country: international debt, international portfolio flows, and FDI. FDI and portfolio investment are both a form of international equity investment, i.e., they represent shares of ownership of foreign firms, unlike private and public debt that is pledged to be returned at the end of the life of the international loan that generates it. The difference between FDI and portfolio equity, instead, is that the latter lacks the objective of influencing the management decisions of the controlled firm, and as such is generally more liquid and more short term.

Table 1**U.S. Outward FDI by Destination Country in 2006, Stocks and Flows**

Country	Stock of U.S. FDI holdings in other countries (percent)	U.S. outward FDI by country (percent)
U.K.	15.29	6.88
Netherlands	11.43	18.42
Canada	9.37	3.67
Bermuda	5.48	8.53
Switzerland	4.69	5.07
Japan	3.76	4.56
Luxembourg	3.85	7.91
Mexico	3.39	3.96
Germany	3.92	2.42
Other	38.82	38.59
Total	100	100

Bermuda, and Luxembourg, which offer particularly favorable fiscal regimes, although the final investment is likely to refer to an affiliate located in a third country.

Besides the problems with data reporting, the definitions used by statistical agencies may differ from the legal treatment of multinational firms in international treaties, such as the ones managed by the World Trade Organization (WTO) or NAFTA, that aim at reducing legal barriers to FDI. Although many efforts in this direction have been undertaken over the years, no encompassing multilateral treaty with the aim of setting standards for the liberalization of multinational activity exists, with two partial exceptions. The WTO treaties include (i) some specific measures aimed at the liberalization of FDI in the service sector in the General Agreement on Trade in Services (GATS) and (ii) some provisions meant to prevent members' actions aimed at restricting trade-related FDI in the Trade-Related Investment Measures (TRIMs) Agreement. No agreement exists to regulate FDI in the manufacturing sector.

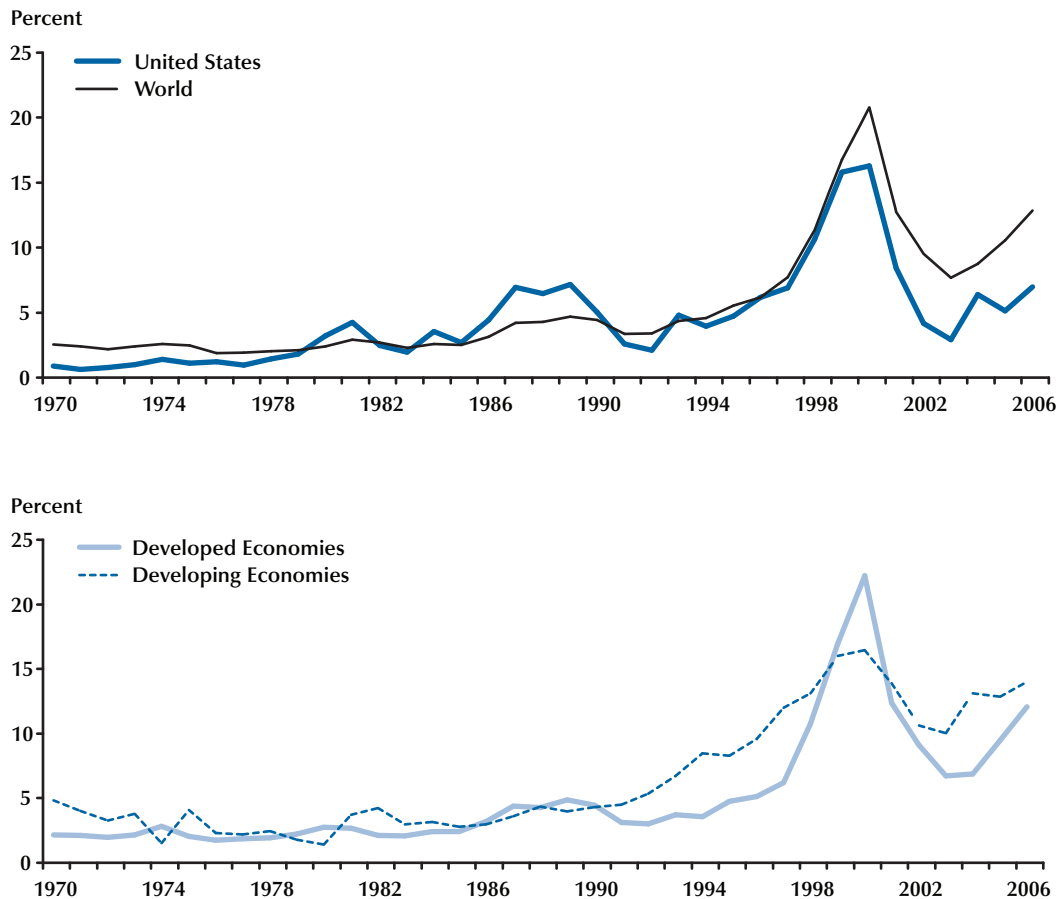
Measurement and Relevance

Typically available macroeconomic time series for FDI include the nominal value of the flows in or out of a country (inward and outward) and

stock values. Both measures have problems that sometimes undermine the cross-country comparability of the series, especially because statistical agencies of different countries may use different definitions of FDI. For example, in Estonia, the 10 percent benchmark for equity ownership suggested by the International Monetary Fund/Organisation for Economic Co-operation and Development (IMF/OECD) definition of FDI has been applied only since the beginning of 2000, whereas the previous threshold was 20 percent. Prior to 1997, Poland used a criterion of "effective voice in management" that might not have amounted to 10 percent or more ownership—a criterion they later adopted in compliance with international standards (IMF, 2003). In both cases, the series are clearly not comparable before and after the change in methodology, and large changes in their levels may either be a purely statistical artifact or have an economic basis to them.

A second problem with datasets available from international organizations, such as the ones available from the IMF, the World Bank, and the UNCTAD, is that they often have missing data points, particularly for developing countries.

A third issue with the use of aggregate data in studying FDI is that the records may not capture a part of the investment in the foreign project. The fact that multinational activities can be financed

Figure 2**Inward FDI Flows as a Percentage of Gross Fixed Capital Formation**

SOURCE: UNCTAD.

using local or foreign financial markets implies that measures of FDI flows and stock that capture only the foreign financing of the projects provide a potentially distorted measure of the extent of multinational activity across countries, as we discuss in the next paragraph. When foreign owners raise capital in the host country by issuing bonds or shares, no international capital flow is recorded in the balance of payment data of the source and the host country, and hence it does not show up in FDI statistics (Marin and Schnitzer, 2006). Therefore, using the flow of FDI might lead to incorrect inference, as part of the capital used to finance the multinational activity might be raised locally

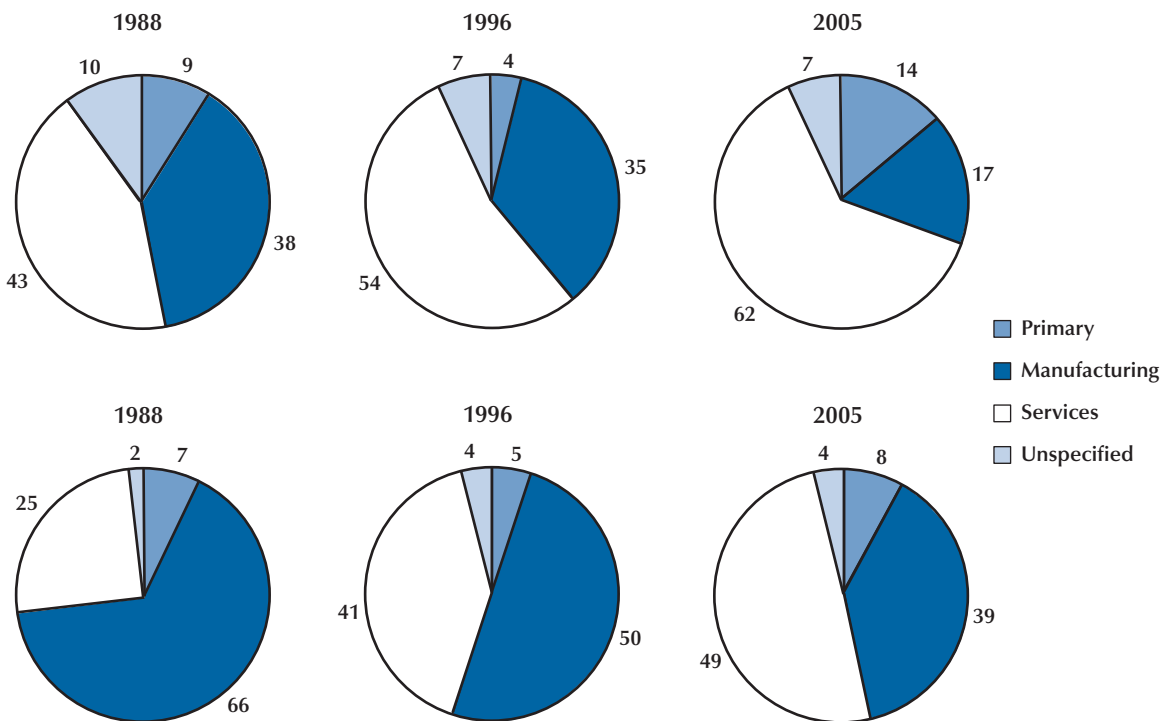
and hence not be recorded as an international capital flow in the balance of payments.

A common way to gauge the relevance of FDI is by comparing FDI with domestic investment. Figure 2 plots inward FDI flows as a percentage of gross fixed capital formation for developed and developing countries, as well as the world and the United States. The time series show clearly that these ratios were basically flat during the 1970s and part of the 1980s and subsequently started growing during the 1980s, eventually crossing the threshold of 5 percent.

Looking at groups of countries hides large variability in individual countries and regional

Figure 3

FDI Inflows in Developed (top) and Developing (bottom) Countries (percent)



SOURCE: UNCTAD.

experiences, even during these early years. Beginning in the mid-1980s, the ratio started to grow quite steeply and later accelerated in the mid-1990s to reach a peak at the end of the 1990s.⁷

After the large correction of equity prices of the late 1990s and the early 2000s recession, the flows of FDI as a share of gross fixed capital formation returned to the levels of the late 1980s. More recently, however, the slowdown reached a turning point and the ratio is back to a level between 5 and 10 percent for the world as a whole. It remains to be seen whether the 2008 credit contraction in OECD countries will mark another turning point.⁸

⁷ Part of the steep increase in the 1990s is due to the wave of cross-border mergers and acquisitions across OECD countries at the time.

⁸ Contessi, De Pace, and Francis (2008) show that industrial countries' outward and inward FDI is significantly procyclical, whereas emerging countries' inward FDI is significantly countercyclical.

The figures prompt two additional observations. First, the cyclical nature of FDI flows emerges quite clearly for the United States and developed countries, with at least three (long) boom-bust cycles with a first peak around 1980, a second peak around 1988, and a third peak in 2000-01. The second observation is that developing countries stand out because they experienced steady trend growth until 2000, rather than boom-bust cycles; moreover, in these countries the early 2000s correction is much less pronounced than in developed countries and certainly much less than in the United States. Inward flows to developing countries now make up a larger percentage of total inward investment when compared with developed countries.

Industrial and developing countries differ in other dimensions regarding FDI. In Figure 3 we compare the industry composition of FDI in indus-

trial and emerging countries for three years, 1988, 1996, and 2005. Looking at both developed and developing countries one can see how the share of services in total flows has increased quite substantially—by almost 50 percent in fact—whereas the share of manufacturing has dropped to 50 percent of what it was in 1988 in industrial countries. This shift is likely due to the rise of the service sector; the elimination of many restrictions to foreign entry and ownership, particularly in the banking and telecommunications sector; and the commitments contained in the GATS agreement and other regional agreements.⁹ The pie charts in Figure 3 also reveal that the share of manufacturing FDI is consistently larger for developing countries than for developed countries, although less so as time goes by. Finally, this shift is also likely to be a response to comparative advantage and the emergence of vertical fragmentation of manufacturing production over the past decades.

Do the Aggregate Data Show Evidence of a Positive Relationship Between FDI and Growth?

Output is assumed to be produced according to the following production function:

$$(1) \quad Y_j = Z_j f(K_j^D, K_j^I, L_j^D, L_j^I, M_j, \dots),$$

where final output of country j in time t (subscript omitted) is a function of a set of inputs, such as potentially different domestic and foreign capital (K_j^D, K_j^I), domestic and possibly foreign labor (L_j^D, L_j^I), intermediate inputs (M_j), and other factors are combined using a technology $f(\cdot)$ common across firms and scaled by a total factor productivity parameter Z_j to produce Y_j units of real output.

Researchers who use aggregate data normally postulate that the aggregate production function is a Cobb-Douglas production function. Equation (1) then has the functional form

$$(2) \quad Y_j = Z_j (K_j^I)^{\beta_0} (L_j^I)^{\varphi} \underbrace{(X_j)^{\gamma}}_{(K_j^D, L_j^D, M_j, \dots)}$$

and is transformed in logarithmic form and studied in first differences. The final regression estimate is

$$(3) \quad \dot{Y}_j = \alpha_j + \beta_0 FDI_j + \beta_1 (FDI_j \times x_j) + \gamma X_j + \varepsilon_j,$$

where α_j is a constant, the matrix X_j contains a number of non-FDI factors assumed to affect growth and varies greatly across studies, and the interaction term ($FDI_j \times x_j$) takes care of possible interactions between FDI and other regressors, that is, other non-FDI factors that are assumed to affect growth and contained in X_j .¹⁰ Researchers then test whether the estimated coefficient $\hat{\beta}_0$ is positive and significant in the regression.

The most common and somewhat natural way to measure growth in these studies is to use real per capita GDP growth. Measures of FDI may vary, but gross FDI inflows as a share of GDP, FDI inflows per capita, or multinational sales are common choices. FDI stocks are more troublesome not only because time series of flows date back to the 1970s at the earliest, making stocks series difficult to reconstruct for a sufficient number of countries, but especially because the values of firms and of the FDI stock change over time, introducing a substantive discrepancy between the originally recorded book value and the market value.

The matrix X_j is a set of control variables and the specification leaves room for interaction terms between FDI_j and one or multiple variables in X_j . The latter may include various measures of initial per capita income, average years of schooling of the working population, government size, inflation, openness to trade, black market premium,¹¹ private credit, and so on. Loosely speaking, regression analysis including only the FDI_j term usually reveals nonsignificant effects of FDI_j alone through β_0 , while regression analysis including an inter-

⁹ For example, many provisions aimed at completing the European Single Market or NAFTA's Chapter 11 contain specific FDI provisions that protect firms and individuals investing in Mexico, Canada, and the United States.

¹⁰ Estimation techniques vary but usually range from ordinary least squares to dynamic panels.

¹¹ This is the percentage differential between the black market and the official exchange rate.

action term with one of the variables in X_j reveals a positive relationship.

Various panel and cross-sectional studies show the importance of different complementary variables. Balasubramanyam, Salisu, and Sapford (1996) use exports as a measure of openness to trade; Borensztein, De Gregorio, and Lee (1998) a proxy for human capital; and Alfaro et al. (2004) various proxies for financial development.¹² They all find that FDI_j has a positive impact only if an additional variable is interacted with it. Hence, the more the country is financially developed, open to trade, or endowed with human capital, the more FDI increases growth. Unfortunately, the most sophisticated in this group of studies, Carkovic and Levine (2005), demonstrates the lack of a robust positive correlation between FDI and growth once the temporal dimension of international data is exploited using panel data.¹³

There are at least two important caveats that might affect the reading of such results. The first more general caveat is common to cross-country growth regressions—that is, these studies are plagued by a multiplicity of issues of parameter heterogeneity, outliers, omitted variables, model uncertainty, measurement error, and endogeneity, as eloquently discussed in Rodrik (2005). Inference based on results that do not discuss the potential biases in these studies should be taken with a grain of salt. The second caveat was raised in Blonigen and Wang (2005), who argue that pooling rich and poor countries without distinguishing between levels of development (i.e., assuming the same β_0 for all countries and hence underestimating cross-country heterogeneity) leads to incorrect inference, an argument that squares nicely with the evidence discussed in the studies just cited.¹⁴ In particular, they show that

it is not the use of panel technique, but the pooling of countries with different levels of wealth, that makes the evidence of a positive relationship between FDI stocks and growth disappear. This is consistent with a previous study by Blomström, Lipsey, and Zejan (1994), which found no significant effect of foreign capital on economic growth of lower-income developing countries (as opposed to relatively richer developing countries), suggesting that the relationship may be positive for “sufficiently wealthy” countries.

In Figures 4 and 5, we plot the ratio of the inward FDI stock to GDP in 1990 and 2000 on the horizontal axis; on the vertical axis, we plot real GDP growth for the 1990-2000 period and for the 2000-05 period for the largest groups of countries for which these data are available in each of the periods.¹⁵ The scatter diagrams in the bottom part of each graph do not distinguish among different levels of development, whereas the top graphs do, in a way that isolates industrial countries (squares) and emerging and transition countries (triangles) from other less-developed countries (dots). Although we make no attempt at formal inference, we think that the picture alone suggests that the relationship between FDI and growth that we could infer from these relationships is quite different depending on the composition of the group of countries and the time period under scrutiny.

What do we learn from these studies? The positive relationship between FDI—whether flows, flows to GDP, or stocks—and growth emerges in macroeconomic data, when “some other factor” is present. However, the presence of FDI is likely to be affected by the presence of these factors, making the issue of endogeneity hard to resolve. If FDI has a positive impact on economic growth, it increases market size, and a larger market size attracts more FDI. Hence, FDI and growth are interdependent in a nontrivial way that needs to be somehow addressed in the econometric analysis, which has made two possible solutions available for similar problems: The first, the use of random-

¹² Borensztein, De Gregorio, and Lee (1998) average data over 10-year periods, because annual flows might not have a discernible effect on growth over longer periods.

