Federal Reserve Bank
of Chicago

Third Quarter 2005

Economic perspectives

Economic perspectives

Economic Perspectives is published by the Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

Economic Perspectives and other Bank publications are available on the World Wide Web at http:/www.chicagofed.org.

chicagofed.org

ISSN 0164-0682

Contents

Third Quarter 2005, Volume XXIX, Issue 3

Determinants of supplier plant location: Evidence from the auto industry

Thomas Klier

Introduction and summary

The auto industry in the United States directly employs over 1 million workers, and is so large that gross motor vehicle output alone represents more than 3 percent of the U.S. economy. In discussing its fortunes, however, we often focus on the assembly segment of the industry. Assembly-related activities represent only the most visible part of this industry, the tip of the iceberg, if you will. Below the waterline lies the entire supply structure that ultimately feeds into the assembly line, at the end of which rolls off a car or light truck. That part of the industry, which encompasses everything from inputs such as steel coils to the subassembly of entire vehicle interiors, is larger, both by count of plants and employment, than the assembly part of the industry.[1] Yet our understanding of the auto supplier industry is quite limited, mostly due to the noisiness of the publicly available data for that sector.[2]

From numerous trade and business press stories, we know that the way auto suppliers relate to their assembly customers has fundamentally changed over the last 20 years. The main driver was the arrival of lean manufacturing, a production system aimed at the elimination of waste in every area of production including product design, supplier networks, and factory management, in North America during the early 1980s. Since then, lean manufacturing production techniques have become standard practice for auto assembly as well as the largest supplier companies. Some auto assemblers even operate “supplier support organizations” in order to transfer technology and knowledge to improve the efficiency of operations at their suppliers. Furthermore, assemblers no longer interact directly with most of their suppliers. The number of independent supplier plants assembly companies work with *directly* has fallen greatly during the last ten years to 15 years. In turn, many suppliers now supply primarily other supplier plants. At the same time, the Big Three automakers, notably Ford and General Motors (GM), have increased the share of parts they procure from outside their company. For example, both Ford and GM spun off many of their own parts plants as independent companies several years ago. In addition, the remaining assembler-owned parts plants have experienced rather dramatic job reductions over the last few years (Klier, 2005). Finally, this industry, like most manufacturing industries, has become noticeably more international. As producers of cars and light trucks pursue a global manufacturing footprint, their main suppliers need to be able to meet the needs of the assemblers globally (Roland Berger, 2004).

In estimating models of supplier plant location, this article contributes to the current discussion of the changing geography in the U.S. auto industry. The ongoing loss of market share by the domestically headquartered producers to foreign-headquartered producers of vehicles, both through imports as well as production in the U.S., raises important questions about the location trends for the industry (Klier, 2005).[3] Between the first quarter of 2000 and the first quarter of 2005, the U.S. share of light-vehicle sales by Big Three nameplates has fallen from 67.9 percent to 57.8 percent. While some of that market share loss is attributable to a rise in imports, most of it is explained by increased U.S. production of foreign-headquartered assembly companies. This matters for the geography of this industry as most of these “new domestic” assembly plants in North America tend to be located farther south than the assembly plants of the traditional domestic

Thomas Klier is a senior economist in the Economic Research Department at the Federal Reserve Bank of Chicago. *The author would like to thank Jeff Campbell, Craig Furfine, and Dan McMillen for helpful comments and Cole Bolton, Anna Gacia, Joanna Karasewicz, Paul Ma, and Alexei Zelenev for excellent research assistance.*

producers. In fact, the assembly plants opened most recently, such as the Honda plant in Lincoln, Alabama, and the Nissan plant in Canton, Mississippi, have been situated in the most southern area of the auto region. As the geography of the auto sector continues to change, one wonders whether Detroit can continue to be the hub of this industry over the medium-term horizon.[4] The public policy issues of a changing location pattern in the auto sector are considerable as the traditional auto states are struggling with this southward shift of auto production and related economic activities.[5] For example, Michigan is currently suffering from its heavy exposure to the domestic auto and parts makers. In her 2005 State of the State address, Michigan Governor Jennifer Granholm proposed a sizable bond issue to attract and retain jobs in the state. The business press reported recently that Michigan is heavily recruiting Toyota to locate one of two currently proposed assembly facilities in the state (Hakim, 2005).

This article utilizes detailed plant-based data on the U.S. auto supplier industry. After describing the spatial properties of this data, I estimate two simple models of plant location.[6] I find the auto industry to be strongly spatially concentrated. The core of the auto region is densely packed with plants, reaching from Michigan up into Ontario, west to Chicago, and south to northern Alabama and into the Carolinas. The states within the auto region show variations along a number of dimensions. For example, the northern half of the auto region is more densely populated by domestic supplier plants[7] whereas foreign plants are more concentrated in the southern half. That pattern is not surprising as it replicates the regional distribution of assembly facilities. Union plants are concentrated in Michigan, Indiana, and Ontario. Larger plants, however, tend to be located farther away from Detroit. A plant-level model of employment shows that plants located farther from Detroit tend to have larger employment, as do tier 1 (discussed in detail later in the text) and foreign-owned plants. In addition, I find plant size to vary by type of part produced. Modeling plant location choices of recently opened supplier plants at the county level consistently finds the presence of an interstate highway to be significantly related to plants locating in such counties. In addition, the size of the market, as measured by the number of assembly plants within a day's drive (approximately 450 miles) from a county, is positively related to the number of recently opened plants in a county.

Literature review

Economic interest in agglomeration issues goes back to at least Alfred Marshall (1920); for more recent research, see Krugman (1991) and Ellison and Glaeser (1997).

Regarding the question of what drives the geography of the auto industry, a number of studies address the reconcentration of *assembly* plants in the Midwest, a development which started in the mid-1970s. Rubenstein (1992) attributes this to the demise of the branch plant system, which was based on producing identical models in plants located close to population centers. The subsequent reconcentration of assembly plants in the heart of the country was driven by an increase in the choice of models available to the consumer that far outpaced the growth of the market, resulting in much reduced production runs per model. As a result, individual models tend to support only a single assembly plant. That plant is then best located in the heart of the country, as the final product has to be shipped all over the country from that one production location.

Geographic trends in the *supplier* industry have followed a different pattern. While this part of the auto industry has remained remarkably concentrated in the Midwest since the industry's beginning over 100 years ago, it has experienced a migration of mostly labor-intensive parts to the southern U.S. and Mexico for some time. For example, in 2002, 73 percent of all wiring harnesses—gatherings of electrical wires terminating in a central plug that distribute electricity in a car to operate the turn signals, brake lights, etc.—"consumed" in the U.S. were imported, 82.7 percent of which were produced in Mexico.

There is evidence that, within the auto region, assembly and supplier plants want to locate in proximity to one another (see Smith and Florida, 1994, for a model for Japanese-affiliated manufacturing establishments in auto-related industries). State of the art supply chain management requires most supplier plants to be located within a day's drive from the assembly plant customer (see Klier, 1999, and 2005). And so, supplier networks of individual assembly plants are of a regional nature, as the existing transportation infrastructure allows for reliable on-time delivery of products (see Woodward, 1992, and Smith and Florida, 1994, for the importance of highway transportation).

Yet, as the auto industry continues to be very highly concentrated across space, the geographic extension of its core region has changed. No longer reaching eastward from Detroit to Pennsylvania and New York, it now is defined in a marked north–south direction, extending from Detroit to Kentucky and Tennessee and beyond with fingers reaching north into Canada and south into Mexico. In other words, the core auto region has pivoted around Detroit over several decades. During the last few years this development has gained greater attention as the old-line auto states have been losing production and employment to the southern end of the auto corridor. The changing fortunes of domestic and foreign assembly plant customers appear to be profoundly reshaping the regional distribution of supplier employment (Klier, 2005).

How to measure the auto supplier industry?

Overview of the supplier industry

For the purpose of this article, auto suppliers are companies that supply light-vehicle assembly companies.[8] Among them, one can distinguish the following categories: suppliers that deal directly with the assembly company and those that deal primarily with other suppliers. The first category is commonly referred to as tier 1 suppliers, while the other category is referred to as tier 2 suppliers. The number of tier 1 suppliers has been shrinking over the last decade, as assemblers have been reducing the number of companies they do business with directly. At the same time, that segment of the supplier industry has been subject to a series of mergers and acquisitions. Finally, there are a number of tier 1 parts operations that are owned and operated by the assemblers themselves, such as engine and stamping facilities. These are generally referred to as *captive suppliers*. A number of years ago the two largest U.S. assemblers decided to spin off the majority of their captive parts operations. In 1999, GM spun off most of its captive plants as Delphi, which instantly became the largest independent tier 1 auto parts supplier. One year later, Ford Motor Company divested a large number of its captive plants as a separate company called Visteon. It then became the second largest independent parts supplier in North America.[9] Table 1 lists the 15 largest auto supplier companies as ranked by the industry weekly *Automotive News* in 2003 based on sales in North America. The 50 largest suppliers on that list each have global sales exceeding $1 billion, amounting to a total of about $285 billion. If one classifies these companies based on the location of their headquarters, the following pattern emerges: 53 percent of the 150 largest suppliers represent companies based in one of the NAFTA (North American Free Trade Agreement) countries, 20 percent are from Japan, and the remaining 27 percent are from Europe. This illustrates the degree of global competition present in this industry.

Plant-level data

The analysis of auto supplier plants presented in this article is based on data acquired from ELM International, a Michigan-based vendor. While not designed with research applications in mind, the ELM database is intended to cover auto supplier companies and their plants in North America.[10] The database provides 3,542 plant-level records. Included is information on a plant's

TABLE 1

Largest auto supplier companies, 2003

			OEM automotive parts sales ($ bn.)	
Rank	**Company name**	**HQ in**	**North America**	**Worldwide**
1	Delphi Corp.	U.S.	19.5	25.5
2	Visteon Corp.	U.S.	11.1	16.9
3	Lear Corp.	U.S.	9.4	14.4
4	Magna International	CDN	9.1	12.4
5	Johnson Controls Inc.	U.S.	8.0	13.7
6	Dana Corp.	U.S.	5.5	7.3
7	Robert Bosch Corp.	GER	5.0	19.1
8	TRW Automotive Inc.	U.S.	4.6	9.9
9	Denso International America Inc.	J	3.9	15.3
10	ThyssenKrupp Automotive AG	GER	3.7	6.2
11	American Axle	U.S.	3.5	3.5
12	Collins & Aikman	U.S.	2.9	3.9
13	DuPont Automotive	U.S.	2.8	5.4
14	Continental AG	GER	2.3	5.6
15	Yazaki North America	J	2.2	5.8
			93.5	164.9

Note: OEM is original equipment manufacturer; CDN is Canada; GER is Germany; and J is Japan.
Source: *Automotive News*, available at www.autonews.com/datacenter.cms?dataCenterId=129, by subscription.

address, employment, parts produced, customer(s), union status, as well as square footage. In order to clean up the data for research purposes, several operations were performed. First, records were cross-checked with state manufacturing directories to obtain information on the plant's age.[11] We also appended information on the nationality of the company to the record of each plant from the ELM company-level data.[12] Plants of supplier companies listed in the 2003 *Automotive News* "top 150 automotive suppliers list" were coded with the companies' ranks in that listing. Information on captive parts plants was also checked with Harbour (2003). For all the *Automotive News* top 150 companies, the accuracy and completeness of ELM's plant listings—that is, the number of plants as well as their location—was crosschecked with the companies' websites when possible.[13] Overall, that resulted in a net addition of 335 records. Finally, the accuracy of the employment for the largest plants (employment greater 2,000) was also checked with company websites or phone calls. After this preparation the data consists of 3,877 observations of auto supplier plants located in the U.S. and Canada (see table 2).[14] To my knowledge, this may well be the most accurate plant-level description of the North American auto supplier industry currently available.

Table 2 summarizes the supplier plant data for the U.S. and Canada along several dimensions. Of the 3,877 plants more than half are characterized as lower tier suppliers. That is, they primarily do business with other supplier companies. These plants tend to be smaller (their average employment is 241) than tier 1 suppliers (average employment of 388), which make up 42 percent of all plants. Captive suppliers, while small in numbers, represent by far the largest plants. Their average employment is above 1,000. Of the three groups, captive plants tend to be located closest to Detroit. The union variable covers only 83 percent of all plants; 25 percent are unionized, while 58 percent are not. Unionized plants have larger employment and are located closer to Detroit than nonunion plants. As for ownership, just under 80 percent of supplier plants are part of a company that has its headquarters in the U.S., Canada, or Mexico. "Foreign" plants are larger and are located farther away from Detroit than "domestic" plants. Finally, a quarter of the plants appears to be single-establishment firms.[15] These plants show the lowest average employment of all groups listed in table 2.

Spatial characteristics of the auto supplier industry

This plant-level data allows a fairly detailed description of the spatial properties of the auto supplier industry. Figure 1 shows the distribution of auto supplier plants. It represents all 3,877 U.S. and Canadian plants in the data set, aggregated to the zip code level of detail. The symbols representing supplier plants are scaled to convey the spatial density of plant locations.

The most interesting feature of the map is the high degree of clustering exhibited by this industry. It is self-evident that southern Michigan represents the hub of the North American auto sector.[16] The core region of this industry extends from that area west to Chicago, northeast to Toronto, and south to Tennessee and arguably into northern Mississippi, Alabama, Georgia, and the Carolinas.[17] Pennsylvania represents the link between the heart of the industry in the Midwest and a cluster on the East Coast. West of the Mississippi the country is mostly empty of auto supplier activity except for a thinly populated band that extends from eastern Texas and northern Louisiana

TABLE 2

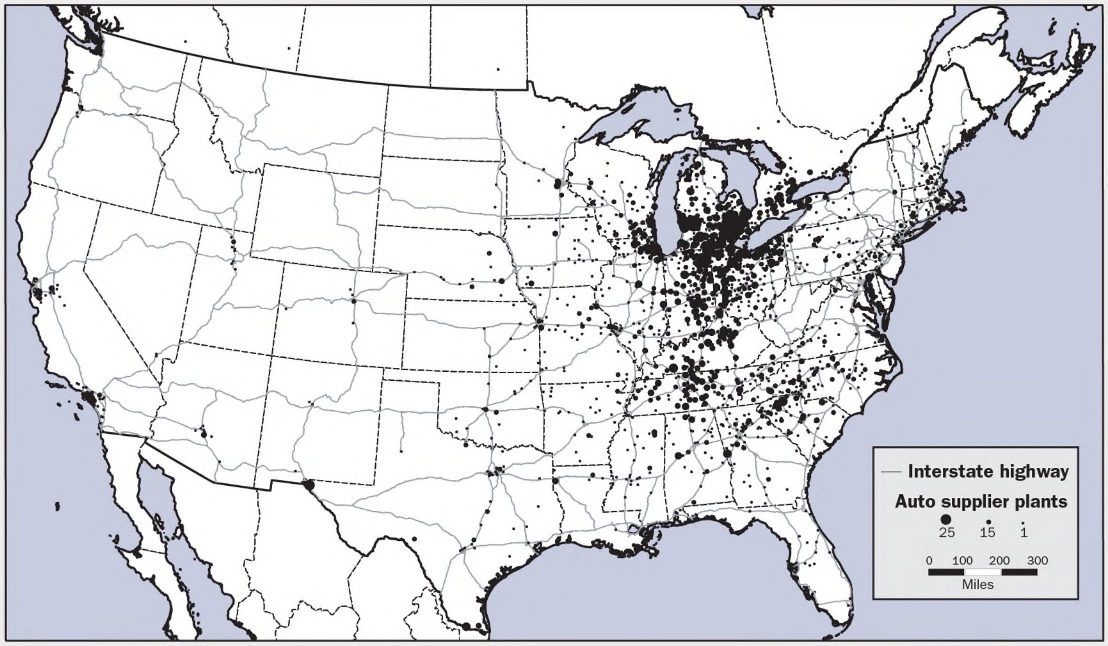
Supplier data summary, U.S. and Canada, 2003

	% of plants	% of employment	Average employment	Median distance to Detroit (miles)
Tier 1 suppliers	41.7	49.5	388	253
Captive suppliers	2.7	9.5	1,153	136
Lower tier suppliers	55.6	40.9	241	218
Union	25.3	38.0	491	180
Nonunion	58.1	52.0	293	256
Domestic	79.2	77.3	319	210
Foreign	20.8	22.7	357	309
Single plant	24.0	17.0	236	198
Multiplant	76.0	83.0	400	247
All	100	100	327	237

Note: Based upon 3,877 observations at auto supplier plants.

FIGURE 1

Distribution of auto supplier plants



north to Nebraska and Iowa and into Minnesota. Other than that, one can observe two clusters in California, one in the Bay area and the other in the L.A. basin. Finally, Utah, Colorado, Arizona, and New Mexico are home to small localized clusters, and the border between Texas and Mexico shows centers of activity around El Paso and Laredo/Brownsville. These are related to border crossings that link the Mexico-based maquiladora plants to the U.S.-based suppliers.[18]

Table 3 provides further detail on the *distribution of plants and employment* in the auto supplier industry. The information is first summarized by the four Census regions plus Canada (see panel A). The bottom panel of the table provides an alterantive breakdown of the data, focusing on the two halves of the auto corridor. Column 2 shows that 90.1 percent of all 3,877 plants are located in the Midwest, South, or Canada. Michigan alone is home to 22.5 percent of all auto supplier *plants*, followed by Ohio (11.6 percent) and Ontario (10.7 percent). The auto corridor as a group represents just under 79 percent of all auto supplier plants in the U.S. and Canada. Columns 3–8 of table 3 provide three different breakdowns of the location of auto supplier plants.

Grouping supplier plants by *nationality of company*, one can see that the auto corridor consists of two halves: The northern end shows a higher concentration of domestic plants (64.7 percent) and lower concentration of foreign-owned plants (46.7 percent) than overall. Likewise, the southern end shows a much higher concentration of foreign-owned supplier plants (33.7 percent) and a smaller share of domestics (13.8 percent). In addition, 21.5 percent of domestic automotive supplier plants in the U.S. and Canada (and 19.6 percent of foreign ones) are located outside the auto corridor. The share of foreign supplier plants located at the southern end of the auto corridor is 2.4 times as large as the share of domestic plants. This pattern suggests an influence of the location of the primary customer on the supplier plant location (Klier, 1999, and Smith and Florida, 1994). The median distance of foreign-owned supplier plants to Detroit is 309 miles, noticeably larger than the 210 miles for domestic supplier plants (see table 2).[19] One can argue that in setting up operations in North America, foreign suppliers choose locations close to foreign-owned assembly plants, which presumably were their prime customers at that time.

The *tier status* of a supplier plant is measured by its inclusion in *Automotive News*' top 150 supplier companies list. That is a somewhat arbitrary yet plausible way to define which plants are tier 1 plants. In essence, it assumes that all of the large supplier companies' plants deal directly with assembly plants. Since captive

TABLE 3

Distribution of plants and employment by region, 2003

A. By Census region

	Plant count							Employment count						
	All	Domestic	Foreign	Tier 1 and captives	Others	Union	Nonunion	All	Domestic	Foreign	Tier 1 and captives	Others	Union	Nonunion
Observations	3,877	3,072	805	1,811	2,066	980	2,259	1,268,135	980,381	287,754	848,378	419,757	484,708	659,817
Region	%	%	%	%	%	%	%	%	%	%	%	%	%	%
Midwest	54.3	57	44.2	53	55.5	61.2	54.2	52.7	56.2	40.7	58.2	41.5	66.1	45.9
Northeast	6.7	7	5.3	4.8	8.2	9	5.6	8	8.9	5.1	5.9	12.3	9	7.1
South	24.3	20.2	39.9	27.7	21.3	13.3	27.7	24.3	19	42.5	22.3	28.2	14.1	32.3
West	3.2	2.9	4.2	2.3	4	1.3	3.9	4.6	4.5	4.8	2.6	8.6	1	7.4
Canada	11.5	12.9	6.3	12.1	10.9	15.2	8.6	10.4	11.4	6.9	10.9	9.3	9.9	7.4
Total	100	100	100	100	100	100	100	100	100	100	100	100	100	100

B. By auto corridor location

Region	%	%	%	%	%	%	%	%	%	%	%	%	%	%
Auto corridor NORTH	60.9	64.7	46.7	59.4	62.3	72	58.2	58	62.5	43	64.6	44.9	72.2	48.1
Auto corridor SOUTH	17.9	13.8	33.7	21.1	14.9	7.8	21.4	18.6	13.4	36.2	17.8	20.1	9.1	25.9
rest of US/CDN	21.2	21.5	19.6	19.5	22.8	20.2	20.4	23.4	24.1	20.8	17.6	35	18.7	26
Sum	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Notes: Seventeen percent of plants have no information on their union status. Therefore, this comparison (columns 6, 7, 13, and 14) only applies to 83 percent of the records.
States not listed do not have automotive supplier plants located in them.
Midwest: IA, IL, IN, KS, MI, MN, MO, NE, OH, SD, WI
Northeast: CT, MA, ME, NH, NJ, NY, PA, RI, VT
South: AL, AR, DE, FL, GA, KY, LA, MD, MS, NC, OK, SC, TN, TX, VA, WV
West: AZ, CA, CO, NM, NV, OR, UT, WA
Auto corridor North: IL, IN, MI, OH, Ontario, WI
Auto corridor South: AL, GA, KY, MS, NC, SC, TN
Source: *Automotive News*, available at www.autonews.com/datacenter.cms?dataCenterId=129, by subscription.

suppliers tend to interact directly with assembly plants, they are grouped with tier 1 plants in table 3. While generally very similar in their regional distribution, tier 1/captive plants are more prevalent in the South and less so in the Northeast.

Table 3 also shows a disproportionate concentration of unionized supplier plants in the Midwest and Ontario.[20] Nonunionized plants, on the other hand, are concentrated in the South where many states have right to work laws. Within the auto corridor, this split shows very strongly. Seventy-two percent of all union plants are found in the northern end of the auto corridor. Correspondingly, they are quite rare in the southern end (7.8 percent of all unionized plants versus 21.4 percent of all nonunionized plants).

The location of *employment*, shown in columns 9–15, resembles the location of plants, column 2, very closely in the aggregate. The auto corridor is home to 76.6 percent of the industry's employment and 78.8 percent of its plants. At a more disaggregate level, table 3 reveals a regional difference in the geography of plants and employment, indicating that plants located in the northern end of the auto corridor tend to have, on average, fewer employees. For example, employment at foreign-owned plants is noticeably more concentrated in the southern half of the auto corridor than employment at domestic plants. The foreign-owned plants located in the south also tend to be disproportionately large, as measured by employment. They represent 33.7 percent of all plants, yet 36.2 percent of all employment in the sector. In contrast, both domestic and foreign-owned plants located in the northern half are disproportionately smaller; that is, they represent a smaller share of industry employment than of plants. However, that pattern does not apply to unionized plants. For example, Michigan is home to 26.9 percent of unionized plants and 29.1 percent of employment at unionized plants.

Formal analysis of employment and plant distribution

This section reports on two formal models to estimate the location of employment as well as plant distribution. The idea is to formally test what underlies the observed agglomeration in the auto supplier industry. The models utilize data on U.S. plant locations only. Table 4 lists the summary statistics for both the plant-level as well as the county-level models reported.

First, we regress *plant*-level employment on a number of plant-level characteristics that the detailed database allows us to draw on. The model also uses a number of variables that are measured at the county level, such as the presence of an interstate highway. The model incorporates that information only for counties in which plants are actually located. That explains why the mean of the interstate highway variable is 0.78 in the plant-level model: 78 percent of plants are located in counties that are reached by an interstate highway.

The geography of plants is measured by two different variables. *DISTANCE* measures the straight-line distance between the centroid of the zip code in which the supplier plant is located and the centroid of the zip code for downtown Detroit.[21] Detroit seems an obvious spatial reference point as it is clearly the hub of this industry. *VDISTANCE* measures distance to Detroit only in the north–south direction. In addition, the following set of plant characteristics is included in the model. A set of dummy variables indicating if the plant is part of a single plant company; if it is part of one of the largest 150 supplier companies;[22] if it is an assembler-owned supplier plant (*CAPTIVE*); if it is unionized;[23] and if its headquarter operations are located outside North America. In addition, a group of dummy variables controls for what subsystem of the car the plant's output feeds into (table 5, p. 10).[24] Finally, the model includes a control variable for counties in right-to-work states as well as a couple interactive terms of the plant control variables.

Table 6 (p. 11) reports the results of three different specifications and the variables used in constructing each of them. A simple model (specification 1) can explain about 20 percent of the variation in the dependent variable. In addition, the model identifies a statistically significant relationship between the plant-level employment and tier status as well as nationality of headquarters: Plants of tier 1 supplier companies as well as plants of foreign-headquartered companies are found to have larger employment. The presence of unions in a supplier plant is only related to larger plant employment if the plant is either captive or part of a tier 1 supplier company. That is to say, unionized plants are larger than others only if they are either tier 1 or captive plants. Specification 2 controls for what the supplier plants are producing by distinguishing 8 major subsystems of a car. Employment at plants producing parts for chassis (such as tires), body, engine electrical (which includes the electronics components suppliers), and engine attached (often referred to as air and fuel handling) is consistently found to be larger than that of the control group, plants that produce generic parts. Finally, specification 3 controls for a number of county-level characteristics that might influence plant location decisions, such as the degree of local work force education, transportation infrastructure, as well as the presence of other

TABLE 4				
Descriptive statistics				
		County-level model		
	Plant-level model	**All new plants**	**All new domestic**	**All new foreign**
Employment	359.922 (473.248)			
Share of young supplier plants		0.042 0.162		
Share of domestic young suppliers			0.0229 0.114	
Share of foreign young suppliers				0.019 0.111
Log employment	5.35 (1.052)			
Distance to Detroit (miles)	361.933 (388.950)	456.174 205.216	456.174 205.216	456.174 205.216
Vertical distance to Detroit (miles)	203.768 (220.904)			
Single plant company	0.257			
Plant part of top 150 supplier	0.363			
Plant is captive	0.024			
Plant is unionized	0.262			
Company headquarters outside North America	0.206			
Right-to-work state	0.237	0.467	0.467	0.467
Interaction top 150 and unionized	0.106			
Interaction captive and unionized	0.019			
Parts for body (%)	0.142 (0.297)			
Parts for chassis (%)	0.199 (0.329)			
Parts for drivetrain (%)	0.039 (0.144)			
Parts for engine attached (%)	0.103 (0.249)			
Parts for engine electrical (%)	0.071 (0.225)			
Parts for engine (%)	0.093 (0.238)			
Parts for interior (%)	0.149 (0.312)			
Generic parts (%)	0.186 (0.335)			
Presence of interstate highway	0.787 (0.411)	0.506 (0.50)	0.506 (0.50)	0.506 (0.50)
Share of employment in manufacturing	25.536 (8.218)	23.807 (9.93)	23.807 (9.93)	23.807 (9.93)
High school education (%)	0.74 (0.082)	0.672 (0.105)	0.672 (0.105)	0.672 (0.105)
Population in 1990 (million)	0.515 (1.092)	0.093 (0.227)	0.093 (0.227)	0.093 (0.227)
No. of supplier plants in county	19.355 (31.025)	1.335 (4.818)		
No. of domestic supplier plants in county			1.072 (4.328)	1.072 (4.328)
No. of foreign supplier plants in county			0.263 (0.804)	0.263 (0.804)
No. of assembly plants within 450 miles	37.113 (16.074)	31.223 (16.197)	31.223 (16.197)	31.223 (16.197)
No. of domestic assembly plants in county			22.842 (13.523)	22.842 (13.523)
No. of foreign assembly plants in county			8.381 (3.693)	8.381 (3.693)
No. of observations	3,097	1,607	1,607	1,607

Note: Standard deviations are in parentheses for continuous variables.

TABLE 5

Parts classification

Major subsystem	ELM subsystem	Frequency of parts listed (%)	
Engine		27	
Engine proper	Engine		11
Engine electrical			
	Ignition systems		1
	Electronic supply		1
	Electronics		3
Engine attached			
	Engine cooling		2
	Climate control		3
	Fuel systems		4
	Exhaust systems		2
Chassis		20	
	Chassis electrical		6
	Chassis systems		2
	Suspension		3
	Steering		3
	Braking		4
	Wheels and tires		2
Interior		15	
	Interior body		14
	Passenger restraints		1
Body		16	
	Body glass		2
	Body components		14
Drivetrain	Drivetrain	5	
Generic	Generic	16	
		100	

Source: ELM and author's calculations.

supplier and assembly companies. However, the county-level variables do not add to the plant-level model of employment (table 6).

Next, I estimate a model of plant location at the *county* level (table 7, p. 12). The dependent variable is the share of supplier plants in a county that opened recently.[25] As the underlying data is cross-sectional in nature, it seems prudent to focus on location decisions of more recently established plants.[26] Going back much further in time could introduce survivor bias to the model. The premise is that county characteristics matter in plant location decisions. The model accounts for the presence of existing assembly and supplier plants to capture possible agglomeration effects within the auto industry.

The number of assembly plants located within 450 miles of a county's centroid measures the size of the market available to a supplier locating in that county. That is an important reference point as the ability to deliver reliably within a day is a key requirement of the just-in-time production system. The distance of 450 miles corresponds to an industry rule of being able to deliver within a day's drive. The model also includes a measure of how many suppliers had previously located in a county to account for agglomeration effects. Finally, the set of county-level controls used in specification 3 of the plant-level model (table 6) is included in the county-level model as well. Table 7 reports the results that utilize information for all counties east of the Mississippi to capture the region of the country most densely populated by the auto industry.[27]

Across all specifications estimated, the presence of an interstate highway in a county is consistently associated with a higher share of recently opened supplier plants in that county. In addition, the size of the market for suppliers, as measured by the number of assembly plants within a day's drive from a county, is related to suppliers choosing a county. Specifications 2 and 3 distinguish domestic and foreign plants, both for the dependent as well as the independent agglomeration variables. It turns out that only the presence of foreign assembly plants within a 450 mile radius is significantly related to the incidence of both domestic and foreign "young" supplier plants locating in a county.

Simulation of policy effects

Based on the model results presented in table 7, I perform two simple simulation exercises. The idea is to elicit from the model what the estimated response in the distribution of supplier plants would be to a simulated change in the location of an assembly plant. First, assume that Tennessee has one less light-vehicle assembly plant and Michigan has one more. I assume Spring Hill as the location of the plant in Tennessee, and Grand Rapids for the fictional plant in Michigan. Subsequently, I re-calibrated the variable that measures the number of assembly plants located within a 450-mile radius of each county. To that reconfigured variable and all the others in the model, the estimated coefficients as reported in table 7 were subsequently applied. In doing so one performs what is referred to as an "out-of-sample" forecast. In essence, one can simulate what would happen to the distribution of young supplier plants if Grand Rapids had an assembly plant and Spring Hill did not. Constraining the estimation to result in a zero sum redistribution of supplier plants, the following result emerges. The three states of Michigan, Indiana, and Ohio would increase their count of supplier plants that opened

TABLE 6

Estimation of plant employment

Variable	Specification 1	Specification 2	Specification 3
Distance to Detroit	0.113**	0.097**	0.107**
	(0.027)	(0.027)	(0.046)
Vertical distance to Detroit	–0.095	–0.112	–0.144
	(0.067)	(0.067)	(0.075)
Single plant company	–5.370	2.270	4.470
	(19.850)	(19.927)	(20.022)
Top 150 supplier	152.368**	149.414**	147.093**
	(20.823)	(21.312)	(21.356)
Captive supplier	169.406	204.883*	204.998*
	(108.186)	(108.325)	(108.376)
Unionized plant	21.976	25.07	25.253
	(23.711)	(23.634)	(23.654)
Headquarters outside North America	79.872**	59.633**	56.298**
	(19.685)	(19.816)	(20.002)
Right-to-work state	49.263*	49.432*	42.641
	(28.268)	(28.245)	(32.975)
Top 150 supplier and unionized	293.919**	281.682**	284.626**
	(36.616)	(36.471)	(36.544)
Captive supplier and unionized	952.425**	926.215**	937.641**
	(123.098)	(121.933)	(122.275)
Chassis %		205.226**	199.212**
		(29.870)	(29.977)
Drivetrain %		90.164	90.000
		(56.584)	(56.590)
Interior %		18.102	11.047
		(30.334)	(30.473)
Body %		56.473*	52.878*
		(31.771)	(31.815)
Engine %		50.999	41.566
		(38.084)	(38.295)
Engine electrical %		304.689**	303.297**
		(38.824)	(38.885)
Engine attached %		141.791**	135.461**
		(35.394)	(35.537)
Presence of interstate highway			29.881
			(20.828)
Manufacturing employment (%)			2.016*
			(1.145)
High school education (%)			–0.897
			(1.342)
Population in 1990			–1.24.970
			(924.818)
No. of supplier plants in county			–0.546
			(0.336)
No. of assembly plants within 450 miles			–0.016
			(1.034)
Constant	193.497**	114.932**	127.432
	(16.850)	(23.081)	(134.086)
No. of observations	3,097	3,050	3,050
R squared	0.19	0.22	0.22

**Significant at the 5% level.
*Significant at the 10% level.
Note: Standard errors are in parentheses.

TABLE 7

Supplier plant locations between 1994 and 2003

	All	Domestic only	Foreign only
No. assembly plants w/450 miles	0.001**		
	(0.00)		
No. domestic assembly plants w/450 miles		–0.001	0
		(0.001)	0
No. foreign assembly plants w/450 miles		0.004**	0.004**
		(0.001)	(0.001)
No. existing supplier plants	0		
	(0.001)		
No. existing domestic suppliers		0	–0.001
		(0.001)	0
No. existing foreign suppliers		0.003	0.006
		(0.004)	(0.004)
Interstate highway	0.03**	0.012**	0.014**
	(0.009)	(0.006)	–0.006
Right to work state	0.019	–0.005	0.007
	(0.012)	(0.009)	(0.009)
Share of manuf. employment	0.001**	0.001	0
	0.000	0	0
Percent high school ed.	0	0.001*	0
	(0.001)	0	0
Population, 1990	0.03	0.027*	0.011
	(0.021)	(0.015)	(0.015)
Distance to Detroit	0	0	0
	0.000	0.000	0.000
Constant	–0.62	–0.033	–0.018
	(0.065)	(0.046)	(0.045)
Observations	1,607	1,607	1,607
R squared	0.03	0.02	0.02

**Significant at the 5% level.
*Significant at the 10% level.
Notes: Standard errors are in parentheses. Observations: 1,607. Model is estimated for all counties east of the Mississippi.

between 1995 and 2003 by 42, from 122 to 164. The three states of Kentucky, Tennessee, and Alabama would see their count of young supplier plants fall by 37, from 65 to 28. The simulated redistribution represents about 14 percent of all young supplier plants opened during the last 10 years. That represents a significant impact.[28]

A second experiment consisted allocating a *foreign* assembly plant in Michigan (again, Grand Rapids), instead of Spartanburg, South Carolina, and estimating the effect on the distribution of foreign-owned young supplier plants (there were 107 of them that opened between 1995 and 2003). Michigan, Indiana, and Ohio would gain young foreign suppliers. The count for the three states would increase by 27 from 30 to 57. By the same token, South Carolina and the surrounding auto corridor states North Carolina, Kentucky, Tennessee, Alabama, and Georgia would have received fewer recently opened foreign suppliers: Their plant count of foreign young would go down by 26 from 57 to 31.[29] According to this simulation, placing one foreign assembly plant into Michigan instead of South Carolina would affect the location of a quarter of all foreign supplier plants opened between 1995 and 2003.

Conclusion

This study set out with the intent to shed more light on the geography of the auto parts sector which is far less understood than that of the auto assembly sector of the auto industry. The analysis of a rich plant-level data set with records of almost 3,800 auto supplier plants located in the U.S. and Canada shows an industry that is very spatially concentrated. Today Detroit remains the center of a highly clustered auto region that extends north–south from Michigan, reaching up into Ontario, west to Chicago, and south to northern Alabama and into the Carolinas. While the analysis is purely cross-sectional, it reveals a surprising amount of variation in the location pattern exhibited along a number of dimensions. The study confirms the north–south split within the auto region by nationality of plant: Plants of domestically headquartered suppliers are concentrated in the northern end of the auto corridor and plants of foreign-headquartered suppliers are concentrated in the southern end. Overall, employment and plants are distributed quite similarly.

A plant-level model of employment shows that plants located farther from Detroit tend to have greater employment, as do tier 1 and foreign-owned plants. In addition, we find plant size to vary by type of part produced. A simple model of recent supplier plant openings at the county-level points out the importance of regional transportation infrastructure. The presence of interstate highway access in a county is consistently related to a higher share of recently located supplier plants. Furthermore, the number of assembly plant customers reachable within a day's drive is also related to supplier location choices. This finding points to the continued importance of agglomeration in this industry.

A policy simulation asks what the effect of a change in the location of one assembly plant would be on the geography of recent supplier plant openings. Two different simulations are presented, one moving an assembly plant from Tennessee to Michigan, the other moving a foreign assembly plant from South Carolina to Michigan. Both suggest a sizable regional effect on the location of supplier plants. A number of them would have located closer to the "new" location of the assembly plant as they need to be within 450 miles of their assembly plant customers.

NOTES

[1]U.S. motor vehicle parts employment is about four times as large as employment in motor vehicle assembly.

[2]Many different manufacturing sectors contribute to the production of vehicles and at the same time supply non-automotive customers. Furthermore, the census data on shipments do not distinguish between producers of parts for the aftermarket and the original equipment market. The 2002 *Census of Manufacturing*, however, reports the cost of materials used in U.S. light-vehicle assembly plants at $152.5 billion. That measure includes imported parts.

[3]In addition, factors such as the continuing consolidation and internationalization within the supplier industry also affect its spatial structure.

[4]The northern end of the auto corridor is home to over half of all light-vehicle assembly plants in the U.S., 81 percent of these are Big Three facilities. Conversely, the southern end of the auto region is home to about 20 percent of all light-vehicle assembly plants; half of these are foreign producer facilities. Testa, Klier, and Mattoon (2005) identify such a regional shift as the most likely structural threat to the Midwest's economy.

[5]See the speech of Michigan's Governor Granholm from August 4, 2004, in which she outlines a framework on how Michigan should respond to the current challenges facing its most important manufacturing sector. See also McAlinden and Hill (2003).

[6]The role of the border is not addressed in this article. Post 9/11, elevated national security concerns have exacerbated demands on the already strained border infrastructure between the U.S. and Canada, potentially affecting plant location decisions in an industry that continues to be very tightly integrated and has straddled both sides of the border for many years (see Simon, 2004, and Klier and Testa, 2002).

[7]"Domestic" refers to supplier companies which are headquartered in either the U.S., Canada, or Mexico, "foreign" to companies headquartered elsewhere.

[8]The term light vehicles refers to passenger cars and light trucks, which include minivans and sport utility vehicles.

[9]See White (2005) on the recent restructuring of the original agreement between Ford and Visteon.

[10]Data are available at the plant and company level. However, plants producing primarily for the aftermarket are not part of database, nor are plants that produce raw materials, such as steel and paint. The ELM data were purchased at the end of 2003. The database is continuously updated by the vendor.

[11]Plants for which no matching records were found were contacted by phone.

[12]Based on the location of company headquarters, the article distinguishes North American (U.S.-, Canadian-, or Mexican-owned plants), Japanese, as well as other foreign-owned plants.

[13]Thanks to my colleague Jim Rubenstein who shared his plant-level data for the 150 largest supplier companies.

[14]Mexican data are available for 601 plants, but have not yet been scrutinized to the same extent.

[15]I construct that variable from the database, utilizing plant names and company information. It is possible that some of these single-plant companies have plants that are not included in the database.

[16]A map of employment, instead of plant count, looks virtually identical.

[17]Based on the shape of the core auto region, I define the "auto corridor" to be the states and Canadian provinces that represent the contiguous north–south cluster visible in figure 1. They are Alabama, Georgia, Illinois, Indiana, Kentucky, Michigan, Mississippi, North Carolina, Ohio, Ontario, South Carolina, Tennessee, and Wisconsin. Mississippi and Alabama are included as they recently received new assembly plants.

[18]Maquiladora plants in northern Mexico were established by the 1965 Border Industrialization Program. This program allowed U.S. companies to assemble products in Mexico destined for export elsewhere. Later companies from other countries also established such plants near the northern Mexico border.

[19]Of all domestic assembly plants operating in the U.S, 38 percent are located within 100 miles of Detroit. The corresponding figure for foreign-owned assembly plants is only 7 percent.

[20]Note that 17 percent of plants have no information on their union status. Therefore, this comparison (see columns 6 and 7) only applies to 83 percent of the records.

[21]The geographic coordinates for the zip code centroids come from the Maptitude GIS program. The distance between the two sets of coordinates is given by the following formula: $acos(\sin(la1)*\sin(las) + \cos(la1)*\cos(la2)*\cos(lo2 - lo1))*6370*.62$, where $la1$ and $lo1$ are the latitude and longitude (in radians) of the zip code centroid of the supplier plant and $la2$ and $lo2$ are the coordinates for the zip code centroid of downtown Detroit.

[22]As explained earlier, tier 1 suppliers are the ones that interact directly with the assembler. One would have to know the identity of a supplier's customer plants in order to identify that group. The top 150 variable tries to proxy for that relationship in the absence of such detailed customer information. The underlying assumption is that the vast majority of tier 1 suppliers happen to be large companies.

[23]In the estimation we treat plants with unknown union status as not unionized. Based on size and location these plants are very similar to plants identified as nonunion.

[24]The ELM data provide information on what parts an individual plant produces in a very detailed way. Unfortunately, it does not provide the distribution of actual output across the various parts. The ELM parts classification system distinguishes 20 subsystems in a car (table 5). Altogether, it identifies 492 individual parts. Utilizing the relative frequency of the detailed parts listed for each plant, we converted this information on what each plant produces into a more aggregate system that distinguishes only 8 subsystems. They are body, chassis, drivetrain, engine attached (such as the exhaust system), engine electrical, engine proper, generic parts, as well as interior parts. The subsystem variables measure the share of individual parts codes in each of these by plant.

[25]A small downside of utilizing the information on plant age is that it is missing for 19 percent of the data. However, there seems to be no relation between that and the location of plants. For a slightly different treatment of such an estimation, see Klier, Ma, and McMillen (2004).

[26]Table 7 reports results for supplier pants that were not older than 10 years in 2003 (1994-2003). Estimating the model for a smaller set of "young" plants, the ones that opened between 1999 and 2003, yields robust results.

[27]Estimating the county-level model for the auto corridor only as well as for the entire U.S. produces robust results.

[28]To test for robustness of this exercise, I performed the same experiment on the model that estimates the location determinants for all supplier plants that opened between 1999 and 2003. The resulting redistribution of suppliers, while different in absolute numbers, represents a relative change of a similar order of magnitude as described above.

[29]That result is found to be robust when basing it on the locations of foreign supplier plants that opened since 1999 instead.

REFERENCES

Ellison, Glenn, and Edward L. Glaeser, 1997, "Geographic concentration in U.S. manufacturing industries: A dartboard approach," *Journal of Political Economy*. Vol. 105, No. 5, pp. 889–927.

Granholm, Jennifer M., 2004, "Automotive futures: Made in Michigan," speech given at a conference in Traverse City, MI, August 4.

Hakim, Danny, 2005, "Taking down the 'No foreign cars' signs in Michigan," *New York Times*, March 3.

Harbour and Associates, 2003, *The Harbour Report—North America 2003*, Troy, MI.

Klier, Thomas, 2005, "Caution ahead—Challenges to the Midwest's role in the auto industry," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, February, No. 211.

__________, 1999, "Agglomeration in the U.S. auto supplier industry," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 23, No. 1, pp. 18–34.

Klier, Thomas, Paul Ma, and Dan McMillen, 2004, "Comparing location decisions of domestic and foreign auto supplier plants," Federal Reserve Bank of Chicago, working paper, No. WP–2004–27.

Klier, Thomas, and William Testa, 2002, "Linkages across the border—The Great Lakes economy," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, July, No. 179b.

Krugman, Paul, 1991, *Geography and Trade*, Cambridge, MA: MIT Press.

Marshall, Alfred, 1920, *Principles of Economics*, London: McMillan.

McAlinden, Sean P., and Kim Hill, 2003, *The Market Renewal of Major Automotive Manufacturing Facilities in Traditional Automotive Communities*, Ann Arbor, MI: Automotive Communities Program, Center for Automotive Research, August.

Roland Berger Strategy Consultants, 2004, *The Odyssey of the Auto Industry—Suppliers' Changing Manufacturing Footprint*, Munich, Germany.

Rubenstein, James M., 1992, *The Changing U.S. Auto Industry—A Geographical Analysis*, London: Routledge.

Shirouzu, Norihiko, 2004, "Chain reaction—Big Three's outsourcing plan: Make parts suppliers do it," *Wall Street Journal*, June 10, p. A1.

Simon, Bernard, 2004, "Wheels of trade seize up at the world's busiest border," *Financial Times*, August 3, p. 3.

Smith, Donald, and Richard Florida, 1994, "Agglomeration and industrial location: An econometric analysis of Japanese-affiliated manufacturing establishments in automotive-related industries," *Journal of Urban Economics*, Vol. 36, No. 1, pp. 23–41.

Testa, William, Thomas Klier, and Richard Mattoon, 2005, "Challenges and prospects for Midwest manufacturing," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, March, No. 212b.

White, Joseph B., 2005, "Ford to pay up to $1.8 billion on Visteon," *Wall Street Journal*, May 26, p. A3.

Woodward, Douglas, 1992, "Locational determinants of Japanese manufacturing start-ups in the United States," *Southern Economic Journal*, Vol. 58, No. 3, pp. 690–708.

Switching primary federal regulators: Is it beneficial for U.S. banks?

Richard J. Rosen

Introduction and summary

In the United States, commercial banks can select among three primary federal regulators. A bank chooses a chartering agency and decides whether it will be a Federal Reserve System (Fed) member, thereby selecting its regulatory authority. A nationally chartered bank is regulated by the Office of the Comptroller of the Currency (OCC). If it is a Fed member, a state-chartered bank has the Fed as its primary federal regulator; otherwise, it is overseen by the Federal Deposit Insurance Company (FDIC).[1] By choosing its charter and deciding whether to be a Fed member, a bank effectively selects its regulator.

This article explores how banks use their option to select a regulator. Specifically, I examine banks that switch from one regulator to another. Is the ability to switch regulators a positive aspect of our current system? I offer some insight into this issue by examining whether banks benefit from switching and how switching affects social welfare. This study helps shed light on the behavior of regulators and the efficacy of the current system of multiple regulators. There has been debate about whether regulators, when setting policies, act in the public interest or not. This article builds on Rosen (2003), where I focused on whether the regulatory competition was beneficial or destructive. Competition could spur useful innovation or regulatory flexibility, thereby allowing banks to benefit without reducing social welfare. It could also be a "race for the bottom"—or a "competition for laxity," to use former Federal Reserve Chairman Arthur Burns's term—if regulators try to attract banks by easing restrictions on unsafe or unsound practices. The evidence presented here is not consistent with a race for the bottom, while there is some evidence of beneficial competition. In general, a bank's return either stays the same or increases after it switches regulators, while its risk of failure does not rise.

While most banks never switch regulators, the aggregate number of switchers is not small. Over 10 percent of banks switched regulators at least once during the period 1977–2003. I compare banks that switch with others that do not in an attempt to learn why banks switch. I find that, prior to changing regulators, switchers have approximately the same return on assets as other banks, and switchers are somewhat riskier. Small banks are less likely than large banks to switch regulators, but this is largely due to the fact that small banks are less likely to be in a bank holding company. Nonlead banks that are in a holding company are more likely to switch than either lead (largest) banks in a holding company or banks not in a holding company.

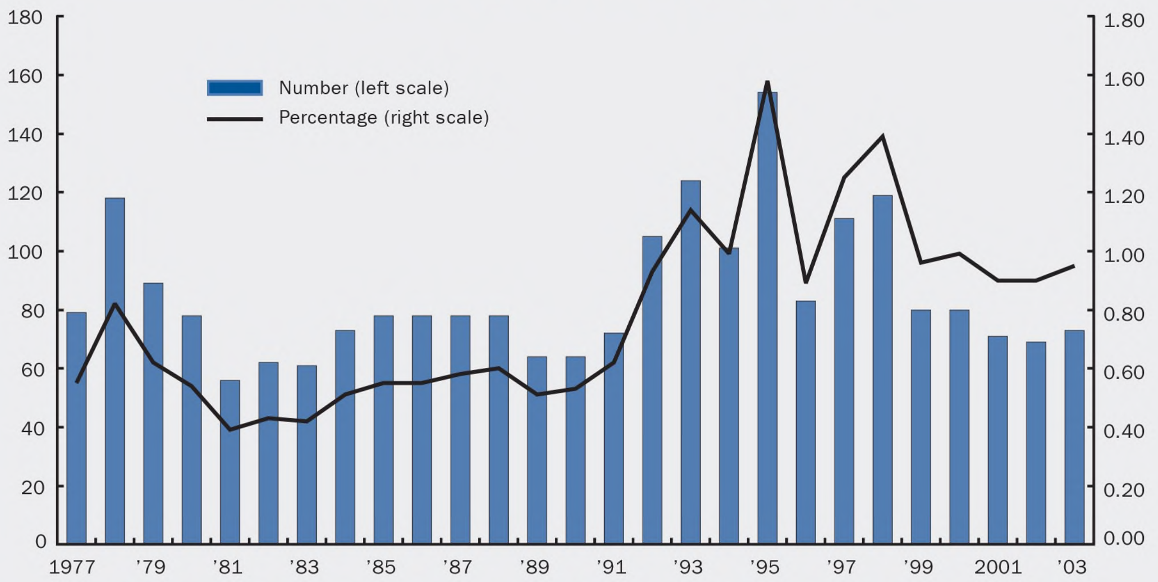
The effect of a switch on return and risk can indicate whether switches are beneficial. I find that banks generally increase their return when they switch regulators. There is little significant impact of a switch on risk. Banks tend to reduce their equity-to-asset ratio following a switch, but more inclusive measures of risk, such as the bank failure rate, point toward no increase in risk. An increase in return with no significant increase in risk is evidence consistent with beneficial competition. However, the aggregate results hide differences over the sample period in the performance of banks that switch.

The percentage of banks switching varies throughout my sample period—rising in the late 1970s, then falling to a lower rate in the 1980s, before rising again in the 1990s (see figure 1). There are many reasons why banks switch regulators, some of which may explain part of the pattern of switching over time. A switch

Richard J. Rosen is a senior economist and economic advisor at the Federal Reserve Bank of Chicago. *The author would like to thank Tara Rice, Craig Furfine, and the participants in a workshop at the Chicago Fed for their helpful comments.*

FIGURE 1

Banks that switch primary federal regulators, 1977–2003



Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

may be prompted by changes in the structure of a banking organization, issues relating to the interaction between a banking organization and its regulators, or a shift in strategy sought by a banking organization. The sample period I explore—1977–2003—was one of major changes in banking, both in the structure of the industry and in the regulatory framework under which it operated. I explore whether the characteristics of banks that switch regulators vary over time, perhaps indicating changing motivations for switching. I find that prior to 1992, switching regulators has little impact on overall risk and return. However, switches in the latter part of the sample, specifically 1992–2003, have a significant impact on performance. Banks that switch in this period show an increase in return without a commensurate increase in risk, as would be expected if there is beneficial competition. Note that the post-1991 period is also when the rate of switching is at its highest.

The plan of the article is as follows. First, I provide an overview of when banks switch primary regulators. Next, I review the arguments for and against a system in which banks can choose among multiple regulators. Then, I examine the characteristics of banks that switch primary regulators and present an analysis of how switches affect performance, including failure probabilities.

The pattern of banks switching primary regulators

Banks have been switching primary regulatory agencies for many years (Scott, 1977, documents switches from 1950 to 1974). I examine switches that occurred from 1977 to 2003, a period that covers major changes in banking and bank regulation. I identify the year a bank changes primary regulators from the *Reports of Income and Condition* (call reports). Table 1 gives an overview of the banks that switched primary regulators. As table 1 shows, there were 2,298 switches during the sample period, an average of 85 per year. Over the sample period, 10.8 percent of banks left their respective regulators at least once (0.7 percent of banks switched more than once). Table 1 also provides a breakdown of switches based on the size of the bank. The smallest banks were the least likely to switch.

The pattern of banks switching regulators can be partially explained by regulatory changes. In 1980, the Depository Institutions Deregulation and Monetary Control Act (DIDMCA) was passed. Prior to DIDMCA, there were important differences among regulators. For example, reserve requirements (the funds a bank must hold against specified deposit liabilities) depended on whether a bank was a member of the Federal Reserve System. DIDMCA leveled the playing field for all banks, regardless of membership in the Federal Reserve System. It is possible that many of the regulatory

TABLE 1

Banks that switch primary federal regulators, 1977–2003

Year	All switching banks: Number of banks	All switching banks: Percentage of banks	Total assets less than $1 billion: Number of banks	Total assets less than $1 billion: Percentage of banks	Total assets between $1 billion and $10 billion: Number of banks	Total assets between $1 billion and $10 billion: Percentage of banks	Total assets greater than $10 billion: Number of banks	Total assets greater than $10 billion: Percentage of banks
1977	79	0.55	79	0.55	0	0.00	0	0.00
1978	118	0.82	118	0.83	0	0.00	0	0.00
1979	89	0.62	89	0.63	0	0.00	0	0.00
1980	78	0.54	75	0.53	2	1.15	1	5.56
1981	56	0.39	55	0.39	1	0.54	0	0.00
1982	62	0.43	57	0.40	5	2.38	0	0.00
1983	61	0.42	58	0.41	3	1.29	0	0.00
1984	73	0.51	70	0.50	3	1.18	0	0.00
1985	78	0.55	77	0.55	1	0.35	0	0.00
1986	78	0.55	72	0.52	6	1.97	0	0.00
1987	78	0.58	76	0.58	2	0.63	0	0.00
1988	78	0.60	77	0.61	1	0.31	0	0.00
1989	64	0.51	61	0.50	3	0.97	0	0.00
1990	64	0.53	61	0.52	2	0.66	1	2.22
1991	72	0.62	68	0.60	4	1.37	0	0.00
1992	105	0.93	90	0.82	14	4.33	1	1.96
1993	124	1.14	111	1.06	12	3.74	1	1.82
1994	101	0.99	96	0.97	4	1.35	1	1.75
1995	154	1.58	140	1.50	12	3.88	2	2.90
1996	83	0.89	78	0.87	3	1.03	2	2.99
1997	111	1.25	101	1.18	10	3.79	0	0.00
1998	119	1.39	115	1.40	4	1.42	0	0.00
1999	80	0.96	76	0.95	4	1.46	0	0.00
2000	80	0.99	70	0.90	9	3.32	1	1.43
2001	71	0.90	62	0.82	7	2.50	2	3.03
2002	69	0.90	63	0.86	6	2.11	0	0.00
2003	73	0.95	63	0.87	8	2.46	2	2.63
Total	2,298	0.73	2,158	0.70	126	1.76	14	1.18

Note: Size classes are based on total assets in 2003 dollars.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition,* Washington, DC.

switches that occurred prior to and immediately after passage of DIDMCA were related to the changes instituted by the act rather than any actions of the regulators.

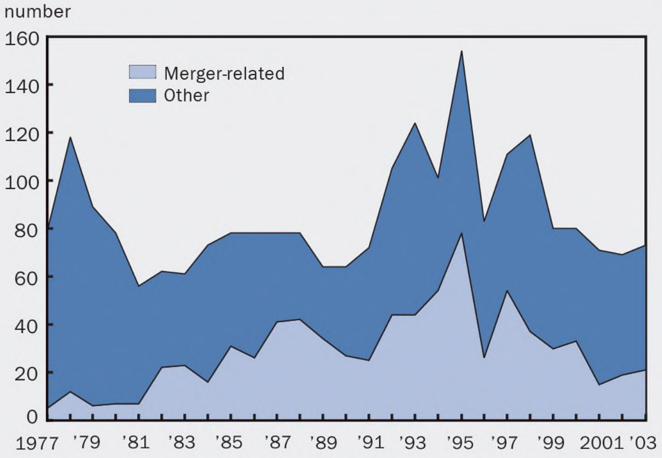
During the 1980s, states gradually reduced their restrictions on interstate and intrastate expansion (Amel, 1991; Amel and Starr-McCluer, 2002). This may have prompted the merger wave in the 1980s and could have led to some of the switches of primary regulators during that decade. In addition, the Riegle–Neal Act of 1994 removed the restrictions on interstate branching. This act was phased in over the next few years as states gradually adopted its provisions (Dick, 2006).

Merger activity, and switches associated with mergers, varied significantly over the sample period. Figure 2 gives the number of merger-related and other switches by year. I define a bank as having switched because of a merger if it switches its primary regulator in the year of its merger or the following year. If banks with different primary regulators merge, the newly formed bank will have to choose one of the two regulators. Following a merger, if the acquiring bank changes from its pre-merger regulator to the target bank's regulator, then I record this as a switch of primary regulators for the newly formed bank. A total of 779 banks switched regulators following a merger, one-third of all switches.

In the main analysis of the article that follows, I do not include banks that have recently merged. An objective of this article is to examine whether the ability of banks to switch regulators is a valuable option. This is difficult to do with switches following mergers for at least two reasons. The first is that, as noted above, banks with different primary regulators are forced to choose one. This leads to a different—and likely, lower—threshold for switching regulators than for non-merging banks. It is possible that the inclusion of

FIGURE 2

Merger-related and other switches



Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

banks that switch regulators concurrent with a merger will bias the results toward finding no impact from switching. The second reason for dropping merger-related switches is perhaps more important. A merger can significantly affect the reported return and risk for a bank. Costs related to the integration of the merging banks can depress the return for several years. Also, the decision to participate in a merger may be related to return and risk. Banks may be more likely to be merger targets when their return has been declining or when their risk has been increasing. This may bias the before-and-after comparison of return and risk for banks that merge.

Bank mergers affect not just the merging banks, but also other banks. The bank merger waves in the 1980s and the 1990s increased the average size of a bank dramatically. However, these waves had a much smaller effect on local market competition, with the average market concentration index essentially unchanged. These changes affected competition in local markets (see, for example, Berger, Udell, and Rosen, 2005). Some switches of regulators may have been partially in response to these repercussions of bank consolidation. These switches are included in the sample, since they do not suffer from the drawbacks noted in the previous paragraph.

DIDMCA and the merger waves may have induced some of the changes in my sample period. However, there are many switches that cannot be explained purely by regulatory changes or industry consolidation. In the next section, I examine additional possible explanations for switches of primary regulators.

Are multiple regulators beneficial?

There has been a debate over the best regulatory structure for a long time (see Rosen, 2003, for some examples). This section briefly explores why banks switch regulators and discusses some concerns about the current regulatory system, as well as some of its benefits.

When bank managers are asked why they change primary regulators, they generally respond in one of three ways. These managers claim that a bank switches because it can gain additional powers (as Chase Manhattan Bank did when it changed the primary regulator of its Delaware bank in 1990); save on regulatory compliance costs (as Chase Manhattan Bank did after its merger with Chemical Bank in 1995); or expand more easily nationwide (as HSBC USA did when it changed its charter in 2004). Broadly speaking, regulation at the three agencies—the OCC, the Fed, and the FDIC—and among the states (for state-chartered banks) is similar.[2] But for some banks, the differences among regulators might be important enough to induce a switch. During part of the sample period, for example, the insurance powers granted to banks varied among regulators. To conduct certain insurance activities, a bank might have needed to switch regulators. Thus, Chase Manhattan switched the regulator of its Delaware bank to allow it to sell insurance.

Some switches might be prompted because of the costs of regulation, which are both indirect and direct. The indirect costs include managerial and legal costs involved in meeting with bank examiners and making required reports. Indirect costs also involve the opportunity costs of restrictions on portfolio choices imposed on banks by regulators, such as reserve requirements and expedited funds availability.[3] There is no reason to believe that there are systematic differences in the indirect costs that banks would face at the different agencies. However, there are differences in direct costs. Both the OCC and the FDIC charge for bank exams, but the Fed does not. This may seem to give the Fed a cost advantage, but examination of state-chartered banks is shared with state regulators, who charge for their exams. Still, there can be cost differences among regulators. This may induce some switches if the OCC, the FDIC, or some states change the cost of exams (or if, because of competition in banking, a bank feels it has to squeeze out additional cost savings). Cost considerations may have also prompted some holding companies to simplify their regulatory structures.

From a social perspective, some question whether having multiple regulators is a good idea. There are several potential drawbacks to the current regulatory system. At minimum, having multiple regulators introduces complications. For instance, when J. P. Morgan Chase merged with Bank One in 2004, J. P. Morgan Chase Bank had a state charter and Bank One had a national charter. As part of the merger process, J. P. Morgan Chase had to decide which charter to adopt (and if it chose a state charter, whether to become a Fed member). This took time and resources that would not have been necessary if there were only a single bank regulator. Moreover, when it selected a national charter, the bank's former state regulator had to shift its personnel and pricing to account for the loss of a major bank. These costs may not be large, but they are certainly present.[4]

A potentially more serious issue is that regulators might not always act in the social interest. Stigler (1971) points out that regulators can be captured by the firms they cover because those inside a particular industry care a lot more about the regulators' decisions than outsiders do. As a result, they may choose policies that benefit banks rather than the public.

Related to this concern, the literature on regulatory structure explores a "race for the bottom" among regulatory agencies. In the 1970s, then-Fed Chairman Arthur Burns commented that he feared destructive competition among regulators for banks (their customers, in a sense). He brought up the possibility of what he called a "competition for laxity," a scenario in which banks would relax regulation to capture market share (see Scott, 1977). Since the budget of an agency depends in part on the number and size of the firms it regulates, regulators might compete against each other by offering lenient treatment in order to attract firms. When Chase Manhattan Bank elected to have a state rather than a national charter, subsequent to its merger with Chemical Bank in 1995, the OCC lost fees amounting to 2 percent of its budget. Similarly, when its successor, J. P. Morgan Chase Bank, returned to a national charter after its merger with Bank One in 2004, the New York Banking Department (the state regulatory agency) lost 27 percent of its revenues. If either agency was concerned with maximizing its budget, it would have an incentive to remove burdens on banks to keep them from switching.[5]

A race for the bottom could allow banks to manipulate the system. That is, banks might choose their primary federal regulator (and potentially, their state of incorporation, and thereby, their state regulator) to take actions that benefit the bank but are not in the public interest. An example of this would be a bank that switched regulators in order to adopt a new, risky strategy (or to hide risks it was already taking). The risk could increase the exposure of the deposit insurance fund. It is important to note that a bank can only switch to a new regulator if that regulator approves. Thus, regulators have the ability to block switches of this kind.