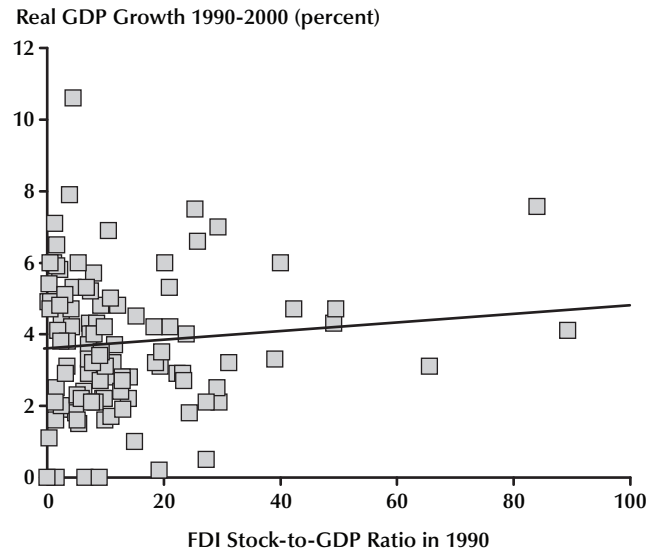
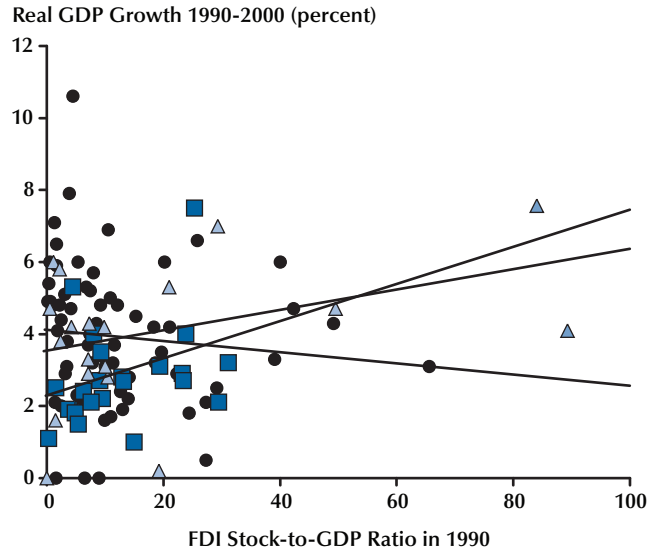
¹³ Carkovic and Levine (2005) use 72 countries over the period 1960-95. They average data over non-overlapping, 5-year periods, so that there are seven observations per country (1961-65, 1966-70, and so on).

¹⁴ Results in Contessi, De Pace, and Francis (2008), for example, echo this argument: They show that FDI inflows are clearly procyclical in developed countries but countercyclical in emerging countries and discuss how pooling all countries together would not reveal any pattern of cyclicity.

¹⁵ We obtained GDP growth data from the 2007 World Development Indicators of the World Bank and the stock of inward FDI to GDP from the “External Wealth of Nations—Mark II” dataset that has been publicly released by Lane and Milesi Feretti (2007).

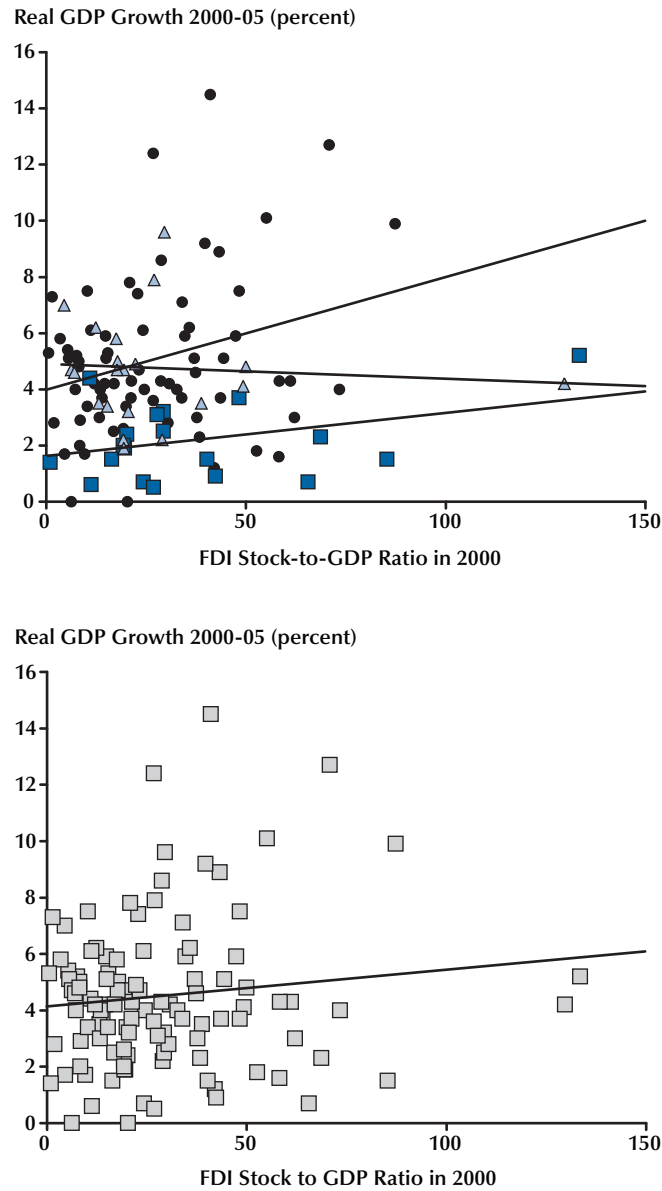
Figure 4

FDI Stock-to-GDP Ratio and Real GDP Growth (1990-2000)



NOTE: In the top panel, triangles represent emerging economies, squares represent industrial countries, and dots represent all other countries. In the bottom panel, all countries are pooled together. The straight line is a linear trendline for each group.
 SOURCE: 2007 World Development Indicators of the World Bank and “External Wealth of Nations–Mark II” dataset (Lane and Milesi Ferreti, 2007).

Figure 5
FDI Stock-to-GDP Ratio and Real GDP Growth (2000-05)



NOTE: In the top panel, triangles represent emerging economies, squares represent industrial countries, and dots represent all other countries. In the bottom panel, all countries are pooled together. The straight line is a linear trendline for each group.

SOURCE: 2007 World Development Indicators of the World Bank and “External Wealth of Nations–Mark II” dataset (Lane and Milesi Ferreti, 2007).

ized trials, is clearly impossible in growth regressions. However, Carkovic and Levine (2005) argue that the second, the use of more sophisticated panel techniques, can substantially alleviate the problem of endogeneity.¹⁶

In the discussion of Carkovic and Levine (2005) and Blonigen and Wang (2005), Melitz (2005) puts forward the possibility that certain types of FDI ventures known in the literature as *export platform* FDI may have a potentially larger effect on growth. Hence, export-oriented countries may benefit more from FDI than import-substituting countries, an idea originally discussed in Bhagwati (1988) and tested explicitly in Balasubramanyam, Salisu, and Sapford (1996). For example, the Volkswagen automobile plant in Puebla, Mexico, is the only assembly unit of this multinational in North America but meets the demands of the three NAFTA countries. Although the motivations of this venture are both cost saving through lower input costs and access to the sizeable North American consumer markets, the factory benefits the Mexican economy more than it would if its production were sold only in Mexico.

A part of the literature that uses aggregate data has focused on regional, rather than county, growth. The methods used are very similar to studies we just described: Authors attempt to disentangle the various variables affecting productivity growth to single out the individual effect that can be attributed to foreign investment. Although cross-country models attempt to measure the “average” effects that FDI has on country-wide economic outcomes as a whole, they cannot measure how certain subnational units might be affected. Regional growth studies deal with very specialized locales and can focus on the effects of foreign investment in the particular community where multinational firms establish their operations. In particular, although evidence suggests foreign investment increases wages and growth, it is unclear whether certain areas benefit suffi-

ciently to outweigh the costs incurred, particularly in terms of incentives to locate in the area. Figlio and Blonigen (2000) and Ford, Rork, and Elmslie (2008) find that foreign investment has more positive effects than domestic investment, but with substantial costs to local communities in terms of reduction of budget expenditures.

Foreign investment’s worst effect appears to be the crowding out of domestic investment. Figlio and Blonigen (2000) use a detailed county-level panel dataset from South Carolina to compare foreign manufacturing firms with domestic manufacturing firms using share of employment. They estimate the following model:

$$(4) \quad w_{jkt} = \alpha_0 L_{jkt} + \alpha_1 l_{jkt}^f + \delta \gamma_{kt} + \delta_j,$$

where j and t are, respectively, counties and years for each industry k . In the model, w_{jkt} is the average annual wage, L_{jkt} is total manufacturing employment, l_{jkt}^f is the level of employees in foreign-owned establishments, and γ_{kt} and δ_j are unobserved county-specific differences in wages and time-varying effects, respectively.¹⁷ The authors also estimate a similar model that replaces wages with budget expenditures. They find that a marginal new foreign manufacturing job has seven times the effect on wages as does a new domestic job, but also that the new foreign job is associated with twelve times the revenue reduction and eight times the expenditure reduction. Therefore, the findings suggest that although new foreign investment can have a positive effect on wages, the welfare effect on the local community will depend on the magnitude of benefits versus the costs of investment in the community; and the costs can be nontrivial.

Finally, a study by Ford, Rork, and Elmslie (2008), similar to the panel studies described above, adds an interaction term of FDI with human capital and finds that FDI has a greater impact on growth per capita than domestic investment in U.S. states, conditional on a minimum human capital threshold, exactly as occurs in studies that use country data. This implies that FDI is posi-

¹⁶ Carkovic and Levine (2005) argue that instrumental variables and the use of dynamic panels estimated by generalized method of moments can help in this context. Given the nontechnical level of our article, we invite the interested reader to consult this article for further details.

¹⁷ Notice that the total manufacturing employment measure, however, does not exclude the level of employees in foreign-owned “greenfield” plants, a feature that might introduce a bias in the estimates.

tively related to states' growth only when a sufficient level of human capital is present and that only states with a comparatively well-trained workforce can take advantage of the presence of foreign technology and should try to attract FDI.

DOES THE OUTPUT OF THE MULTINATIONAL SECTOR EXHIBIT HIGHER LABOR PRODUCTIVITY?

There is important information contained in aggregate data about FDI and labor productivity, as studied in Corrado, Lengermann, and Slifman (forthcoming) for the United States and Criscuolo (2005) for a group of OECD countries.

Corrado, Lengermann, and Slifman (forthcoming) focus on labor productivity in the United States. The United States is a particularly interesting country in that productivity has increased at a particularly high average annual rate of 1.84 percent between 1977 and 2006, faster than in most other industrial countries, at least since 1995. Moreover, the United States is both a major recipient of FDI from abroad and an especially major source of FDI for the rest of the world. The authors separate U.S. gross domestic product and productivity growth into that produced by exclusively domestic firms and that produced by the so-called multinational sector (i.e., the "industry" composed of foreign firms with U.S. production units and U.S. firms with foreign production units—not only foreign firms in the United States). Labor productivity estimates are then calculated by sector: In each year, productivity levels are defined as real value added per total hours worked of all persons. Private multinational nonfarm, nonfinancial firms (again, not just foreign producers in the United States) contribute only 40 percent of the output of nonfinancial corporations but more than 75 percent of the increase in labor productivity between 1977 and 2000. Moreover, all of this new productivity in nonfinancial corporate sectors in the late 1990s can be traced back to multinationals (Figure 6 and Table 2).

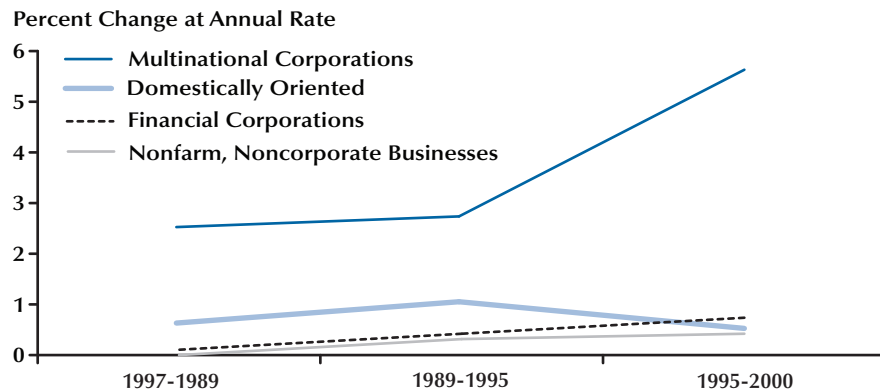
Does the evidence for the United States carry over to a broader set of industrial countries? Criscuolo (2005) evaluates the direct contribution

of the *affiliates* of foreign multinational firms to labor productivity for a large set of OECD host countries and shows that the productivity advantage of affiliates of multinational firms varies greatly across countries in both the manufacturing and service sectors.¹⁸ In the United States, France, and Sweden, foreign affiliates have approximately the same level of measured labor productivity as domestic firms. In Spain, Hungary, and the United Kingdom, foreign affiliates are twice as productive and this advantage appears to be markedly larger in low-tech sectors, such as food products, beverages, tobacco, textile and garment, leather, and footwear. Two possible explanations for the lack of affiliates' advantage in the United States, France, and Sweden are that domestic firms in these countries are at the technological frontier in many sectors and that domestic firms' productivity might have increased as a response to early trade openness and foreign competition. Hence, the data might be revealing that openness to competition induces survival and development of more-productive firms that in turn are more likely to become multinationals and contribute to their home country's productivity growth—as the Corrado, Lengermann, and Slifman (forthcoming) study would suggest. In Criscuolo (2005), however, the contribution to total productivity growth is quite heterogeneous, ranging from 32 percent in the United States to 42 percent in Finland, 164 percent in the Czech Republic, and an astonishing 251 percent in Norway (the highest).¹⁹ Her study is consistent with Corrado, Lengermann, and Slifman (forthcoming) because they define the multinational sector as the sum of foreign affiliates in the United States and U.S. firms that are multinational and control affiliates in other countries.

Interestingly, Criscuolo (2005) also shows how in a given time period, for example one year, the contribution to labor productivity growth can be decomposed into two components: a within effect,

¹⁸ The sources of data are three OECD maintained datasets: the STAN productivity database, the AFA (Activity of Foreign Affiliates) database, and the FATS (Foreign Affiliates' Trade in Services) statistics. The time period considered is 1995-2001.

¹⁹ Contributions larger than 100 percent are explained by a steep increase in the presence of affiliates with a higher labor productivity or by negative labor productivity growth of domestic firms, as in the United Kingdom.

Figure 6**Growth of U.S. Labor Productivity**

SOURCE: Corrado, Lengermann, and Slifman (forthcoming).

attributed to existing multinationals located in a country, and a between effect, attributed to the increase in the share of the employment of multinational affiliates. In her article, the within component drives the results in the United States, Hungary, and the Netherlands, while in all other countries the between component is the main contributor. The author also shows that the within effect is stronger in high-tech sectors, which are presumably information technology-intensive, which is a result that is consistent with the findings in microdata, as in Bloom, Sadun, and Van Reenen (2007).

Finally, a different dimension of disaggregation, industry composition, might be important in understanding the FDI contribution to growth. Based on an industry-level analysis, Fillat Castejón and Wörz (2006) argue that disaggregating flows by industry conveys important information because industrial specialization and composition are different across countries and this may affect the ability of a country as a whole to take advantage of the multinationals' technology. Even at the industry level, and in line with previous studies, Fillat Castejón and Wörz (2006) find that an additional factor needs to be included in the cross-country or cross-industry regression. Their analysis confirms previous analysis in terms of

Table 2**Composite of U.S. GDP, 1997 and 2002 (%)**

	1977	2002
Multinational corporations	25.5	26.2
Domestically oriented	45	39.3
Financial corporations	4.6	9.2
Noncorporate businesses	25	25.3

SOURCE: Corrado, Lengermann, and Slifman (forthcoming).

requiring “other factors”—namely, export orientation and domestic investment in this case—to find a positive relationship between FDI and growth in the data.

WHAT DO AGGREGATE DATA HIDE?

As Criscuolo (2005) shows, disaggregating the data helps illuminate the effects of multinational affiliates' activity on productivity and ultimately on growth.

Imagine dividing the population of firms in the host economy into a foreign multinational component (indexed by I) and a purely domestic

component (indexed by D). Then total final output, Y_j , could be also reinterpreted as the sum of two components produced by domestic and foreign firms, $Y_j = Y_j^D + Y_j^I$. In each time period a total of N_j^D domestic firms and N_j^I affiliates of multinational firms produce and sell locally Y_j^D and Y_j^I , and perhaps a share of these firms sells abroad as exporters. (In this context, the distinction can be neglected without loss of generality.) One of the issues with estimation of aggregate data models like equation (3) is that variables referring to foreign firms are not excluded from the aggregate and the dependent variable is the growth rate of *total* output. If we distinguish domestic from foreign firms and the real output they produce, then we are really considering two production functions and observing total output as the sum of the output of the domestic firms and the multinational firms:

$$(5) \quad Y_j = Y_j^D + Y_j^I$$

$$(6) \quad Y_j^D = Z_j^D f(K_j^D, X_j^D)$$

$$(7) \quad Y_j^I = Z_j^I f(K_j^I, X_j^I).$$

There are at least two dimensions along which microeconomic heterogeneity affects econometric analyses that use aggregate data.