On the other hand, having multiple regulators offers potential benefits. A single regulator might have less incentive to allow banks to undertake new powers or to use new products. There is a natural tendency for regulators to be risk averse, since they are assigned blame for anything that goes wrong, but may not be recognized for permitting beneficial changes. Potentially beneficial changes that one regulator views as too risky might be adopted by another regulator. In addition, having multiple regulators allows for some specialization. Tiebout (1956) presents a model of public goods provision by local communities that has often been modified to examine other regulatory issues. The Tiebout framework can be used to show that under certain conditions (including when there are no externalities and there is costless mobility), regulatory competition leads to optimal standards setting. Different localities can offer distinct menus of public goods, with each individual choosing the menu best suited for that individual (referred to as Tiebout sorting). This model underlies the arguments for local control of securities regulation (Romano, 1998), antitrust enforcement (Easterbrook and Fischel, 1991), and environmental policy (Revesz, 2000). These papers also claim that the benefits of competition among local agencies eliminate (or should eliminate) a race for the bottom.

Connected to Tiebout sorting, another reason that banks might switch is that regulatory enforcement may differ among agencies. There may be an explicit policy shift at a particular agency. For example, in 1991, Federal Reserve Chairman Alan Greenspan was worried that examiners were contributing to a "credit crunch" by requiring banks to hold too much capital against loans. This was interpreted by some as a signal for examiners to relax enforcement. This could have encouraged banks to switch to the Fed from the other agencies.[6]

An additional complication to this analysis is that a bank regulatory agency is essentially a collection of examiners. Unlike regulators in many other areas, examiners in banking frequently make subjective decisions about the banks they visit.[7] Berger, Kyle, and Scalise (2000) review examiner and regulatory agency discretion when monitoring banks. Examiners go into a bank to evaluate its risk. Based on this assessment, the examiners decide whether the bank's reserve for loan losses is sufficient, and then they assign a strength rating—the CAMELS (capital, asset quality, management, earnings, liquidity, and sensitivity) rating—to the bank. If a bank wants to change its portfolio, its examiners must decide how to react. The examiners can either accede to the change or make it costly for the bank by requesting a higher loan loss reserve (resulting in a charge against income) or by giving the bank a lower CAMELS rating (resulting in greater regulatory costs for the bank). Thus, to an extent, examiners can decide how costly it is for a bank to add risk. Having multiple regulators and the ability to switch among them allows the bank to escape examiners that the bank feels are out of line.

One potential problem that a bank might have is that its examiners can exploit the discretion they have when assessing the bank to serve their own ends. Some examiners may be interested in leading a "quiet life" (Rosen, 2003).[8] That is, they may want to get by with as little work and as little career risk as possible. To get a quiet life, some examiners might prefer to regulate banks with portfolios that are as simple as possible to evaluate.

There is another reason why examiners may put up roadblocks to change by banks. Regulatory behavior may be influenced by a desire to avoid criticism from groups other than the firms that examiners assess. Importantly, Congress and public interest groups may criticize ex post actions that were proper ex ante (as Kane, 1989, argues they did early in the savings and loan crisis in the 1980s). This gives regulatory agencies and, by extension, examiners an incentive to avoid actions that could increase the risk of bank failure. Fear of criticism may induce risk aversion on the part of examiners who want a quiet life.

Whether having the ability to switch regulators leads to beneficial competition or a race for the bottom can be tested by examining which banks switch and how switching affects the performance of these banks. The key here is to decide which switches are "beneficial" and which are not. A beneficial switch allows a bank to move to a better risk–return tradeoff without increasing societal risk. I used the risk of bank failure as a proxy for societal risk. Banks are overseen by government agencies for many reasons. For instance, banks are regulated in order to maintain a smoothly operating payments system and to confirm that their deposits are insured. Both of these objectives imply that regulators want to limit excessive risk-taking by banks, which should limit bank failures. A race for the bottom might work this way: Regulators could allow banks that switch to increase societal risk without a compensatory increase in return. Bank managers or shareholders could profit from this, but only by taking advantage of the deposit insurance system. Beneficial competition among regulators, on the other hand, would allow banks to move to a better risk–return trade-off without increasing failure probabilities.[9] Note that these tests are sufficient to indicate beneficial competition or a race for the bottom, but there are other factors that may not be figured in. Beneficial competition can help all banks, not just those that switch. I cannot directly test for this, but the increase in bank profits and decrease in bank failures over the past 15 years are consistent with beneficial competition—and not a race for the bottom. Still, since these trends are also a function of macroeconomic factors, this is at best weak evidence.

Characteristics of banks that switch regulators

To evaluate banks that switch regulators, I need measures of return and risk. Return is easy to measure. I use the return on assets (*ROA*), but its results are similar to other measures, such as the return on equity. Unfortunately, there is no simple inclusive measure of risk. I use direct and indirect risk evaluations. The direct measure of risk I use is a failure prediction model. As noted above, bank failures can reduce the smooth operating of the payments system and increase losses to the deposit insurance fund. Thus, if a regulator allows banks that switch to take actions that increase their failure probabilities, this suggests a race for the bottom. To attain a second estimate of failure probability and to determine how risk changes relative to return, I use four accounting ratios that capture different aspects of risk. The most direct is the Sharpe ratio, which is the ratio of ROA to the standard deviation of the ROA (again, the results are similar to the return on equity). To construct this measure for year t, I use the ROA for year t as the numerator. The denominator is the standard deviation of the semiannual ROA (expressed as an annual return) for all the periods from year $t-4$ to year t for which return data exists. I keep all observations with at least two years of return data as of year t. Even with ten semiannual periods, I do not have a very precise measure of risk. Still, while noisy, the ratio of ROA to its standard deviation does give a picture of the risk–return trade-off.

I also use other accounting measures of risk traditionally used to evaluate banks. The equity-to-asset ratio (*EQUITY/ASSET*) is a measure of leverage, with higher values indicating lower risk, since equity

TABLE 2

Performance of banks prior to switching regulators

	Banks that switch			Banks that never switch			
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	Test of difference of means (p value)
ROA	0.93	1.00	0.65	0.93	1.03	0.78	0.842
SHARPE RATIO[a]	0.84	0.89	0.61	0.87	0.93	0.59	0.151
EQUITY/ASSET	8.69	8.11	2.54	9.18	8.59	2.80	0.000***
LOAN/ASSET	55.93	57.02	13.79	53.92	55.08	13.99	0.000***
CHRG/LOAN	0.43	0.19	0.87	0.56	0.22	1.70	0.000***
DEP/LIAB	95.50	98.13	7.98	96.58	98.42	6.14	0.000***
LOG ASSETS	7.88	7.85	0.51	7.77	7.73	0.48	0.000***
NONHC BANK	0.34	0	0.47	0.44	0	0.50	0.000***
LEADBANK HC	0.41	0	0.49	0.44	0	0.50	0.024**
NONLEAD SREG	0.10	0	0.30	0.07	0	0.26	0.003***
NONLEAD DREG	0.16	0	0.36	0.05	0	0.21	0.000***
OCC	0.43	0	0.49	0.30	0	0.46	0.000***
FED	0.16	0	0.36	0.07	0	0.26	0.000***
FDIC	0.42	0	0.49	0.63	1	0.48	0.000***
Observations		1,246			231,948		

[a]The Sharpe ratio only includes banks with at least two years of data.
**Significant at 5 percent level.
***Significant at 1 percent level.
Notes: Banks that switch regulators include all banks that switch regulators, except those that switch in the year of or year following a merger. Variable definitions are given in the text. The data are year-end (except for *ROA*, which is for the full year) for the period 1977–2003. Data for switchers are from the year prior to a switch. The variables *ROA, EQUITY/ASSET, LOAN/ASSET, CHRG/LOAN*, and *DEP/LIAB* are expressed as percentages. All other variables, except *LOG ASSETS*, are expressed as ratios.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

offers a cushion against failure. The loan-to-asset ratio (*LOAN/ASSET*) is likely to be correlated with risk as well. Loans are among the riskiest assets on bank balance sheets. A bank with more loans, all else being equal, is more likely to fail. However, loans can vary significantly in risk. To measure the riskiness of a loan portfolio, I use the charge-off-to-loan ratio (*CHRG/LOAN*).[10] This ratio reflects expected losses on loans made in the past. A riskier loan portfolio, all else being equal, has higher charge-offs. Charge-offs can also reflect bad luck, poor management, or investments in risky but predictable loans (for example, some credit card loans). To capture risk differences on the liability side, I use the ratio of deposits to liabilities (*DEP/LIAB*). Deposits are a more stable source of funding than other liabilities, such as loans from other banks. Results based on these ratios should be viewed with caution, since they may be associated with changes in productivity as well as risk.

The loan-to-asset ratio and the charge-off-to-loan ratio can also be viewed as proxies for the workloads of bank examiners. Examiners have to spend more effort when reviewing loans than other assets, and they have to spend even more effort when reviewing nonperforming loans than other loans. If examiners desire a quiet life, they prefer banks to have nonloan assets, such as cash and government securities, and they are inclined toward banks that do not issue loans with a high probability of becoming nonperforming.

To assess whether a switch of primary federal regulators is beneficial, it is useful to know what leads a bank to switch regulators. To do this, I use a simple model to predict which banks will switch regulators as a function of the return and risk characteristics of the banks. The dependent variable, *SWITCH*, is a dummy that takes the value 1 in year t if a bank switches regulators in the year $t + 1$. The model is:

1) *SWITCH* = *f*(*ROA*, *SHARPE RATIO*, *EQUITY/ASSET*, *LOAN/ASSET*, *CHRG/LOAN*, *DEP/LIAB*, *control variables*).

When analyzing the data, I drop banks in any year that they are in the top or bottom 1 percent of *ROA*, *EQUITY/ASSET*, *LOAN/ASSET*, or *DEP/LIAB*.

To examine whether banks that switch regulators are different from other banks, it is important to control for reasons unrelated to return and risk that might lead a bank to shift its primary regulator. Table 1 shows that small banks are disproportionately less likely than larger banks to switch regulators. For this reason, I control for bank size using the log of total assets (*LOG ASSETS*). Structural considerations may play a role in the decision to switch. I control for holding company status, using dummies for whether the bank is the lead bank in a holding company (*LEADBANK HC*), or whether it is a non-lead bank within a holding company that has the same (*NONLEAD SREG*) or different (*NONLEAD DREG*) charter than the lead bank. Banks not in a holding company (*NONHC BANK*) compose the excluded category. This allows us to test for switches that reduce the number of regulators to which a holding company reports. There may also be other differences across primary regulators. To control for this, I include dummies for whether a bank is regulated by the Federal Reserve or the FDIC at the end of year $t - 1$ (the OCC is the excluded category). Finally, I include year dummies to control for systemic changes, such as changes in overall levels of return and risk in the industry as a whole.

Table 2 reports summary statistics for the return, risk, and control variables. Banks that switch regulators have a similar return and Sharpe ratio to other banks. There are differences between the two groups in the other risk measures. The equity-to-asset, loan-to-asset, and deposit-to-liability ratios all indicate higher risk for switchers than for other banks that have not switched, but switchers have a lower charge-off-to-loan ratio. However, I need to account for correlations among these variables and patterns in the proportion and type of banks that switch. I do this using a regression framework.

Equation 1 is estimated using a logistic regression. The results of the regression are reported in the first column of table 3. Consistent with the univariate statistics, the coefficient on *ROA* is not statistically significantly different from zero. So, I cannot use a bank's return to predict whether it will switch regulators. Most of the risk variables, on the other hand, are significant and can help predict which banks will switch. Banks with a lower Sharpe ratio, more leverage, and a lower deposit-to-liability ratio—all indicators of higher risk—are more likely to switch. Pointing to a trend in the other direction, banks with fewer charge-offs, signaling less risk, are also more likely to switch.[11]

As figure 1 and table 1 show, the proportion of banks that switch regulators varies over time. It is possible that the strength of banks varies along with switching intensity. To test this, I divide my sample period into two smaller periods. The early period includes all switches from 1977 to 1991. This covers the implementation of DIDMCA and the lesser wave of mergers in the 1980s.[12] The late period includes all switches from 1992 to 2003. This includes the peak of bank consolidation. This is also the time when the proportion of banks that switch regulators is largest.

The second and third columns of table 3 present regression results for the two smaller periods. There are differences across the two periods in the magnitude and statistical significance of the return and risk variables. For example, the coefficients on the Sharpe ratio and the charge-off-to-loan ratio are larger and statistically significant only in the late period. Still, the overall pattern is similar. Return is not a predictor of switching in either period, and banks that switch look somewhat riskier in every dimension except their level of charge-offs.

The control variables differ in important ways across the two periods. In the early period, 1977–91, small banks, all else being equal, are more likely to switch regulators. This is reversed in the late period, 1992–2003, when large banks are more likely to switch. Overall, banks that are not the lead bank in a holding company are more likely to switch than either lead banks or banks not in a holding company. Consistent with a desire to simplify the regulatory structure of their respective holding companies, non-lead banks that have different charters than their lead banks switch more often in both periods. Non-lead banks with the same regulator as their lead banks are only more likely to switch in the early period. This may reflect banks switching to exploit differences among regulators in the types of investments allowed, such as insurance activities. These differences tended to be larger in the early period than in the late period, especially once the Financial Modernization Act (also known as the Gramm–Leach–Bliley Act) was passed in 1999. The regulatory dummies also provide some interesting insights. Banks are more likely to switch from the Fed than either the OCC (the omitted regulator) or the FDIC in the early period. In the late period, however, banks under the Fed are less likely to switch than those under the OCC—and as likely as those under the FDIC. Over the entire period, banks under the FDIC are the least likely to switch.

The results suggest that banks that switch are different from those banks that do not. They also suggest that these differences depend on when the banks switch. These findings do not help determine whether there is a race for the bottom or beneficial competition, but they point out the importance of controlling for why and when banks switch, as well as other bank characteristics.

TABLE 3

Probability that bank will switch regulators in the next year

	Full sample	Early period (1977–91)	Late period (1992–2003)
ROA	0.020	0.031	–0.007
	(0.451)	(0.369)	(0.872)
SHARPE RATIO	–0.079	–0.056	–0.126
	(0.035)**	(0.233)	(0.018)**
EQUITY/ASSET	–0.023	–0.016	–0.032
	(0.000)***	(0.033)**	(0.000)***
LOAN/ASSET	0.0005	–0.001	0.002
	(0.564)	(0.380)	(0.048)**
CHRG/LOAN	–0.036	–0.024	–0.059
	(0.011)**	(0.121)	(0.038)**
DEP/LIAB	–0.003	–0.005	–0.001
	(0.081)*	(0.039)**	(0.619)
LOG ASSETS	0.005	–0.067	0.102
	(0.864)	(0.068)*	(0.008)***
LEADBANK HC	0.019	0.050	–0.036
	(0.483)	(0.159)	(0.369)
NONLEAD SREG	0.121	0.190	–0.005
	(0.002)***	(0.000)***	(0.933)
NONLEAD DREG	0.489	0.577	0.347
	(0.000)***	(0.000)***	(0.000)***
FED	0.079	0.215	–0.160
	(0.017)**	(0.000)***	(0.004)***
FDIC	–0.274	–0.356	–0.184
	(0.000)***	(0.000)***	(0.000)***
Pseudo-R^2	0.044	0.060	0.027
Observations	243,714	165,268	78,446

*Significant at 10 percent level.
**Significant at 5 percent level.
***Significant at 1 percent level.
Notes: The data are from 1977 to 2003, with year dummies not shown. The dependent variable is a dummy for whether a bank switches regulators in the next calendar year. Other variable definitions are given in the text. Robust p values adjusted for cluster effects are in parentheses.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

Performance of banks that switch regulators

In this section, I examine the change in performance at banks that switch regulators, comparing return and risk before and after a switch. This allows us to address two issues. The first is whether switching is good for banks and the second is whether it is good for society. Beneficial competition implies that banks can benefit from switching while the probability of bank failure (our proxy for social welfare) does not increase. If switching allows banks to take actions that increase the risk of bank failure, then that is evidence consistent with a race for the bottom. The section is divided into two parts. I examine the accounting measures of performance in the first part and then a failure prediction model in the second part.

Accounting measures of performance

To examine how performance changes preceding and following a switch of regulators, I use the following model:

2) *Performance* = *f*(*Pre-change indicators*, *Post-change indicators*, *Control variables*),

where *performance* is measured using our return and risk variables. The model is estimated for the entire sample of banks, not just banks that switched. This allows us to compare changes in performance at banks that switch with otherwise similar banks that have not.

A priori, there is no reason to believe that the changes induced by a switch of regulators should be immediately reflected in the performance. For this reason, I look over five-year periods before and after a switch. This allows a long enough time before a switch to see whether there was some change in a bank's performance that might prompt a switch. It also allows a long enough time after a switch to ensure that all the changes that result from it are reflected in the accounting data I examine. For banks that switch regulators, I use dummy variables for pre- and post-switch periods as well as a trend variable. Let *DUMMY PRE*, the pre-switch dummy, equal 1 for each of the five years prior to a switch (year $t-5$ to $t-1$ for a switch in year t) and equal 0 otherwise. Similarly, let *DUMMY POST*, the post-switch dummy, equal 1 for each of the five years following a switch (year $t+1$ to $t+5$ for a switch in year t) and equal 0 otherwise. For banks that never switch, both *DUMMY PRE* and *DUMMY POST* equal 0. I set the trend variables so that they are increasing in time. For banks that switch in year t, let *TREND PRE* take the value 1 in year $t-5$, 2 in year $t-4$, and so on until it has the value 5 in year $t-1$. For other years and other banks, it equals 0. Similarly, define *TREND POST* as taking the value 1 in year $t+1$, 2 in year $t+2$, and so on until it has the value 5 in year $t + 5$ for switchers, and the value 0 otherwise.

The control variables are similar to those in the prediction model in the previous section. The risk choices a bank makes affect its return and vice versa. Thus, I include risk and return variables as controls (excluding the performance measure being estimated). Results are qualitatively similar without these controls. I also include the structural controls from the prediction model. These cover the holding company status and regulator of a bank. Finally, I use the log of total assets as a control, since larger banks are more diversified, all else being equal, and year dummies to control for systemic changes.

Table 4 presents the results of regressions using equation 2 for the risk and return measures. Panel A gives the regression coefficients. Some but not all of the trend and dummy variables are significant. What I am most concerned with is the net change following a switch. For example, the post-switch trend is significant for the ROA, but the post-switch is not. What does this say about the net change in ROA? To get an idea of how important these changes are, it is necessary to combine the trend and dummy variables. For example, five years prior to a switch, the average bank has an ROA that is 0.016 percentage points below that of an otherwise similar bank that never switches ($0.016 = -0.018 + 0.002 \times 1$). By the year before the switch, ROA is 0.008 percentage points below that of an otherwise similar bank that never switches ($0.008 = -0.018 + 0.002 \times 5$ with rounding), indicating an increase of 0.008 percentage points in ROA in the four years before a bank switches. The increasing return after a switch is such that five years after a switch, the average bank has an ROA that is 0.083 percentage points above that of an otherwise similar bank that never switches ($0.083 = 0.008 + 0.015 \times 5$ with rounding). Panel B of table 4 presents the estimated changes for the years before and after a switch.

The results for the period prior to a switch indicate that banks are changing their balance sheets significantly prior to a switch. Leverage increases as the equity-to-asset ratio falls. Banks are also shedding loans. If examiners want a quiet life, then changes such as these may make them unhappy. This may, in turn, make it more probable that a bank will switch regulators.

The results in panel B of table 4 show that return rises significantly in the five years after a switch. They also provide evidence on the accounting risk measures. Overall, the picture on risk changes before and after a switch is mixed. The key factors are that the Sharpe ratio is unchanged but the equity-to-asset ratio decreases heading into and following a switch.

To get an idea of how the risk and return changes compare after a switch, I use the data in panel B of table 4 to compare the percentage change in the equity-to-asset ratio (the risk measure that increases) to the percentage change in ROA. The percentage change is measured by dividing the change by the pre-switch mean and is given in the final row of the table. From the year prior to a switch to five years after the switch, ROA is estimated to increase by 0.091 percentage points, 9.8 percent of the average ROA prior to the switch. Over a similar period, the equity-to-asset ratio is estimated to decrease by 0.463 percentage points, 5.3 percent of the average ratio prior to a switch. Thus, return increases by a larger fraction than the risk (as measured by the accounting variables) increases. This, in combination with no significant change in the Sharpe ratio, suggests that banks do better following a switch and provides no evidence that social risk increases.

Recall that the regression results in table 3 show that the factors that lead banks to switch regulators have changed over time. It makes sense, then, to see whether the performance of banks before and after a switch differs over time. To do this, let *EARLY* be a dummy variable that takes the value 1 if a bank switches regulators between 1977 and 1991, and let *LATE* be a dummy variable that takes the value 1 if a bank switches regulators between 1992 and 2003. I create a series of eight interaction variables using these dummies. Each interaction variable is the product of one of the period dummies and either *TREND PRE*, *DUMMY PRE*, *TREND POST*, or *DUMMY POST*. Using these variables, I estimate equation 2. Rather than presenting the entire regression results, table 5 gives the estimated changes relative to otherwise similar banks that have not switched in the accounting return and risk measures for the three periods, mirroring panel B of table 4.[13] It is clear from the table that the effect of switching on return appears only in the late period. Return increases significantly in the late period but changes little in the early period. The effect of a switch on risk is mixed in both periods but is driven by different factors in each. In the early period, the Sharpe ratio is unchanged and charge-offs decrease, signaling no change or a decrease in risk. But, the equity-to-asset ratio increases, indicating higher risk. In the late period, on the other hand, the Sharpe ratio increases (in large part due to the increase in ROA), while charge-offs also increase. The only constant is that banks add leverage following a switch in both periods. Also, while the coefficient on the change after a switch in the late period is positive for the charge-off-to-loan ratio regression, a deeper examination of the data (not shown) indicates that the positive coefficient reflects

TABLE 4

Performance regressions

A. Regression coefficients

	Return on assets	Sharpe ratio	Equity-to-asset ratio	Loan-to-asset ratio	Charge-off-to-loan ratio	Deposit-to-liability ratio
TREND PRE	0.002	–0.002	–0.079	–0.502	–0.004	–0.067
	(0.724)	(0.746)	(0.000)***	(0.000)***	(0.578)	(0.254)
DUMMY PRE	–0.018	–0.018	0.021	2.724	–0.008	–0.019
	(0.533)	(0.441)	(0.832)	(0.000)***	(0.805)	(0.937)
TREND POST	0.015	0.002	–0.097	–0.222	0.005	0.102
	(0.045)**	(0.671)	(0.000)***	(0.102)	(0.690)	(0.129)
DUMMY POST	0.008	–0.006	–0.353	0.938	–0.047	–0.349
	(0.739)	(0.779)	(0.000)***	(0.044)**	(0.145)	(0.199)
ROA			1.467	–0.026	–0.672	0.415
			(0.000)***	(0.855)	(0.000)***	(0.000)***
SHARPE RATIO			–0.617	–1.410	–0.234	–0.014
			(0.000)***	(0.000)***	(0.000)***	(0.797)
EQUITY/ASSET	0.071	0.033		–1.310	0.027	0.019
	(0.000)***	(0.000)***		(0.000)***	(0.000)***	(0.212)
LOAN/ASSET	–0.003	–0.002	–0.041		–0.004	0.002
	(0.000)***	(0.000)***	(0.000)***		(0.000)***	(0.527)
CHRG/LOAN	–0.138	–0.096	0.058	–0.279		0.007
	(0.066)*	(0.066)*	(0.092)*	(0.021)**		(0.660)
DEP/LIAB	0.007	0.003	0.003	0.014	0.001	
	(0.000)***	(0.000)***	(0.219)	(0.524)	(0.669)	
LOG ASSETS	0.267	0.164	–1.159	2.348	0.077	–5.152
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
LEADBANK HC	0.163	0.068	–0.913	0.932	0.074	–0.453
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
NONLEAD SREG	0.059	0.013	–0.991	2.225	–0.001	–1.168
	(0.000)***	(0.107)	(0.000)***	(0.000)***	(0.905)	(0.000)***
NONLEAD DREG	0.065	–0.002	–0.926	1.611	0.031	–0.782
	(0.000)***	(0.810)	(0.000)***	(0.000)***	(0.533)	(0.000)***
FED	0.001	0.018	0.243	2.056	0.002	–0.790
	(0.956)	(0.036)**	(0.000)***	(0.000)***	(0.962)	(0.000)***
FDIC	0.103	0.063	0.056	1.698	0.062	–0.069
	(0.000)***	(0.000)***	(0.072)*	(0.000)***	(0.000)***	(0.334)
Observations	253,291	249,988	249,988	249,988	249,988	249,988
R^2	0.245	0.188	0.312	0.159	0.167	0.212

B. Estimated changes in accounting variables

	Return on assets	Sharpe ratio	Equity-to-asset ratio	Loan-to-asset ratio	Charge-off-to-loan ratio	Deposit-to-liability ratio
Change from 5 years prior to switch to 1 year prior to switch	0.008 (0.724)	–0.008 (0.746)	–0.315 (0.000)***	–2.009 (0.000)***	–0.017 (0.578)	–0.268 (0.254)
Change from 1 year prior to switch to 1 year after switch	0.029 (0.195)	0.025 (0.197)	–0.076 (0.000)***	–0.503 (0.223)	–0.013 (0.629)	0.107 (0.638)
Change from 1 year prior to 5 years after switch	0.091 (0.001)***	0.034 (0.139)	–0.463 (0.000)***	–0.396 (0.500)	0.006 (0.897)	0.517 (0.072)*
Change from 1 year prior to 5 years after switch divided by sample mean	0.098	0.040	–0.053	–0.007	0.014	0.005

*Significant at 10 percent level.
**Significant at 5 percent level.
***Significant at 1 percent level.
Notes: The data are from 1977 to 2003, with year dummies not shown. Variable definitions are given in the text. For both panels A and B, robust p values adjusted for cluster effects are in parentheses.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition,* Washington, DC.

TABLE 5

Performance by periods, early (1977–91) and late (1992–2003)

	Return on assets		Sharpe ratio		Equity-to-asset ratio	
	Early	Late	Early	Late	Early	Late
Change from 5 years prior to switch to 1 year prior to switch	0.005 (0.894)	0.025 (0.553)	–0.033 (0.319)	0.046 (0.161)	–0.299 (0.008)***	–0.235 (0.132)
Change from 1 year prior to switch to 1 year after switch	–0.046 (0.177)	0.115 (0.000)***	–0.027 (0.363)	0.081 (0.001)***	–0.038 (0.673)	–0.116 (0.331)
Change from 1 year prior to 5 years after switch	–0.012 (0.774)	0.211 (0.000)***	–0.024 (0.490)	0.101 (0.000)***	–0.324 (0.005)***	–0.629 (0.000)***
Change from 1 year prior to 5 years after switch divided by sample mean	0.014	0.205	0.029	0.117	0.039	0.068

	Loan-to-asset ratio		Charge-off-to-loan ratio		Deposit-to-liability ratio	
	Early	Late	Early	Late	Early	Late
Change from 5 years prior to switch to 1 year prior to switch	–2.811 (0.000)***	–1.019 (0.189)	–0.017 (0.690)	–0.017 (0.704)	–0.439 (0.137)	0.057 (0.894)
Change from 1 year prior to switch to 1 year after switch	0.579 (0.324)	0.299 (0.642)	–0.417 (0.066)*	0.053 (0.085)*	–0.243 (0.885)	–0.211 (0.557)
Change from 1 year prior to 5 years after switch	0.623 (0.403)	–1.634 (0.078)*	–0.147 (0.003)***	0.186 (0.013)**	0.578 (0.094)*	0.419 (0.401)
Change from 1 year prior to 5 years after switch divided by sample mean	0.012	0.027	0.258	0.736	0.006	0.004

*Significant at 10 percent level.
**Significant at 5 percent level.
***Significant at 1 percent level.
Notes: The results are based on regressions of equation 1 with interaction terms between the period dummies and the pre- and post-switch dummies and trend variables. Each regression has 249,988 observations. The change variables are calculated based on the coefficients on the interaction involving the pre- and post-switch dummies and trend variables. Variable definitions are given in the text. Robust p values adjusted for cluster effects are given in parentheses.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

a large decline in charge-offs at banks that do not switch rather than an increase at banks that switch.

The results in table 5 point out characteristics of switchers and switching in the two periods, 1977–91 and 1992–2003. Prior to a switch, there is not much difference between the banks that switch in the early and late periods. Return is flat, and the equity-to-asset and loan-to-asset ratios are decreasing, although the change in the two ratios is only significant in the early period.

The major differences between the two periods are in the change in return and risk at banks that switch. The most important is that return increases after a switch only in the late period. In the early period, the average change in return is statistically and economically insignificant. The findings for risk are mixed in both periods. In the early period, there is no change in the Sharpe ratio following a switch. This may reflect the balancing of higher leverage and lower charge-offs. In the late period, on the other hand, the Sharpe ratio is increasing following a switch, indicating a reduction in risk. However, leverage and charge-offs are increasing, signifying higher risk.

The evidence using the accounting data is consistent with beneficial competition, but only in the post-1991 period. Return increases after a switch in the late period, but not in the early period. The results for risk are mixed, but there is no strong indication of higher risk. In the late period, the best measure of risk—the Sharpe ratio—signals a reduction in risk after a switch. These findings are indicative of beneficial competition among regulatory agencies. However, before drawing stronger conclusions, I need to examine the direct measure of failure probabilities.

Failure probability model

The accounting data present a mixed picture of how switching primary federal regulators affects risk. From a social perspective, the critical issue is whether the changes in risk promote bank failure. To directly examine this, I use a failure prediction model.

I use two approaches to determine whether switching regulators makes a bank more likely to fail than if it had not switched. First, I estimate a failure prediction model with a dummy for whether a bank has recently switched regulators. Since so few banks fail in any given year (approximately 0.5 percent per year), I look at three- and five-year horizons to minimize noise in the model. Let *FAIL DUMMY X* be a variable that takes the value 1 in year t if a bank fails prior to the end of year $t + x$, where x is either 3 years or 5 years. In my sample, an average of 1.5 percent of banks fail over a three-year horizon and 2.3 percent fail over a five-year horizon. For banks that switch regulators, I include just data for the years following the switch because a bank's decision to switch is only observed if it survives long enough to complete the switch. To include switches in the failure prediction model, let *SWITCH* be a dummy variable that takes the value 1 if a bank has switched regulators within the past three years, and the value 0 otherwise.[14] I also interact *SWITCH* with the period dummies.

For the prediction model, failure is assumed to depend on the accounting return and risk measures used earlier, as well as the log of total assets (since larger banks are more diversified) and year dummies to capture systemic movements in failure probabilities:

3) *FAIL DUMMY X* = *f*(*SWITCH, LOG ASSETS, ROA, SHARPE RATIO, EQUITY/ASSET, LOAN/ASSET, CHRG/LOAN, DEP/LIAB, year dummies*).

I estimate the model two ways. First, to establish a baseline, I only include observations for banks that never switch. Then, I include all banks. The model is estimated over the years 1977–2001 to allow at least three years after a switch for banks to potentially fail (since I have failure data through 2004).

In the analysis of the accounting data, I dropped outliers because they often have a disproportionate effect on regression results. In this section, on the other hand, all observations are included except banks with negative equity (since these have effectively failed already). This is because it is precisely the outlier banks, at least those in the lower tail, that are most likely to fail in the near term. Excluding the outliers pushes the results more toward switches reducing the probability of failure, although, for the most part, the differences are not statistically significant.

Table 6 presents the results of estimating equation 3 using a logistic regression. The signs of the coefficients on the control variables are consistent with expectations. Increasing either size or return decreases failure probability, while increasing risk has the opposite effect. The coefficient on *SWITCH* is statistically insignificant for both the three- and five-year failure windows. This is not consistent with the hypothesis that, all else being equal, a bank that has recently switched regulators is more likely to fail than an otherwise similar bank that has never switched.

The final column of table 6 includes the interaction terms between *SWITCH* and the time dummies. In this regression, the coefficient on *SWITCH LATE* is positive and significant. The positive coefficient on *SWITCH LATE* is consistent with the hypothesis that, all else being equal, a bank that switched regulators in the late period is more likely to fail than an otherwise similar bank that has never switched.

The careful wording in the last sentences of the previous two paragraphs reflects an assumption implicit in the failure prediction model (equation 3). The model assumes that a switching bank would have the same risk–return profile whether or not it had switched. In essence, it rules out the possibility that a bank is able to, or chooses to, change its portfolio precisely because it has switched regulators. For example, a regulator involved in a race for the bottom might attract new banks by allowing those banks to greatly increase leverage (that is, decrease their equity-to-asset ratio) after they switched to its oversight. If banks that switched increased leverage, they would be more likely to fail. However, if these banks failed at the rate that otherwise similar banks *with their new level of leverage* failed, then the coefficients on the switch dummies in equation 3 would not be significantly positive. Related to this proposition, if regulatory specialization allows banks that switch regulators to increase return and reduce their failure rate, but the failure rate is still above that at otherwise similar banks *with their new ROA*, then the coefficients on the switch dummies in equation 3 would be significantly positive. Since ROA increases for banks that switch regulators in the late period, this means that the significant positive coefficient on *SWITCH LATE* does not necessarily imply that there is a race for the bottom in that period.

A second approach is to assume that a bank would have kept its pre-switch risk–return profile had it not changed regulators. By taking this approach, I can then examine whether a switching bank has a higher

TABLE 6

Predicted failure probabilities

	(1) FAIL DUMMY 3	(2) FAIL DUMMY 3	(3) FAIL DUMMY 5	(4) FAIL DUMMY 5	(5) FAIL DUMMY 5
SWITCH		–0.087		–0.014	
		(0.543)		(0.895)	
SWITCH EARLY					–0.074
					(0.513)
SWITCH LATE					0.674
					(0.040)**
LOG ASSETS	–0.422	–0.429	–0.456	–0.464	–0.464
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
ROA	–0.095	–0.091	–0.092	–0.087	–0.087
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
SHARPE RATIO	–0.817	–0.823	–0.676	–0.681	–0.681
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
EQUITY/ASSET	–0.318	–0.319	–0.227	–0.230	–0.230
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
LOAN/ASSET	0.059	0.060	0.064	0.064	0.064
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
CHRG/LOAN	–0.011	–0.010	0.001	0.001	0.001
	(0.266)	(0.278)	(0.263)	(0.265)	(0.267)
DEP/LIAB	0.003	0.004	0.004	0.004	0.004
	(0.332)	(0.297)	(0.155)	(0.154)	(0.154)
Observations	225,066	228,980	225,066	228,980	228,980
Pseudo-R^2	0.365	0.364	0.297	0.296	0.296

*Significant at 10 percent level.
**Significant at 5 percent level.
***Significant at 1 percent level.
Notes: The regression is estimated for 1977 to 2001, with year dummies not shown. The logistic regressions in columns 1 and 3 include all banks that never switch primary federal regulators. The logistic regressions in the other columns include banks that never switch plus banks that have switched regulators in the previous six years (excluding the year of the switch). Variable definitions are given in the text. Robust p values adjusted for cluster effects are given in parentheses.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

failure rate after its change than its steadfast counterparts with similar pre-switching profiles. To do this, I compare the predicted failure probability of the bank in the year it switches with the actual failure rate. To get the predicted failure probability, I use the five-year failure rate model estimated over banks that never switch regulators (that is, the model with coefficients reported in column 3 of Table 6). Table 7 gives the predicted and actual failure rates for all switches, broken down by the time of the switch and the type of switch (both merger-related and otherwise). There is no statistically or economically significant difference between the predicted and actual failure rates. Specifically, the failure rate is not higher for banks that switch regulators in the late period, even if the switches do not occur after a merger. This is consistent with the positive coefficient on *SWITCH LATE* in table 6 arising because banks that switch in the late period have lower failure rates than if they had not switched, but not as low as do banks with their new level of return.

Switches do not appear to increase failure risk. Using a simple failure prediction model, I have shown that for most switching banks, their post-switch failure rate is the same as that of otherwise similar banks. The one exception is found among banks that switch regulators after 1991. These banks fail at a higher rate than otherwise similar banks. However, the failure prediction model does not compare switchers to banks that are otherwise similar to the switchers prior to their changing regulators. In particular, in the late period, return increases for banks after a switch. Thus, the "higher failure rate" may be above that for banks with the new, high ROA, but it is lower than for banks with the pre-switch ROA. To test this, I have compared

TABLE 7

Predicted and actual failure rates for banks that switch regulators

	Predicted failure rate over the next five years using equation 3	Actual failures over the next five years	p value for test of difference between predicted and actual failure rate
Both periods (1977–2001)	1.82% (4.81)	1.46% (12.01)	0.242
Early period (1977–91)	2.99 (6.04)	2.46 (15.50)	0.311
Late period (1992–2001)	0.30 (1.38)	0.17 (4.10)	0.459

Notes: Failure rates over the next five years for banks that switch regulators as of the end of the year of the switch. The predicted failure rate is based on the coefficient for regression reported in column 3 of table 6. The standard deviations of the predicted and actual failure rates are in parentheses.
Source: Data from Federal Deposit Insurance Corporation, 1977–2003, *Reports of Income and Condition*, Washington, DC.

the actual failure rate to the level predicted in the year of a switch. I have found that the actual failure rate is no higher than the predicted rate, even for switches after 1991. This implies that switches in regulators do not increase the level of bank failures.

Robustness

The focus of this article is on changes of primary federal regulators. There are two potential alternative approaches to analyzing changes among banks that I address here. The first one involves an approach in which the choice of a national versus state charter is emphasized, without regard to the further choice of taking membership in the Federal Reserve System (for banks that elect state charters).[15] Using a switch of charters rather than a switch of primary federal regulators in the analysis does not change the qualitative results. When I replicate the performance regressions in table 4 or the failure prediction model in table 6 for changes of charters rather than changes of primary federal regulators, the same coefficients are significant at the 5 percent confidence level.

A second approach takes into account that for state-chartered banks, regulation is shared between federal regulators and state regulators. To control for the effect of state regulators, I add state dummies for banks with a state charter. The qualitative results are unchanged. Examining results on a state-by-state basis, there are not enough switches to obtain meaningful results, even for the largest states.

The choice of periods is motivated by changes in regulation and the pattern of banks that switch. To test the impact of the division, I run the regression (equation 1) with a separate set of switching trends and dummies for each year in which a bank might switch. I focus on the change in return between the year prior to a switch and five years after a switch. This analysis shows a distinct break between 1991 and 1992, with the change in performance mixed for changes prior to 1992, but consistently positive thereafter. This suggests that the break between the early and late periods is set correctly and is important.

In the main analysis, I exclude switches that might be related to a merger. As discussed earlier, roughly one-third of all switches are in the year of a merger or the following year. Because the threshold for switching following a merger is different than for switching at other times (and due to accounting issues), I dropped merger-related switches from the main sample. When I examine merger-related switches, the post-switch changes are qualitatively similar to those for switches at banks that did not merge in the period before and after the switch. There is an increase in return, but only in the late period, 1992–2003, and there is no unambiguous indication of an increase in either accounting or failure risk. Prior to the switch, however, there are differences in performance for merger-related and other switches. Heading into a merger-related switch, return is decreasing. This may be related to reasons behind the merger (including accounting issues)—and not to reasons behind the switch. Still, for the purposes of this article, the key is that the post-switch performance is similar for the two types of switchers.

Conclusion

This article has attempted to shed some light on the effects of having multiple regulatory agencies in commercial banking. I have studied the performance of banks that switch their primary federal regulators as an indication of whether there is beneficial competition or a race for the bottom among agencies. Whether banks are able to increase return without increasing risk following a switch constitutes my test for beneficial competition. A race for the bottom would be evidenced by an increase in the failure rate of banks that switch, especially if there is no compensatory increase in return. Overall, I find evidence of beneficial competition instead of a race for the bottom, since return rises and failure rates remain effectively unchanged. However, this masks important differences over time.

The reasons for switching regulators may have changed over time. My sample includes banks that switched between 1977 and 2003, a period of massive changes in banking and bank regulation. I divide the sample into two smaller periods. The early period, 1977–1991, combines two time spans—one marked by the passage of the Depository Institutions Deregulation and Monetary Control Act (DIDMCA) in 1980, the other notable for the initial lessening of prohibitions on interstate banking in the 1980s. Switches in the late 1970s and early 1980s may be a response to DIDMCA or to pre-DIDMCA differences among regulators. Switches in the 1980s through 1991 may reflect banks adjusting to their new competitive environment, although the rate of switching during this period was the lowest in my sample. Finally, in the late period, 1992–2003, prohibitions on interstate banking and on mergers between banks and other financial firms were essentially eliminated. Perhaps because of these changes, there was again a major merger wave in banking.

I find that switches in the early part of my sample—those prior to 1992—had little impact on bank performance. Return did not change significantly following a switch, and there was no unambiguous effect on accounting risk. Moreover, the evidence suggests that bank failure rates did not increase as the result of switches.

My results imply that banks switching regulators in the late part of my sample, 1992–2003, increased return without a rise in bank failures. This is evidence of beneficial competition among regulators, and supports the hypothesis that there is specialization among them. Interestingly, starting in 1992, there was an increase in the rate of regulatory switching that lasted through at least 2003. It is possible that the increase in switches was associated with the onset of this type of beneficial competition.

Finally, note that this analysis is intrinsically limited to looking at one aspect of regulatory competition. While I find evidence of beneficial competition only in the post-1991 period, that should not be taken to imply that other types of beneficial competition did not exist throughout my sample period.

NOTES

[1]Regulatory authority for state-chartered banks is shared with the appropriate state chartering agencies. Unless otherwise stated, when I refer to a bank's "regulator" (or "primary regulator"), I mean its primary federal regulator.

[2]Butler and Macey (1988) point out that differences among regulators are not very large due in part to the use of federal supremacy laws. In essence, federal regulators impose their rules on state-chartered banks through direct regulation or by making federal deposit insurance conditional on accepting certain rules.

[3]Elliehausen (1998) gives estimates of the cost of regulation that range between 5 percent and 15 percent of non-interest expense, or between 2 percentage points and 6 percentage points of return on equity.

[4]It typically takes between 15 days and 30 days to change primary regulators. This time is necessary to get approval from the new regulator. The approval process can be longer if the new regulator chooses to do an exam prior to approving a new applicant; however, this is not generally done for banks that are financially strong and well managed.

[5]Another potential drawback of having multiple regulatory agencies is that the agencies may respond to their constituencies but ignore externalities. When externalities are important, control by local agencies may lead to too little regulation (Baumol and Oates, 1988; Stewart, 1992). As an example, for many years Britain did not control sulfur emissions from its power plants because prevailing winds blew them offshore, with most of the damage being felt in continental Europe (Lomas, 1988). I do not examine this here, since this sort of externality is not a big problem in banking.

[6]Greenspan spoke in October 1991. Later that year, Treasury Secretary Nicholas Brady made similar remarks. The OCC is part of the Treasury Department.

[7]In other industries, interpretation of regulations most frequently occurs at the agency level. There is literature that studies whether regulatory agencies act as Congress wants them to (see, for example, Libecap, 1996).

[8]Berger and Hannan (1998) talk about the desire of bankers for a quiet life.

[9]It is also possible to test the source of beneficial competition, but this is beyond the scope of this article. See Rosen (2003).

[10]Results using the ratio of nonperforming loans to total loans are more likely to indicate a reduction of risk after a switch than those using the charge-off-to-loan ratio are. Nonetheless, I use charge-offs rather than nonperforming loans since data on nonperforming loans are not available for the entire sample period.

[11]Whalen (2002) finds lower return and higher risk at banks that change charters.

[12]The early period actually comprises two different subperiods, one marked by the passage of the DIDMCA, the other notable for the1980s merger wave. Switching activity in the DIDMCA subperiod was higher than during the bulk of the 1980s. However, there was no economically important difference in the relative performance of banks that switched in either subperiod. Thus, to simplify the exposition, I combined my findings from the two subperiods.

[13]Recall that these are calculated by considering changes to the pre- and post-switch trend and dummy variables only.

[14]I use the three years following a switch as the base years (and thus, look at failures for either the first six or eight years after a switch). The reason to restrict how long after a switch I examine is that, eventually, one cannot attribute a failure to be the direct result of a switch. However, looking out further after a switch does not change the qualitative results.

[15]Whalen (2002) also examines banks that change their charters; however, that paper does not examine post-change performance indicators.

Amel, Dean F., 1991, "State laws affecting commercial bank branching, multibank holding company expansion, and interstate banking," Board of Governors of the Federal Reserve System, working paper.

Amel, Dean F., and Martha Starr-McCluer, 2002, "Market definition in banking: Recent evidence," *Antitrust Bulletin*, Vol. 47, No. 1, Spring, pp. 63–89.

Baumol, William J., and Wallace E. Oates, 1988, *The Theory of Environmental Policy*, Cambridge, UK: Cambridge University Press.

Berger, Allen N., and Timothy H. Hannan, 1998, "The efficiency cost of market power in the banking industry: A test of the 'quiet life' and related hypotheses," *Review of Economics and Statistics,* Vol. 80, No. 3, August, pp. 454–465.

Berger, Allen N., Margaret K. Kyle, and Joseph M. Scalise, 2000, "Did U.S. bank supervisors get tougher during the credit crunch? Did they get easier during the lending boom? Did it matter to bank lending?," in *Prudential Supervision: What Works and What Doesn't,* Frederic M. Mishkin (ed.), Chicago: University of Chicago Press, pp. 301–349.

Berger, Allen N., Gregory F. Udell, and Richard J. Rosen, 2005, "Does market size structure affect competition? The case of small business lending," Indiana University, working paper.

Butler, Henry N., and Jonathan R. Macey, 1988, "The myth of competition in the dual banking system," *Cornell Law Review*, Vol. 73, May, pp. 677–718.

Dick, Astrid A., 2006, "Nationwide branching and its impact on market structure, quality, and bank performance," *Journal of Business,* Vol. 79, No. 2, forthcoming.

Easterbrook, Frank H., and Daniel R. Fischel, 1991, *The Economic Structure of Corporate Law*, Cambridge, MA: Harvard University Press.

Elliehausen, Gregory, 1998, "The cost of bank regulation: A review of the evidence," Board of Governors of the Federal Reserve System, staff study, No. 171.

Kane, Edward J., 1989, *The S&L Insurance Mess: How Did It Happen?*, Washington, DC: Urban Institute Press.

Libecap, Gary D. (ed.), 1996, *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth: Reinventing Government and the Problem of Bureaucracy*, Vol. 7, Greenwich, CT: JAI Press.

Lomas, Owen, 1988, "Environmental protection, economic conflict, and the European Community," *McGill Law Journal,* Vol. 33, No. 3, pp. 506–539.

Revesz, Richard L., 2000, "Federalism and regulation: Extrapolating from the analysis of environmental regulation in the United States," *Journal of International Economic Law,* Vol. 3, No. 2, June, pp. 219–233.

Romano, Roberta, 1998, "Empowering investors: A market approach to securities regulation," *Yale Law Journal,* Vol. 107, No. 8, June, pp. 2359–2430.

Rosen, Richard J., 2003, "Is three a crowd? Competition among regulators in banking," *Journal of Money, Credit, and Banking,* Vol. 35, No. 6, part 1, December, pp. 967–998.

Scott, Kenneth E., 1977, "The dual banking system: Model of competition in regulation," *Stanford Law Review*, Vol. 30, pp. 1–49.

Stewart, Richard B., 1992, "Environmental law in the United States and the European Community: Spillovers, cooperation, rivalry, institutions," *University of Chicago Legal Forum*, pp. 41–80.

Stigler, George J., 1971, "The economics theory of regulation," *Bell Journal of Economics,* Vol. 2, No. 1, pp. 3–21.

Tiebout, Charles, 1956, "A pure theory of local expenditures," *Journal of Political Economy,* Vol. 64, No. 5, October, pp. 416–424.

Whalen, Gary, 2002, "Charter flips by national banks," Office of the Comptroller of the Currency, working paper, No. WP2002-1, June.

Financial constraints and entrepreneurship: Evidence from the Thai financial crisis

Anna L. Paulson and Robert M. Townsend

Introduction and summary

Poorly functioning financial markets can limit entry of new firms and lead to inefficient production in existing firms. Small-scale entrepreneurs that have limited access to formal financial markets may be particularly affected by financial constraints. Despite this, small entrepreneurial firms are an important source of innovation, jobs, and economic growth in both developed and developing countries. In the U.S., 44 percent of the private work force is employed in small firms, which account for approximately 50 percent of non-farm gross domestic product (GDP).[1] Striking similarities exist between small firms in the U.S. and those in developing countries. In Thailand, for example, small firms employ 60 percent of the work force and account for approximately 50 percent of GDP.[2] Investment from banks and other formal financial institutions is typically limited in small firms. Thus, in both the U.S. and Thailand, two-thirds of the initial investment in small firms comes from savings and funds from family and friends.[3]

Outside investment in small firms may be limited for a number of reasons, including the difficulty of providing credible information to investors about the expected profitability of a planned investment project or the entrepreneurial skill of a potential borrower. This type of problem is typically called asymmetric information. In addition, the provision of a loan may reduce the incentives for an entrepreneur to exert the necessary effort to make a project successful, since the profits of a successful project will have to be shared with investors. This type of problem is called moral hazard. Asymmetric information and moral hazard are concerns in both developed and developing economies. However, these problems are likely to be acute in developing economies where financial markets are less efficient.

When financial markets are less developed, entrepreneurial activity may also be vulnerable to events like the Asian Financial Crisis. This crisis began in July 1997 when the Thai government abandoned its policy of pegging the value of Thailand's currency, the baht, to a basket of developed countries' currencies heavily weighted to the U.S. dollar. The Asian Financial Crisis led to widespread turmoil in international financial markets and to recessions in many Asian countries. In the wake of the crisis, the Thai economy entered a period of marked contraction. In 1997 Thailand's GDP fell 1.5 percent, and in 1998 it fell 11 percent.[4]

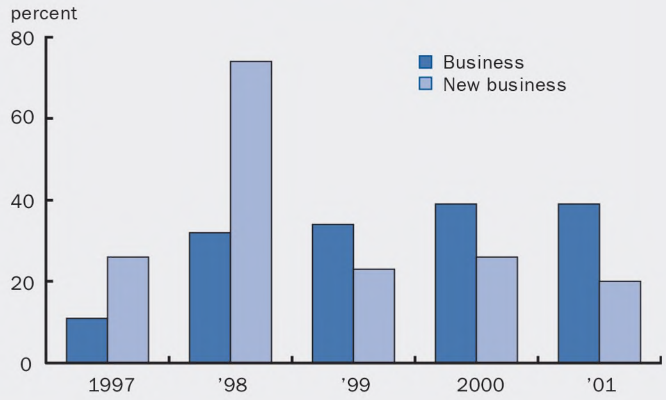
At the same time, entrepreneurial activity in Thailand *increased*. In the 12 months following the onset of the crisis, data from a survey we conducted reveal that the number of business households more than doubled (see figure 1). In the spring of 1997, approximately 11 percent of survey households operated a business. One year later, the percentage had tripled, with more than 30 percent of the survey households operating a business. By studying entrepreneurial activity in Thailand before, during, and after the financial crisis, we can enhance our understanding of entrepreneurship and financial constraints generally, and improve our understanding of the role of small businesses during a period of economic contraction.

We use new longitudinal data from rural and semi-urban Thailand to examine the factors that influence entrepreneurial activity in the pre-crisis and crisis periods. The data cover an interval from the spring of 1997 to the spring of 2001, so we are also able to gain

Anna L. Paulson is a senior economist at the Federal Reserve Bank of Chicago. Robert M. Townsend is the Charles E. Merriam Distinguished Service Professor of the Department of Economics at the University of Chicago and a consultant to the Federal Reserve Bank of Chicago. *The authors wish to thank Kristin Butcher, Craig Furfine, Xavi Giné, Ellen Rissman, and Alicia Williams for helpful comments, and Shirley Chiu for excellent research assistance. They are also grateful to Sombat Sakuntasathien for making the data collection possible and to the National Science Foundation and the National Institutes of Health for funding.*

FIGURE 1

Thai business households and new business households, by year



some insight into the post-crisis period. We are particularly interested in entrepreneurial activity during the crisis period.

Before the crisis, we find that wealthier households are more likely to start businesses and that they invest more in these businesses than their less wealthy counterparts (Paulson and Townsend, 2004). During the crisis, however, the positive correlation between entrepreneurial activity and wealth disappears. These findings are robust to the inclusion of various control variables, alternative functional form assumptions, and various techniques for controlling for the endogeneity of wealth. The traditional explanation of these findings would be that financial markets were inefficient prior to the Asian Financial Crisis, but effectively allocated capital to entrepreneurial activities during the crisis.

However, this interpretation strains credulity, given the major weaknesses of the Thai financial system revealed by the crisis itself. Restricting our attention to the operation of financial markets in rural and semi-urban areas, where the survey takes place, we find it difficult to imagine that imperfections in these financial markets were somehow alleviated during the crisis period.

Instead, we argue that rising unemployment and falling real wages during the crisis led to changes in the types of people who started businesses—and in the types of businesses they started. For instance, businesses that were initiated at the height of the financial crisis required only a *median* of 1,250 baht (approximately $50) in start-up capital.[5] The median initial investment in businesses that were started prior to the crisis was 36,750 baht (approximately $1,470). To put these figures into context, note that median annual income in Thailand in the year before the crisis was 40,000 baht ($1,600) for non-business households and 100,000 baht ($4,000) for business households.

In this article, we provide some insights into how rural and semi-urban households in Thailand coped with the financial crisis. The results of this article also underscore the importance of carefully controlling for changes in the returns to non-entrepreneurial activities, notably labor market conditions, in studying the determinants of entrepreneurial activity more generally. These findings help us to understand, for example, increases in self-employment observed in the U.S. during the recession that ended in November 2001.

The rest of this article is organized as follows. First, we discuss some of the relevant related literature. Then, we provide more background on the impact of the Thai financial crisis, detail the financial environment in the survey areas, and describe the longitudinal data that we analyze. Next, we use regression analysis to examine the role of financial constraints in explaining patterns of entrepreneurship before, during, and after the crisis. Finally, we consider how to interpret these findings in the light of other trends in entrepreneurial characteristics over the 1997–2001 period.

Related literature

If financial constraints were not important, then potential entrepreneurs would make the decision to start a business based solely on the expected profitability of the planned endeavor. If necessary, they would be able to get outside financing to start the project, and their own wealth would not be a significant factor in whether the business was started. When financial constraints are important, however, outside financing may be unavailable or insufficient. Wealthier households will be more likely to start a business than poorer ones under these conditions.

Holtz-Eakin, Joulfaian, and Rosen (1994) use data from tax records in the U.S. to examine the reduced-form relationship between inheritance and entrepreneurship, and conclude that financial constraints are important. Using U.S. data from the *National Longitudinal Survey of Youth* (NLSY), Evans and Jovanovic (1989) draw the same conclusion in their structural study of the impact of wealth on career choices. On the other hand, Hurst and Lusardi (2004) find no evidence that entrepreneurial activity in the U.S. is affected by financial constraints when they allow for a non-linear relationship between wealth and entrepreneurship.

In work that is particularly relevant to this article, Rissman (2003) and Aaronson, Rissman, and Sullivan (2004) point to the importance of taking into account labor market conditions when analyzing the decision to be self-employed. Rissman (2003) models self-employment as an alternative to unemployment, suggesting that self-employment is countercyclical. This conclusion is supported by her analysis of U.S. data from the NLSY. Aaronson, Rissman, and Sullivan (2004) also find some evidence of countercyclical self-employment in the U.S. in their analysis of *Current Population Survey* data. They find that higher rates of unemployment are associated with higher rates of self-employment. They attribute recent increases in self-employment to weak labor market conditions during the recession ending in November 2001.

The operation of existing businesses will also be affected by the entrepreneur's wealth when financial constraints are present. In particular, financial constraints may prevent entrepreneurs from investing the optimal amount in their businesses. If financial constraints did not exist, then entrepreneurs would be able to make up the shortfall between their own funds and the profit-maximizing level of investment by borrowing. In this situation, entrepreneurial investment and entrepreneurial wealth would be independent of one another. When there are financial constraints, however, entrepreneurs may be unable to borrow, or only be able to borrow a limited amount. In this case, wealthier entrepreneurs will be able to invest more in their own businesses, since they are less dependent on the availability of outside financing.

Fazzari, Hubbard, and Petersen (1988) explore this implication of financial constraints in a sample of publicly traded manufacturing firms in the U.S. and show that investment is sensitive to cash flows for some firms. In their two studies, Petersen and Rajan (1994, 1995) hypothesize that banking relationships increase small businesses' access to credit by overcoming information problems that would otherwise constrain the availability of credit to them. Their analysis of data collected by the Small Business Administration (SBA) suggests that banking relationships do indeed play this role for small firms. In contrast, McKenzie and Woodruff (2003) use semi-parametric techniques to show that returns on investment do *not* increase with investment in a sample of small Mexican firms, as one would expect if financial constraints were important.

A number of other theoretical studies, relying on a wide variety of assumptions about how financial markets operate, imply a positive relationship between entrepreneurship and wealth and between investment and wealth.[6] Paulson, Townsend, and Karaivanov (2005) show that moral hazard concerns limit entrepreneurial activity in Thailand in the period leading up to the Asian Financial Crisis.

Background and data

Thai financial crisis

The initial repercussions of the Thai financial crisis were felt in large urban areas, especially in Bangkok, where many construction workers were laid off. Total unemployment increased from an annual rate of 1.1 percent in 1996 to 3.4 percent in 1998, and wages and hours worked fell as well.[7] By some measures, rural areas were particularly hard hit. In these areas, unemployment increased from 3 percent to 8 percent. In the poor northeastern region, real earnings fell by 8 percent.[8] Workers with little education were particularly vulnerable. Real earnings fell 13–20 percent among those who had, at most, completed primary school. Prices also rose during this period, with the Consumer Price Index increasing by 14 percent from 1996 to 1998. From 1998 to 2001, annual inflation in Thailand averaged 1.2 percent.[9]

The overall poverty rate in Thailand increased 24 percent from 1996 to 1999, from 17 percent to 21 percent.[10] However, increases in poverty were not uniform across the country. In the Northeast, for example, rural poverty rates increased nearly 40 percent, going from 28 percent to 39 percent. In the Central region, rural poverty actually decreased from 13 percent to 12 percent from 1996 to 1999. However, urban poverty in the Central region increased nearly 9 percent, going from 6.96 percent to 7.59 percent.

Financial environment

The formal financial sector in Thailand provides two main sources of funding for households in rural and semi-urban areas: the Bank for Agriculture and Agricultural Cooperatives (BAAC) and commercial banks.[11] Of these two, the BAAC is much more active in rural areas. Ninety-five percent of northeastern Thai villages and 89 percent of Central Thai villages had at least one BAAC borrower in 1994. The BAAC offers two types of loans. One is a standard collateralized loan, and the other requires no formal collateral and is secured instead through a joint liability agreement with a group of farmers who all belong to a BAAC group.

While the bulk of the BAAC's loans are uncollateralized, these loans tend to be small, and the majority of funds are lent through collateralized loans. Commercial banks are active lenders in 41 percent of Thai villages. However, commercial bank borrowers tend to be concentrated in the relatively prosperous Central region, where 50 percent of villages have at least one commercial bank borrower. In contrast, only 31 percent of northeastern villages have a commercial bank borrower. Commercial bank loans are almost always secured with a land title. In addition to these formal sector lenders, there are a number of quasi-formal institutions that offer savings and lending services to villagers: village savings and lending institutions and rice banks. It is also common for households to borrow from relatives and neighbors and moneylenders. Often households will borrow from several sources to finance one investment project.

Survey data

The data that we analyze were derived from our own ongoing socioeconomic study in Thailand, which is funded by the U.S. National Institutes of Health and the National Science Foundation. The initial survey of households, village financial institutions, and village key informants was completed in May 1997. It covers regions at the doorstep of Bangkok as well as in the relatively poor Northeast. The data provide a wealth of pre-financial crisis data from 2,880 households, 606 small businesses, 192 villages, 161 local financial institutions, 262 borrowing groups of the BAAC, and soil samples from 1,880 agricultural plots. A subset of these households was included in an ongoing longitudinal survey, which takes place between March and May of each year. The data we analyze cover the period from 1997 to 2001 and include 960 households.

The study focuses on four Thai provinces that were chosen because of the availability of retrospective data from the Thai *Socio-Economic Survey* (SES). These provinces are emblematic of two distinct regions of Thailand: rural and semi-urban households living in the Central region, close to Bangkok, and more obviously rural households living in the semi-arid and much poorer northeastern region. The Central region is wealthier and more developed than the Northeast.

In each province, four geographic areas, called tambons, were chosen at random. Each tambon includes approximately ten villages. In each sample tambon, four villages were chosen at random.[12] Fifteen households were randomly selected from each sample village. Overall, the data include five years of information for 960 households (4 provinces × 4 tambons × 4 villages × 15 households) from 64 Thai villages (4 provinces × 4 tambons × 4 villages).

The data include survey year and retrospective information on wealth (household, agricultural, business, and financial); occupational history (transitions to and from farm work, wage work, and entrepreneurship); and access to and use of a wide variety of formal and informal financial institutions (commercial banks, agricultural banks, village lending institutions, and moneylenders, as well as friends, family, and business associates). The data also provide detailed information on household demographics, entrepreneurial activities, and education. The retrospective data on wealth and interactions with financial institutions help us to disentangle the effects of running a business from the forces that make it possible to start a business in the first place.