First, there may be important aggregation effects that might mirror microeconomic heterogeneity in terms of *directly* measurable productivity advantages ($Z_j^I > Z_j^D$). If there is a large heterogeneity of multinational firms across countries and over time, and cross-country or panel studies that do not account for it, the average impact of FDI on growth may have little economic meaning. Contessi (2009a,b) discusses how thinking of multinational firms as heterogeneous agents in the sense of Ghironi and Melitz (2005) may help to reconcile the evidence on superior productivity leadership with the aggregate evidence discussed in this paper. Ghironi and Melitz (2005) assume that a firm’s labor productivity is the product of an idiosyncratic time-invariant productivity level and a time-varying aggregate productivity level that is common across all firms

in the economy.²⁰ The idiosyncratic productivity level can be interpreted as the management’s ability to appropriate and scale up (or down) the existing technology available to all firms. If one assumes the existence of fixed costs to export (as in Ghironi and Melitz, 2005) or to engage in multinational production (as in Contessi, 2009b), then some firms will export, some firms will be multinational, and the participation of firms in international markets evolves over time with entry and exit dynamics.

If firms incur a fixed cost to become multinationals, and either this cost decreases over time because of FDI liberalization or the attractiveness of foreign markets increases due to their growth, then entrants sort according to their own productivity. The most productive firms become multinationals earlier, while firms that are relatively less productive enter host economies later, a fact documented by Yeaple (2008) using firm-level data for U.S. multinationals. Specifically, Yeaple (2008) finds that a 10 percent increase in a host country’s GDP is associated with a 5.4 percent increase in the number of U.S. firms owning an affiliate there, but also with a 1.4 percent decrease in the average productivity of the entrants. Entry of relatively less-productive firms then reduces the average productivity of the “multinational sector” in the host economy, exactly while the stock of FDI increases. Hence, while we expect countries with larger stocks of FDI to grow faster, the contribution of multinational firms to the host countries’ productivity level, and growth, decreases.

Second, there may be relevant *positive externalities (spillovers)* that regressions based on aggregate data are likely to miss. The presence of spillovers implies that the mere presence of a foreign firm increases the productivity of domestic firms—or, in our discussion, $\partial Z_j^D / \partial N_t^I > 0$. There are at least three groups of intra-industry spillovers. Javorcik (2008) contains an updated survey of the literature on spillovers from FDI and an interesting discussion of the reasons why their presence is difficult to measure and appears to vary greatly across countries.

²⁰ By idiosyncratic productivity, we mean firm-specific labor or total factor productivity that is not common to other firms.

The demonstration effect is the first type of positive agglomeration externality: Although foreign firms have a strong incentive to protect their firm-specific product or process knowledge, domestic companies can still observe their practices to some extent and learn about new technology, marketing techniques, and product development. Second, domestic firms may hire workers who were employed, and hence trained, by multinationals in the past; in this case, spillovers occur through workers' mobility. Third, intra-industry spillovers may also occur through the so-called competition effect: Domestic firms must increase their productivity to compete with the foreign firms and survive their competition, while nonprofitable firms are likely eliminated from the market.

There are various ways in which the presence of spillovers from FDI may alter our reading of growth regressions that study FDI, as discussed in Hale and Long (forthcoming). On the one hand, if aggregate output measurements include the output of both the domestic sector and the output of foreign multinationals in the country, then the impact of FDI will always be positive ($\hat{\beta}_0 > 0$), even if domestic productivity is not affected by the foreign presence, just because the output measurement already incorporates the multinational output growth that will always positively correlate to multinational presence. Hence, one would want to use only domestic output as a dependent variable.

On the other hand, estimating the presence of spillovers using data that focus on domestic variables is conducive to a second type of bias. A large part of recorded FDI consists of acquisition, rather than new establishments. Hence, a selection bias problem, called *cherry-picking* in the literature, has to be dealt with in formal analyses. If acquired affiliates are picked from the upper tail of the productivity distribution (firms with high productivity *ex ante*), the post-acquisition distribution is truncated from above. This means that the larger the number of acquisitions, the lower the share of domestically owned firms with relatively lower productivity left in the population of "domestic" firms. If one estimates a regression such as the one in equation

(3), then the effect of FDI_t will be *negative* ($\hat{\beta}_0 < 0$). Therefore, the researcher should estimate a model with sample selection in the tradition of Heckman (1979) using firm- or plant-level data.

These issues are likely to be at least in part responsible for the lack of evidence of a relationship between FDI and growth that we discussed earlier in this article.

CONCLUSION

The contributions that multinational firms make toward economic growth of the host economies have been studied extensively, but little consensus has emerged as to whether FDI is boon or bane for a country as a whole. Quite simply, the evidence is as mixed now as it was when Rodrik (1999) wrote the line quoted at the beginning of this article. Lacking unambiguous empirical evidence, it is difficult to formulate solid expectations on how proposed FDI policies will affect the entry of foreign firms. Current empirical evidence provides little guidance as to whether one should support or oppose policies. As we have discussed in this article, studies that use a growth regression approach and aggregate data are not likely to help researchers sort out the growth effect of FDI because of methodological problems and huge heterogeneity hidden by the data. The prior 10 years of research have confirmed that the (aggregate) evidence is still sobering.

However, a large body of empirical research that uses firm- and plant-level data has documented that multinational firms and their affiliates (compared with domestic firms) are larger, are more capital intensive, make more abundant use of skilled workers, invest more in physical and intangible capital, and pay higher wages (Barba Navaretti and Venables, 2004). Because the evidence based on microdata shows that firms investing and producing in foreign countries have superior productivity at home, foreign affiliates should also enjoy a productivity advantage compared with local firms in the host economy. Indeed, we have discussed some of the evidence that reveals such an effect in aggregated data.

Recent developments in the use of microdata to study aggregate productivity such as the ones

proposed by Petrin and Levinsohn (2005) might prove to be key in overcoming the methodological difficulties and the problems arising with the use of aggregate data; ultimately, these developments may increase our understanding of this relationship between FDI, productivity, and growth.

REFERENCES

- Aitken, Brian and Harrison, Ann. "Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela." *American Economic Review*, June 1999, 89(3), pp. 605-18.
- Alfaro, Laura; Chanda, Areendam; Kalemli-Ozcan, Sebnem and Sayek, Selin. "FDI and Economic Growth: The Role of Local Financial Markets." *Journal of International Economics*, October 2004, 64, pp. 89-112.
- Balasubramanyam, V.N.; Salisu, M. and Sapford, David. "Foreign Direct Investment and Growth in EP and IS Countries." *Economic Journal*, January 1996, 106(434), pp. 92-105.
- Barba Navaretti, Giorgio and Venables, Anthony. *Multinational Firms in the World Economy*. Princeton, NJ: Princeton University Press, 2004.
- Bhagwati, J.N. "Protectionism," in *The Ohlin Lectures*. Cambridge, MA: MIT Press, 1988.
- Blomström, Magnus; Lipsey, Robert E. and Zejan, Mario. "What Explains Developing Country Growth?" NBER Working Paper No. 4132, National Bureau of Economic Research, 1994.
- Blonigen, Bruce A. and Wang, Miao. "Inappropriate Pooling of Wealthy and Poor Countries in Empirical FDI Studies," in Theodore H. Moran, Edward D. Graham, and Magnus Blomström, eds., *Does Foreign Direct Investment Promote Development?* Washington, DC: Institute for International Economics, 2005, pp. 221-44.
- Bloom, Nicholas; Sadun, Raffaella and Van Reenen, John. "Americans Do I.T. Better: US Multinationals and the Productivity Miracle." NBER Working Paper No. 13085, National Bureau of Economic Research, 2007.
- Borensztein, Eduardo; De Gregorio, Jose and Lee, Jong-Wha. "How Does Foreign Direct Investment Affect Economic Growth?" *Journal of International Economics*, June 1998, 45(1), pp. 115-35.
- Boyd, John H. and Smith, Bruce D. "Intermediation and the Equilibrium Allocation of Investment Capital: Implications for Economic Development." *Journal of Monetary Economics*, December 1992, 30(3), pp. 409-32.
- Carkovic, Maria and Levine, Ross. "Does Foreign Direct Investment Accelerate Economic Growth?" in Theodore H. Moran, Edward D. Graham, and Magnus Blomström, eds., *Does Foreign Direct Investment Promote Development?* Washington, DC: Institute for International Economics, 2005, pp. 195-220.
- Contessi, Silvio. "FDI, Productivity, and Growth: Can We Reconcile Microeconomic and Aggregate Evidence?" Unpublished manuscript, 2009a.
- Contessi, Silvio. "What Happens When WalMart Comes to Your Country? Multinational Firms' Entry, Productivity, and Inefficiency." Unpublished manuscript, 2009b.
- Contessi, Silvio; De Pace, Pierangelo and Francis, Johanna. "The Cyclical Properties of Disaggregated Capital Flows." Working Paper No. 2008-041, Federal Reserve Bank of St. Louis, November 2008; <http://research.stlouisfed.org/wp2008/2008-041.pdf>.
- Corrado, Carol; Lengermann, Paul and Slifman, Larry. "The Contribution of MNCs to US Productivity Growth 1977-2000," in M. Reinsdorf and M. Slaughter, eds., *International Flows of Invisibles: Trade in Services and Intangibles in the Era of Globalization*. Chicago: University of Chicago Press and NBER, forthcoming.
- Coughlin, Cletus. "Foreign-Owned Companies in the United States: Malign or Benign?" *Federal Reserve Bank of St. Louis Review*, May/June 1992, 74(3), pp. 17-31. (Reprinted in Robert W. Kolb, ed., *International Finance Reader*. Boulder, CO: Kolb Publishing, 1993.)

- Criscuolo, Chiara. "Foreign Affiliates in OECD Economies: Presence, Performance, and Contribution to Host Countries." *OECD Economic Studies*, 2005, 41(2), pp. 109-39.
- Figlio, David and Blonigen, Bruce. "The Effects of Direct Foreign Investment on Local Communities." *Journal of Urban Economics*, September 2000, 48(2), pp. 338-63.
- Fillat Castejón, Carmen and Wörz, Julia. "Good or Bad? The Influence of FDI on Output Growth: An Industry-level Analysis." WIIW Working Paper No. 38, Vienna Institute for International Economic Studies, 2006.
- Findlay, Ronald. "Relative Backwardness, Direct Foreign Investment, and the Transfer of Technology: A Simple Dynamic Model." *Quarterly Journal of Economics*, February 1978, 92(1), pp. 1-16.
- Ford, Timothy; Rork, Jonathan and Elmslie, Bruce. "Foreign Direct Investment, Economic Growth, and the Human Capital Threshold: Evidence from US States." *Review of International Economics*, February 2008, 16(1), pp. 96-113.
- Ghironi, Fabio and Melitz, Marc. "International Trade and Macroeconomic Dynamics with Heterogeneous Firms." *Quarterly Journal of Economics*, August 2005, 120, pp. 865-915.
- Hale, Galina and Long, Cheryl. "Are There Productivity Spillovers from Foreign Direct Investment in China?" *Pacific Economic Review*, forthcoming.
- Heckman, James J. "Sample Selection Bias as a Specification Error." *Econometrica*, January 1979, 47(1), pp. 153-61.
- Helpman, Elhanan; Melitz, Marc J. and Yeaple, Stephen R. "Export Versus FDI with Heterogeneous Firms." *American Economic Review*, March 2004, 94(1), pp. 300-16.
- Ihrig, Jane and Marquez, Jaime. "Modeling Direct Investment Valuation Adjustments and Estimating Quarterly." International Finance Working Paper No. 857, Board of Governors of the Federal Reserve System, 2006.
- International Monetary Fund. "Direct Investment Methodology: Latest SIMSDI Metadata: Individual Country." 2003; www.imf.org/external/np/sta/di/country.asp.
- Javorcik, Beata S. "Can Survey Evidence Shed Light on Spillovers from Foreign Direct Investment?" *World Bank Research Observer*, 2008, 23(2), pp. 139-59.
- Lane, Philip and Milesi Ferretti, Gian-Maria. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities." *Journal of International Economics*, November 2007, 73(2), pp. 223-50.
- Marin, Dalia and Schnitzer, Monika. "When Is FDI a Capital Flow?" CEPR Discussion Paper No. 5755, Centre for Economic Policy Research, July 2006.
- Melitz, Marc J. "Comment," in Theodore H. Moran, Edward D. Graham, and Magnus Blomström, eds., *Does Foreign Direct Investment Promote Development?* Washington, DC: Institute for International Economics, 2005.
- OECD. *OECD Benchmark Definition of Foreign Direct Investment*. Fourth Edition. Paris: OECD, 2008.
- Petrin, Amil and Levinsohn, James. "Measuring Aggregate Productivity Growth Using Plant-Level Data." NBER Working Paper No. 11887, National Bureau of Economic Research, 2005.
- Rivera-Batiz, Francisco and Rivera-Batiz, Luis. "The Effects of Direct Foreign Investment in the Presence of Increasing Returns Due to Specialization." *Journal of Development Economics*, November 1991, 34(1-2), pp. 287-307.
- Rodrik, Dani. *The New Global Economy and Developing Countries: Making Openness Work* (Policy Essay 24). Baltimore, MD: Johns Hopkins University Press, 1999.
- Rodrik, Dani. "Why We Learn Nothing from Regressing Economic Growth on Policies." Unpublished manuscript, Harvard University, March 2005.

Contessi and Weinberger

Yeaple, Stephen. "Firm Heterogeneity and the Structure of U.S. Multinational Activity: An Empirical Analysis." NBER Working Paper No. 14072, National Bureau of Economic Research, 2008.



Quick Exits of Subprime Mortgages

Yuliya S. Demyanyk

All holders of mortgage contracts, regardless of type, have three options: keep their payments current, prepay (usually through refinancing), or default on the loan. The latter two options terminate the loan. The termination rates of subprime mortgages that originated each year from 2001 through 2006 are surprisingly similar: about 20, 50, and 80 percent, respectively, at one, two, and three years after origination. For loans originated when house prices appreciated the most, terminations were dominated by prepayments. For loans originated when the housing market slowed, defaults dominated. The similarity of the loan termination rates for all vintages in the sample suggests that subprime mortgage loans were intended to be “bridge” (i.e., temporary) loans. In addition, between 2001 and 2006, the number of terminated subprime purchase-money loans (loans used to purchase rather than refinance a house) outweighed the estimated number of first-time-homebuyers with subprime mortgages. The effect of the subprime lending on the increase of homeownership in the United States—a potentially positive outcome of subprime mortgages—most likely has been overstated. (JEL D12, G1, G21)

Federal Reserve Bank of St. Louis *Review*, March/April 2009, 91(2), pp. 79-93.

The subprime mortgage market boomed between 2001 and 2006 and began to collapse in 2007.¹ Initial signs of the collapse were poor performance and even default of loans,² often within months of their origination: The delinquency, default, and foreclosure rates of subprime loans that were originated in 2006 and 2007 were three times higher than in earlier years. In 2008, the sub-

prime securitized market froze completely and essentially died.

Researchers, policymakers, journalists, and other individuals have offered many explanations for the collapse of the subprime mortgage market, including mortgage interest rate resets, fraud, poor underwriting, discrimination, a housing market slowdown, and deterioration of loan quality (due to unobserved or unexplored borrower information). The negative *consequences* of this market's collapse are well known and well publicized. The effects include foreclosures and defaults, impaired credit histories for borrowers, reduced housing values, destabilized neighborhoods as a result of vacant properties, and an overall slowdown in many segments of the economy.

But aside from these pitfalls, did subprime lending have any benefits, even if they were much

¹ The term “subprime,” at times used imprecisely, essentially can describe (i) borrowers with a low credit score, history of delinquency or bankruptcy, or poor employment history; (ii) lenders specializing in high-cost loans and selling fewer loans to government-sponsored enterprises; (iii) securities that encompass a subprime loan (the most- to least-risky of which are subprime, Alt-A, and prime); and (iv) certain mortgages (e.g., 2/28 or 3/27 “hybrid” mortgages) generally not available in the prime market.

² Here, “default” is used to indicate the protracted failure to meet the terms of a mortgage loan agreement, ending in foreclosure; “early default” is defined later in the paper.

Yuliya S. Demyanyk is a senior research economist at the Federal Reserve Bank of Cleveland. The data analysis for this article was conducted when she was an economist in the Banking Supervision and Regulation Division of the Federal Reserve Bank of St. Louis.

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

less obvious than the problems? Anecdotal evidence suggests that the subprime market, with its easier mortgage financing, may have promoted U.S. homeownership. The rationale is that, even if default rates are about 20 percent for the most recent vintage of subprime mortgages, 80 percent of subprime borrowers are still making their monthly payments. Given this view, the financial innovation that spawned subprime lending may have promoted homeownership, and thus the majority of borrowers *benefited* because they most likely would not have qualified for mortgages under terms in the prime market.

This paper attempts to analyze whether borrowers intended to keep their subprime mortgages long enough to substantiate an increase in homeownership or planned a quick exit strategy at origination, using subprime loans as bridge financing to speculate on house prices (i.e., quickly sell the house for a profit after its value increases).

“Exit” from a subprime mortgage can take two forms: prepayment or default. In this study, a mortgage loan is considered “prepaid” if a borrower has either paid the mortgage loan in full or refinanced it within a certain period after the loan was originated. A mortgage loan is “in default” if (i) a borrower has missed more than two mortgage payments, (ii) the property is in the process of foreclosure (after more missed payments), (iii) the property is “real-estate owned” (i.e., has been taken over by the lender as part of the loan termination³ process) within a certain period of time after origination, or (iv) the borrower defaults on the contract (“walks away”).

The paper is organized as follows. First, it briefly describes the evolution of the U.S. subprime mortgage market, the crisis, and some of the earlier research that analyzes factors associated with loan termination (exit from the market). Second, it outlines the empirical analysis of explanatory factors of prepayment, default, and termination (prepayment and default combined) within two years of loan origination; it further compares the number of prepaid and defaulted loans per year within two years of origination.

³ In this paper, the terms “exit” and “termination” are used interchangeably.

Third, it points to the quick termination of subprime loans, indicating that these loans must have been designed and intended to be temporary and their existence most likely did not contribute to increased homeownership rates in the United States between 2001 and 2006.

SUBPRIME MORTGAGE CRISIS: HIGH DEFAULT RATES

The boom and subsequent collapse of the subprime mortgage market has drawn the attention of numerous researchers and policymakers. This analysis of delinquencies and foreclosures is not new. For example, Von Furstenberg and Green (1974) analyzed the causes of mortgage delinquencies, apart from foreclosures and defaults, for mortgages originated between 1961 and 1972. They refer to and confirm findings published as early as 1969 and 1970 (by Von Furstenberg) that such factors as high loan-to-value (LTV) ratios (or equity-to-value ratios) and low borrower income are important determinants of mortgage default, *ceteris paribus*. Thus, these findings were known some three decades before these subprime issues unfolded, before very large LTVs were deemed “acceptable,” and so-called no-income, no-documentation, no-asset mortgage loans were introduced.

In a more recent, but precrisis analysis, Cutts and Van Order (2005) suggest that several economic models can, in fact, explain the main characteristics of the subprime market. In particular, “option-based” models are consistent with pricing and loan characteristics of subprime mortgages (for example, improving a borrower’s credit score makes refinancing more likely); “separating equilibrium” models sort borrowers into prime and subprime markets through signaling mechanisms; and “adverse selection” models are consistent with the choice between the lower costs of the secondary market and the information advantages of the primary market. However, many issues were and still are beyond fundamental and conventional economic modeling. For instance, Demyanyk (2008) shows that the Fair Isaac and Company (FICO) credit score failed to predict the

subprime mortgage crisis, even though it is one of the most important determinants of serious delinquency and foreclosure in mortgage lending.⁴

Pennington-Cross and Chomsisengphet (2007) studied a sample of subprime securitized loans—first-lien, fixed-rate, homeowner-occupied—that originated between 1996 and 2003. The authors note that borrowers with subprime mortgages are more likely to cash-out refinance compared with those with prime mortgages.⁵ Moreover, subprime borrowers seem to substitute mortgage debt for credit card debt and auto loans: They tend to refinance their mortgages when interest rates on credit cards and auto financing rise. Analyzing the performance of subprime loans, the authors observed that cash-out refinances tend to default and prepay less frequently than non-cash-out refinances. Demyanyk and Van Hemert (2008) observed that cash-out refinances between 2001 and 2007 tended to default less frequently than even purchase-money mortgages.

Demyanyk and Van Hemert (2008) were among the first to analyze the subprime mortgage crisis in detail. Using loan-level data, they first showed that—contrary to popular belief—the subprime crisis of 2007 was not confined to a particular market segment, such as loans with mortgage rates scheduled to increase or no-documentation loans. Instead, it was a (subprime) marketwide phenomenon. Second, they identified factors most likely to be associated with a larger probability that a subprime mortgage loan would become seriously delinquent: FICO credit score, combined LTV (CLTV) ratio, mortgage interest rate, and house price appreciation between the period of loan origination and the loan-performance evaluation. These factors were not sufficiently different in the crisis years (2006 and 2007) than in the earlier years and thus do not entirely explain the crisis, its magnitude, or its timing. Even house price appreciation does not explain—by itself or in a combination with other factors (a phenome-

non called risk layering)—why the subprime crisis was so rapid and large.

Demyanyk and Van Hemert (2008) also showed the presence of nonmeasurable risk in these mortgage contracts and the increased risk over time. More specifically, they first adjusted mortgage performance for values of observable borrowers' characteristics at origination (e.g., credit scores, LTV ratios, debt-to-income ratios), loan characteristics (e.g., fixed-rate mortgage [FRM] or hybrid mortgage, if homeowner-occupied, presence of prepayment penalty clause), and macroeconomic conditions (e.g., change in unemployment, household income, house price appreciation since origination). Second, they calculated the adjusted performance of the loans for all vintage/loan age combinations in their sample; this exercise revealed that the market has worsened each year, monotonically and dramatically, since 2001. In other words, the crisis did not emerge suddenly in 2007 or 2008. It had been brewing for at least six years prior.

Even though this scenario and time frame are not readily observable by looking at the data—a statistical exercise is needed to see the deterioration of the subprime market—Demyanyk and Van Hemert (2008) show that securitizers, those who mostly dictated mortgage rates in the market, were to some extent aware of this gradual deterioration. The decline in loan quality was monotonic but not equally spread among different types of borrowers. Over time, loans with high LTV ratios had higher adjusted delinquency, foreclosure, and defaults rates. Securitizers started to link mortgage interest rates to LTV ratios; obviously, they did not do so enough. Loan quality deteriorated while loan riskiness increased every year from 2001 to 2007; however, the price of risk—the subprime-prime markup—in fact, declined. The combination of increasing loan riskiness and decreasing prices was not sustainable. In 2008, the market collapsed and massive foreclosures, bank failures, and a credit crunch followed.

Haughwout, Peach, and Tracy (2008) took the analysis by Demyanyk and Van Hemert (2008) a step further and analyzed early defaults of subprime mortgages. “Early default” is defined as either delinquency (missed payments) for more

⁴ For a more detailed discussion of delinquency and foreclosure determinants, see Demyanyk and Van Hemert (2008).

⁵ A term “cash-out” refinance refers to a situation when a borrower refinances an existing mortgage loan into a larger one, taking cash out. This, by definition, means that a borrower is extracting the equity from the house.

than 60 days or foreclosure within the first year after origination. Haughwout, Peace, and Tracy (2008) confirm the finding of Demyanyk and Van Hemert (2008) that, although credit/lending standards are important determinants of early default, they alone cannot explain the timing and the magnitude of the crisis in 2007 and 2008. They also confirm that, even if depreciation in house prices is an important determinant of increased delinquencies and foreclosures in the immediate precrisis years, a large portion of the increase in serious delinquencies remains unexplained. On the other hand, Keys et al. (2008) found that (observed) lending standards in the subprime mortgage market did deteriorate; and the main driving force of the deterioration was the securitization of those loans.

In their analysis of the subprime crisis, Mian and Sufi (2008) suggest that securitization of mortgage assets may have increased the supply of credit in geographic areas that had relatively more mortgage application rejections a decade before the crisis (in 1996); such credit allowed more home purchases and thereby could have led to the rapid increases in house prices between 2001 and 2005. When housing values started declining, between 2005 and 2007, defaults followed.

Gerardi, Shapiro, and Willen (2007), using a unique dataset covering the homeownership experience in Massachusetts between 1989 and 2007, found that homeownership that began with a subprime mortgage ended in foreclosure 20 percent of the time; importantly, this number is about six times larger than a corresponding share of homeowners who started with prime mortgages. Foote et al. (2008) find that, based on the same dataset, almost half of residential foreclosures are concentrated in subprime mortgages, even if the subprime mortgage was a refinance of a prime loan.

Foote, Gerardi, and Willen (2008) argue that even though borrowers facing negative equity in their houses are more likely to default, they may not default in the absence of an idiosyncratic shock, such as illness, divorce, or the loss of a job. Also, borrowers need to consider if the cost of default—which includes the cost of renting after the default—outweighs a potential (future)

benefit from home equity, should the home price increase in the future. In other words, negative equity is a necessary but not a sufficient condition for default.

EMPIRICAL ANALYSIS OF LOAN TERMINATIONS

A simple logit model was used to calculate the impact of a set of explanatory factors—such as borrower and loan characteristics and house price appreciation in the area surrounding the property—on the probability of either prepayment or default. According to the estimated results, the main factors affecting the probability of prepayment within two years of origination are (i) house price appreciation (pre-origination and post-origination), (ii) the presence of prepayment penalties, (iii) the resetting structure of mortgage rates (as with hybrid mortgages), and (iv) the CLTV ratio, which measures the amount of equity in the house. The main factors affecting the probability of default within two years of origination are (i) the FICO credit score, (ii) the CLTV ratio, (iii) the mortgage rate, and (iv) post-origination house price appreciation. Notably, the credit score affects only the likelihood of default, not prepayment; and pre-origination house price appreciation affects only prepayment, not default. Borrowers with hybrid mortgages do tend to prepay and default more often than those with FRMs (see Demyanyk and Van Hemert, 2008, for supporting evidence); however, *ceteris paribus*, the sole fact that a mortgage loan is a hybrid is not a strong predictor of default.

The factors that most affect prepayments and defaults were not substantially different in the precrisis years, with the exception of house price appreciation. For loans originated in 2003 and 2004, high house price appreciation is the main contributing factor for high prepayment rates. For loans originated in 2005 and 2006, low house price appreciation is the main contributor for the high default rates. Although house price depreciation is the main contributing factor, it is not the sole explanation for the magnitude of the crisis: The default rates are higher than what can be explained by housing market factors alone.

Borrowers' options to prepay or default on their mortgages have been analyzed in the context of the pricing of mortgage contracts for decades. Deng, Quigley, and Van Order (2000) provide an extensive literature review describing earlier analysis of prepayment only, default only, and default and prepayment as joint options. The authors theoretically unify several economic models to analyze prepayment and default options considered by borrowers simultaneously and empirically test this model on a sample of fixed-rate, fully amortized loans that originated between 1976 and 1983 and observed until the first quarter of 1992. All these loans were purchased by Freddie Mac. Even though the loans were made and their performance evaluated long before subprime issues emerged, the implications of this research are important: The authors found evidence of the interdependence of the decisions to prepay (akin to exercising a call option) or default (akin to exercising a put option). Forecasts that ignore this interdependence can lead to serious errors in estimating the default risk. For a related analysis, see Pennington-Cross and Chomsisengphet (2007).

The following logit regression model is estimated to analyze a random sample of subprime securitized loans (between 2001 and 2006) as a cross section:

$$\text{Probability}(Z) = \Phi(\beta'X),$$

where Z is either prepayment or default on (and thus exit from) a subprime mortgage loan within 24 months of origination; $\Phi(x) = 1/(1 + \exp(-x))$ is the logit function; $x = \beta'X$; X is the vector of explanatory variables; and β is the vector of regression coefficients.

The explanatory factors used in the analysis are the FICO credit score, a dummy variable indicating whether full documentation was provided at origination, a dummy variable indicating whether a prepayment penalty is present, the debt-to-income ratio (back-end), a dummy variable indicating whether a debt-to-income ratio is not provided, the mortgage interest rate, a dummy variable indicating whether a borrower is an investor, a dummy variable indicating whether a mortgage was a refinance at origination, the origi-

nation amount, the CLTV ratio, a margin for hybrid loans, a dummy variable indicating whether a mortgage is a hybrid, a dummy variable indicating whether a mortgage is an adjustable rate-mortgage (ARM, nonhybrid), a dummy variable indicating whether a mortgage is a balloon, post-origination house price appreciation (from loan origination up to the point of loan performance evaluation, up to three years later), and pre-origination house price appreciation (from two years before origination up to origination).

When to evaluate loan performance (within two years of origination) was a choice driven mainly by two factors: the FICO credit score and the popularity of hybrid mortgages in the sample. The FICO credit score, as with any credit score, measures the creditworthiness of individuals or businesses. Lenders/securitizers use these scores to estimate the likelihood of eventual delinquency or default. By design, the higher the credit score, the less likely it is that a borrower will miss payments or go into default on a loan within one or two years after the score has been calculated (Demyanyk, 2008). The prevalence of hybrid mortgages is also important. More than half of subprime securitized mortgage loans are ARMs, and almost all are so-called hybrid contract types, which means they carry a fixed interest rate for an initial period (usually two or three years) after which the rate resets. Starting the analysis at two (or three) years after origination eliminates the effect on these loans of mortgage rates resetting into a mostly larger market-driven rate plus a margin. (See Demyanyk and Gopalan, 2007, for a more detailed description and definitions.)

DATA AND VARIABLE DEFINITIONS

Loan-level data used for the analysis are provided by the First American CoreLogic LoanPerformance database, as of July 2008. In the dataset, loan, borrower, and property characteristics are provided for about half of all U.S. subprime mortgages. All loans in this dataset have been securitized. According to the *Mortgage Market Statistical Annual* (2008), securitization

rates are as follows: 60.7 percent (2001), 63.0 percent (2002), 67.5 percent (2003), 62.6 percent (2004), 67.7 percent (2005), 67.6 percent (2006), 74.2 percent (2007), and 77.3 percent (first six months of 2008). Among all subprime mortgages, the portion securitized ranged from 54 percent in 2001 to 75 percent in 2006. For the empirical analysis of this study, only first-lien subprime mortgages are used. The variables used in the analysis are defined as follows:

Cash-out: A dummy variable that equals 1 if the mortgage loan is a cash-out refinancing loan at origination and 0 otherwise.

CLTV ratio: The combined mortgage values of all liens divided by the value of the house at loan origination.

Debt-to-income ratio: The back-end debt-to-income ratio; it is defined as total monthly debt payments divided by gross monthly income at origination. A higher debt-to-income ratio (i.e., a higher degree of indebtedness) makes it harder for a borrower to make the monthly mortgage payment.

Default: A dummy variable that equals 1 if (i) the borrower has missed more than two monthly mortgage payments, (ii) the borrower has defaulted on the loan (with the foreclosure procedure finalized), or (iii) the property is in foreclosure or is real-estate owned (taken over by the lender) within the first two years of origination; the variable takes a value of 0 otherwise.

Documentation: A dummy variable that equals 1 if full documentation on the loan is provided and 0 otherwise.

FICO score: The FICO credit score at origination. The FICO score was recommended for use in mortgage lending by Fannie Mae and Freddie Mac in 1995 as a measure of borrowers' creditworthiness. The higher the FICO score, the less likely a borrower will default on a loan within about two years of loan origination. Given the nature of FICO scores, it is expected that a relationship will be found between borrowers' scores and the incidence of default and foreclosure during the subprime mortgage crisis.

Investor: A dummy variable that equals 1 if the borrower is an investor and does not owner-occupy the property and 0 otherwise.

Margin: The additional percentage points for an ARM or hybrid mortgage over an index interest rate, usually the six-month LIBOR rate, applicable after the first interest rate reset. The higher the margin, the higher the interest rate after the reset, which increases the monthly mortgage payments.

Missing debt-to-income: A dummy variable that equals 1 if the back-end debt-to-income ratio was not provided in the data (reported as 0); the variable takes a value of 0 otherwise. In the data, the debt-to-income value was not reported for approximately 30 percent of loans.

Mortgage rate: The initial interest rate as of the first payment date. A higher interest rate makes monthly mortgage payments larger and, therefore, can make it more difficult for a borrower to make timely monthly mortgage payments.

Origination amount: The size of the mortgage loan. Loan size can affect the size of a monthly mortgage payment: The larger the loan, the larger the monthly payment, and the harder it can be for a borrower to make those payments in a timely manner. Also, a borrower's creditworthiness can affect the size of the loan: Less-risky borrowers may be expected to get larger loans. Which of the two effects is dominant is an empirical question addressed later in this study.

Post-origination house price appreciation: The metropolitan statistical area (MSA)-level house price appreciation from the time of loan origination to the time the performance of the loan is evaluated. Appreciation is measured as a ratio of the house price indexes reported by the Office of Federal Housing Enterprise Oversight (now the Federal Housing Finance Agency) for the two corresponding periods.

Pre-origination house price appreciation: The MSA-level house price appreciation two years before mortgage origination and origination period.

Prepayment: A dummy variable that equals 1 if a borrower has either paid off or refinanced a mortgage loan within two years of origination; the variable takes a value of 0 otherwise.

Prepayment penalty: A dummy variable that equals 1 if a prepayment penalty is associated with a loan and 0 otherwise.

Product type: Major types in the subprime mortgage market include FRMs, hybrid mortgages, ARMs, and balloons. Three dummy variables for the latter three are included in the regression analysis; the magnitude of their impact therefore should be interpreted as the effect on the probability of prepayment, default, or exit relative to an FRM. The FRM is chosen as a benchmark because FRMs show the smallest expected and realized probability of default.

Termination: A dummy variable that equals 1 if a borrower has either defaulted or prepaid the mortgage loan within two years of origination; the variable takes a value of 0 otherwise.

EXPLANATORY FACTORS OF PREPAYMENT, DEFAULT, AND EXIT

Prepayment

House price appreciation occurring within two years of origination has the largest impact on the probability of a borrower to prepay or refinance a loan (see Table 1, column 1). An increase in house price appreciation of 1 standard deviation (SD) above its mean is associated with a 13-percentage-point increase in the likelihood that a loan will be prepaid, *ceteris paribus*. If house prices in the area appreciated 1 SD above the mean two years before origination, there is a 7-percentage-point increase in the likelihood a loan will be prepaid. This perhaps indicates that individuals build their expectations about future home values based on immediate past values (or the past trends).

Borrowers with hybrid mortgages tend to prepay more often; all other factors being the same,

if a loan is a hybrid and has a mortgage rate scheduled to reset in two or three years, the probability of prepayment increases by about 5.5 percentage points. Loan originators and securitizers must have been aware of this pattern; and so, to compensate for the expected losses of interest payments (payments borrowers never make if they prepay the loan before the end of the term), they imposed prepayment penalties on about 70 percent of subprime securitized mortgages. The prepayment penalty factor has its expected effect on the probability of prepayment: It decreases it—specifically, by about 6 percent within two years of origination.