Because these data provide rich and detailed information about both the firm and the entrepreneurial household, as well as information on financial intermediaries, they are particularly well designed for studying the relationship between entrepreneurship and the financial system. Economic theory emphasizes that both firm and entrepreneurial characteristics are important in determining the supply and demand for credit. In many studies the available data force a focus on either the firm or the entrepreneur, but do not allow both to be treated with equal thoroughness.[13]

Business characteristics

In this section we highlight some of the key features of the data that are important for this article. The businesses we study are quite varied and include shops and restaurants, trading activities, raising shrimp or livestock, and the provision of construction or transportation services. We rely on household reports on whether its members ran a business except in the case of shrimp and fish farming. All of these activities are treated as businesses. It is quite common for households to run a business in addition to working for wages and farming, usually rice. Most business households run only a single business and rely very heavily on family workers. Only 10 percent of the businesses paid anyone for work during the year prior to the survey.

While there are many different types of businesses, shrimp and/or fish raising, shops, and trade account for most of the businesses. These categories account for 65 percent of businesses founded prior to the crisis, 60 percent founded in the year of the crisis, and 39 percent founded in the immediate post-crisis period. The distribution of business types within these categories changes substantially following the crisis. Trade accounts for 17 percent of all businesses that were started in the five years before the crisis. However, 47 percent of the businesses that were founded in the year of the crisis were in trade. The trade category includes retail and wholesale trading activities, ranging from selling desserts in a local market to selling gasoline to shops and gas stations.

There is substantial variation in initial investment in new businesses over time, as we alluded to in the introduction (see table 1). The median initial investment in a business founded prior to the crisis, between 1992 and 1997, is 36,747 baht. The median initial investment in a business that began at the height of the crisis in 1998 is 1,350 baht. The median initial investment in a trading business was 52,533 baht prior to the crisis, just 793 baht in the year of the crisis, and zero in the three years following the crisis. For all the major business types, median initial investment is substantially lower for businesses founded during the first year of the crisis and afterwards compared with businesses founded between 1992 and 1997.

Households rely heavily on savings (either in the form of cash or through asset sales) to fund initial investment in their businesses. Approximately 60 percent of the total initial investment in household businesses that were founded between 1992 and 1997 comes from savings. Loans from commercial banks account for about 9 percent of total business investment, and BAAC loans account for another 7 percent. In the Northeast, the BAAC plays a larger role compared with commercial banks, and in the Central region, the opposite is true. In the crisis and post-crisis periods, when investment is lower, the importance of credit for funding initial investment in the business declines.

In some of the empirical work, we control for participation in formal and informal financial markets by business and non-business households. We group formal and informal financial institutions into six categories. The first, formal financial institutions, includes commercial banks, finance companies, insurance companies, and national employee credit unions, such as the Teachers Credit Union. The second, village institutions and organizations, is made up of production credit groups (PCGs),[14] rice and buffalo banks, and village poor and elderly funds. Formal loans from the BAAC, the Agricultural Cooperative, and local farmers' groups are included in the third group, agricultural organizations. BAAC customers whose loans are secured through joint liability arrangements make up the fourth group. Moneylenders and rotating savings and credit associations (ROSCAs) make up the fifth and sixth groups, respectively. Households were asked to report when they became a customer or member of each organization. Hence, we are able to look at the influence of participation in these organizations prior to starting a business, as distinct from becoming a client of an institution because of the business.

Because households were asked to report when they acquired household and agricultural assets and land, the data provide measures of past wealth as well as current wealth. In the empirical work, which we discuss in the next section, we examine the relationship between past wealth (that is, wealth prior to starting a business) and entrepreneurship. This allows us to avoid some problems of endogeneity that are likely to plague current wealth measures, since current wealth reflects both the resources available to start a business for potential entrepreneurs and the past profitability of a business for current entrepreneurs. Because we can measure wealth before a business was founded, we can isolate the resources available to start a business.

For the time being, however, our interest is in current rather than past wealth. Panel A of figure 2 describes the trend in median wealth in real 1997 Thai baht for business and non-business households over the years 1997–2001. Business households are wealthier than their non-business counterparts over the entire span, and all households experience modest declines in wealth during the crisis. Between 2000 and 2001, median wealth increases for all households, with

TABLE 1

Thai business types and median initial investment

	Pre-crisis		Crisis		Post-crisis	
Business types	**Percent**	**Median inv.**	**Percent**	**Median inv.**	**Percent**	**Median inv.**
Shrimp and/or fish	19	42,027	6	37,800	10	14,745
Shop	29	26,595	7	10,366	4	5,362
Retail and wholesale trade	17	52,533	47	793	25	0
Other	35	78,626	40	5,166	61	0
All	100	36,747	100	1,350	100	0
Sample size	102		208		213	

Notes: Pre-crisis refers to businesses that were started between 1992 and 1997 and were still in operation in 1997. Crisis refers to businesses that were started in 1998 and were still in operation in 2001. Post-crisis refers to businesses that were started between 1999 and 2001 and were still in operation in 2001. Median initial investment (median inv.) is in real 1997 Thai baht.

increases being more dramatic for business households compared with non-business households.

In figure 2, we compare important characteristics of business and non-business households from 1997 to 2001. Prior to the crisis, the heads of business households were more educated than the heads of non-business households (see figure 2, panel B). Business household heads had almost 4.8 years of schooling compared with 3.9 years for non-business household heads. Table 2 provides further details on the distribution of education (and other variables) for business and non-business households. While 61 percent of business and non-business household heads had completed four years of school in 1997, 23 percent of business household heads had additional education compared with just 13 percent of non-business household heads.[15] During the crisis, the gap in education between business and non-business households narrowed substantially, indicating that individuals who started businesses during the crisis were less educated than those who started businesses prior to the crisis. Among households that started businesses in 1999, for example, 35 percent of household heads had less than four years of schooling (see table 2, panel B).

We see a similar pattern with age (see figure 2, panel C). The heads of business households tend to be younger than the heads of non-business households. Before the crisis, they are almost three years younger. However, this gap virtually disappears during the crisis. This indicates that the people who founded businesses during the crisis were significantly older than the individuals who founded businesses prior to the financial crisis.

In panel D of figure 2, we examine trends in household size for business and non-business households. Here we see a different pattern. Business households tend to be larger than non-business households, and the difference increases between 1997 and 2001. There are two potential explanations for this trend, both of them related to urban migrants returning to rural and semi-urban areas in the wake of the crisis. One possibility is that existing business households were more likely to be joined by family members who had migrated prior to the crisis. Another possibility is that urban migrants were more likely to rejoin households that did not have a business prior to the crisis, and these migrants spurred the creation of businesses during the crisis.

Panel E of figure 2 reports on trends in median income (net of expenses for business and farm activities) for business and non-business households.[16] Business households have higher median income than the non-business households over the 1997–2001 period. However, while non-business income drops modestly during the crisis, business income decreases significantly with the onset of the crisis. In 1997 median business income is nearly 90,000 baht, and in 1998 it is just 65,000 baht. As before, there are two potential factors that lie behind this decline. Businesses in operation prior to the crisis may have experienced a dramatic drop in income during the crisis. In addition, businesses started during the crisis may simply generate less income than those started before the crisis. We return to which of these factors is likely to be more important later in this article.

In panel F of figure 2, we examine trends in median expenditure for business and non-business households. Expenditure provides a measure of both current welfare and also reflects expectations about future economic conditions. Households that expect crisis conditions to continue are likely to curtail their expenditures more than households that expect the crisis to be resolved relatively quickly. Median expenditure is higher for business households compared with non-business households throughout the 1997–2001 period, and expenditure decreases from 1997 to 2000 and then increases in 2001 for all households. However, business households experience a sharper decline in expenditure from 1997 to 1998 than non-business households, potentially driven by the entry of new households into this category. By 2001, median non-business household expenditure exceeds pre-crisis levels. For business households, median expenditure in 2001 is still lower than it was in 1997.

Before moving on to discuss the results of a more formal analysis of the role of financial markets before, during, and after the crisis period, it is useful to review the observations that we would like to be able to account for:

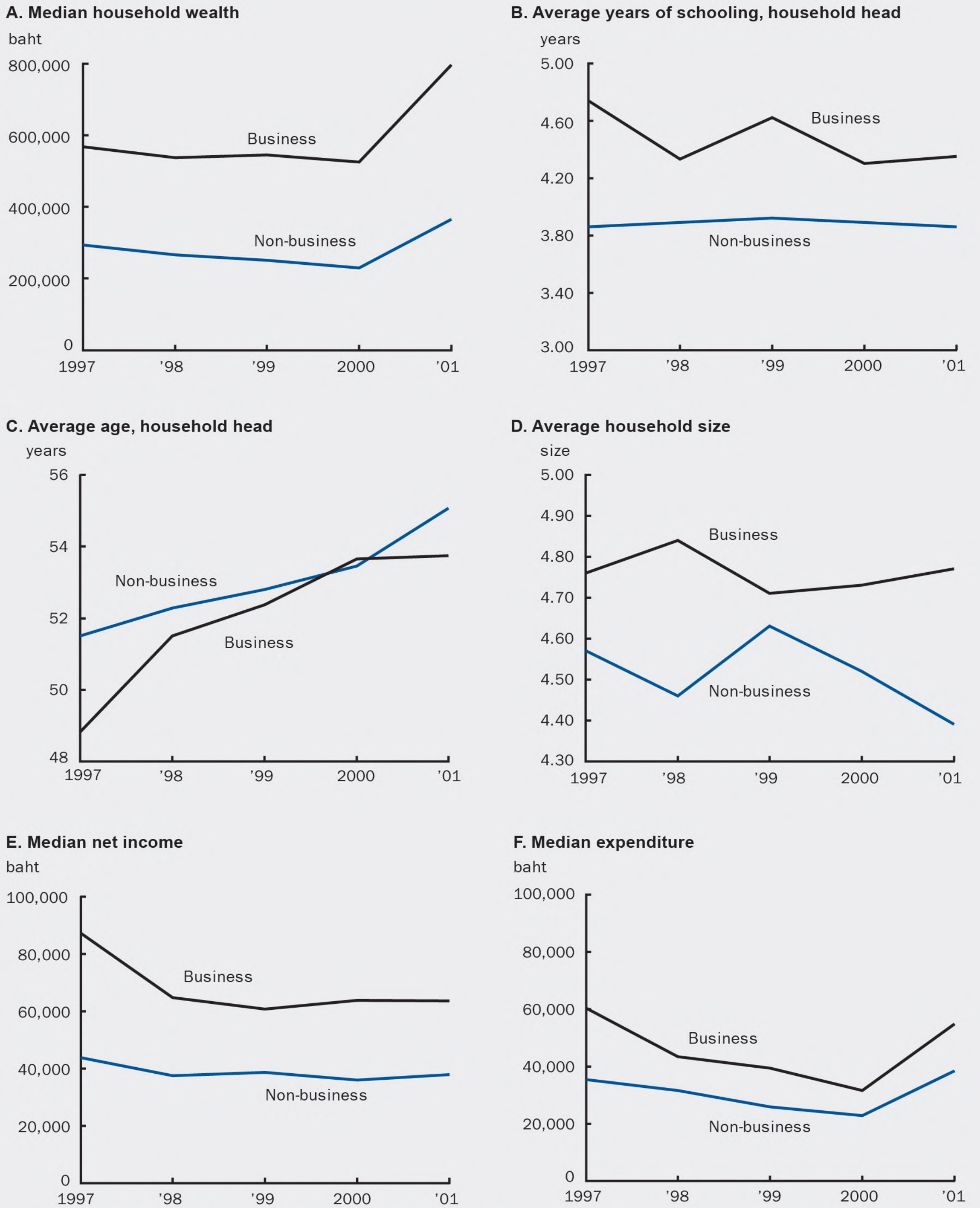
- The percentage of business households nearly tripled during the crisis.
- Businesses started during the crisis tend to have very low or even no initial investment.
- The heads of households who established businesses during or after the crisis tend to be less educated and older than the heads of households with businesses already in operation prior to the crisis.
- Business households have higher wealth, net income, and expenditure compared with non-business households, although the gap between business and non-business households narrows during the crisis period.

Evidence of financial constraints

In this section, we consider the evidence that financial market imperfections played a role in shaping patterns of entrepreneurship before, during, and after

FIGURE 2

Characteristics of Thai business households vs. non-business households



Notes: In panels A, E, and F, amounts are in real 1997 Thai baht. In panel B, years refer to years of schooling completed by head of household. In panel C, years refer to the age of head of household. In panel D, size refers to the number of individuals that make up a household.

TABLE 2

Thai household characteristics

A. Non-business households, by year

	1997	1998	1999	2000	2001
Age of head	51.51	52.28	52.80	53.45	55.08
	(13.45)	(13.71)	(13.58)	(13.84)	(13.44)
Years of schooling— head					
Average	3.86	3.88	3.92	3.89	3.86
	(2.81)	(2.84)	(2.89)	(2.87)	(2.79)
0–3 years (percent)	26	24	24	25	24
4 years (percent)	61	63	64	62	64
5–16 years (percent)	13	13	13	14	12
No. of adult males in household	1.42	1.39	1.43	1.38	1.38
	(0.94)	(0.84)	(0.91)	(0.91)	(0.91)
No. of adult females in household	1.55	1.49	1.51	1.49	1.49
	(0.78)	(0.73)	(0.73)	(0.75)	(0.73)
No. of children (< 18 years) in household	1.60	1.58	1.69	1.64	1.52
	(1.24)	(1.20)	(1.25)	(1.26)	(1.22)
Mean past wealth (in 000s)	803	945	360,000	1,140,000	20,400
	(3,217)	(3,615)	(5,630,000)	(25,100,000)	(428,000)
Median past wealth (in 000s)	135	254	270	244	237
No. of observations	790	607	547	492	479

B. Business households, by year business was started

	1992–97	1998	1999	2000	2001	1999–2001
Age of head	48.79	52.37	53.22	55.16	53.07	53.95
	(14.89)	(13.18)	(13.99)	(12.69)	(12.76)	(13.11)
Years of schooling— head						
Average	4.74	4.18	3.74	4.15	3.97	3.97
	(3.35)	(2.98)	(3.04)	(3.01)	(2.93)	(2.99)
0–3 years (percent)	16	23	35	19	28	26
4 years (percent)	61	62	52	71	54	60
5–16 years (percent)	23	16	14	11	18	14
No. of adult males in household	1.46	1.56	1.39	1.44	1.61	1.47
	(0.88)	(1.01)	(0.83)	(0.78)	(0.97)	(0.86)
No. of adult females in household	1.55	1.63	1.45	1.59	1.52	1.53
	(0.77)	(0.76)	(0.61)	(0.68)	(0.67)	(0.66)
No. of children (< 18 years) in household	1.75	1.67	1.30	1.52	1.69	1.50
	(1.20)	(1.22)	(1.00)	(1.12)	(1.26)	(1.13)
Mean past wealth (in 000s)	1,479	1,196	1,432	110,000	3,853	45,500
	(2,994)	(2,817)	(3,383)	(1,000,000)	(23,700)	(634,000)
Median past wealth (in 000s)	258	414	398	325	319	328
No. of observations	102	208	67	85	61	213

Notes: Standard deviations are in parentheses. For 1998 through 2001, two rows— mean past wealth and median past wealth— refer to wealth in real 1997 Thai baht in the year prior to the year the business started. For example, for the column headed 2000, past wealth is the value of wealth in 1999, expressed in real 1997 Thai baht. However, for the column headed 1997 in panel A, past wealth is the value of wealth in 1991, expressed in real 1997 Thai baht. And for the column headed 1992–97 in panel B, past wealth is the value of wealth in 1991, expressed in real 1997 Thai baht. In panel B, for 1998 through 2001, the figures describe businesses that were started in that given year and were still in operation in 2001; the column headed 1992–97 describes businesses that were started between 1992 and 1997 and were still in operation in 1997.

the financial crisis. We examine the implications of financial constraints for business start-ups and for initial investment in new businesses.

In the analysis, we divide household businesses into three groups:

1) Pre-crisis businesses: businesses founded between 1992 and 1997, still in operation in 1997;
2) Crisis businesses: businesses founded in 1998, still in operation in 2001; and
3) Post-crisis businesses: businesses founded between 1999 and 2001, still in operation in 2001.

For ease of exposition, we label the third group "post-crisis," but we do not mean to imply that the impact of the Thai financial crisis was limited to 1998. We concentrate on businesses that survived for some period because of the design of the 1997 survey. The 1997 survey identifies businesses that were in operation at the time of the survey—that is, businesses that were started at some point in the past and were still in operation in 1997. We restrict our attention to businesses that were started in the five years prior to this survey. To make sure that we are looking at roughly comparable businesses after 1997, the analysis excludes businesses that were started in 1998 but failed between 1998 and 2001 and businesses that were started between 1999 and 2001 and were not in operation in 2001. Of the businesses that were founded at the height of the crisis in 1998, 63 percent were still in operation in 2001.

To examine the importance of financial constraints, we focus on two key relationships. The first is the relationship between the likelihood that a household starts a business and household wealth prior to the time that the business was founded. The second is the relationship between the initial investment in the business and household wealth prior to the time that the business was founded. If financial constraints are important, we expect that business start-ups will be sensitive to the wealth of potential entrepreneurs and that wealthier entrepreneurs will invest more in their businesses.[17]

In order to evaluate the implications of financial constraints, we need to come up with appropriate measures of entrepreneurial talent and wealth. The proxy we use for entrepreneurial talent is education. While education is certainly not a perfect indicator of entrepreneurial talent, it is likely to be positively related to business skill. In Paulson, Townsend, and Karaivanov (2005), we show that, at least for Thailand, formal education seems to be strongly associated with business skill.

The appropriate wealth variable is wealth at the time the decision is made to start a business. For the pre-crisis analysis, we use wealth six years prior to the 1997 survey as an empirical counterpart to this variable. We exclude households with businesses that were founded prior to 1992 from the analysis. For the crisis and post-crisis periods, we measure wealth in the year before the business was started. The items that are included in the wealth variable are: the value of household and agricultural assets and land. We do not include the value of any business assets that the household may have owned prior to starting a business.

By using past, rather than current wealth, and by excluding business assets acquired before the business was started, we hope to avoid issues of endogeneity: Wealthier people are more likely to start businesses, and business owners have higher earnings than wage workers, which allow business owners to become even richer. In this scenario, current wealth captures both the cause and the effect of having been able to start a business in the past.

Wealth and the likelihood of starting a business

In table 3, we estimate probit models of who becomes an entrepreneur for the three periods. The first set of results in this table reports on the pre-crisis findings. The dependent variable is equal to one if the household runs a business in 1997 that was founded between 1992 and 1997 and zero if the household does not have a business in 1997.[18] The second set of results reports on the crisis findings, where the dependent variable is equal to one if the household starts a business in 1998 that survives until 2001, and it is equal to zero otherwise. The post-crisis findings are found in the third set of results, and the dependent variable in this regression is equal to one if the household has a business in operation in 2001, which was founded between 1999 and 2001, and it is equal to zero otherwise. The figures reported in the table indicate the marginal effect of an infinitesimal change in each continuous variable on the probability of starting a business. For dummy variables, we report the impact of changing the variable in question from zero to one.

In addition to wealth prior to starting a business, the explanatory variables include characteristics of the household head that may be indicators of business talent—age, age squared, and years of schooling. There are also variables that control for the amount of household labor that is available—the number of adult males, adult females, and children under the age of 18 living in the household.[19]

We control for credit market availability by including measures of whether the household was a member or customer of various financial institutions in the past. Like the labor supply variables, we include

TABLE 3

Probit estimates of Thai business start-ups

	Pre-crisis		Crisis		Post-crisis	
	dF/dx	z-statistic	dF/dx	z-statistic	dF/dx	z-statistic
Age of head	–0.0127	–2.36	–0.0003	–0.02	0.0062	0.93
Age of head squared	0.0001	2.14	–0.0000	–0.08	–0.0000	–0.83
Years of schooling– head	0.0097	2.46	0.0086	1.25	0.0079	1.88
No. of adult males in household	0.0135	1.12	0.0311	1.63	0.0217	1.72
No. of adult females in household	0.0055	0.37	0.0662	2.68	–0.0077	–0.47
No. of children (< 18 years) in household	0.0030	0.34	–0.0014	–0.08	0.0021	0.22
Past wealth	0.0226	2.53	0.0318	1.11	–0.0040	–0.85
Past wealth squared	–0.0008	–1.77	–0.0022	–0.77	0.0000	0.85
Past member or customer of						
Formal financial institutions[a]	0.0135	0.44	–0.0128	–0.33	0.0098	0.43
Village institutions/organizations[a]	–0.0398	–1.12	–0.0320	–0.76	0.0096	0.39
Agricultural lenders[a]	0.0332	1.11	–0.0033	–0.08	0.0158	0.67
BAAC groups[a]	–0.0009	–0.03	0.0749	1.70	–0.0086	–0.34
Moneylenders[a]	–0.0160	–0.28	0.0143	0.27	0.0404	1.21
Pseudo R-squared (%)	12.94		14.67		17.00	
Log likelihood	–268.58		–244.27		–212.70	
No. of observations	824		514		472	

[a]Dummy variables.

Notes: Pre-crisis refers to businesses that were started between 1992 and 1997 and were still in operation in 1997. Crisis refers to businesses that were started in 1998 and were still in operation in 2001. Post-crisis refers to businesses that were started between 1999 and 2001 and were still in operation in 2001. For dummy variables, dF/dx represents the change in probability when the dummy variable goes from zero to one. For all other variables, dF/dx is the change in probability from an infinitesimal change in the independent variable in question. Past wealth is made up of the value of household assets, agricultural assets, and land. The coefficient on past wealth in the table is the actual one $\times 10^6$. The coefficient on past wealth squared is the actual one $\times 10^{12}$. Sixteen geographic controls are also included (tambons).

these variables so that we can appropriately interpret the coefficient of the wealth variable. In order to separate the impact of the availability of a particular credit institution in the local area from the impact of being a client of the institution, the estimates also include controls for each of the tambons that were sampled. The tambon controls are meant to capture geographic variations in the supply of credit along with other important characteristics, such as infrastructure and the size of the market. The inclusion of the tambon controls means that the credit market variables provide an indication of the average probability that patrons of the various institutions will start businesses, relative to the probability that households in a particular tambon will start businesses.

During the pre-crisis period, the likelihood that a household starts a business is positively related to preexisting wealth. In particular, the coefficients reported in the first set of results imply that a 1,000,000 baht ($40,000) increase in wealth would be associated with a 2.3 percentage point increase in the likelihood of starting a business.[20] This is an increase of 21 percent above the observed percentage of households that have started a business in the past five years. The coefficient on wealth squared is significant, although very small, suggesting that the impact of wealth on starting a business decreases as wealth increases.

In contrast to the pre-crisis findings, during the crisis and post-crisis periods, there is no statistically significant relationship between wealth and the likelihood of starting a business. This suggests that the importance of financial constraints declines during the crisis and post-crisis periods.

Table 3 estimates also reflect trends in the difference between the characteristics of business and nonbusiness households over the crisis period, described previously. Prior to the crisis, older household heads are significantly less likely to start a business. During and after the crisis, there is no significant relationship between the age of the household head and the likelihood of starting a business. More education is associated with a greater likelihood of starting a business prior to the crisis, but has no significant impact on business start-ups during the crisis. Larger households, as captured by the number of adult males and females, are more likely to start businesses during and after the crisis. These variables have no significant impact on the likelihood of starting a business prior to the crisis. Business talent appears to have been more important prior to the crisis than during the crisis, and the availability of household labor seems to be more important during the crisis than before the crisis.

In general, access to credit, as measured by past patronage of the various financial institutions, does not seem to play an important role in business start-ups before, during, or after the crisis. With one exception, the variables that control for access to credit are insignificant. During the crisis, however, households that had a prior relationship with the BAAC, in the form of a joint liability borrowing arrangement, are 7.5 percentage points more likely to start a business than those without prior ties to the BAAC. This corresponds to nearly a 30 percent increase in the likelihood of starting a business during the crisis period.

Wealth and initial business investment

In table 4, we examine the relationship between initial business investment and preexisting household wealth for pre-crisis, crisis, and post-crisis businesses. In these regressions, the log of initial business investment plus one is regressed on household wealth prior to the period when the business was started. In panel A, the sample includes only businesses with positive initial investment. In panel B, the sample is augmented with businesses that began with zero initial investment. When we restrict the sample to businesses with positive initial investment, as we do in panel A, it makes it more difficult to find *no* relationship between investment and wealth.

In addition to household wealth, these regressions also include the same household controls discussed earlier.[21] For businesses with positive initial investment, higher levels of wealth prior to starting a business are associated with greater initial business investment prior to the crisis and after the crisis but not during the crisis (see table 4, panel A). An increase in past wealth of 1,000,000 baht is associated with an increase in investment of 46 percent prior to the crisis. These

TABLE 4

Regression estimates of log initial Thai business investment

A. Businesses with initial investment greater than zero

	Pre-crisis		Crisis		Post-crisis	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Age of head	−0.0346	−0.37	0.0524	0.40	−0.2004	−1.58
Age of head squared	0.0000	0.06	−0.0007	−0.56	0.0018	1.55
Years of schooling– head	0.0914	1.54	0.2669	3.57	0.0278	0.41
No. of adult males in household	0.2145	0.96	−0.3217	−1.27	0.2084	0.70
No. of adult females in household	0.7075	2.57	0.8533	2.46	0.0865	0.24
No. of children (< 18 years) in household	−0.1862	−1.07	−0.1154	−0.50	0.1042	0.55
Lag wealth	0.3930	2.16	0.0754	0.46	0.2120	4.12
Lag wealth squared	−0.0156	−1.68	0.0007	0.11	−0.0000	−4.12
Constant	10.3572	4.28	6.4398	1.83	12.9643	3.82
Adjusted R-squared (%)	19.67		10.98		16.13	
No. of observations	69		131		95	

B. All businesses

	Pre-crisis		Crisis		Post-crisis	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Age of head	0.2688	1.07	0.0288	0.15	−0.3250	−1.68
Age of head squared	−0.0027	−1.15	−0.0005	−0.27	0.0028	1.63
Years of schooling– head	0.0198	0.12	0.2668	2.23	0.3569	3.18
No. of adult males in household	0.6307	1.02	0.2218	0.58	0.1204	0.33
No. of adult females in household	0.7356	0.99	0.2362	0.46	0.2049	0.40
No. of children (< 18 years) in household	−0.8216	−1.90	0.1276	0.42	0.4046	1.38
Lag wealth	−0.8890	−1.05	0.2720	1.03	0.0055	0.22
Lag wealth squared	0.0212	0.17	−0.1150	−0.11	−0.0000	−0.21
Constant	1.3736	0.22	3.3881	0.65	10.3189	1.97
Adjusted R-squared (%)	8.89		2.02		6.37	
No. of observations	99		206		214	

Notes: Pre-crisis refers to businesses that were started between 1992 and 1997 and were still in operation in 1997. Crisis refers to businesses that were started in 1998 and were still in operation in 2001. Post-crisis refers to businesses that were started between 1999 and 2001 and were still in operation in 2001. Lag wealth is made up of the value of household assets, agricultural assets, and land in the year prior to starting a business. The coefficient on lag wealth is the actual one $\times 10^6$. The coefficient on lag wealth squared is the actual one $\times 10^{12}$. The dependent variable is the natural log of initial investment plus one. In panel A, only businesses with non-zero initial investment are included. In panel B, all businesses, regardless of initial investment, are included.

findings suggest that financial market imperfections restrict investment levels prior to the crisis and after the crisis but not during the crisis itself.

After the height of the crisis in 1998, the importance of financial constraints on investment levels appears to return, at least for businesses with positive initial investment. For these businesses, an increase in past wealth of 1,000,000 baht is associated with an increase in investment of 26 percent. Interestingly, during the crisis more educated business household heads invest significantly more in their businesses. There is some evidence that this is also the case prior to the crisis, but the size and the significance of the coefficient on schooling is smaller.

When we include businesses that begin with zero initial investment (see table 4, panel B), we find no relationship between initial business investment and past wealth before, during, or after the crisis.[22] Education is a strong predictor of initial business investment during the crisis and post-crisis periods according to these estimates, although the magnitude of the effect is fairly small. An additional year of schooling is associated with an increase in initial investment of 1.3 to 1.4 baht. Keep in mind, however, that 37 percent of the crisis businesses and 56 percent of the post-crisis businesses had zero initial investment.

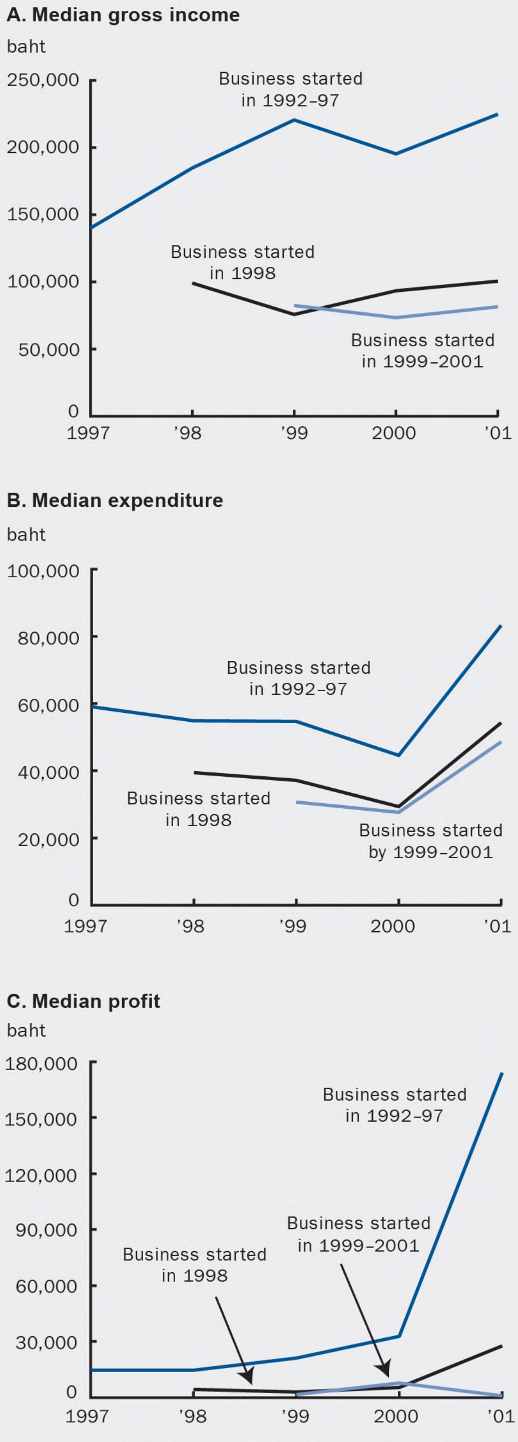
Overall, the relationship between investment and past wealth suggests that financial constraints led to underinvestment in existing businesses prior to the crisis, and possibly after the crisis, but did not place important restrictions on business investment during the crisis.