The mortgage rate at origination plays an important role as well: The higher the rate, the higher the chance a loan will be prepaid within its first two years. The marginal effect of the mortgage rate is approximately 5 percentage points.

A loan's purpose at origination also affects prepayment. If a mortgage is originated to refinance an existing mortgage, it is more likely to be refinanced again after two years or less, compared with home purchase (purchase-money) loans.

Also, the smaller the down payment at origination, the less likely a borrower is to prepay or refinance a loan within two years of origination. In unfavorable economic circumstances, such as a housing market slowdown or job loss, *ceteris paribus*, a borrower would be expected to default rather than refinance a mortgage that had little equity.

The more expensive a property was at origination, the more likely its mortgage will be refinanced or prepaid. A larger origination amount is associated with larger monthly mortgage payment. The greater incentive to refinance more expensive properties may be a desire to lower monthly payments or a need to extract cash to cope with those (larger) monthly payments.

Default

The marginal effects of individual factors on the probability of default are listed in column 2 of Table 1. Four major factors seem to most affect the probability of default two years after origination: post-origination house price appreciation, FICO credit score, CLTV ratio, and the mortgage

Table 1**Impact of Individual Factors on the Probability of Prepayment, Default, or Exit within Two Years of Mortgage Loan Origination (2001-06)**

Explanatory factor	Prepayment	Default	Exit
FICO credit score	0.19*	-3.28***	-4.11***
If full documentation is provided (dummy)	0.38***	-1.31***	-1.21***
If prepayment penalty is present (dummy)	-6.27***	0.65***	-5.29***
Debt-to-income ratio (back end)	1.58***	1.28***	3.12***
If debt-to-income ratio is not provided (dummy)	1.17***	1.01***	2.28***
Mortgage interest rate	5.23**	2.27***	7.76***
If an investor (dummy)	-1.05***	0.93***	0.00
If a mortgage is for refinancing at origination (dummy)	2.68***	-1.08***	0.73***
Origination amount	3.03***	0.75***	4.16***
Combined loan-to-value ratio	-4.24***	4.34***	-0.89***
Margin for hybrid loans	0.46***	0.85***	2.26***
If a hybrid (dummy)	5.53***	0.36***	4.30***
If an ARM (dummy)	1.60***	0.05	1.64***
If a balloon (dummy)	0.72***	0.51***	1.48***
Post-origination house price appreciation	13.28***	-4.29***	7.31***
Pre-origination house price appreciation	7.31***	-0.46***	6.39***

NOTE: A mortgage loan is considered “prepaid” if a borrower has either prepaid or refinanced a mortgage loan within a certain period after loan origination. A mortgage loan is considered in “default” if a borrower has defaulted on a loan or has missed more than two mortgage payments or the property is in the process of foreclosure or is real-estate owned (i.e., is likely to default) within 2 years of origination. “Exit” from a subprime mortgage is either prepayment or default. The reported results are the marginal effects of each variable i calculated as follows:

$$MEFF_i = \Phi(\beta'X + \beta_i\sigma_i) - \Phi(\beta'X),$$

where $\Phi(\beta'X)$ is the likelihood that event Z will occur; Z is either prepayment (column 1), default (column 2), or exit (column 3) from a subprime mortgage loan within two years of origination; $\Phi(\cdot)$ is the logistic function; X is the vector of explanatory variables, σ_i is the standard deviation of variable i , and β is the vector of regression coefficients.

*, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

interest rate. This finding is consistent with the results obtained by Demyanyk and Van Hemert (2008), who estimated the effects of those factors on the probability of serious delinquency one year after origination. According to the estimates, a 1 SD increase in the FICO credit score, ceteris paribus, is associated with a decrease in a probability of default by 3.3 percentage points. Note that the credit score has almost no explanatory power for prepayment but is a critical factor in explaining defaults.

According to the estimates, a 1 SD increase in house value appreciation measured at the MSA-

level is associated with a 4.3-percentage-point decrease in the likelihood of default; the effect on prepayments and refinancing is about three times larger and has the opposite sign as expected. The difference in the absolute values of the marginal effects reflects an asymmetry in how equity affects different actions taken by the borrower. An increase in appreciation increases the probability of prepayment much more than it decreases the probability of default. Pre-origination house price appreciation, even though it has an economically significant impact on prepayments, has almost no effect on defaults.

The CLTV ratio's effect on default is comparable in magnitude (but opposite in sign) to its effect on prepayment. Less equity in the house, or a larger LTV ratio, is associated with an increased probability of default but decreased probability of prepayment. In both cases, the marginal effect is about 4.3 percentage points.

The mortgage interest rate has a marginal effect on the probability of default of 2.3 percentage points; recall that in the case of refinancing it is about double that number. This evidence seems to indicate that a high mortgage rate gives borrowers incentives to exit the mortgage through either prepayment or default.

Exit

Column 3 of Table 1 reports the estimates of the logit regression with the "exit" being a dependent variable; that is, each factor is being analyzed for its impact on prepayment and default combined. According to the estimates, the factors that have a significant effect on either prepayment or default have a significant impact on both of these options combined.⁶ The only exception is the CLTV ratio, where the effects on prepayment and default cancel each other in a joint regression.

ANNUAL FACTOR CONTRIBUTION TO PREPAYMENT AND DEFAULT

Through the boom and the subsequent bust of the subprime mortgage market, almost half of the subprime loan borrowers in the sample terminated their original mortgages through prepayment or default. The shares of prepayment and default among the terminated loans, however, varied by the vintage of those loans. For example, Figure 1 shows that the largest rates of prepayment within two years of origination were observed for loans

originated in 2002-04.⁷ This section attempts to empirically answer the following question: What observable factors, individually or in combination, can explain changes in prepayment and default ratios?

This study uses a method similar to the one developed by Demyanyk and Van Hemert (2008) to measure the extent each factor explains the likelihood of prepayment or default for different mortgage vintages. Specifically, for each year Y in the sample, the impact of each explanatory variable i is calculated as the difference between the logit function Φ where, for one variable i , the overall mean is substituted by its mean value in year Y (the values of all other variables remain at their overall mean values) and the logit function where all variables are at their overall mean values. More formally, the annual factor contribution (AFC_i^Y) for prepayment or default of each variable i and year Y is calculated by

$$AFC_i^Y = \Phi\left(\beta' \bar{X} + \beta_i \left(\bar{X}_i^{year Y} - \bar{X}_i\right)\right) - \Phi\left(\beta' \bar{X}\right),$$

where $\Phi(\beta' \bar{X})$ is the likelihood that prepayment or default will occur within 24 months of mortgage loan origination, $\Phi(\cdot)$ is the logistic function, X is the vector of explanatory variables, and β is the vector of regression coefficients.

As shown in Figure 1, within two years of origination, loans originated in 2001 had delinquency and default rates almost as high as loans originated in 2005. Column 1 of Table 2 shows the contribution of each factor for this origination year plus prior and subsequent house price appreciation.

Table 2 also shows how low FICO credit scores, high mortgage interest rates, and relatively low house price appreciation within two years of origination contributed to high default rates for the 2001 vintage loans. The mortgage interest rate continued to be a factor in defaults for vintage 2002 loans but was of a much smaller magnitude.

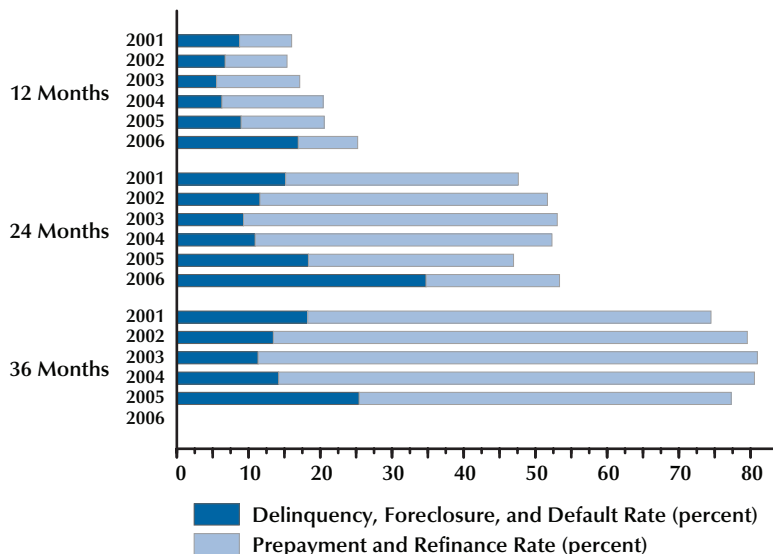
For 2003 and 2004 vintage loans, only post-origination house price appreciation (fast and

⁶ Deng, Quigley, and Van Order (2000) and Pennington-Cross and Chomsisengphet (2007) analyze the determinants of mortgage termination empirically, using a maximum likelihood framework analogous to the one used in the current study. However, a simpler approach has been undertaken here. Instead of a multinomial logit model (as in the study by Pennington-Cross and Chomsisengphet, 2007) or hazard functions (as in Deng, Quigley, and Van Order, 2000, or Demyanyk and Van Hemert, 2008), a simple logit function is estimated in this study for each of the outcomes of a loan termination.

⁷ The rates for all subprime loans in the sample (originated as both refinancings and purchase-money) are remarkably similar to those documented in Figure 1.

Figure 1

Termination of Subprime Purchase-Money Loans Within 12, 24, and 36 Months of Origination by Origination Year



NOTE: All loans used for this figure were securitized, originated as purchase-money, are first-lien mortgages, and have the borrower and loan characteristics reported in the data.

SOURCE: Author’s calculations based on FirstAmerican CoreLogic LoanPerformance loan-level dataset, as of July 2008.

positive) contributed to low default rates; defaults were substituted by prepayment and refinancing options exercised by borrowers, as discussed below in greater detail.

For 2005 and 2006 vintage loans, the only factor that contributed to higher default rates than those in all other years in the sample was post-origination house price depreciation. For these loans, house price appreciation contributed 2.6 and 7.5 percentage points, respectively, in 2005 and 2006 to the increase in the default rates two years after origination. However, the default rates for those loans were in fact about 20 to 30 percent, much higher than the rates explained by house price appreciation alone.

As shown in Table 3, column 1, the main contributing factor for high refinance rates within two years of origination for 2001 vintage loans was a high mortgage interest rate; its value accounted for 6.3 percentage points of the average prepayment rate. Post-origination and pre-origination

house price appreciation contributed negatively to prepayment rates: 4 and 3.4 percentage points, respectively. A somewhat important factor was the CLTV ratio prevailing in the market. In 2001, its value at origination contributed to a 1.2-percentage-point larger probability of prepayment two years later.

The value of the prevailing mortgage interest rate for loans that originated in 2002 was again the most important contributor to explaining prepayment rates. However, the impact of this factor (see Table 3, column 2) is much smaller compared with its effect on loans that originated in 2001. The important contribution of post-origination house price appreciation is no longer present, as it was with the 2001 vintage loans, and the contribution of the CLTV ratio has decreased.

For 2003 and 2004 vintage loans, the primary contributing factor to high prepayment rates was the house price appreciation that took place between the origination period and the subsequent

Table 2**Annual Factor Contribution to Mortgage Loan Default (2001-06)**

Explanatory factor	2001	2002	2003	2004	2005	2006
FICO credit score	1.03	0.52	-0.15	-0.14	-0.32	-0.09
If full documentation is provided (dummy)	-0.31	-0.11	-0.03	0.02	0.10	0.13
If prepayment penalty is present (dummy)	0.05	0.04	0.01	0.00	-0.01	-0.03
Debt-to-income ratio (back end)	-0.18	-0.23	-0.02	0.06	-0.01	0.30
If debt-to-income ratio is not provided (dummy)	0.10	0.17	-0.01	-0.06	0.03	-0.19
Mortgage interest rate	2.77	1.07	-0.40	-0.87	-0.58	0.66
If an investor (dummy)	0.00	0.00	0.00	0.00	0.00	0.00
If a mortgage is for refinancing at origination (dummy)	-0.09	-0.05	-0.07	-0.03	0.06	0.10
Origination amount	-0.30	-0.20	-0.07	0.01	0.14	0.19
Combined loan-to-value ratio	-0.96	-0.79	-0.31	0.10	0.46	0.69
Margin for hybrid loans	-0.17	0.01	-0.12	0.02	0.08	0.09
If a hybrid (dummy)	-0.07	-0.01	-0.04	0.04	0.05	-0.08
If an ARM (dummy)	0.00	0.00	0.00	0.00	0.00	0.00
If a balloon (dummy)	0.04	-0.06	-0.09	-0.11	-0.02	0.40
Post-origination house price appreciation	2.07	-0.20	-3.00	-2.22	2.63	7.51
Pre-origination house price appreciation	0.23	0.22	0.22	0.05	-0.19	-0.20

NOTE: The annual factor contribution is

$$AFC_i^Y = \Phi(\beta' \bar{X} + \beta_i(\bar{X}_i^{\text{year } Y} - \bar{X}_i)) - \Phi(\beta' \bar{X}),$$

where for each year Y , the impact of each explanatory variable i (first column) is calculated as the difference between the logit function Φ , where, for one variable i , the overall mean is substituted by its mean value in year Y (all other variables remain at their overall mean values) and the logit function where all variables are at their overall mean values.

$\Phi(\beta' \bar{X})$ is the likelihood that default will occur within two years of mortgage loan origination, $\Phi(\cdot)$ is the logistic function, X is the vector of explanatory variables, and β is the vector of regression coefficients.

A mortgage loan is considered in default if a borrower has defaulted on a loan or has missed more than two mortgage payments, property is in the process of foreclosure, or is real-estate owned (i.e., is likely to default).

two years. For 2003 vintage loans, a diminishing factor was the pre-origination house price appreciation, which contributed to the decline in the prepayment rates. For the 2004 vintage loans, the mortgage interest rate also diminished prepayment incentives for subprime borrowers.

For 2005 and 2006 vintage loans, the sole contributing factor for the prepayment and refinance rate, again, was house price appreciation. However, because the housing market slowdown reversed the trend and house prices depreciated, the contribution was of the opposite sign compared with earlier years. With all other factors equal, pre-origination house price appreciation contributed positively, tending to increase refi-

nance rates; however, post-origination housing values declined and the lower refinance rates prevailed. In other words, the door to refinancing opportunity was closed by declining housing prices and refinancing was largely overtaken by defaults in the termination rates of subprime mortgages.

Quick Exits

Surprisingly, almost every other loan exited the subprime market (in one way or another) within two years of origination. Moreover, just 30 to 40 percent of all subprime loans in the sample were purchase-money (used to purchase rather

Table 3**Annual Factor Contribution to Mortgage Loan Prepayment (2001-06)**

Explanatory factor	2001	2002	2003	2004	2005	2006
FICO credit score	-0.05	-0.02	0.01	0.01	0.01	0.00
If full documentation is provided (dummy)	0.09	0.03	0.01	0.00	-0.03	-0.03
If prepayment penalty is present (dummy)	-0.53	-0.39	-0.14	0.00	0.12	0.28
Debt-to-income ratio (back end)	-0.25	-0.30	-0.03	0.08	-0.01	0.39
If debt-to-income ratio is not provided (dummy)	0.12	0.20	-0.01	-0.07	0.04	-0.24
Mortgage interest rate	6.26	2.57	-1.02	-2.24	-1.48	1.60
If an investor (dummy)	0.00	0.01	0.00	0.00	0.00	0.00
If a mortgage is for refinancing at origination (dummy)	0.21	0.13	0.16	0.08	-0.13	-0.24
Origination amount	-1.26	-0.82	-0.31	0.05	0.56	0.77
Combined loan-to-value ratio	1.21	0.99	0.37	-0.12	-0.54	-0.79
Margin for hybrid loans	-0.09	0.01	-0.07	0.01	0.04	0.05
If a hybrid (dummy)	-1.10	-0.19	-0.59	0.65	0.78	-1.28
If an ARM (dummy)	0.02	0.01	-0.04	-0.02	0.01	0.00
If a balloon (dummy)	0.05	-0.08	-0.14	-0.16	-0.02	0.57
Post-origination house price appreciation	-4.12	0.45	8.32	5.79	-5.09	-11.66
Pre-origination house price appreciation	-3.39	-3.27	-3.30	-0.81	2.88	3.15

NOTE: See first note to Table 2.

A mortgage loan is considered "prepaid" if a borrower has either paid off or refinanced a mortgage loan within 2 years of origination.

than refinance a house). The remaining borrowers refinanced their existing homes, and refinances do not contribute to an increase in homeownership.

Jaffee (2008) summarized research that analyzed what went wrong with the subprime market that could cause the crisis and what went right—potential benefits from subprime lending that might offset consequences of the subprime crisis. Jaffee calculated that the subprime mortgage market funded approximately 5 million home purchases between 2000 and 2006, with slightly more than 1 million loans to first-time homebuyers. Jaffee suggests that the subprime mortgage market had at least one benefit to the economy: the increase in homeownership.

However, as shown in Figure 1, for all purchase-money mortgage loans originated between 2001 and 2006, between 15 and 25 percent were terminated in the first year, about 50 percent in the first 2 years, and 80 percent in the

first three years. For all origination years, of only first-lien, home-purchase (purchase-money) mortgages that were securitized and for which reliable data were provided, more than 600,000 loans were terminated within the first year after origination. Within two years, approximately 1.9 million loans were terminated. Among the terminated loans, about 1 million were seriously delinquent or in default; the remaining million were refinanced or prepaid. For subprime mortgages, the data seem to suggest that the number of foreclosed homes, with mortgages funding the home purchases, already *exceeds* the estimated number of first-time homebuyers with subprime mortgages.

The number of prepaid and refinanced properties is less informative because the data do not provide the after-prepayment outcome of the mortgages. A refinanced loan can be either a new subprime loan that follows the original path described above (a borrower would either default

or prepay again) or a prime loan (which borrowers can also default on or prepay). Given the degree of uncertainty on this issue, no inference based on the number of prepaid loans is made here.

Even if borrowers refinanced their initial subprime loans into more stable subprime or prime mortgages (those observed in the data before prepayment or refinance), the 80 percent termination rate within the first three years after origination would indicate that the initial boom in subprime lending could have, at most, *accelerated* growth of homeownership, even if temporarily. In other words, in a hypothetical “success” example, if a borrower took out a subprime loan in 2001, say as a first-time homebuyer, and then refinanced into a better loan in 2004, the same borrower most likely could have skipped the subprime step and become a first-time homebuyer in 2004, starting with a more stable loan and avoiding high interest rate payments and prepayment penalties. Given the impossibility of knowing when any first-time homebuyer who used a subprime mortgage would have become a homeowner with a prime loan, if ever, the data do not support the argument that subprime mortgages increased homeownership.

Given that the percentages of terminated loans in the sample are almost the same for all loan vintages (origination years), one can infer that subprime loans rarely were expected or intended to last much longer than three years. Lenders must have known that these loans were temporary (i.e., it would be impossible to collect sufficient interest payments to cover loan origination costs). Therefore, prepayment penalties were imposed, high interest rates and fees were charged, and complicated loan modifications were designed. (As well, the securitization structure is very complex, rendering individual loan modifications almost impossible.) In addition, borrowers must have been planning to use subprime mortgages for so-called bridge financing. If subprime borrowers were planning a quick exit from the very beginning, then these loans were risky not only from a credit-risk perspective but also from the standpoint of interest rate risk (would rates go up?) and liquidity risk (would there be a possibility to refinance?). Given these risks, lenders and

investors could experience much higher losses than expected purely on the basis of credit risk. In hindsight, we know that the risks did materialize and the losses did skyrocket.

CONCLUSION

The subprime mortgage crisis of 2007 resulted in a massive wave of foreclosures and serious delinquencies, a large proportion of which consisted of mortgages originated in 2006 and 2007. Much of the debate among researchers and policymakers involves causes, consequences, and remedies for these early defaults and foreclosures. Still unexplained, however, is the temporary nature of subprime loans. This study shows that loans that originated in any year from 2001 to 2006 generally had a life of less than three years. In fact, almost half of these loans exited the market through either prepayment or default within the first two years after origination; about 80 percent of them did so within three years.

Even though mortgage termination rates have been remarkably similar for all origination years evaluated one, two, or three years after origination, the split between default and prepayment rates varied. There is a J shape in the graphed representation of defaults for origination years 2001 to 2006. The trough of the pattern corresponds to the years 2003 and 2004, when the housing market was booming. When default rates are small, refinancing rates are high. When the trend in the housing market reversed, refinancing became impossible and defaults took their place.

The evidence in this paper is consistent with that reported by Demyanyk and van Hemert (2008), who explain that the crisis—the unusually high default rates among 2006 and 2007 vintage loans—did not occur because these loans were in some respects much worse than all loans that originated earlier. Subprime mortgages were very risky all along; however, their true riskiness was hidden by rapid house price appreciation, allowing mortgage termination by refinancing/prepayment to take place. When prepayment became very costly (with zero or negative equity in the house increasing the closing costs of a refinancing), defaults took their place.

Demyanyk

The results in this paper also suggest that subprime lending did not increase homeownership: The number of defaults in a limited sample (about 50 percent) of subprime purchase-money mortgages within two years of origination is almost equal to the estimated number of first-time homebuyers who took subprime mortgages. If the data for the rest of the market were available, the number of defaults would no doubt be even greater.

Several questions remain and require further attention. First, the available data do not help identify what happened to loans that were terminated but did not end in default (i.e., prepaid or refinanced loans). Mortgages originated for refinancing tend to be refinanced again within a couple of years and tend to default as well. If more comprehensive data become available, further analysis on the homeownership policy discussion may be fruitful. Foote et al. (2008) raise the same question and explain the difficulty in answering it. Second, several studies indicate that most of the materialized risks associated with subprime mortgage lending had been neither observable nor measurable (e.g., the credit score did not predict likelihood of default; see Demyanyk and Van Hemert, 2008, and Haughwout, Peach, and Tracy, 2008). Little is known about these risks except that they existed and increased over time. More sophisticated models and comprehensive data are needed to answer these questions.

REFERENCES

- Cutts, Amy C. and Van Order, Robert A. "On the Economics of Subprime Lending." *Journal of Real Estate Finance and Economics*, March 2005, 30(2), pp. 167-96.
- Dell'Ariccia, Giovanni; Igan, Deniz and Laeven, Luc. "Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market." Working Paper IMF WP/08/106, International Monetary Fund, April 2008; www.imf.org/external/pubs/ft/wp/2008/wp08106.pdf.
- Demyanyk, Yuliya. "Did Credit Scores Predict the Subprime Crisis?" Federal Reserve Bank of St. Louis *Regional Economist*, October 2008, pp. 12-13; <http://stlouisfed.org/publications/re/2008/d/pages/mortgage.html>.
- Demyanyk, Yuliya and Gopalan, Yadav. "Subprime ARMs: Popular Loans, Poor Performance." Federal Reserve Bank of St. Louis *Bridges*, Spring 2007, pp. 4-5; <http://stlouisfed.org/publications/br/2007/a/pages/2-article.html>.
- Demyanyk, Yuliya and Van Hemert, Otto. "Understanding the Subprime Mortgage Crisis." *Review of Financial Studies*, forthcoming; <http://ssrn.com/abstract=1020396> (posted December 5, 2008).
- Deng, Yongheng; Quigley, John M. and Van Order, Robert A. "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options." *Econometrica*, March 2000, 68(2), pp. 275-308.
- Foote, Christopher L.; Gerardi, Kristopher; Goette, Lorenz F. and Willen, Paul. "Subprime Facts: What (We Think) We Know about the Subprime Crisis and What We Don't." Public Policy Discussion Paper No. 08-2, Federal Reserve Bank of Boston, May 30, 2008; www.bos.frb.org/economic/ppdp/2008/ppdp0802.pdf.
- Foote, Christopher; Gerardi, Kristopher and Willen, Paul. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics*, September 2008, 64(2), pp. 234-45.
- Gerardi, Kristopher; Shapiro, Adam and Willen, Paul. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures." Working Paper 07-15, Federal Reserve Bank of Boston, May 4, 2008; www.bos.frb.org/economic/wp/wp2007/wp0715.pdf.
- Haughwout, Andrew; Peach, Richard and Tracy, Joseph. "Juvenile Delinquent Mortgages: Bad Credit or Bad Economy?" Staff Report No. 341, Federal Reserve Bank of New York, August 2008; www.newyorkfed.org/research/staff_reports/sr341.pdf.
- Inside Mortgage Finance Publications. *Mortgage Yearbook First Half 2008 Special Edition*. Bethesda, MD: Inside Mortgage Finance Publications, 2008.

Jaffee, Dwight, M. "The U.S. Subprime Mortgage Crisis: Issues Raised and Lessons Learned." Working Paper No. 28, Commission on Growth and Development and the World Bank, 2008; <http://faculty.haas.berkeley.edu/jaffee/Papers/DJSubprimeMay5.pdf>.

Keys, Benjamin J.; Mukherjee, Tanmoy; Seru, Amit and Vig, Vikrant. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." Athens Meetings Paper, European Finance Association, December 2008; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1093137.

Mayer, Christopher and Pence, Karen. "Subprime Mortgages: What, Where, and to Whom?" Finance and Economics Discussion Series Working Paper 2008-29, Federal Reserve Board, December 7, 2007; www.federalreserve.gov/pubs/feds/2008/200829/200829pap.pdf.

Mian, Atif and Sufi, Amir. "The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis." NBER Working Paper 13936, National Bureau of Economic Research, April 2008; www.nber.org/papers/w13936.

Pennington-Cross, Anthony and Chomsisengphet, Souphala. "Subprime Refinancing: Equity Extraction and Mortgage Termination." *Real Estate Economics*, 2007, 35(2), pp. 233-63.

Von Furstenberg, George M. and Green, R. Jeffery. "Home Mortgage Delinquencies: A Cohort Analysis." *Journal of Finance*, December 1974, 29(5), pp. 1545-48.

Von Furstenberg, George M. "Default Risk on FHA-Insured Home Mortgages." *Journal of Finance*, June 1969, 24(3), pp. 459-77.

Von Furstenberg, George M. "Risk Structures and the Distribution of Benefits within the FHA Mortgage Insurance Program," *Journal of Money, Credit, and Banking*, August 1970, 2(3), pp. 303-22.



Firm Volatility and Credit: A Macroeconomic Analysis

Leo Kaas

This paper examines a tractable real business cycle model with idiosyncratic productivity shocks and binding credit constraints on entrepreneurs. The model shows how firm volatility increases in combination with credit market development. It further generates the observed comovement of credit and firm volatility with output at business cycle frequencies in response to aggregate productivity shocks. (JEL E32, E44, O16)

Federal Reserve Bank of St Louis *Review*, March/April 2009, 91(2), pp. 95-106.

Parallel to the decline in macroeconomic volatility over the past decades (see Blanchard and Simon, 2001, and Stock and Watson, 2002), there is some evidence that volatility at the firm level has increased during the same period. For the United States, such evidence is available for idiosyncratic stock returns (Campbell et al., 2001), as well as for employment, sales, and investment. Comin and Mulani (2004) and Comin and Philippon (2005) document similar results for other countries.¹

There are different explanations for an increase in firm volatility. One is that deregulation and intensified global competition force firms to adjust prices and business strategies faster. Another is that financial development leads to more risk-taking by entrepreneurs or facilitates leverage, which both could potentially drive up firm volatility. Indeed, Comin and Philippon (2005) find some support for both hypotheses. They also show that the increase in firm volatility is driven neither by the growing share of small firms in the sample nor

changes in firm ownership, including merger and acquisition activities.

This paper puts the link between financial development and firm volatility in a macroeconomic perspective. To this end, I develop a tractable real business cycle model with idiosyncratic productivity shocks and collateral-based borrowing constraints. Productive entrepreneurs borrow up to the value of their collateral. Because their capital return exceeds the capital cost, the leveraged return on equity exceeds the equity return of less-productive entrepreneurs. An increase in credit market development relaxes borrowing constraint and increases leverage, and thereby also the spread between internal rates of return across firms. As a result, firm growth rates become more volatile.

Another implication of my model is that both credit market development and firm volatility respond positively to an aggregate productivity shock. Higher productivity raises the value of pledgeable assets, thus softening credit constraints and leverage. Hence, both the volume of firm credit and firm-level volatility are procyclical. In the following section, I demonstrate that such

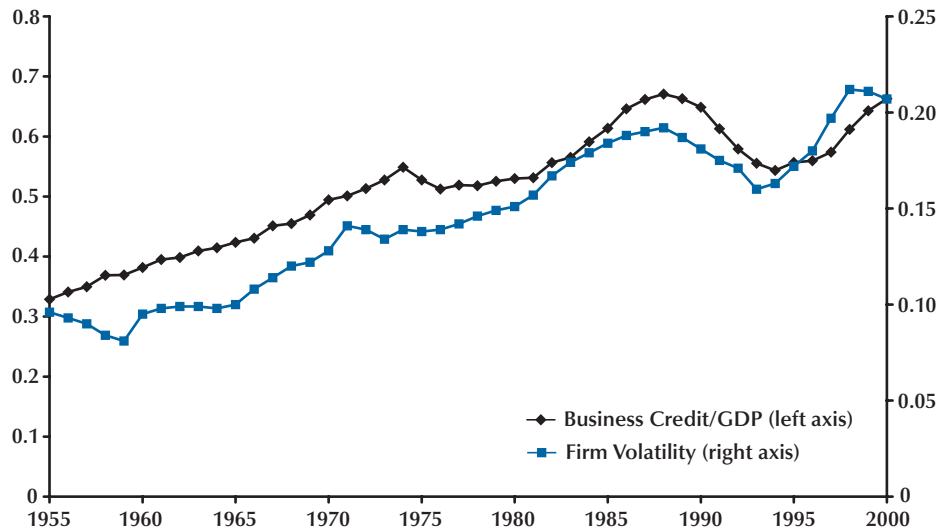
¹ Davis et al. (2006) demonstrate, however, that firm-level employment volatility has increased only for publicly traded firms and not for privately held firms.

Leo Kaas is a professor of economics at the University of Konstanz and was a visiting scholar at the Federal Reserve Bank of St. Louis when this paper was written. The author thanks Costas Azariadis and Carlos Garriga for helpful comments and the German Research Foundation for financial support (grant No. KA 1519/3).

© 2009, The Federal Reserve Bank of St. Louis. The views expressed in this article are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

Figure 1

Business Credit (Share of GDP) and Firm Volatility (Annual U.S. Data 1955-2000)



procyclicality is indeed observable in postwar U.S. data. In the quantitative section, I match the model to the U.S. business cycle and show that it replicates reasonably well the comovements among the three key variables of output, credit, and firm volatility. However, the amplification of firm volatility is twice as large as in the data, and its cross-correlations with output and credit are somewhat too low. The model can also be used to investigate the effects of a financial crisis; in particular, a severe crisis where collateral value drops temporarily by 20 percent features a decline of gross domestic product (GDP) below trend by 3.5 percent.

By adopting collateral-based borrowing constraints in combination with logarithmic utility and Cobb-Douglas production technologies, my model is essentially a variation of the approach of Kiyotaki (1998) and Kiyotaki and Moore (2008), who also develop tractable business cycle models with binding credit constraints. Other theoretical contributions on idiosyncratic production risk and finance constraints in dynamic equilibrium models are those of Hopenhayn and Vereshchagina (2003) and Meh and Quadrini (2006). But while

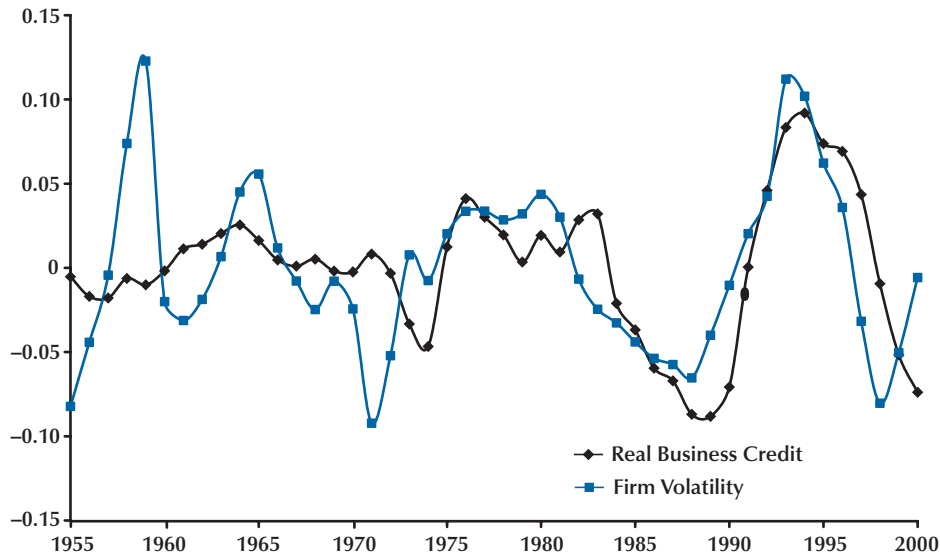
they examine risk-taking in incomplete-market environments, the effect of borrowing constraints on firm leverage is the driving force of this paper. Further, the model of this paper has closed-form solutions, which make its basic mechanics particularly clear.

THE EVIDENCE

For the purpose of this paper, the appropriate measure of credit market development is the share of business credit in GDP, where “business credit” includes all credit market debt owed by nonfinancial firms, including corporations and noncorporations.² Figure 1 illustrates the substantial financial deepening during the period 1955 to 2000; as a share of GDP, business credit roughly doubled. Real business credit actually increased by a factor of 8.9.³

² See Board of Governors of the Federal Reserve (2008).

³ Real business credit is defined as business credit divided by the GDP deflator. Over the same horizon, by comparison, real household debt increased by a factor of 9.5, real government debt by a factor of 4.9, and real credit market debt of the financial sector by a factor of 99.5.

Figure 2**Detrended Real Business Credit and Firm Volatility (Annual U.S. Data 1955-2000)**

NOTE: Both variables are reported in logs as deviations from a Hodrick-Prescott trend with smoothing parameter 100.

Figure 1 also shows the increase in firm volatility during the same period. I use the “median sales volatility” reported in Comin and Philippon (2005, table 1) as a volatility measure. Specifically, Comin and Philippon calculate a rolling standard deviation (SD) of sales growth for nearly every firm in the Compustat database; the median of the cross section then measures firm volatility at every point in time. It becomes evident from the figure that firm volatility also doubled between 1955 and 2000. Of note, firm volatility is quite dispersed in the cross section. For example, in the 1990s, sales growth volatility was below 0.1 for 25 percent of firms and above 0.3 for another 25 percent of firms (see Figure 2 in Comin and Philippon, 2005).

Figure 1 further suggests that business credit and firm volatility are positively correlated at the business cycle frequency. To see this more clearly, note that Figure 2 shows the detrended time series of real business credit and of firm volatility, where the trend is a Hodrick-Prescott (H-P) filter with $\lambda = 100$ and the cyclical components are reported as log deviations from trend. Particularly since

Table 1**Descriptive Statistics (Annual U.S. Data 1955-2000)**

	Output	Credit	Volatility
Standard deviation	0.021	0.042	0.050
Annual autocorrelation	0.546	0.826	0.675
Correlation matrix			
Output	1	0.460	0.257
Credit	—	1	0.567
Volatility	—	—	1

NOTE: All variables are reported in logs as deviations from an H-P trend with smoothing parameter 100.

the 1970s, the two cycles are closely synchronized and the percentage deviations from trend are similar in magnitude.