Business performance

In figure 3, we examine the performance of the three groups of business households from 1997 to 2001. We examine three indicators of business household success: gross income, expenditure, and profit (panels A, B, and C, respectively). Figure 3 underscores the emerging picture that households that start businesses during and after the crisis are different along important dimensions from households that were running businesses when the crisis hit. Gross income, expenditure, and profit are all much higher for households that were already running a business at the time of the crisis compared with households that started a business during or after the crisis. Businesses founded in the post-crisis period have notably lower profits (figure 3, panel C). One potential explanation for this finding is that households with more entrepreneurial talent started businesses earlier—either before the crisis or during the crisis. The businesses that were founded in the post-crisis period may be operated by

FIGURE 3

Performance of Thai business households



Note: All are measured in real 1997 Thai baht.

relatively untalented individuals, and hence have very low profits.

These patterns suggest that the narrowing gap between business and non-business households—in terms of wealth, net income, and expenditure (figure 2, panels A, E, and F, respectively)—is primarily due to the entry of new businesses with lower income and expenditure during and after the crisis rather than a weakening of the economic status of existing businesses. Note, in particular, that the income of households that had businesses at the time of the crisis went *up* from 1997 to 1999 at the height of the crisis (figure 3, panel A).

Conclusion

Beginning with the observation that the number of household businesses in rural and semi-urban Thailand nearly tripled in the wake of the Thai financial crisis, we describe and analyze a number of important features of pre-crisis, crisis, and post-crisis businesses. In particular, we show that businesses started during and after the Thai financial crisis are more similar to non-business households than households that started businesses prior to the crisis. Prior to the crisis, business start-ups and initial investment are significantly related to past household wealth. However, during the crisis, business start-ups and initial investment are unaffected by household wealth. In addition, crisis and post-crisis businesses are characterized by low initial investment.

During the post-crisis period, business start-ups are unaffected by wealth, but initial business investment (for businesses with non-zero investment) is increasing with wealth. Recall that the median business founded during the post-crisis period has zero initial investment. Profits are highest for businesses started prior to the crisis and lowest for businesses started during the post-crisis period. Compared with businesses started during and after the crisis, pre-crisis businesses appear to recover faster and more sharply.

Financial market imperfections seem to restrict business start-ups and investment prior to the crisis but not during the crisis. What might account for this finding? It seems plausible to rule out improvements in financial markets as an explanation, since the crisis itself suggests that Thai financial markets are (or at least were) quite fragile. The key to understanding the apparent lack of financial constraints during the crisis period in Thailand—and how financial constraints have an impact on entrepreneurial activity more generally—is to consider the alternative occupations available to households.

The model of Evans and Jovanovic (1989) provides a useful framework for understanding the increase in business activity during the Thai financial crisis and over the business cycle. Their model implies that when wages fall, more businesses will be started as the returns to entrepreneurial activity exceed wages for more households. In addition, this model implies that the new businesses will tend to be capitalized at lower levels and be run by less talented entrepreneurs. We see evidence of this in the data—crisis and post-crisis investment levels are very low, profits are also low, and the household heads that founded crisis and post-crisis businesses are also less educated than those that founded businesses prior to the crisis. We can reconcile the facts we have described above by understanding how falling wages affect both who finds entrepreneurial activity profitable and how much they invest in business activity.

As alternatives to business employment worsened during the Thai financial crisis, households began businesses because their wage employment options deteriorated. Low capital business opportunities that were unattractive prior to the crisis looked good during the crisis. Note that business investment during the crisis period generated lower profits than pre-crisis investment. Despite the finding that business start-ups and investment are insensitive to wealth during the crisis, there was no improvement in financial markets during this period. Instead, typical business investment during the financial crisis was so low that credit was not required.

This article's findings underscore the general importance of taking into account economic conditions at the time a business is founded in order to account for firm investment and profitability. This insight extends to both developed and developing countries, and applies to dramatic events like the Thai financial crisis, as well as to more modest business cycle type variation in economic conditions.

NOTES

[1]Small Business Administration (SBA) statistics are drawn from the U.S. Bureau of the Census and *Current Population Survey* data. According to the SBA, small firms are defined as manufacturing firms with fewer than 500 employees and non-manufacturing firms with less than $5 million in annual sales.

[2]APEC (Asia-Pacific Economic Cooperation) Center for Technology Exchange and Training for Small and Medium Enterprises. Small Thai firms include manufacturing and service firms with 50 or fewer employees; wholesale trade firms with 25 or fewer employees; and retail trading operations with 15 or fewer employees. Medium-sized firms may have up to 200, 50, and 30 employees in each of these categories, respectively.

[3]This is determined from Bitler, Robb, and Wolken (2001) and calculations from the authors' survey from Thailand.

[4]In the years leading up to the crisis, the Thai economy had grown rapidly. From 1980 to 1995, real per capita GDP had grown 8 percent per year. Following the crisis, the Thai economy recovered somewhat, and real per capita GDP growth averaged 3 percent per year from 1999 to 2001 (World Bank, World Development Indicators).

[5]Throughout this article, monetary values are reported in real 1997 Thai baht. Prior to the devaluation in July 1997, 25 Thai baht equaled 1 U.S. dollar (25 baht = $1).

[6]For example, these implications are shared by a model where there is no credit (Lloyd-Ellis and Bernhardt, 2000), a model where credit is exogenously limited to be a fixed multiple of household wealth (Evans and Jovanovic, 1989), and a model where credit is allocated as the optimal solution to an information-constrained moral hazard problem (Aghion and Bolton, 1997). They are also consistent with the asymmetric information framework emphasized by Fazzari, Hubbard, and Petersen (1988, 2000).

[7]Unemployed individuals are those who are currently not working but are actively looking for work (World Bank, World Development Indicators).

[8]See The World Bank Group (2000).

[9]Prior to the crisis, inflation in Thailand was determined by inflation in the currencies to which the Thai baht was pegged; this means that price increases in Thailand largely mimicked those of the U.S.

[10]All poverty rate figures are reported in Thailand Development Research Institute (2003) and are based on calculations from the Thai National Statistics Office, *Socio-Economic Survey* (SES) data. The poverty rate is defined as the percentage of people in a given region living below the poverty line for that region.

[11]This section is based on the authors' observations and discussions with BAAC officials as well as on data from the Community Development Department of the Thai Ministry of the Interior that cover 60,000 Thai villages every other year from 1988 through 1994.

[12]Each village is a distinct political entity with an elected headman or woman, very much like a mayor.

[13]For example, the *National Longitudinal Survey of Youth*, analyzed by Evans and Jovanovic (1989), has detailed information on the self-employed, but very sparse information on the businesses they run. The Small Business Administration (SBA) data analyzed by Petersen and Rajan (1994, 1995) provide a wealth of details about the firm but very little information about the entrepreneur.

[14]These are village-run savings institutions where members pledge to save a certain amount and interest earnings are determined by the profitability of the whole institution for the year. A sizable fraction of PCGs offer loans, which are secured by savings, as well.

[15]Four years of schooling was the statutory minimum at the time most of the sample's household heads were in school.

[16]In each survey year, households were asked to report on income and expenditure for the 12 months prior to the survey. Thus, for the survey year 1997, income and expenditure figures cover the period from the spring of 1996 to the spring of 1997.

[17]We explain why financial constraints generate these predictions in the Related literature section.

[18]Households with businesses that are operating in 1997 but were founded prior to 1992 are eliminated from the analysis.

[19]In table 2, these variables are summarized in panel A for non-business households, by year, and in panel B for business households, by the year the business was started.

[20]A 1,000,000 baht increase in wealth corresponds to doubling the current wealth of the median business household in 1997 and tripling the wealth of the median non-business household.

[21]Because the sample sizes are smaller here, we do not control for past use of financial institutions and geographic location.

[22]We have experimented with different statistical models and gotten qualitatively similar results. For example, we have estimated probit models where 0 corresponds to zero initial investment and 1 corresponds to positive initial investment and ordered probit models where 0 corresponds to zero initial investment, 1 corresponds to initial investment of less than 10,000 baht, and 2 corresponds to initial investment greater than 10,000 baht.

REFERENCES

Aaronson, Daniel, Ellen R. Rissman, and Daniel G. Sullivan, 2004, "Assessing the jobless recovery," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 28, No. 2, pp. 2–20.

Aghion, Phillippe, and Patrick Bolton, 1997, "A trickle-down theory of growth and development," *Review of Economic Studies*, Vol. 64, No. 2, April, pp. 151–172.

Binford, Michael, Tae Jeong Lee, and Robert Townsend, 2001, "Sampling design for an integrated socioeconomic and ecologic survey using satellite remote sensing and ordination," University of Chicago, manuscript.

Bitler, Marianne P., Alicia M. Robb, and John D. Wolken, 2001, "Financial services used by small businesses: Evidence from the 1998 *Survey of Small Business Finance*," *Federal Reserve Bulletin*, Board of Governors of the Federal Reserve System, April, pp. 183–205.

Evans, David S., and Boyan Jovanovic, 1989, "An estimated model of entrepreneurial choice under liquidity constraints," *Journal of Political Economy*, Vol. 97, No. 4, August, pp. 808–827.

Fazzari, Stephen M., R. Glenn Hubbard, and Bruce C. Petersen, 2000, "Investment-cash flow sensitivities are useful: A comment on Kaplan and Zingales," *Quarterly Journal of Economics*, Vol. 115, No. 2, pp. 695–705.

__________, 1988, "Financing constraints and corporate investment," *Brookings Papers on Economic Activity*, Vol. 1, pp. 141–195.

Greenwood, Jeremy, and Boyan Jovanovic, 1990, "Financial development, growth, and the distribution of income," *Journal of Political Economy*, Vol. 98, No. 5, pp. 1076–1107.

Holtz-Eakin, Douglas, David Joulfaian, and Harvey S. Rosen, 1994, "Sticking it out: Entrepreneurial survival and liquidity constraints," *Journal of Political Economy*, Vol. 102, No. 1, February, pp. 53–75.

Hurst, Erik, and Annamaria Lusardi, 2004, "Liquidity constraints, household wealth, and entrepreneurship," *Journal of Political Economy,* Vol. 112, No. 2, April, pp. 319–347.

Kaboski, Joseph, and Robert Townsend, 1998, "Borrowing and lending in semi-urban and rural Thailand," University of Chicago, manuscript.

Lerner, Josh, 1999, "The government as venture capitalist: The long-run impact of the SBIR program," *Journal of Business*, Vol. 72, No. 3, July, pp. 285–318.

Lloyd-Ellis, Huw, and Dan Bernhardt, 2000, "Enterprise, inequality, and economic development," *Review of Economic Studies*, Vol. 67, No. 1, pp. 147–168.

McKenzie, David, and Christopher Woodruff, 2003, "Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises," Stanford University, manuscript.

Paulson, Anna L., and Robert M. Townsend, 2004, "Entrepreneurship and financial constraints in Thailand," *Journal of Corporate Finance*, Vol. 10, No. 2, pp. 229–262.

Paulson, Anna L., Robert M. Townsend, and Alexander K. Karaivanov, 2005, "Distinguishing limited liability from moral hazard in a model of entrepreneurship," Federal Reserve Bank of Chicago, working paper, revised April 2005, No. WP-2003-06.

Petersen, Mitchell A., and Raghuram G. Rajan, 1995, "The effect of credit market competition on lending relationships," *Quarterly Journal of Economics*, Vol. 110, No. 2, May, pp. 407–443.

__________, 1994, "The benefits of lender relationships: Evidence from small business data," *Journal of Finance*, Vol. 49, No. 1, pp. 3–37.

Rissman, Ellen R., 2003, "Self-employment as an alternative to unemployment," Federal Reserve Bank of Chicago, working paper, No. WP-2003-34.

Thailand Development Research Institute, 2003, "Thailand Economic Information Kit: 2003," Bangkok, report.

Townsend, Robert M., with Anna L. Paulson, Sombat Sakuntasathien, Tae Jeong Lee, and Michael Binford, 1997, "Questionnaire design and data collection for NICHD grant 'Risk, insurance, and the family' and NSF grants," University of Chicago.

World Bank Group, The, 2000, "Thai workers and the crisis," *Thailand Social Monitor*, July.

Seasonal monetary policy

Marcelo Veracierto

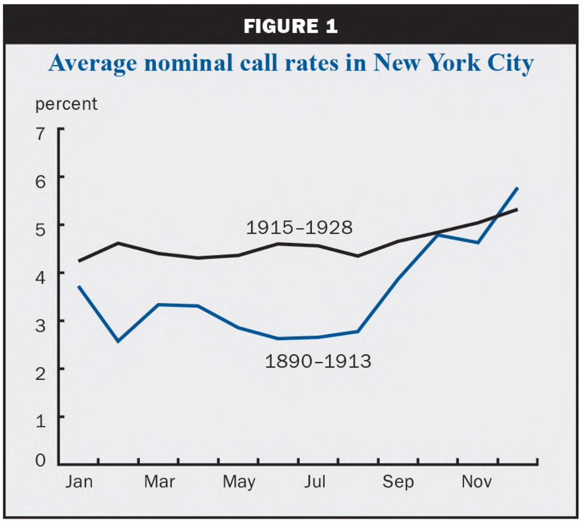
Introduction and summary

It is widely known that economic activity does not evolve smoothly over the course of a year, but that it varies systematically across the different seasons. This is not surprising: Weather is an important factor in many sectors of production. While agriculture is an obvious example, construction is another important activity affected by weather: No doubt, it is much harder to build a house in Chicago during the winter months than during the rest of the year. Institutional arrangements also lead to seasonal fluctuations in economic activity. For instance, a disproportionate fraction of American families take vacations during the summer months partly because they coincide with school recess. Another example is Christmas, which sharply increases retail activity during the last month of the year. While most modern discussion about monetary policy centers on what is the best policy to follow over booms and recessions, very little is said about what is the best policy to follow across different seasons. However, this has not always been the case. The evolution of U.S. monetary institutions and, in particular, the creation of the Federal Reserve System have been partly guided by this discussion.

Before the creation of the Federal Reserve System in 1914, the U.S. monetary system was commonly criticized for its alleged "inelasticity" in responding to fluctuations in the demand for credit. While some of these fluctuations were associated with business cycles and bank panics, an important part of them were the result of regular seasonal fluctuations in economic activity. As a matter of fact, in those days it was common for the U.S. economy to go through recurrent periods of monetary tightness during the fall crop-moving and Christmas seasons (September through December). To illustrate this, it suffices to consider the seasonal pattern for short-term interest rates. The reason is that, to the extent that the end-of-year increase in the demand for credit was not matched by a comparable increase in money supply, the short-term interest rates would have to increase. A classic source for the seasonal behavior of interest rates is Kemmerer (1910), who reported the seasonal weekly pattern for average interest rates on call loans in the New York Stock Exchange between 1890 and 1908. Indeed, Kemmerer showed a strong seasonal pattern: He reported that the call rate decreased quite rapidly from 7.38 percent during the last week of the year to 2.50 percent during the last week of January. Moreover, after a long period of relative stability, the call rate increased from 3.04 percent during the first week of September to reach a peak of 7.38 percent during the last week of the year.

To use the words of Friedman and Schwartz (1963, p. 292): "That seasonal movement was very much in the minds of the founders of the (Federal Reserve) System and was an important source of their belief in the need for an 'elastic' currency." In fact, the creation of the Federal Reserve System in 1914 changed the seasonal behavior of interest rates quite dramatically. Figure 1 shows the average call rate in New York City during the periods 1890–1913 (before the creation of the Fed) and 1915–28 (after the creation of the Fed, but before the Great Depression). For the period before the creation of the Fed, we see the same seasonal pattern that Kemmerer reported in weekly data: Interest rates rising steadily between September and December, and dropping sharply in January. During the period after the creation of the Fed, we see interest rates behaving much more smoothly. We still

Marcelo Veracierto is a senior economist in the Research Department at the Federal Reserve Bank of Chicago.
The author benefited from helpful discussions with Bruce Smith while they were colleagues at Cornell University. He also thanks David Kang and Tina Lam for research assistance.



observe a noticeable increase at the end of the year, but it is small compared to the sharp increases that took place before the creation of the Fed. This type of evidence led Friedman and Schwartz (1963, p. 293) to claim that "the System was almost entirely successful in the stated objective of eliminating seasonal strain."

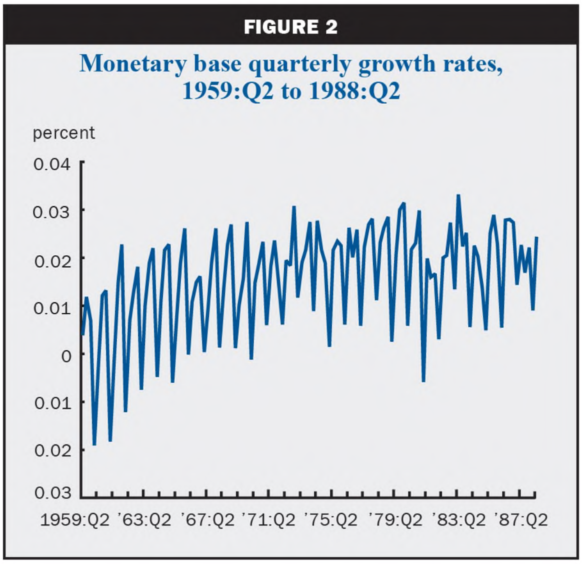
In order to attain such a smooth path for interest rates, the Federal Reserve had to meet the seasonal variations in demand with accommodating expansions and contractions in the supply of high-powered money. Indeed, after presenting supporting evidence, Friedman and Schwartz (1963, p. 294) stated that "the seasonal variation in currency outside the Treasury and Federal Reserve Banks and, we presume, in the total stock of money were decidedly wider in the 1920s than in the earlier periods." In recent times the Federal Reserve has continued to generate large seasonal variations in the quantity of money. Figure 2 reports the seasonally unadjusted monetary base growth rate between 1959:Q2 and 1988:Q2. We see that the monetary base follows a strong seasonal pattern: Its growth rate is relatively low in the first quarter of the year and increases monotonically throughout the rest of the year.

The purpose of this article is to evaluate the consequences of the Federal Reserve following this type of seasonal policy. While smoothing interest rates across seasons was one of the initial objectives of the Federal Reserve System, it is surprising how little work has been done to analyze the associated effects. Would allocations and welfare be significantly different if, instead of following an "elastic" monetary policy across seasons, the Fed would follow more of a "lean against the wind" stance? More precisely, what would be the consequences of following a constant growth rate of money instead of smoothing interest rates across seasons?

The main exercise in this article is to analyze what would the effects be of switching from the seasonal money growth rates that the Fed engineers to a constant growth rate of money. The results in terms of nominal interest rates are quite dramatic. Under a constant money growth rate, the nominal interest rate would be constant during the first three quarters, but would more than double during the last quarter of the year. That is, the pattern for nominal interest rates would resemble the one corresponding to the period before the creation of the Federal Reserve System. On the contrary, under current Federal Reserve policy, most of the seasonal variations in nominal interest rates are eliminated. Despite this, the seasonal monetary policy regime has no important consequences for real allocations: The seasonal patterns for consumption, output, hours worked, and real cash balances are basically the same if the Fed smooths interest rates or if it follows a constant rate of growth of money. As a consequence, the welfare effects of both types of policies are virtually the same.

Smoothing interest rates across the different seasons would have more significant effects if the nominal interest rate targeted were equal to zero at every quarter, that is, if the Federal Reserve followed the celebrated "Friedman rule." Output would increase by 1.1 percent in every season. However, the welfare benefits of switching to the zero interest rates would still be small: only 0.1 percent in terms of consumption.

The rest of the article is organized as follows. The related literature is discussed in the next section. I describe the environment in the third section. The benchmark economy is calibrated to U.S. data in the fourth section. I compare the effects of different



seasonal monetary rules in the fifth section. In the sixth section, I investigate the main source of seasonal fluctuations in the U.S. economy. Three appendices provide all the technical details.

Related literature

This is not the first article to analyze the effects of seasonal monetary policy.[1] Miron (1986) analyzed the problem of a large number of identical banks that take the nominal interest rate as given and must decide how to allocate their deposits into reserves and loans. The banks face a cost function, which depends on their reserve–deposit ratios and on the stochastic realization of a variable called "withdrawals." The model is closed with an exogenous amount of deposits and a demand function for loans that depends negatively on the interest rate and an exogenous activity level. The price level and inflation rate are treated as exogenous. Analyzing this framework, Miron finds that if the Federal Reserve controls the demand for loans (through open market operations) in such a way that equilibrium rates are smoothed across different seasons, banks respond by reducing their seasonal changes in reserve–deposit ratios, which in turn lowers the average costs that the banks face (given the convexity of the cost function). This result is interpreted as a reduction in the likelihood of bank panics. While the paper illustrates that smoothing interest can decrease bank panics, it is hard to assess how plausible the model is, given its highly stylized nature and the lack of quantitative analysis.

Mankiw and Miron (1991) also provide an analysis of seasonal monetary policy, but using an IS–LM framework. After parameterizing the equations to U.S. observations, they use their model to evaluate the benefits of smoothing nominal interest rates across seasons, against the alternative of holding the stock of money constant across seasons. They find, both under "classical" and "Keynesian" assumptions, that holding the stock of money constant would lead to extremely seasonal interest rates: The seasonal amplitude would be about 500 basis points. They also find that, even under extreme Keynesian assumptions about the price level, moving to a constant stock of money regime would have small effects on the seasonal behavior of output.

This article is more closely related to Mankiw and Miron (1991) than to Miron (1986), since it is completely silent on "bank panics." However, a big methodological difference is that it follows a modern dynamic general equilibrium approach instead of an IS-LM analysis. An advantage of this approach is that it allows us to evaluate any welfare benefit of changes in monetary policy. Another advantage is the internal consistency between microeconomic decisions and macroeconomic outcomes. Despite these important differences, this article obtains results that are quite similar to Mankiw and Miron (1991): Switching to a smooth money rule would lead to extremely seasonal nominal interest rates but would have negligible effects on real variables.

The model economy

This article uses a prototype model that has been previously used to evaluate the effects of monetary policy over the business cycle. The model is the one studied by Cooley and Hansen (1995), which introduces a cash-in-advance constraint similar to Lucas and Stokey (1983) into the real business cycle model analyzed by Hansen (1985). An important difference with Cooley and Hansen (1995) is that, instead of having stochastic shocks, this article introduces systematic seasonal changes in preferences, technology, and monetary policy.

The model has a representative agent that likes consuming both a cash good and a credit good, but dislikes working. The household rents labor and capital to a representative firm, which uses them to produce the two consumption goods and investment. The household uses the wage and rental income that it receives from the firm, together with a lump-sum transfer of cash that the agent receives from the government, to purchase consumption goods, investment goods, cash, and bonds. Consumption of the cash good is subject to a cash-in-advance constraint. The cash transfers that the household receives from the government are completely financed by monetary injections.

In this economy the time discount rate, the weight of the cash good in the utility function, the disutility of work, total factor productivity, and the growth rate of money vary deterministically across seasons. Parameter values will be calibrated to reproduce the seasonal fluctuations in consumption, investment, hours worked, real cash balances, and money growth rate observed in U.S. data. Once the model is calibrated to the U.S. seasonal cycles, it will be used to assess the consequences of Federal Reserve monetary policy.

Hereon, a season will be identified with a quarter. For this reason, it will be important to keep track of the year and quarter of the different variables in the model economy. In what follows, $x_{t,s}$ will denote the value of variable x in year t and quarter s, for $s = 1, ..., 4$. To simplify notation, $x_{t,0}$ will be understood to be $x_{t-1,4}$. Similarly, $x_{t,5}$ will refer to $x_{t+1,1}$. A detailed description of the model economy now follows.

The economy is populated by a large number of identical agents. Each agent is endowed with one unit of time every period and has preferences described by the following utility function:

$$1)\quad \sum_{t=0}^{\infty}\beta^{t}\sum_{s=1}^{4}\phi_{s}\left[\alpha_{s}\ln c_{t,s}+\left(1-\alpha_{s}\right)\ln a_{t,s}-\gamma_{s}h_{t,s}\right],$$

where $0 < \beta < 1$ is the annual discount factor, $c_{t,s}$ is consumption of a cash good, $a_{t,s}$ is consumption of a credit good, and $h_{t,s}$ are hours worked. Note that the parameter ϕ_s introduces a seasonal pattern in quarterly discount factors. Similarly, α_s introduces seasonal variations in the desired mix between cash and credit goods, and γ_s introduces variations in the disutility of work effort (that is, on how much agents dislike working as opposed to enjoying leisure).[2]

Output is given by the following production function:

$$y_{t,s}=z_{s}k_{t,s}^{\theta}h_{t,s}^{1-\theta},$$

where $0 < \theta < 1$, $k_{t,s}$ is capital, and $h_{t,s}$ is labor. Note that total factor productivity z_s is assumed to vary across the different seasons.

There is a standard capital accumulation technology given by:

$$2)\quad k_{t,s+1} = (1-\delta)k_{t,s} + i_{t,s},$$

where $0 < \delta < 1$ is the depreciation rate of capital, and $i_{t,s}$ is investment.

Not only are the cash good, $c_{t,s}$, and the consumption credit good, $a_{t,s}$, perfect substitutes in production, but there also is a linear technology to transform consumption goods into investment, $i_{t,s}$. The feasibility condition for output is given by

$$c_{t,s} + a_{t,s} + i_{t,s} \le y_{t,s}.$$

At the beginning of every period there is an asset trading session. Agents enter this session with $m_{t,s}$ units of cash brought from the previous period, principal plus interest payments $(1 + R_{t,s-1})b_{t,s}$ on nominal bonds purchased during the previous period, and current lump-sum cash transfers $T_{t,s}$ received from the government. Agents then acquire nominal bonds $b_{t,s+1}$ (which mature during the following period) and cash balances (which are required to purchase the cash good). Agents do not have access to any further cash balances to purchase the cash good once the asset trading session is over. Therefore, their cash-in-advance constraint is given by

$$3)\quad P_{t,s}c_{t,s} \le m_{t,s} + (1 + R_{t,s-1})b_{t,s} + T_{t,s} - b_{t,s+1},$$

where $P_{t,s}$ is the price of the cash good in terms of money. This constraint will always hold with equality as long as the nominal interest rate is positive in every season.

Aside from this cash-in-advance constraint, households are subject to the following budget constraint:

$$4)\quad a_{t,s}+i_{s}+\frac{m_{t,s+1}}{P_{t,s}}\le w_{t,s}h_{t,s}+r_{t,s}k_{t,s}+\left[\frac{m_{t,s}+\left(1+R_{t,s-1}\right)b_{t,s}+T_{t,s}-b_{t,s+1}}{P_{t,s}}-c_{t,s}\right],$$

where $w_{t,s}$ is the wage rate and $r_{t,s}$ is the rental rate of capital. This constraint states that any cash that was not used to purchase the consumption good or bonds, plus the total earnings from renting labor and capital to the firms, can be used to purchase credit consumption good, $a_{t,s}$, investment goods, i_s, and cash balances to carry into the following period, $m_{t,s+1}$.

The representative firm behaves competitively, taking the wage rate and rental rate of capital as given. The problem of the firm is to maximize profits, which are given by

$$5)\quad z_{s}k_{t,s}^{\theta}h_{t,s}^{1-\theta}-w_{t,s}h_{t,s}-r_{t,s}k_{t,s}.$$

For simplicity, I will assume that government expenditures are equal to zero and that the government doesn't issue bonds. The budget constraint of the government is then given by

$$6)\quad T_{t,s} = M_{t,s+1} - M_{t,s},$$

where $M_{t,s}$ is the aggregate stock of money in circulation. The monetary policy rule is assumed to be as follows:

$$7)\quad M_{t,s+1}=e^{\mu_{s}}M_{t,s}.$$

Observe that the government follows a constant annual growth rate of money rule, but allows the quarterly growth rate to vary in a systematic way across the different seasons.

In a competitive equilibrium: 1) households maximize their utility function (equation 1) subject to the cash-in-advance constraint, (equation 3), the budget constraint (equation 4) and the capital accumulation equation (equation 2); 2) firms maximize profits (equation 5); 3) the government budget constraint (equation 6) is satisfied; 4) the cash market clears

8) $m_{t,s} = M_{t,s}$;

and 5) the bonds market clears

9) $b_{t,s} = 0$.

The formal conditions characterizing a competitive equilibrium are described in appendix A.

Calibration

The rest of the article focuses on stationary competitive equilibria. That is, equilibria in which each real variable (including real cash balances) may take different values across the different seasons, but the seasonal values must be the same across the different years.[3] The purpose of this section is to select policy, preference, and technology parameter values such that the associated stationary competitive equilibria reproduce the seasonal fluctuations observed in the U.S. economy.

The first step in calibrating the model economy is to determine empirical counterparts for its variables. The empirical counterpart for total consumption, $c_{t,s} + a_{t,s}$, is chosen to be consumption of nondurable goods and services. At equilibrium, consumption of the cash good, $c_{t,s}$, is equal to real cash balances, $M_{t,s+1}/P_{t,s}$. Consequently, it is chosen to be the ratio of the monetary base to the Consumer Price Index. Investment, $i_{t,s}$, is in turn associated with fixed private investment plus consumption of durable goods (which entail purchases of capital goods by the households sector). Output, $y_{t,s}$, is then defined as the sum of these consumption and investment components. Finally, the empirical counterpart for hours worked, $h_{t,s}$, is given by the efficiency equivalent hours series constructed by Hansen (1993), which basically weighs the hours worked by individuals by their earnings.

Having determined the empirical counterparts for the different variables, statistical methods can be used to calculate the corresponding seasonal components. In particular, for each real variable, $x_{t,s}$, the following regression was estimated using non-seasonally adjusted time-series data:

10) $$\ln x_{t,s} = \psi_0(4 \times t + s) + \psi_1 d_1 + \psi_2 d_2 + \psi_3 d_3 + \psi_4 + \varepsilon_{t,s},$$

where ψ_0, ψ_1, ψ_2, ψ_3, and ψ_4 are coefficients, $\varepsilon_{t,s}$ is an i.i.d. (independently and identically distributed) normally distributed error with zero mean, and d_s is a dummy variable indicating the quarter (season) of $x_{t,s}$. Observe that the estimated coefficient $\hat{\psi}_0$ provides the quarterly growth rate of the variable. Since all real variables in the model economy are stationary in levels, the seasonal components x_s can then be defined as follows:

11) $$x_s = e^{\hat{\psi}_4 + \hat{\psi}_s}, \text{ for } s = 1, \ldots, 3, \text{ and}$$
$$x_4 = e^{\hat{\psi}_4},$$

where $\hat{\psi}_s$ is the estimated value of y_s, for $s = 1, \ldots, 4$.

Money, $M_{t,s}$, is the only non-stationary variable in the model. However, it is stationary in growth rates. For this reason, the following regression was estimated:

12) $$\frac{M_{t,s+1}}{M_{t,s}} = \psi_1 d_1 + \psi_2 d_2 + \psi_3 d_3 + \psi_4 + \varepsilon_{t,s},$$

where, again, ψ_1, ψ_2, ψ_3, and ψ_4 are coefficients, $\varepsilon_{t,s}$ is an i.i.d. normally distributed error with zero mean, and d_s is a dummy variable indicating the quarter (season) of $M_{t,s}$. The seasonal money growth rates μ_s are then obtained as follows:

13) $$\mu_s = \hat{\psi}_s + \hat{\psi}_4, \text{ for } s = 1, \ldots, 3, \text{ and}$$
$$\mu_4 = \hat{\psi}_4,$$

where $\hat{\psi}_s$ is the estimated value of ψ_s, for $s = 1, \ldots, 4$.