Table 1 summarizes the detrended data of output, real business credit, and firm volatility. Both firm volatility and credit have higher variance than output, and they are positively corre-

lated with contemporaneous output, where the correlation between credit and output is stronger than the one between volatility and output. The correlation coefficient between credit and volatility increases from 0.567 to 0.775 when the period is restricted to the years 1975 to 2000.

THE MODEL

Consider a one-sector growth model with infinitely lived entrepreneurs and workers in which either group is a continuum of mass one. All agents derive logarithmic utility from consumption and discount future utility with the factor $\beta < 1$. All workers supply one unit of labor inelastically. Entrepreneurs do not supply labor; they employ workers and capital to produce output with a Cobb-Douglas technology, which is subject to idiosyncratic productivity shocks. Specifically, entrepreneur i in period t uses capital K_t^i and labor L_t^i to produce output $Y_t^i = A_t^i (K_t^i)^\alpha (L_t^i)^{1-\alpha}$, where the entrepreneur's productivity attains the high level $A_t^i = A_t$ (productive state) with probability π and the lower level $A_t^i = B < A$ (unproductive state) otherwise. These productivity realizations are independent across time and across entrepreneurs. Thus, a fraction π of entrepreneurs is productive in every period. The assumption that productivity states are independent of history simplifies the exposition but can easily be generalized to allow for autocorrelated productivity states. The only complication is that the model must then be augmented by another state variable, which is the share of wealth in the hands of productive entrepreneurs.

Factor productivity at the technology frontier A_t is subject to aggregate productivity shocks. In particular, $\ln A_t$ follows an AR(1) process with coefficient $\rho < 1$, mean $\ln \bar{A}$, and normally distributed shocks with SD σ . The assumption that only the frontier A fluctuates while the inferior technology parameter B is fixed again simplifies the exposition and can be generalized. What is crucial for the results, though, is that B fluctuates less than proportionately with productivity at the frontier.

Output produced in period t becomes available

for entrepreneurs' investment and consumption purposes in the next period. To obtain closed-form solutions, capital fully depreciates within every period. Equivalently, Y_t^i can be interpreted to include both output and undepreciated capital. In the calibration exercise, I use this interpretation and choose the capital share parameter α accordingly.

Each period, all agents have access to a capital market where they can borrow and lend at gross interest rate R_t . Borrowing can be against collateral only, however. Because labor income cannot be collateralized, workers are not permitted to borrow. Further, I show that in any steady state with constrained entrepreneurs, $R < 1/\beta$ holds, which implies that in any stochastic equilibrium near the steady state, workers do not save; hence, workers simply consume their wage income in every period.

Entrepreneurs, in turn, can pledge a fraction $\lambda < 1$ of their output, where the "collateral share" parameter λ plausibly depends both on technological features (e.g., what part of capital is alienable) and on the institutional framework and market environment (e.g., creditors' rights and availability of credit market instruments). Every entrepreneur's principal and interest on debt D_t^i may not exceed the value of collateral. That is, the credit constraint takes the form $R_t D_t^i \leq \lambda Y_t^i$. Credit repayments occur at the beginning of the next period before realization of the next period's productivity.

For any realization of technology shocks $(A_t)_{t \geq 0}$, a *competitive equilibrium* is a list of consumption plans, production plans, and debt positions $(C_t^i, K_t^i, L_t^i, D_t^i)$ for every entrepreneur; consumption plans and debt positions for workers (C_t^w, D_t^w) ; and factor prices for labor and capital (w_t, R_t) such that in every period $t \geq 0$ ⁴:

- (i) $C_t^i, K_t^i, L_t^i, D_t^i$ maximizes entrepreneur i 's expected utility subject to budget and debt constraints; that is, it solves

⁴ In the initial period, $t = 0$, all debt positions are assumed zero, and there is some given distribution of wealth across entrepreneurs.

$$\begin{aligned} & \max E_t \sum_{s \geq t} \beta^{s-t} \ln C_s^i \quad \text{s.t.} \\ & C_s^i + K_s^i - D_s^i = A_{s-1}^i (K_{s-1}^i)^\alpha (L_{s-1}^i)^{1-\alpha} \\ & -w_{s-1} L_{s-1}^i - R_{s-1} D_{s-1}^i, \quad s \geq t, \\ & R_s D_s^i \leq \lambda A_s^i (K_s^i)^\alpha (L_s^i)^{1-\alpha}, \quad s \geq t. \end{aligned}$$

- (ii) C_t^w, D_t^w maximizes workers' expected utility subject to budget and zero debt constraints; that is; it solves

$$\begin{aligned} & \max E_t \sum_{s \geq t} \beta^{s-t} \ln C_s^w \quad \text{s.t.} \\ & C_s^w - D_s^w = w_s - R_{s-1} D_{s-1}^w, \quad s \geq t, \\ & D_s^w \leq 0, \quad s \geq t. \end{aligned}$$

- (iii) Markets for labor and capital clear:

$$\begin{aligned} & \int_0^1 L_t^i di = 1, \\ & \int_0^1 D_t^i di + D_t^w = 0. \end{aligned}$$

The appendix characterizes the solutions to the agents' utility maximization problems. Particularly in the neighborhood of a steady-state equilibrium with binding constraints, workers do not save; hence, $D_t^w = 0$ for all $t \geq 0$. Further, all entrepreneurs save a constant fraction β of their wealth.

Before discussing an equilibrium with binding debt constraints, it is instructive to see how the economy acts when the collateral value λ is large enough. In every period, then, all capital flows to productive entrepreneurs who also hire the total workforce. Because β is the entrepreneurs' savings rate and because total entrepreneur wealth is share α of output, the aggregate capital stock evolves according to $K_{t+1} = \beta \alpha A_t K_t^\alpha$. The model's dynamics thus resemble those in the standard real business cycle model with logarithmic utility, Cobb-Douglas production, and full depreciation.

The following section characterizes equilibrium when productive entrepreneurs are credit constrained and unproductive entrepreneurs do not lend all their capital but also produce. Hence, production is inefficient and the steady-state out-

put level is below the one in the first-best economy. A parameter restriction explained below will ensure that such an equilibrium exists.

EQUILIBRIUM

As is shown in the appendix, all entrepreneurs' capital investments are linear in their equity. Hence, aggregation over entrepreneurs with identical technologies is straightforward, and I write

$$K_t^A = \int_{i:A_t^i=A_t} K_t^i di \quad \text{and} \quad K_t^B = \int_{i:A_t^i=B} K_t^i di$$

to denote aggregate capital investment of productive and unproductive entrepreneurs. L_t^A and L_t^B are similarly defined, and the absence of a super-index indicates an aggregate across all entrepreneurs. Let $k_t^s = K_t^s/L_t^s$, $s = A, B$, denote capital intensities for the two types of entrepreneurs.

Because labor moves freely between employers, the real wage is

$$w_t = A_t (1-\alpha) (k_t^A)^\alpha = B_t (1-\alpha) (k_t^B)^\alpha,$$

which implies that

$$(1) \quad k_t^B = \varphi_t k_t^A \quad \text{with} \quad \varphi_t \equiv (A_t/B_t)^{1/\alpha} > 1.$$

Because labor is perfectly mobile and capital is not, unproductive entrepreneurs operate their technology with a higher capital intensity than productive entrepreneurs. The labor and capital markets are in equilibrium if

$$(2) \quad L_t^A + L_t^B = 1,$$

$$(3) \quad L_t^A k_t^A + L_t^B k_t^B = K_t.$$

Let D_t denote total borrowing of productive entrepreneurs, which equals total lending of unproductive entrepreneurs because workers do not participate in the credit market. Because productive entrepreneurs own πK_t units of the capital stock, their capital input is the sum of equity and debt:

$$(4) \quad L_t^A k_t^A = \pi K_t + D_t.$$

Because the credit constraint binds on each productive entrepreneur, it also holds with equality in the aggregate:

$$(5) \quad R_t D_t = \lambda A_t (k_t^A)^\alpha L_t^A.$$

Unproductive entrepreneurs are indifferent between lending capital at gross return R_t or producing themselves, which leads to the following arbitrage condition:

$$(6) \quad R_t = \alpha B (k_t^B)^{\alpha-1}.$$

From equations (1), (5), and (6), it follows that borrowing is proportional to investment:

$$(7) \quad D_t = \frac{\lambda A_t}{\alpha B} (k_t^A)^\alpha (k_t^B)^{1-\alpha} L_t^A = \frac{\lambda \varphi_t}{\alpha} k_t^A L_t^A;$$

and substitution into equation (4) yields

$$(8) \quad D_t = \frac{\pi \lambda \varphi_t}{\alpha - \lambda \varphi_t} K_t.$$

To ensure that unproductive entrepreneurs produce, their lending may not exceed their capital holdings; that is, D_t must be strictly smaller than $(1 - \pi)K_t$. Together with equation (8), this necessitates $\lambda \varphi_t < (1 - \pi)\alpha$. Because A_t fluctuates around \bar{A} , φ_t fluctuates around $\bar{\varphi} \equiv (\bar{A}/B)^{1/\alpha}$. To guarantee a production-inefficient equilibrium in the neighborhood of the steady state, it must therefore be assumed that

$$(9) \quad \lambda \bar{\varphi} < (1 - \pi)\alpha.$$

The explanation of this condition is as follows. If the collateral share were too large, productive agents would borrow all resources from their unproductive counterparts and production would be efficient. The same would apply if either φ or π were too large: With a large productivity spread, production becomes less attractive than lending for unproductive agents, and a large share of borrowers raises credit above the funds supplied by lenders. Similarly, a too-low capital share would depress the interest rate, driving up the demand for credit above lenders' resources.

Now equations (1), (2), (3), (4), and (7) can be solved for the capital intensity of productive entrepreneurs as follows:

$$k_t^A = C_t K_t \text{ with } C_t \equiv \frac{\alpha - \lambda \varphi_t - \alpha \pi + \alpha \pi \varphi_t}{\varphi_t (\alpha - \lambda \varphi_t)} < 1,$$

where $C_t < 1$ follows from condition (9) when φ_t is close to its steady-state value, so that C_t is close to its steady-state level, $\bar{C} < 1$. This also implies that $k_t^B = \varphi_t C_t K_t > K_t$. Employment is allocated according to

$$L_t^A = \frac{\alpha \pi}{C_t (\alpha - \lambda \varphi_t)} \text{ and } L_t^B = \frac{\alpha (C_t - \pi) - C_t \lambda \varphi_t}{C_t (\alpha - \lambda \varphi_t)},$$

so aggregate output is

$$Y_t = A_t (k_t^A)^\alpha L_t^A + B (k_t^B)^\alpha L_t^B = A_t C_t^\alpha K_t^\alpha,$$

with total factor productivity $A_t C_t^\alpha < A_t$. Because workers earn share $1 - \alpha$ of output and do not save and because all entrepreneurs save share β of their wealth, the aggregate saving rate is $\alpha \beta$. Hence, the capital stock evolves according to

$$K_{t+1} = \alpha \beta A_t C_t^\alpha K_t^\alpha.$$

In the absence of technology shocks, the capital stock converges to its steady-state level:

$$\bar{K} = (\alpha \beta \bar{A} \bar{C}^\alpha)^{1/(1-\alpha)}.$$

Note that the steady-state interest rate is $\bar{R} = 1/(\bar{\varphi} \bar{C} \beta) < 1/\beta$; hence, workers indeed do not save when entrepreneurs are credit constrained.

The steady-state credit share in output is calculated as

$$(10) \quad \frac{D}{Y} = \frac{\alpha \beta D}{K} = \frac{\alpha \beta \pi \lambda \bar{\varphi}}{\alpha - \lambda \bar{\varphi}}.$$

An unproductive entrepreneur's capital grows at rate $\beta \bar{R}$, whereas a productive entrepreneur's capital grows at $\beta \tilde{R} > \beta \bar{R}$, where

$$\tilde{R} = A (k^A)^{\alpha-1} \frac{\alpha (\alpha - \lambda)}{\alpha - \lambda \bar{\varphi}}$$

is the return on equity.⁵ Therefore, the SD of a firm's growth rate in steady state is

$$(11) \quad \sigma = \sqrt{\pi(1-\pi)}\beta(\tilde{R}-\bar{R}) = \frac{\alpha(\bar{\varphi}-1)}{\alpha-\lambda\bar{\varphi}-\alpha\pi+\alpha\pi\bar{\varphi}}.$$

The closed-form expressions (10) and (11) capture the central message of this paper. On the one hand, a rise in λ describes the effect of credit market deepening in this model: When firms are able to pledge more of their assets as collateral, the share of credit in total output expands as shown in expression (10). In tandem with the credit expansion comes a higher firm volatility, as evidenced by equation (11). Relaxed credit limits spur leverage, widening the gap between firm growth rates, $\beta(\tilde{R}-\bar{R})$. On the other hand, a positive technology shock triggers a rise in credit and in firm volatility. The increase of A (relative to B) raises φ , which unambiguously increases D/Y and σ (which is again an implication of inequality (9)). Intuitively, a positive productivity shock boosts the value of collateral and thus the volume of credit. Notably, credit rises more than one-for-one with output, so the share of credit in output also increases. Additionally, the positive technology shock stimulates leverage, which enlarges the spread between firm growth rates, increasing firm volatility.

QUANTITATIVE ANALYSIS

This section explores the quantitative properties of the qualitative results obtained in the previous section: How well does this model explain the observed dynamics of output, business credit, and firm volatility? To calibrate the steady state, I first choose the following five parameters: the capital share, α ; the discount factor, β ; the collateral share, λ ; the mean spread between technologies, $\bar{\varphi}$; and the share of productive entrepreneurs, π . The technology level \bar{A} (and thus B) merely shifts the level of aggregate output and capital but has no impact on the capital-to-output ratio

or on any other relevant economic variables. Therefore, I normalize $\bar{A} = 1$.

Because there is no depreciation in this model, I adjust the capital share to include the value of the undepreciated capital stock. In the following, the term “wealth” refers to GDP plus undepreciated capital, $Y = \text{GDP} + (1 - \delta)K$. With an annual capital-to-GDP ratio of $K/\text{GDP} = 2.7$ and a 5 percent depreciation rate, the wealth-to-GDP ratio is 3.57. With capital income in GDP at one-third, it follows that the capital share in wealth is $\alpha = [0.33 + (0.95 \cdot 2.7)]/3.57 \approx 0.81$. Further, in steady state, $\alpha\beta = K/Y = 2.7/3.57$, which yields $\beta = 0.938$. I choose the collateral share, λ , to match a share of business credit in GDP of around 0.55, the average over the period 1955-2000. As equation (10) gives the credit-to-wealth ratio, the right-hand side of this equation must be equalized to $0.55/3.57$. Given the above choices for α and β , and for any choice of $\bar{\varphi}$ and π , λ is chosen to satisfy this equation. The remaining parameters π and $\bar{\varphi}$ are chosen to match the following two targets: a 3 percent real interest rate ($\bar{R} = 1.03$) and a value of firm volatility (measured by the SD of firm growth) of around 0.14, the average of median firm volatility during 1955-2000. This yields $\bar{\varphi} = 1.13$ and $\pi = 0.08$, which in turn implies that $\lambda = 0.51$. At these parameter values, assumption (9) is satisfied by a wide enough margin. On the other hand, if λ would exceed $\alpha(1 - \pi)/\bar{\varphi} \approx 0.66$, assumption (9) would be violated, in which case all capital would be used at the technology frontier. Although productive entrepreneurs may still be credit constrained,⁶ the model behaves like a standard real business cycle model and the value of λ has no effect on aggregate output.