Table 1 reports the results of estimating equations (equations 10 and 12) using U.S. data. Figure 3 depicts the seasonal components obtained from equations 11 and 13 for the different variables, where the levels of all variables with meaningless units of measurement have been normalized to one during the fourth quarter (Q4). We see that the seasonal fluctuations are extremely large in U.S. data. For instance, the output level, y_s, drops to 0.926 during the first quarter (Q1), only to recover to 0.959 and 0.954 during the second (Q2) and third quarters (Q3), respectively. A similar pattern is followed by consumption, $c_s + a_s$, and investment, i_s. The seasonal pattern for hours, h_s, is also significant, but differs quite considerably from the previous variables: Its lowest level takes place during Q3, when it drops to 0.950. Real cash balances, on the other hand, have a weak seasonal pattern: In Q4, they are only 1 percent larger than during the rest of the year. However, (as was evident from figure 2) the growth rate of money, μ_s, has a strong seasonal pattern: The growth rate is basically zero during Q1, jumps to 1.7 percent during Q2, and rises slowly thereafter

TABLE 1

Regression coefficients

	Ψ_0	Ψ_1	Ψ_2	Ψ_3	Ψ_4
Consumption	.0078222	-.0737527	-.0469182	-.0472672	.9310575
$c_s + a_s$	(71.78)	(-7.02)	(-4.47)	(-4.46)	(93.67)
Real cash balances	.0025632	-.011685	-.0105482	-.0094259	.353
c_s	(18.23)	(-0.86)	(-0.78)	(-0.69)	(27.53)
Investment	.0090536	-.0852452	-.0301924	-.0414874	.0175759
i_s	(31.55)	(-3.08)	(-1.09)	(-1.49)	(0.67)
Output	.0079711	-.0768262	-.0422016	-.0471908	1.297606
y_s	(53.00)	(-5.30)	(-2.91)	(-3.23)	(94.60)
Hours	.0044359	-.0196033	-.0113247	-.051248	.913498
h_s	(64.17)	(-2.94)	(-1.70)	(-7.63)	(144.89)
Money growth rate	N.A.	-.0212816	-.0059991	-.0026545	.022972
μ_s	N.A.	(-11.20)	(-3.18)	(-1.40)	(17.10)

Note: t-statistics are in parenthesis. N.A. indicates not applicable.

reaching 2.0 percent and 2.3 percent during Q3 and Q4, respectively.

Once the seasonal components of the different variables have been determined, parameter values can be selected so that the model economy mimics them quite closely. Appendix C describes this procedure in detail. All calibrated parameter values are depicted in figure 4.

Seasonal monetary policy

While Friedman and Schwartz (1963) acknowledged that “the [Federal Reserve] System was almost entirely successful in the stated objective of eliminating seasonal strain,” they had some doubts about the desirability of this type of policy. On page 295, they give the following qualified statement: “Within the year, there seems little harm and *perhaps* some merit in permitting the stock of money to decline during the summer months and rise in the fall and winter.” At the end of the same paragraph they state “This kind of ‘elasticity’ of the total money stock is *perhaps* desirable.” Friedman (1959, p. 92) takes a much stronger position: “My own tentative conclusion is that it would be preferable to dispense with seasonal adjustments and to adopt the rule that the actual stock of money should grow month by month at the predetermined rate.”

The following question thus arises: Which policy has more merit? Smoothing interest rates across seasons, Friedman’s proposal of following a constant growth rate of money, or some other alternative? The rest of this section explores the different possibilities.

Smooth nominal interest rates

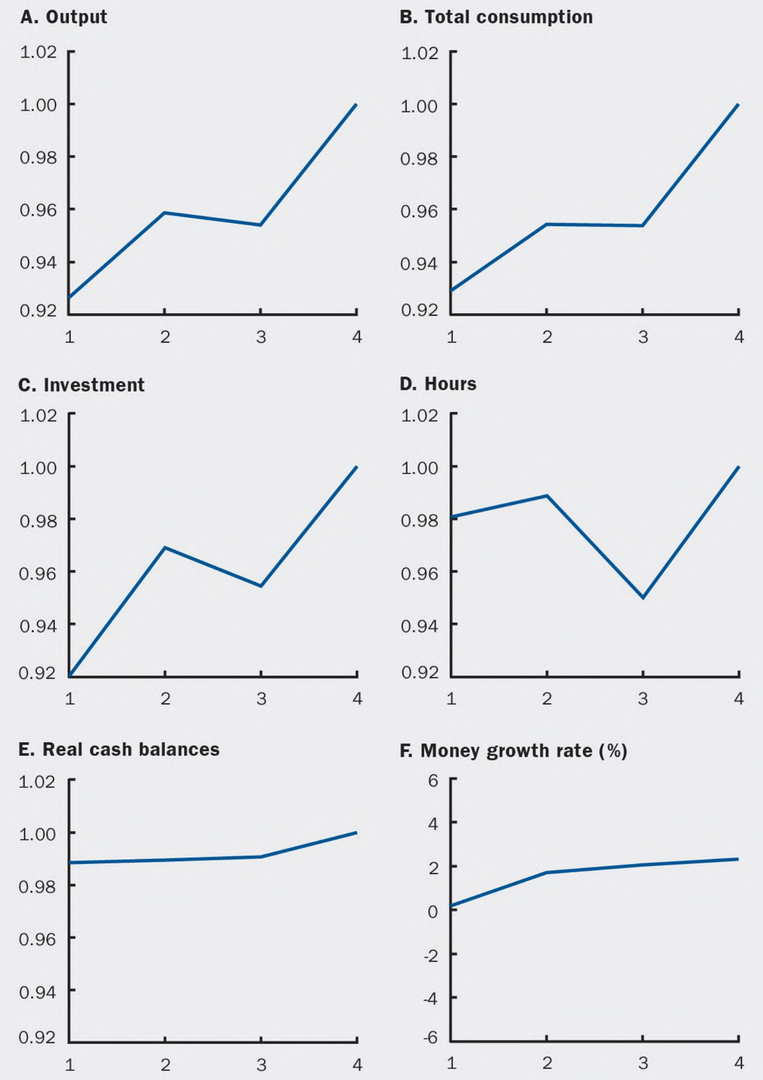
The benchmark economy was calibrated under the actual money growth rates that the U.S. implements across seasons. Figure 4, panel D shows that this policy generates nominal interest rates that are relatively smooth but are not perfectly constant. The first policy question that concerns us is then: What would be the consequences of the Fed changing its actual policy to one of perfectly smoothing interest rates?

To answer this question, I perform the following experiment. I replace the benchmark quarterly money growth rates, μ_s^*, calibrated in the previous section with a seasonal pattern that generates a constant nominal interest rate. The constant interest rate is chosen so that the annual interest rate is the same as in the benchmark economy.[4] The effects of switching to this policy are shown in figure 5. To ease comparisons, benchmark values (corresponding to the economy calibrated in the previous section) are also reported.

Figure 5, panel D, shows the change in interest rates from the benchmark case to the constant interest rate. Observe that the change in interest rates is so small that an almost imperceptible change in monetary growth rates is required to generate it (see figure 5, panel A). With a higher interest rate in the first quarter and a lower interest rate in the third quarter (relative to the benchmark economy), the constant interest rate leads to real cash balances that are somewhat smaller in the first quarter and somewhat larger in the third quarter (figure 5, panel C). This in turn leads to a higher inflation rate in the first quarter and a lower inflation rate in the third quarter (figure 5, panel B). Aside from these changes, we see that the rest of the real variables remain mostly unaffected: The effects on hours, total consumption, investment, and output are negligible. The simulation results thus suggest that the Federal Reserve Bank policy has been quite effective in terms of smoothing interest rates across seasons: Allocations would be basically the same if it completely eliminated any seasonal variations in interest rates.

FIGURE 3

U.S. seasonal patterns



Constant money growth rate

This section evaluates Friedman's recommendation of switching to a constant growth rate of money. To do this, I replace the benchmark quarterly money growth rates calibrated earlier with a constant money growth rate that generates the same annual money growth rate.[5] Figure 6 shows the results.

Figure 6, panel A depicts the constant growth rate of money. We see that, relative to the benchmark case, the growth rate of money is now higher in the first quarter and lower in the third and fourth quarters. The more expansionary monetary policy in the first quarter puts upward pressure on the nominal interest rate during the fourth quarter of the year. Similarly, the more contractionary policy during the third and fourth quarters lower nominal interest rates in the second and third quarter.[6] As a result, the interest rate becomes sharply more seasonal than in the benchmark case. In particular, switching to a constant growth rate of money would make the nominal interest rate constant at about 1.54 percent during the first quarter of the year, but would more than double during the fourth quarter of the year, to 3.34 percent (see figure 6, panel D). Thus, a constant growth rate of money would lead to the same type of increase in fourth quarter nominal interest rates that were observed previous to the creation of the Federal Reserve.

Note that the lower nominal interest rates during the second and third quarters and the higher nominal interest rate during the fourth quarter make real cash balances increase during the second and third quarter and decrease during the fourth. The reason is that the nominal interest rate is the opportunity cost of holding money. The effects on the consumption of cash goods (that is, real cash balances) translate into qualitatively similar effects for total consumption. However, the effects are much smaller in magnitude. Figure 6, panel F and panel I, show that the effects on hours and output are also negligible.

Given the small effects on real allocations, the welfare gains of moving to a constant growth rate of money are equal to zero. We conclude that perfectly smoothing interest rates across seasons or following a constant growth rate of money is irrelevant from a welfare point of view: Real variables are hardly affected.

The Friedman rule

In the two previous subsections, I found that smoothing interest rates or the growth rate of money gives rise to similar outcomes, but this doesn't mean that money does not play a role in this economy. This section shows that allocations can be significantly affected by switching to a zero nominal interest rate across seasons (that is, by implementing the "Friedman rule"). Figure 7, panel A depicts the seasonal money growth rates that are needed to implement the zero nominal interest rule.[7] Since nominal interest rates are rather smooth in the benchmark economy, but at a relatively high level, it is not surprising that this path is basically a downward shift of the benchmark path.

With the zero interest rates, real cash balances increase during each season. The reason is that real cash balances have become uniformly cheaper. This, in turn, translates into an increase in total consumption in each quarter. To satisfy this uniform increase in consumption, hours worked, output, and investment must also increase in every season. The effects are substantial: Output increases by about 1.1 percent in every quarter.

Despite the significant effects on real allocations, the welfare consequences of switching to the Friedman rule are small.[8] Agents should be compensated by having their consumption levels increase by 0.1 percent at every date, to make them indifferent with living in a world where the Fed follows the Friedman rule. The intuition for why the Friedman rule increases welfare is quite straightforward. A positive nominal interest rate makes real cash balances

FIGURE 4

Calibrated values

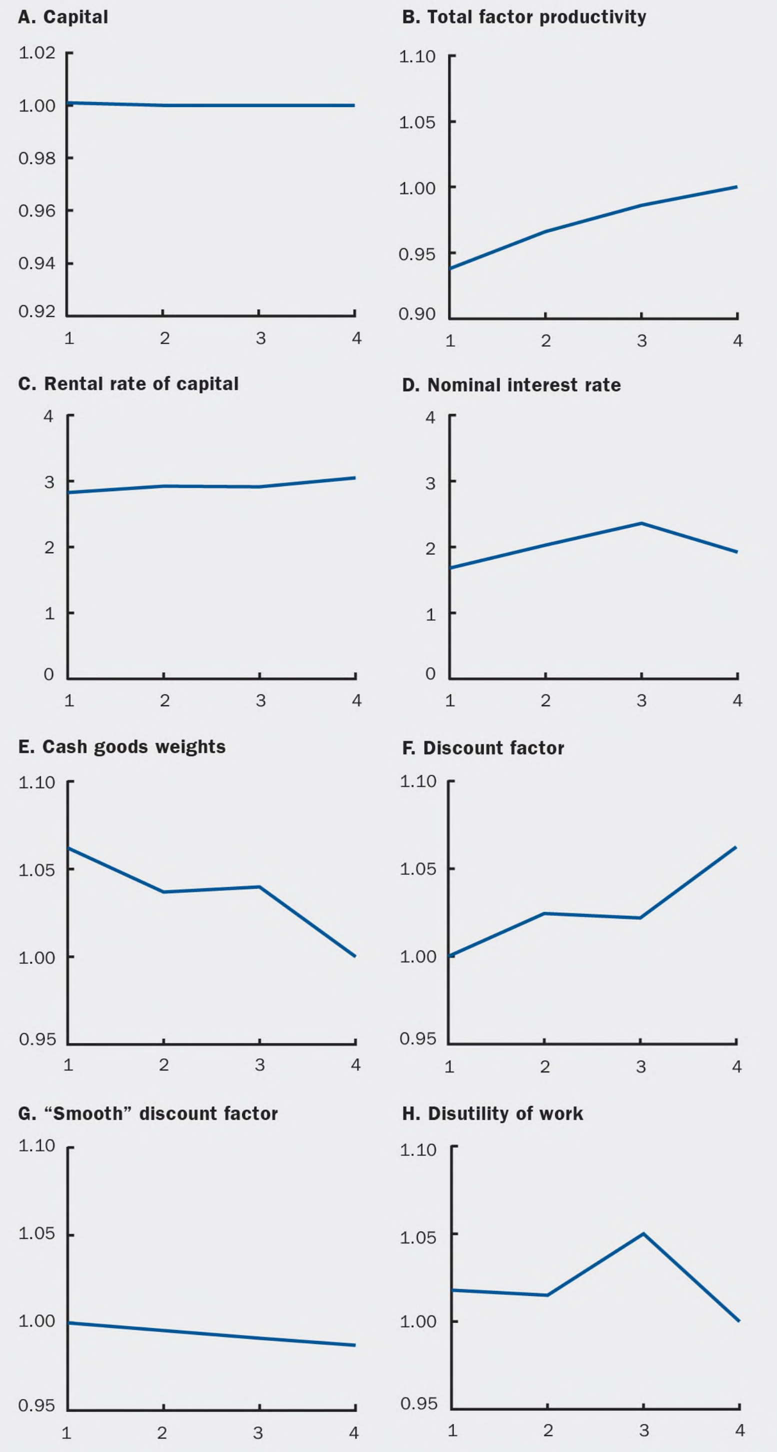
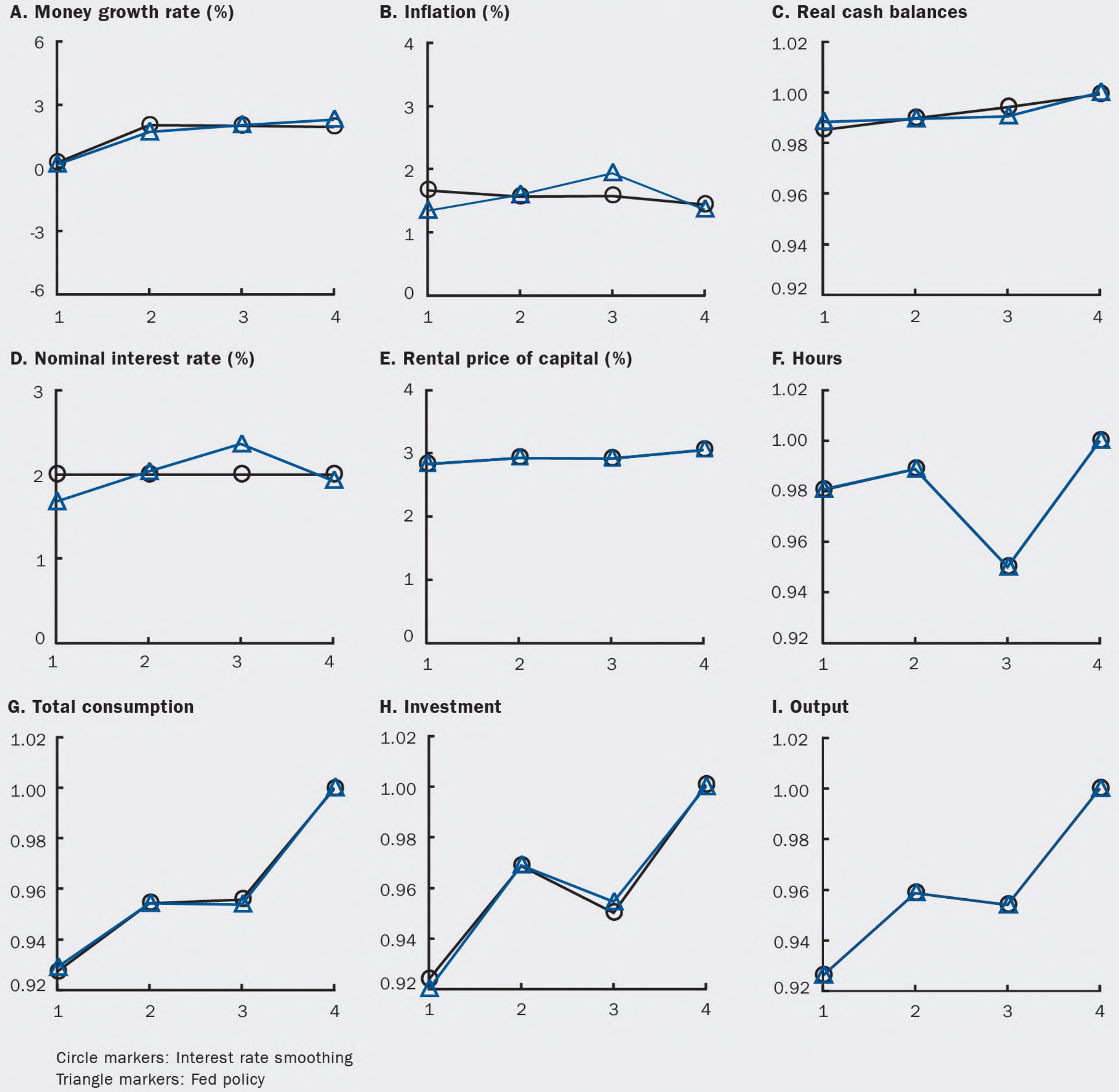


FIGURE 5

Interest rate smoothing



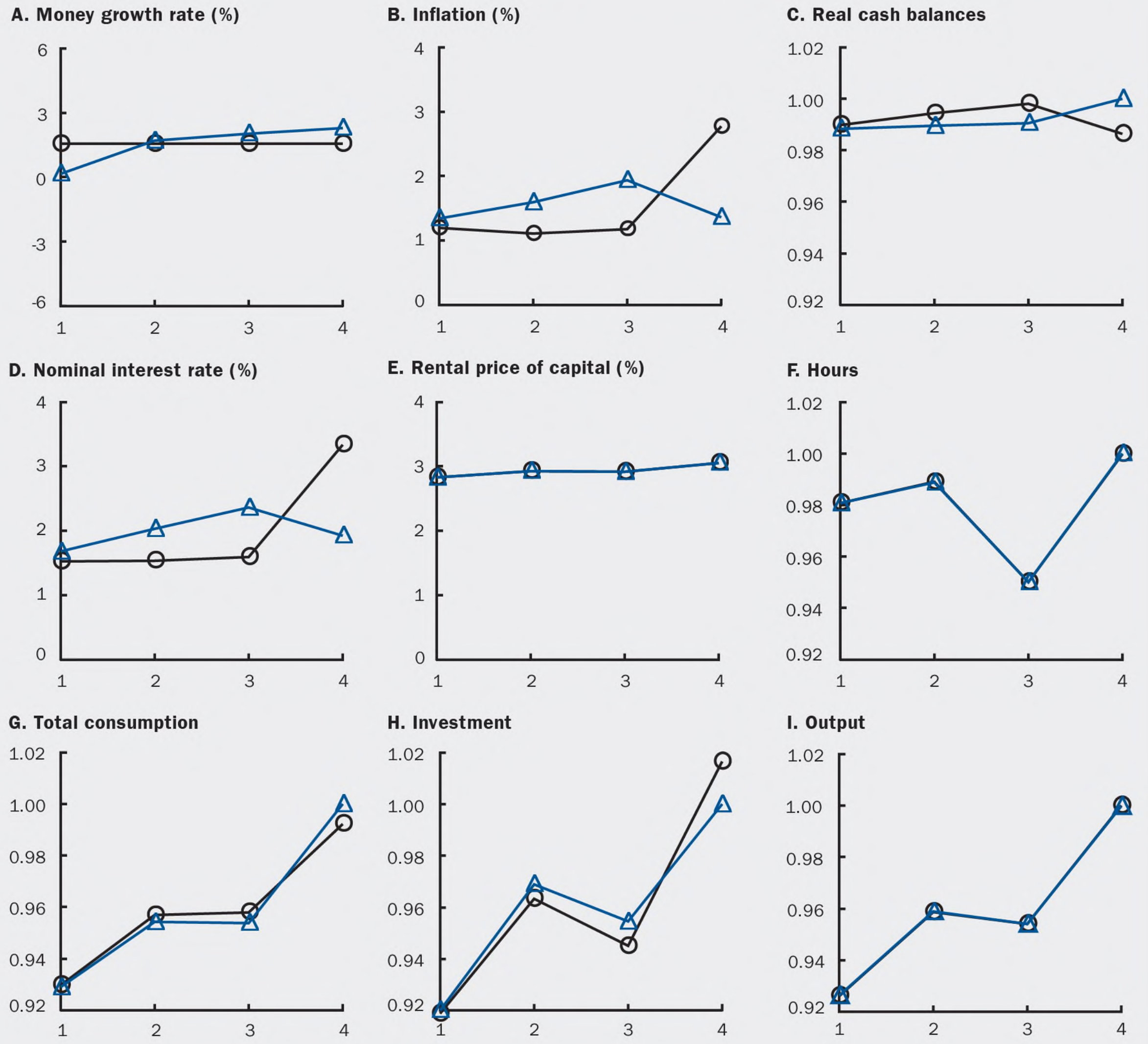
costly, so agents substitute credit goods for cash goods. However, the technological rate of transformation of cash goods to credit goods is equal to one. That is, there are no technological costs for transforming credit goods into cash goods. The only way to make agents internalize that this transformation is really costless is by driving the nominal interest rate to zero. With a zero nominal interest rate, agents are able to choose the optimal mix of credit goods and cash goods in the model economy.

The sources of seasonal fluctuations

The results so far indicate that monetary policy plays a negligible role in seasonal fluctuations. However, I have shown earlier that seasonal fluctuations in the U.S. are quite substantial. An important question that therefore remains is: What is the most important source for U.S. seasonal fluctuations? Since the model has used variations in different parameter values to generate these cycles, it can be used to explore which of these parameters play the most predominant role. This section pursues such analysis.

FIGURE 6

Constant growth rate of money



Circle markers: Constant growth rate of money
Triangle markers: Fed policy

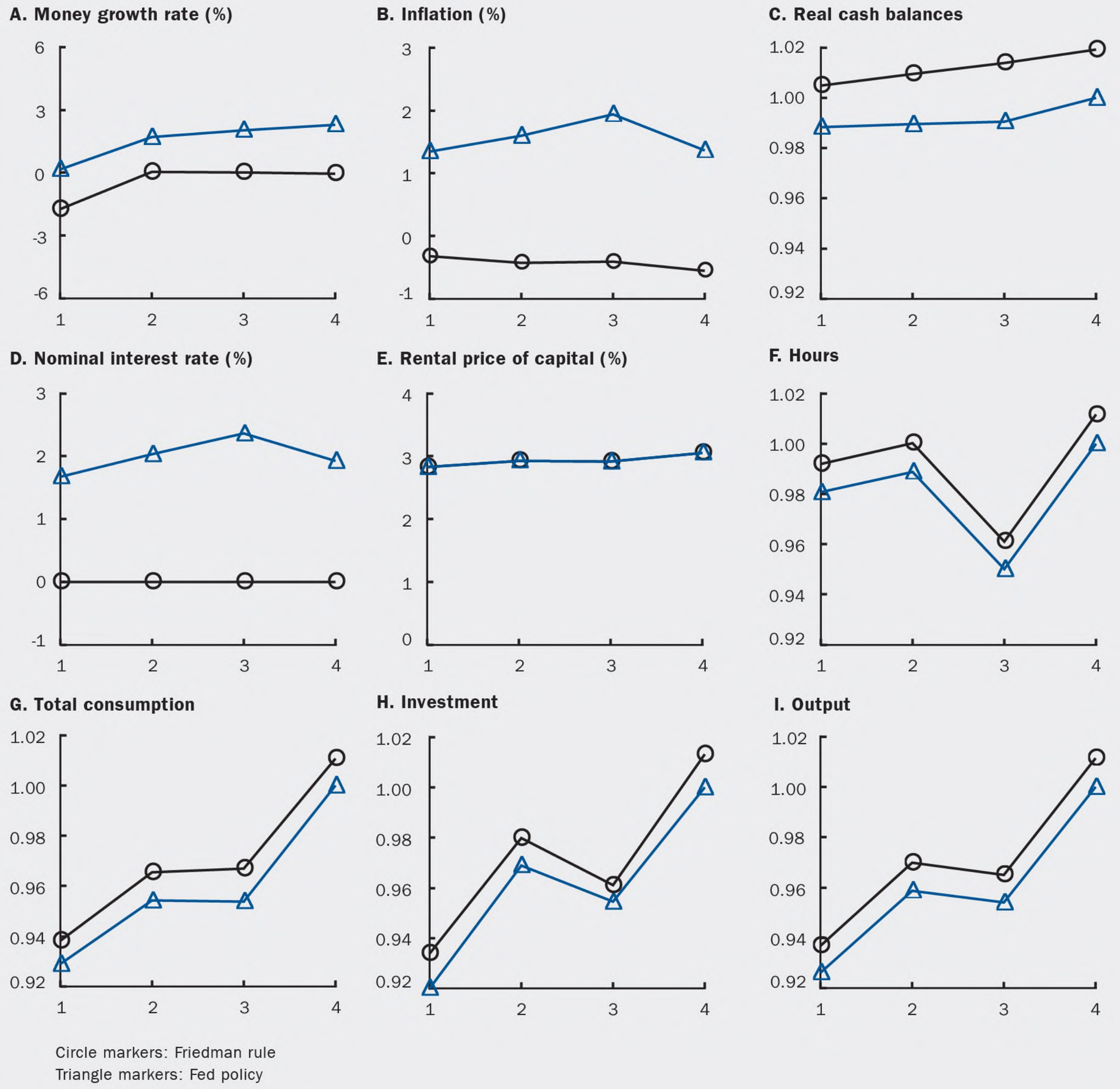
Preference weight on consumption of cash goods (α_s)

Figure 4, panel E shows that the benchmark economy embodies a strong seasonal pattern for the weight, α_s, of cash goods in the utility function. In particular, cash goods are much more valued in the first quarter of the year than in the last. To evaluate what role this plays in U.S. seasonal cycles, I perform the following experiment. I make these weights constant and equal to the cross-seasons average for the benchmark economy. Under the new constant weight, I reset the money growth rates, μ_s, so that the model generates the same seasonal pattern for nominal interest rates as in the U.S. economy. Thus, the Fed's monetary policy together with the rest of the parameter values are kept the same.

Figure 8 shows the results.[9] Removing the seasonal pattern for the α_s weights reduces real cash balances by 2.6 percent in the first quarter and increases them by 3.5 percent in the fourth quarter. But aside from that, the effects on the rest of the variables are negligible. Thus, variations in the velocity of circulation of money are found to play no important role in U.S. seasonal cycles.

FIGURE 7

The Friedman rule



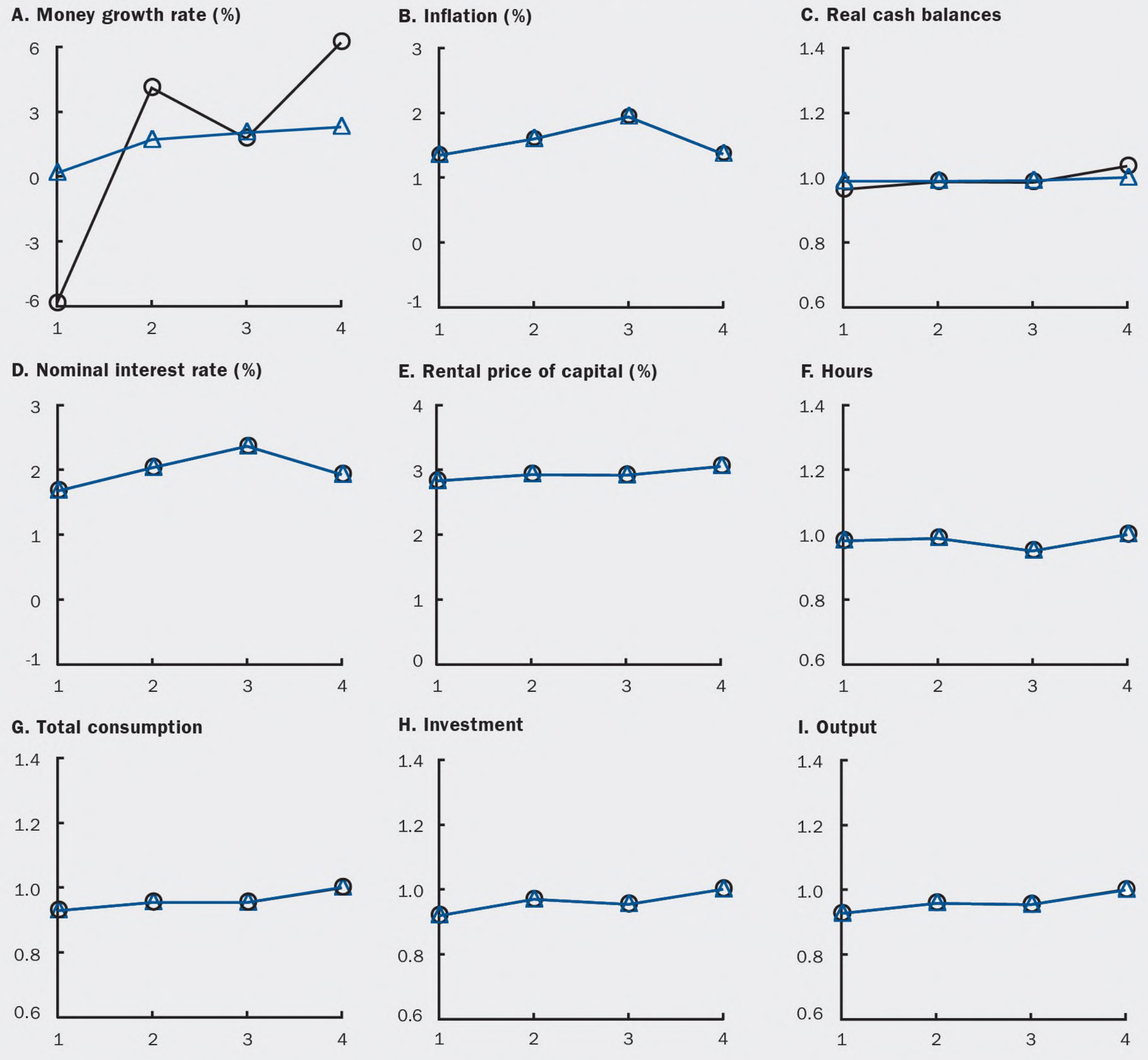
Disutility of work (γ_s)

Figure 4, panel H shows that in the benchmark economy there is a large spike in the disutility of work, γ_s, during the third quarter of the year. To evaluate what role this plays in U.S. seasonal cycles, I make the disutility of work constant and equal to the cross-seasons average for the benchmark economy. Similar to the previous subsection, I reset the money growth rates, μ_s, so that the model generates the same seasonal pattern for nominal interest rates as in the U.S economy.

Figure 9 shows the results. With a constant disutility of work, hours become 7.7 percent higher in the third quarter and 5.2 percent lower in the fourth quarter. The effects on hours worked are reflected on output, which becomes 4.8 percent higher in the third quarter and 3.3 percent lower in the fourth quarter. Given the strong preference for consumption smoothing, all the effects on output are translated into investment while consumption remains unaffected.

FIGURE 8

Smooth weights on consumption of cash goods



Circle markers: Smooth weights on consumption of cash goods
Triangle markers: Fed policy

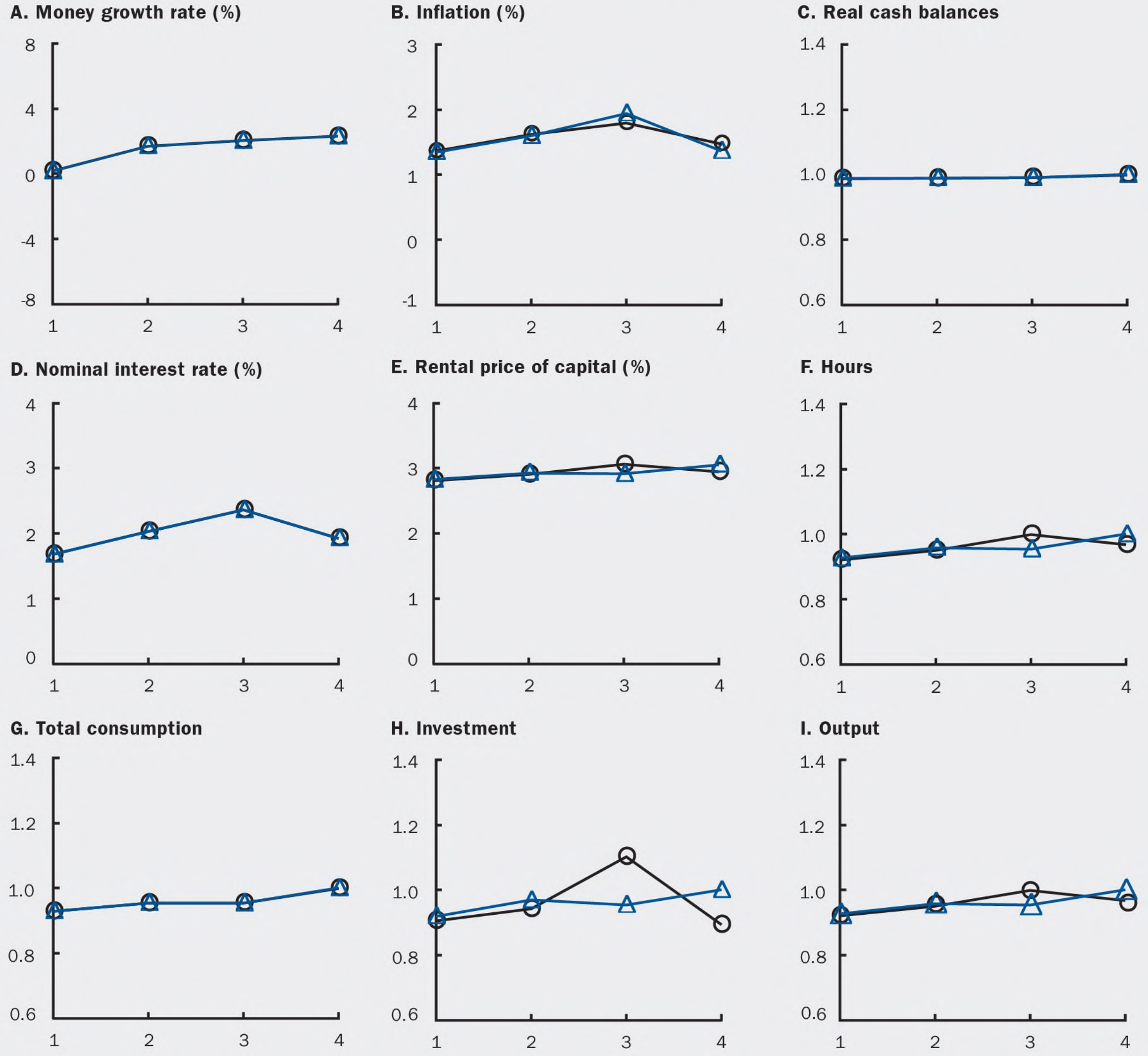
Discount factors (ϕ_s)

Figure 4, panel F shows that the discount factors increase sharply throughout the year. This section evaluates the effects of this exogenous increase by analyzing how the economy would behave if the agent discounted time equally across the seasons, that is, if the discount factors were given by those depicted in figure 4, panel G.[10]

The results are shows in figure 10. Absent the exogenous increase in discount factors throughout the year, consumption would be 3.5 percent higher in the first quarter and 3.7 percent lower in the fourth quarter. This is not surprising since with the increase in discount factors, consumption becomes more heavily weighted in the utility function toward the end of the year. Since nominal interest rates remain unchanged (by construction), the ratio of cash goods to total consumption remains the same as in the benchmark economy. As a consequence, the effects on real cash balances are a mirror of those on total consumption. Note that the smooth discount factors also make work more costly in the first quarter and less costly in the last quarter. As a consequence, hours decrease by 8.5 percent in the first quarter and increase by

FIGURE 9

Smooth disutility of labor



Circle markers: Smooth disutility of labor
Triangle markers: Fed policy

10.9 percent in the last quarter. The qualitative effects on output are the same as for hours, but they have a smaller magnitude. Investment has to decrease by 25.2 percent in the first quarter and increase by 30.4 percent in the fourth quarter to be consistent with the opposite effects on consumption and output.

Thus, exogenous changes in discount factors play a significant role in generating seasonal cycles in the U.S. economy.

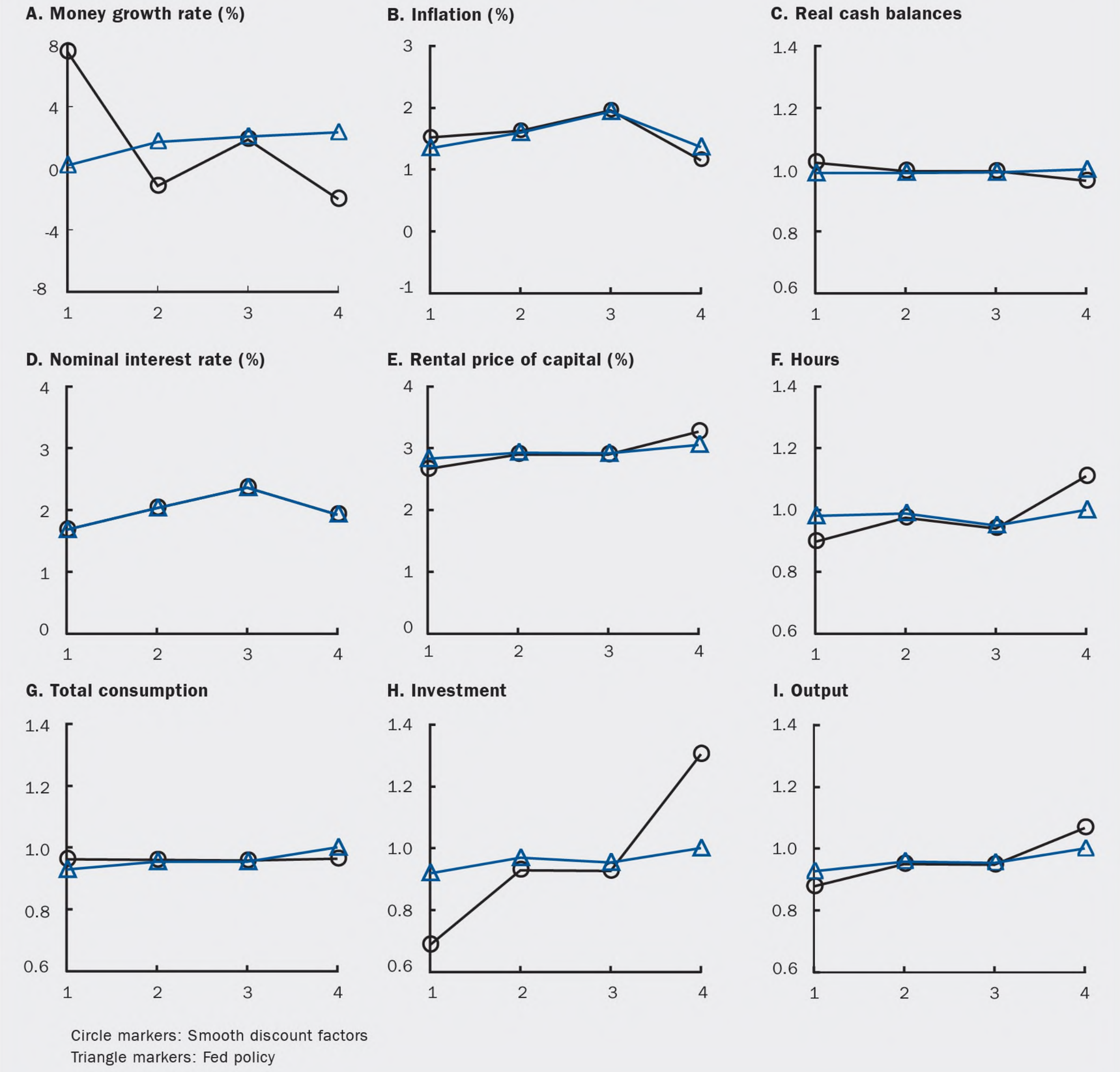
Total factor productivity (z_s)

Figure 4, panel B shows that in the benchmark economy, total factor productivity, z_s, is low in the first quarter and increases continuously throughout the year. This section analyzes the role that this plays in U.S. seasonal cycles by comparing the benchmark economy with one that has a constant total factor productivity.

The results are shown in figure 11. The strong preference for smoothing consumption over time implicit in the utility function (equation 1) means that

FIGURE 10

Smooth discount factors



the seasonal pattern for total consumption and consumption of cash goods remains unaffected by the switch to a constant total factor productivity. All the effects are felt in hours, investment, and output. This is not surprising: Since the productivity of capital is constant (instead of increasing), investment does not need to increase throughout the year. In fact, given the strong seasonal pattern in other parameters (in particular, in discount factors) investment would sharply decrease throughout the year. Since hours enter linearly in the utility function, there are no gains in smoothing them over time. As a result, the sharp decline in investment would be achieved by increasing hours by 9.6 percent during the first quarter and decreasing them by 7.1 percent during the fourth quarter, allowing consumption to remain unchanged.

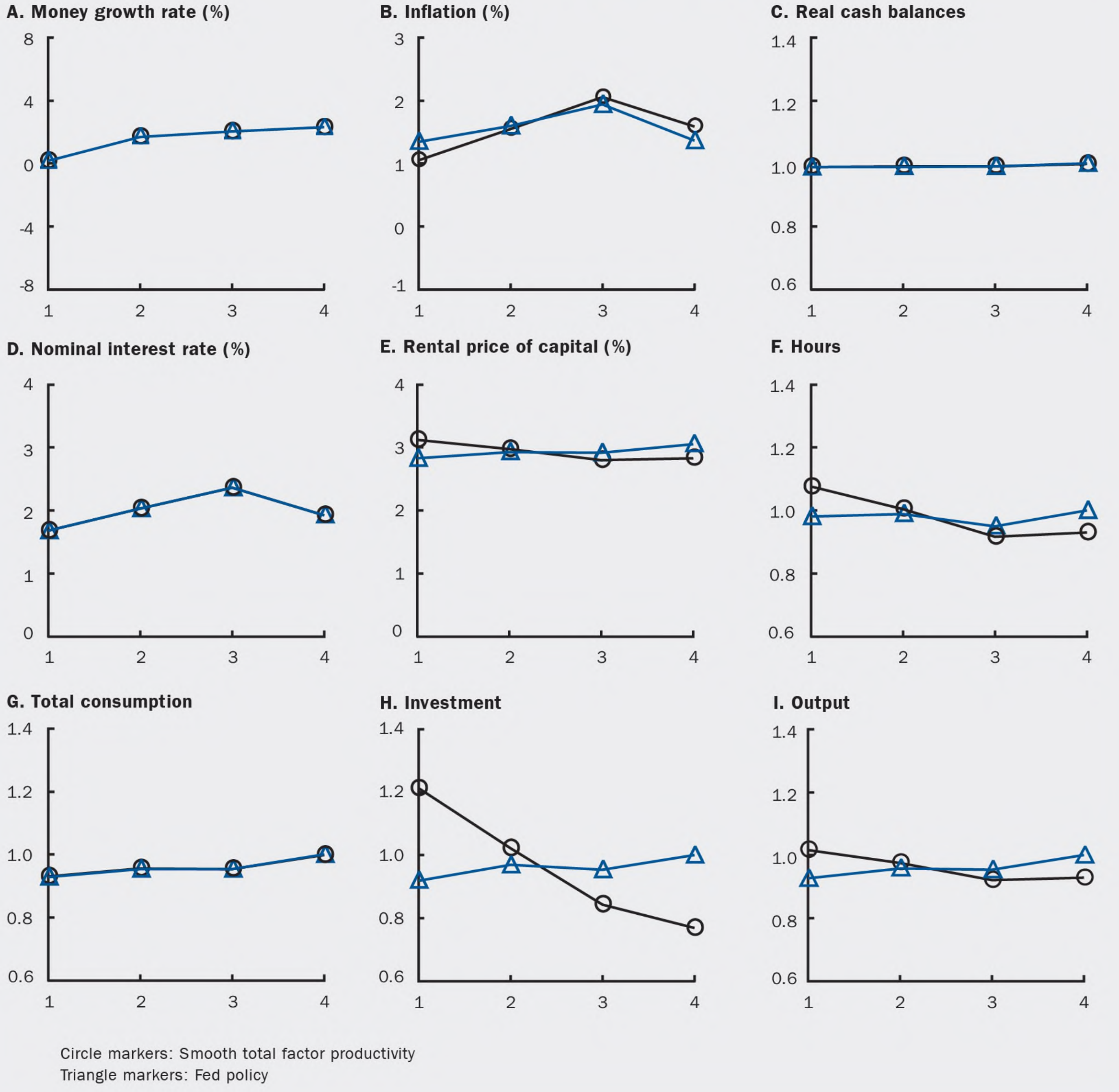
Thus, we see that seasonal variations in total factor productivity play a key role in offsetting the effects of seasonal variations in discount factors that were analyzed in the previous subsection.

Conclusion

In this article, I have used a dynamic general equilibrium cash-in-advance model to study the role

FIGURE 11

Smooth total factor productivity



of monetary policy in U.S. seasonal cycles. I have found that the seasonal monetary policy is largely irrelevant in the model economy: Smoothing interest rates across the seasons and following a constant growth rate of money lead to basically the same real allocations. Only nominal interest rates are significantly affected.

Smoothing interest rates can play a significant role if the level targeted is equal to zero. In particular, following the Friedman rule leads to considerable effects: Output increases by 1.1 percent in every quarter. However, the welfare effects are small: The consumption equivalent benefit of switching to the Friedman rule is only 0.1 percent. Not surprisingly these results are in line with Cooley and Hansen (1995), who evaluated the welfare costs of inflation abstracting from seasonal fluctuations.

I also find that the most important source of seasonal fluctuations in the U.S. economy is exogenous changes in demand, that is in how much agents value consumption over the different seasons. I find a large spike in demand during the last quarter of the year, suggesting that Christmas plays a key role, and a large drop during the first quarter, indicating that people tend to postpone consumption during cold weather. However, seasonal variations in total factor productivity play an important role in offsetting large parts of these effects. Cold weather directly affects activities like construction and agriculture, making total factor productivity hit its lowest values during the first quarter of the year. However, this does not impose much strain on the economy since demand is also the lowest during the first months of the year. After the first quarter, total factor productivity increases steadily to reach its peak during the last quarter of the year, just in time to meet the spike in aggregate demand. In turn, an increase in the value of leisure plays a significant role in flattening the path for hours, output, and investment during the third quarter of the year.

NOTES

[1]The list of papers analyzing seasonal fluctuations is more extensive than the one provided in this section, and includes Braun and Evans (1998) and Krane and Wascher (1999). However, the focus of these papers has been real activity and not monetary policy.

[2]The assumption of linear preferences with respect to labor can be justified on theoretical grounds as in Hansen (1985) and Rogerson (1988).

[3]Appendix B describes the formal conditions that a stationary competitive equilibrium must satisfy.

[4]In particular, let R_s^* be the nominal interest rates corresponding to the benchmark economy (depicted in figure 4, panel D). The constant interest rate, $\overline{R}$ chosen, satisfies the following condition:

$$\left(1+\overline{R}\right)^4 = \left(1+R_1^*\right)\left(1+R_2^*\right)\left(1+R_3^*\right)\left(1+R_4^*\right).$$

The money growth rates, μ_s, that generate this constant interest rate, $\overline{R}$, can be obtained from equations B.8 and B.9.

[5]In particular, let μ_s^* be the growth rates of money corresponding to the benchmark economy (depicted in figure 3, panel F). The constant money growth rate $\overline{\mu}$ satisfies the following condition:

$$4\overline{\mu} = \mu_1^* + \mu_2^* + \mu_3^* + \mu_4^*.$$

[6]Observe from equations B.8 and B.9 that the nominal interest rate R_s is directly related to the growth rate of money in the following quarter, μ_{s+1}.

[7]These growth rates are obtained from equations B.8 and B.9 once the R_s (for $s = 1, ...,4$) are set to zero. Note that, given the seasonal variations in ϕ_s and α_s, these money growth rates associated with the Friedman rule in general will not be constant.

[8]Despite this, the Friedman rule can be shown to be the optimal monetary policy in this environment (from a welfare standpoint).

[9]Observe that the scale for figures 8–11 is different than the scale for figures 5–7, since the effects are much larger in the former set of figures.

[10]Formally, the smooth discount factors, $\overline{\phi}_s$ are given as follows:

$$\overline{\phi}_1 = 1$$
$$\overline{\phi}_2 = \beta^{1/4}$$
$$\overline{\phi}_3 = \overline{\phi}_2^2$$
$$\overline{\phi}_4 = \overline{\phi}_2^3,$$

where β is the annual discount factor in the benchmark economy.

APPENDIX A: FIRST ORDER CONDITIONS

At year t quarter s, the household must be indifferent to two alternatives: 1) using one less unit of the cash available for purchasing the cash good and sacrificing $1/P_{t,s}$ units of the cash good, which entails a loss in marginal utility equal to $\alpha_s \phi_s / c_{t,s}$ per unit, and 2) purchasing one more unit of the bond, obtaining $1 + R_{ts}$ units of cash the following period (as interest payment) that can be used to purchase $1/P_{t,s+1}$ units of the cash good, entailing a gain in marginal utility equal to $\alpha_{s+1} \phi_{s+1} / c_{t,s+1}$ per unit. Thus, the following conditions must hold:

A.1) $$\frac{1}{P_{t,s}} \frac{\alpha_s \phi_s}{c_{t,s}} = \frac{(1+R_{ts})}{P_{t,s+1}} \frac{\alpha_{s+1}\phi_{s+1}}{c_{t,s+1}}, \text{ for } s = 1, ..., 3$$

$$\frac{1}{P_{t,4}} \frac{\alpha_4 \phi_4}{c_{t,4}} = \frac{(1+R_{t4})}{P_{t+1,4}} \frac{\alpha_1 \beta \phi_1}{c_{t+1,1}}.$$

The household must also be indifferent to: 1) purchasing one less unit of the credit good ($a_{t,s}$), which entails a loss in marginal utility equal to $\phi_s(1 - \alpha_s)/a_{t,s}$, and 2) purchasing $1/P_{t,s}$ additional units of end-of-period cash balances that next period can be used to purchase $1/P_{t,s+1}$ units of the cash good, which entails a gain in marginal utility equal to $\phi_{s+1}\alpha_{s+1}/c_{t,s+1}$ per unit. Thus the following conditions must hold:

A.2) $$\frac{1}{P_{t,s}} \frac{\phi_s (1-\alpha_s)}{a_{t,s}} = \frac{1}{P_{t,s+1}} \frac{\phi_{s+1}\alpha_{s+1}}{c_{t,s+1}}, \text{ for } s = 1, ..., 3$$

$$\frac{1}{P_{t,4}} \frac{\phi_4 (1-\alpha_4)}{\alpha_{t,4}} = \frac{1}{P_{t+1,1}} \frac{\beta \phi_1 \alpha_1}{c_{t+1,1}}.$$

The household must also be indifferent to: 1) purchasing one less unit of the credit good ($a_{t,s}$), which entails a loss in marginal utility equal to $\phi_s(1 - \alpha_s)/a_{t,s}$, and 2) purchasing one unit of capital ($k_{t,s+1}$), and renting it to the firm and selling-off the undepreciated portion, obtaining $r_{t,s+1} + 1 - \delta$ units of the credit good the following period, which entails a gain in marginal utility equal to $\phi_{s+1}(1 - \alpha_{s+1})/a_{t,s+1}$ per unit. Thus, the following conditions must hold:

A.3) $$\frac{\phi_s (1-\alpha_s)}{a_{t,s}} = (r_{t,s+1} + 1 - \delta) \frac{\phi_{s+1}(1-\alpha_{s+1})}{a_{t,s+1}}, \text{ for } s = 1, ..., 3$$

$$\frac{\phi_4 (1-\alpha_4)}{a_{t,4}} = (r_{t+1,1} + 1 - \delta) \frac{\beta \phi_1 (1-\alpha_1)}{a_{t+1,1}}.$$

Finally, the household must be indifferent to: 1) working one less unit of time, losing $w_{t,s}$ units of the credit good that the wage rate could buy, which entail a loss in marginal utility equal to $\phi_s(1 - \alpha_s)/a_s$ per unit, and 2) obtaining one more unit of leisure, which entails gain in marginal utility equal to $\phi_s \gamma_s$. Thus, the following conditions must hold:

A.4) $$\frac{(1-\alpha_s)}{a_s} = \gamma_s, \text{ for } s = 1, ..., 4.$$

The conditions characterizing the optimal behavior of the representative firm are much easier to describe. The firm hires labor up to the point where the marginal productivity of labor equals the wage rate

A.5) $$w_{t,s} = z_s k_{t,s}^{\theta} (1-\theta) h_{t,s}^{-\theta}, \text{ for } s = 1, ..., 4,$$

and hires capital up to the point where the marginal productivity of capital equals its rental rate

A.6) $$r_{t,s} = z_s \theta k_{t,s}^{\theta-1} h_{t,s}^{1-\theta}, \text{ for } s = 1, ..., 4.$$

A competitive equilibrium is then a sequence $\{c_{t,s}, a_{t,s}, h_{t,s}, k_{t,s}, m_{t,s}, b_{t,s}, w_{t,s}, r_{t,s}, P_{t,s}, R_{t,s}, T_{t,s}, M_{t,s}\}$ for t = 0, ..., ∞, and s = 1, ..., 4, such that equations 2, 3, 4, 6, 7, 8, 9, A.1, A.2, A.3, A.4, A.5, and A.6 hold.

APPENDIX B: STATIONARY EQUILIBRIA

A stationary equilibrium is a vector (c_s, a_s, i_s, y_s, k_s, h_s, r_s, w_s, R_s), for s = 1, ..., 4, such that the following equations are satisfied:

B.1) $c_s + a_s + i_s = y_s$,

B.2) $k_{s+1} = (1-\delta)k_s + i_s$,

B.3) $y_s = z_s k_s^{\theta} h_s^{1-\theta}$,

B.4) $r_s = \theta \frac{y_s}{k_s}$,

B.5) $w_s = (1-\theta)\frac{y_s}{h_s}$,

B.6) $(1+R_s)\frac{c_s}{c_{s-1}}\frac{1}{e^{\mu_s}} = r_s + 1 - \delta$,

B.7) $\frac{c_s + a_s}{c_s} = \frac{1}{\alpha_s} + \frac{(1-\alpha_s)}{\alpha_s}R_s$,

B.8) $1 = \frac{\phi_{s+1}}{\phi_s}\frac{\alpha_{s+1}}{\alpha_s}\frac{1}{e^{\mu_{s+1}}}(1+R_s)$, (except for $s = 4$),

B.9) $1 = \beta\frac{\phi_1}{\phi_4}\frac{\alpha_1}{\alpha_4}\frac{1}{e^{\mu_1}}(1+R_4)$, and

B.10) $a_s = \left(\frac{1-\alpha_s}{\gamma_s}\right)(1-\theta)\frac{y_s}{h_s}$,

for s = 1, ..., 4.

APPENDIX C: PARAMETERIZATION

This appendix describes the procedure used to calibrate parameter values.

The depreciation rate of capital, δ, is chosen to be 0.025, which is a standard value in the real business cycle literature. The seasonal pattern for the stock of capital, k_s, is then chosen to reproduce the seasonal pattern for investment, i_s, when $\delta = 0.025$. The result is depicted in figure 4, panel A, which shows no significant seasonal variations for the stock of capital, k_s. This result is obtained, despite the strong seasonal pattern in investment, because investment is small relative to the size of capital.

The share of capital in national income is given, at equilibrium, by the curvature parameter θ in the production function. For this reason, θ is chosen to be 0.36, which is the share of capital implicit in the *National Income and Product Accounts*. Given θ, and the seasonal components for capital, k_s, hours, h_s, and output, y_s, the seasonal pattern for total factor productivity, z_s, can be obtained as a residual from the production function. The result is depicted in figure 4, panel B, which shows a strong seasonal pattern: Total factor productivity drops to 0.938 during Q1 and slowly recovers thereafter, reaching 0.966 and 0.986 during Q2 and Q3, respectively.

Given the capital share, θ, the capital–output ratios, k_s/y_s, have direct implications for the rental rate of capital, r_s, in the model economy. Figure 4, panel C shows that this rental rate has a significant seasonal pattern, taking the lowest value during Q1.

The rental rate of capital and the depreciation rate determine the seasonal pattern for the real interest rate in the economy. Considering the seasonal inflation rate pattern implied by real cash balances, c_s, and the money growth rate, μ_s, the nominal interest rates, R_s, can be obtained from a version of the Fisher equation. Figure 4, panel D, shows that the nominal interest rate goes through significant seasonal variations: It ranges from 1.67 percent during Q1 to 2.36 percent during Q3.

The weight of cash goods in the utility function, α_s, is a key determinant of the relation between the nominal interest rate, R_s, and the velocity of circulation of money, $c_s/(c_s + a_s)$, that is, of the demand for money. As a consequence, it was chosen to be consistent with the values for the nominal interest rate, R_s, real cash balances, c_s, and total consumption, $c_s + a_s$, obtained above. The weights, α_s, thus obtained are reported in figure 4, panel E. We see that they have a strong seasonal pattern, the desirability of cash goods being the highest during Q1 and decreasing smoothly throughout the rest of the year.

Given these weights α_s, the discount factors β, ϕ_1, ϕ_2, ϕ_3, and ϕ_4 were selected to be consistent with the nominal interest rates, R_s, and money growth rates, μ_s, reported above. Figure 4, panel F reports that these discount factors have a strong seasonal pattern. To make this clear, figure 4, panel G reports the discount factors that the representative agent should have if it discounted time equally across the seasons. We see that both paths differ quite substantially. In particular, the seasonal pattern for the calibrated values of ϕ_1, ϕ_2, ϕ_3, and ϕ_4 indicate a monotone increase in demand throughout the year, which becomes particularly sharp during Q4.

Finally, the disutility of work parameters, γ_s, are selected to reproduce the seasonal pattern for total hours worked, h_s. The resulting values of γ_s in figure 4, panel H indicate a large increase in the disutility of work during Q3 and a sharp reversal during Q4.

The rest of the appendix describes in detail which equations were used in each stage of the calibration procedure.

The following variables are directly obtained from the data (as described in the model economy section): i_s, c_s, a_s, h_s, and μ_s. Given these variables, model parameters are selected as follows.

1) Set $\delta = 0.025$.
2) Given i_s (for $s = 1, ..., 4$), choose seasonal pattern for k_s, that is consistent with equation B.2.
3) Set $\theta = 0.36$.
4) Given c_s, a_s, and i_s, obtain y_s from equation B.1.
5) Given y_s, k_s, h_s and θ, obtain z_s from equation B.3.
6) Given y_s, k_s and θ, obtain r_s from equation B.4.
7) Given c_s, μ_s, r_s, and δ, obtain R_s from equation B.6.
8) Given R_s, c_s, and a_s, obtain α_s from equation B.7.
9) Given α_s, μ_s, and R_s, set ϕ_1 (this is just a normalization) and obtain ϕ_s, for $s = 2, ..., 4$ and β from equations B.8 and B.9.
10) Given α_s, θ, a_s, y_s, and h_s, get γ_s from B.10.

REFERENCES

Braun, R. A. and C. L. Evans, 1998, "Seasonal Solow residuals and Christmas: A case for labor hoarding and increasing returns," *Journal of Money, Credit, and Banking*, Vol. 30, No. 3, pp. 306–330.

Cooley T. F., and G. D. Hansen, 1995, "Money and the business cycle," in *Frontiers of Business Cycle Research*, T. F. Cooley (ed.), Princeton, NJ: Princeton University Press.

Friedman, M., 1959, *A Program for Monetary Stability*, New York City: Fordham University Press.

Friedman, M., and A. J. Schwartz, 1963, *A Monetary History of the United States, 1867–1960*, Princeton, NJ: Princeton University Press.

Hansen, G., 1993, "The cyclical and secular behavior of the labor input: Comparing efficiency units and hours worked," *Journal of Applied Econometrics*, Vol. 8, No. 1, pp. 71–80.

__________, 1985, "Indivisible labor and the business cycle," *Journal of Monetary Economics*, Vol. 16, No. 3, pp. 309–327.

Kemmerer, E. W., 1910, "Seasonal variations in the relative demand for money and capital in the United States," National Monetary Commission, report.

Krane, S., and W. Wascher, 1999, "The cyclical sensitivity of seasonality in U.S. employment," *Journal of Monetary Economics*, Vol. 44, No. 3, pp. 523–553.

Lucas, R. E., and N. L. Stokey, 1983, "Optimal fiscal and monetary policy in an economy without capital," *Journal of Monetary Economics*, Vol. 12, pp. 55–93.

Mankiw, N. G., and J. A. Miron, 1991, "Should the Fed smooth interest rates? The case of seasonal monetary policy," *Carnegie-Rochester Conference Series on Public Policy*, Vol. 34, pp. 41–69

Miron, J. A., 1986, "Financial panics, the seasonality of the nominal interest rate, and the founding of the Fed," *American Economic Review*, Vol. 76, No 1, pp. 125–140.

Rogerson, R., 1988, "Indivisible labor, lotteries, and equilibrium," *Journal of Monetary Economics*, Vol. 21, No. 1, pp. 3–16.