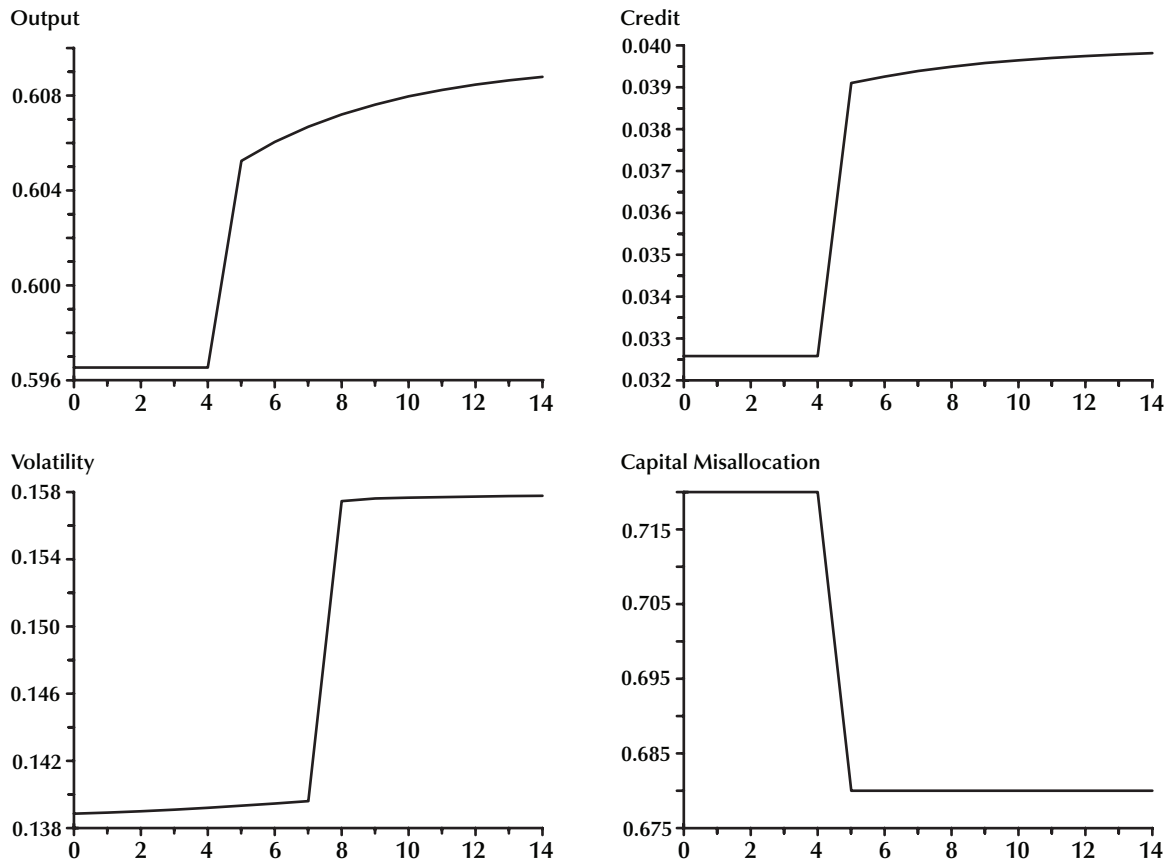
In the stochastic model of this section, I do not compute firm volatility defined over an infinite time horizon—which, in steady state, gives rise to equation (11) for every firm. Instead, I follow the procedure of Comin and Philippon (2005) to calculate rolling SDs of firm growth rates over 10-year time windows. Specifically, at each point

⁵ An entrepreneur with equity E borrows $D = \lambda\varphi E/(\alpha - \lambda\varphi)$ and invests $K^A = \alpha E/(\alpha - \lambda\varphi)$ to earn profit $\pi = A(k^A)^{\alpha-1}K^A - RD - wK^A/k^A = \alpha A(k^A)^{\alpha-1}K^A - \lambda A(k^A)^{\alpha-1}K^A = \bar{R}E$.

⁶ Precisely, when $\lambda < \alpha(1 - \pi) \approx 0.75$, the steady-state interest rate stays below the marginal product of capital of productive entrepreneurs. Although the economy would be production efficient, it is consumption inefficient because idiosyncratic volatility still matters. First-best allocations are attained only when $\lambda > \alpha(1 - \pi)$.

Figure 3

Response to a 5 Percent Permanent Increase in the Collateral Share λ in Period $t = 5$

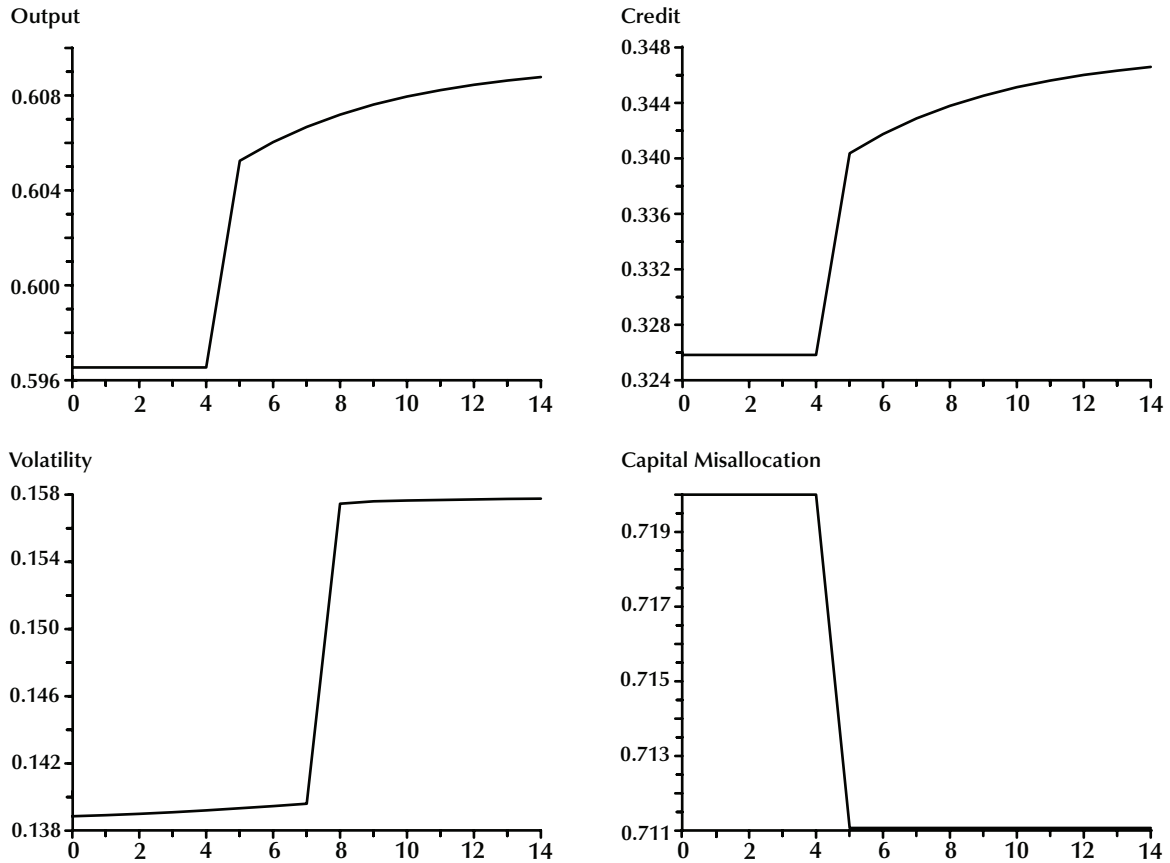


in time t , I bootstrap the distribution of these SDs from 10,000 firms drawing their growth rates from $(\beta R_{t+\tau}, \beta \tilde{R}_{t+\tau})_{\tau=-4}^5$. Then I use the median of this distribution as the volatility measure.

Figure 3 shows the model's response to a permanent increase in the collateral share λ by about 5 percent. Output increases on impact by 1.3 percent, converging to the new steady state, which is more than 2 percent higher. Credit increases by about 23 percent, and volatility increases by 15 percent. The response of volatility is sluggish; also, the response begins four periods before the shock because volatility is constructed using rolling windows that are four periods backward looking and five periods forward looking. The largest adjustment of volatility occurs three

periods after the shock. The lower-right graph shows the "capital misallocation," defined as the share of capital used by unproductive entrepreneurs. Note that only 8 percent of entrepreneurs have access to the technology frontier, but they still use 28 percent of capital when $\lambda = 0.51$. As λ increases to 0.535, 32 percent of capital is used at the technology frontier.

Whereas this experiment shows the effect of a *permanent* rise in collateral value, it also is illuminating to investigate the impact of a *temporary* decline in collateral value as a result of a severe financial crisis. To this end, suppose that the collateral share drops by 20 percent for a period of three years before it returns to its original value. I find that the impact of such a shock is a decline

Figure 4**Response to a 1 Percent Permanent Increase in Productivity A in Period $t = 5$** 

of GDP by 3 to 3.5 percent in the three years of the crisis, the largest occurring in the third year. Because the model has little amplification, output returns to 0.5 percent below its steady-state level in the first year after the crisis. During the three crisis years, credit collapses dramatically: It falls by more than 40 percent, which implies that only 20 percent of capital is used at the frontier technology. Note again that these macroeconomic effects of a decline of λ are due to the misallocation of capital; they would disappear if λ exceeded the threshold level implied by assumption (9). Public policy may attempt to prevent the adverse effects of the credit collapse to some extent, either by restoring collateral value or by the injection of liquidity—for example, by providing unsecured

credit lines. Without analyzing such policies formally, it is clear that they must be of a large scale, given the substantial decline of the credit market.

Figure 4 shows the impulse response to a permanent increase of productivity at the technology frontier by 1 percent. As the value of pledgeable assets rises, credit expands by 6 percent, and the higher leverage leads to an increase of firm volatility by about 15 percent. Output increases by 2 percent, which comes about through two effects: The first is higher productivity at the technology frontier; the second is that capital is now more efficiently allocated as capital misallocation falls from 0.72 to 0.711.

To explore the stochastic model with auto-correlated shocks,

Table 2**Simulation Results with Stochastic Productivity for 10,000 Model Periods**

	Output	Credit	Volatility
Standard deviation	0.022	0.061	0.130
Annual autocorrelation	0.534	0.542	0.643
Correlation matrix			
Output	1	0.980	0.177
Credit	—	1	0.183
Volatility	—	—	1

NOTE: All variables are reported in logs as deviations from an H-P trend with smoothing parameter 100.

$$\ln A_t = \rho \ln A_{t-1} + \sigma \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1),$$

I choose the two parameters ρ and σ to target the SD and autocorrelation of the cyclical component of the model-generated GDP time series. As explained previously, GDP in period t is the difference between Y_t and the undepreciated capital stock, which is $0.95 K_t = 0.95 \alpha \beta Y_{t-1}$. All time series again are detrended with an H-P filter with smoothing parameter 100, and the cyclical components are the log deviations from trend. The result of this exercise is that $\rho = 0.95$ and $\sigma = 0.014$ match the first two moments in Table 1 reasonably well. These and all other model-generated moments from a simulation with 10^5 periods are listed in Table 2. Relative to the data, the amplification of credit is matched reasonably well, whereas the SD of firm volatility is more than twice as large as in the data. Although all contemporaneous correlations have the right sign, the one between output and credit is larger than in the data, whereas the cross-correlations with volatility are too low.

Volatility is strongly amplified because productivity of the inferior technology stays constant. If B were to fluctuate with A according to $B_t/\bar{B} = (A_t/\bar{A})^\gamma$, the SD of firm volatility would halve in value for $\gamma = 0.5$, and it would (counterfactually) become smaller than the SD of output for $\gamma = 1$. Proper calibration of the stochastic dynamics of both A_t and B_t would require matching the time-

series properties of the mean and the SD of the cross-sectional productivity distribution, which is beyond the scope of this paper.

CONCLUSION

This paper has developed a tractable real business cycle model with idiosyncratic productivity shocks and collateral-based credit constraints. Important features of the model are that output is below the efficient level because not all capital is used at the technological frontier and that firm growth rates are volatile. The model allows derivation of closed-form expressions for the dynamics of output and capital, for the volume of credit, and the SD of firm growth rates. It accounts qualitatively for the observed simultaneous long-term increase of the credit-to-output ratio and of firm volatility. Quantitatively, the model is able to generate the correct comovement among output, credit, and firm volatility, although firm volatility is too strongly amplified.

REFERENCES

- Blanchard, Olivier J. and Simon, John A. "The Long and Large Decline in U.S. Output Volatility." *Brookings Papers on Economic Activity*, March 2001, 2001(1), pp. 135-64.
- Campbell, John Y.; Lettau, Martin; Malkiel, Burton G. and Xu, Yexiao. "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk." *Journal of Finance*, February 2001, 56(1), pp. 1-43.
- Comin, Diego and Mulani, Sunil. "Diverging Trends in Macro and Micro Volatility: Facts." NBER Working Paper 10922, National Bureau of Economic Research, November 2004; www.nber.org/papers/w10922.
- Comin, Diego and Philippon, Thomas. "The Rise in Firm-Level Volatility: Causes and Consequences." NBER Working Paper 11388. National Bureau of Economic Research, May 2005; www.nber.org/papers/w11388.

- Davis, Steven, J.; Haltiwanger, John; Jarmin, Ron and Miranda, Javier. "Volatility and Dispersion in Business Growth Rates: Publicly Traded versus Privately Held Firms." NBER Working Paper 12354, National Bureau of Economic Research; June 2006; www.nber.org/papers/w12354.
- Board of Governors of the Federal Reserve System. *Federal Reserve Statistical Release, Z.1. Flow of Funds Accounts of the United States: Flows and Outstandings Third Quarter 2008. Credit Market Debt Outstanding* (December 11, 2008). Washington, DC: Board of Governors, 2008; www.census.gov/compendia/statab/tables/09s1130.xls.
- Hopenhayn, Hugo A. and Vereshchagina, Galina. "Risk Taking by Entrepreneurs." Rochester Center for Economic Research Working Paper 500, University of Rochester, April 2003; http://rcer.econ.rochester.edu/RCERPAPERS/rcer_500.pdf.
- Kiyotaki, Nobuhiro. "Credit and Business Cycles." *Japanese Economic Review*, March 1998, 49(1), pp. 18–35.
- Kiyotaki, Nobuhiro and Moore, John. "Liquidity, Business Cycles, and Monetary Policy." Working Paper, Princeton University, April 2008; www.princeton.edu/~kiyotaki/papers/ChiKM6-1.pdf.
- Meh, Cesaire A. and Quadrini, Vincenzo. "Endogenous Market Incompleteness with Investment Risks." *Journal of Economic Dynamics and Control*, November 2006, 30(11), pp. 2143-65.
- Stock, James H. and Watson, Mark W. "Has the Business Cycle Changed and Why?" NBER Working Paper 9127, National Bureau of Economic Research, August 2002; www.nber.org/papers/w9127.

APPENDIX

This appendix characterizes the solutions to the workers' and entrepreneurs' utility maximization problems. Consider entrepreneurs first, and suppose that productive entrepreneurs are debt constrained, which requires that

$$(A1) \quad R_t < \alpha A_t^{1/\alpha} [(1-\alpha)/w_t]^{(1-\alpha)/\alpha}$$

for all $t \geq 0$. All entrepreneurs hire labor to equalize marginal product to the wage; hence,

$$L_t^i = K_t^i [A_t^i (1-\alpha)/w_t]^{1/\alpha}.$$

Consider first productive entrepreneurs ($A_t^i = A_t$). They borrow up to their debt limit because the marginal product of capital exceeds the interest rate because of equation (A1); hence,

$$D_t^i = \frac{\lambda A_t K_t^i}{R_t} [A_t (1-\alpha)/w_t]^{(1-\alpha)/\alpha}.$$

Let $S_t^i = K_t^i - D_t^i$ denote equity (savings) of the entrepreneur. Wealth at the end of period t is proportional to the borrower's capital investment and also proportional to equity:

$$\begin{aligned}
& A_t (K_t^i)^\alpha (L_t^i)^{1-\alpha} - w_t L_t^i - R_t D_t^i \\
&= (\alpha - \lambda) A_t [A_t (1 - \alpha) / w_t]^{(1-\alpha)/\alpha} K_t^i \\
&= \frac{R_t (\alpha - \lambda)}{R_t A_t^{-1/\alpha} [w_t / (1 - \alpha)]^{(1-\alpha)/\alpha} - \lambda} S_t^i \\
&= \tilde{R}_t S_t^i.
\end{aligned}$$

Here \tilde{R}_t is the return on equity for an entrepreneur who is productive in period t . Consequently, the budget constraint in period $t+1$ reads as $C_{t+1}^i + S_{t+1}^i = \tilde{R}_t S_t^i$. Consider next an unproductive entrepreneur in period t ; that is, $A_t^i = B$. There are two possibilities. First, either the interest rate exceeds the marginal product of capital of these entrepreneurs—in which case they do not produce and their return on savings is simply R_t —or, as is assumed in the main text, the interest rate equals their marginal product of capital,⁷ which requires that

$$R_t = \alpha B^{1/\alpha} [(1 - \alpha) / w_t]^{(1-\alpha)/\alpha}.$$

Again, $S_t^i = K_t^i - D_t^i$ is savings and $-D_t^i > 0$ are financial assets of unproductive entrepreneur i . Wealth at the end of period t is again proportional to savings:

$$\begin{aligned}
& B (K_t^i)^\alpha (L_t^i)^{1-\alpha} - w_t L_t^i - R_t D_t^i = \alpha B (K_t^i)^\alpha (L_t^i)^{1-\alpha} - R_t D_t^i \\
&= R_t (K_t^i - D_t^i) = R_t S_t^i.
\end{aligned}$$

Therefore, any entrepreneur's budget constraint in period $t+1$ reads as $C_{t+1}^i + S_{t+1}^i = R_t^i S_t^i$, where $R_t^i = R_t$ if the entrepreneur is unproductive in period t and $R_t^i = \tilde{R}_t$ if the entrepreneur is productive in t . The Euler equation for entrepreneur i 's utility maximization problem is then

$$\frac{1}{C_t^i} = \beta R_t^i E_t \frac{1}{C_{t+1}^i}.$$

Clearly, constant consumption/saving shares $C_{t+1}^i = (1 - \beta) R_t^i S_t^i$ and $S_{t+1}^i = \beta R_t^i S_t^i$ are the only solution to this equation that satisfy the transversality condition.

For workers, the Euler equations for their problem specified in (ii) of the equilibrium definition is

$$\frac{1}{C_t^w} \geq \beta R_t E_t \frac{1}{C_{t+1}^w}, \quad D_t^w \leq 0,$$

with complementary slackness. In the main text, it is shown that $R_t \beta < 1$ holds in the neighborhood of the steady state. Further, in the neighborhood of the steady state (i.e., small-enough productivity shocks), $w_t \approx \bar{w}$ and $E_t(1/w_{t+1}) \approx 1/\bar{w}$ hold, where \bar{w} is the steady-state wage level. Therefore, it follows from the complementary slackness condition that $D_t^w = 0$ and $C_t^w = w_t$ for all $t \geq 0$.

⁷ The third case—that the interest rate is below the marginal product of capital of all entrepreneurs—is incompatible with equilibrium; then all entrepreneurs would be lenders, whereas workers do not save (as shown below). Hence, the capital market cannot be in equilibrium.