Federal Reserve Bank of Chicago

Ninth International Banking Conference Invitation

Third Quarter 2006

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 2006, Volume XXX, Issue 3

Right before the end: Asset decumulation at the end of life

Eric French, Mariacristina De Nardi, John Bailey Jones, Olesya Baker, and Phil Doctor

Introduction and summary

What happens to people's assets in the period immediately preceding their death? This is an important question for a number of reasons. To begin with, the elderly have a lot of wealth—households whose heads are 65 or older account for more than one-third of U.S. household wealth[1]—and the way in which they manage all this wealth may depend critically on end-of-life events. If, for example, elderly people are afraid of incurring large medical expenses just before they die, they might keep large amounts of assets even in very old age and not run down their assets until their illnesses appear terminal. Moreover, the need to pay for end-of-life expenses should affect the amount of wealth that younger households accumulate to fund their retirement. The issue of whether working households are saving enough has raised enough debate to warrant its own chapter in the current *Economic Report of the President*. This debate cannot be resolved until we learn the magnitude of end-of-life expenses.

In this article, we use data from the *Asset and Health Dynamics of the Oldest Old* (AHEAD), collected by scholars at the University of Michigan, to track the assets and expenses of elderly households in their last years of life. We find that the assets of people who die decline much faster than the assets of people who survive, even after controlling for age, sex, and initial asset levels. For single-person households, average wealth declines by 30 percent in the year preceding death and by 50 percent in the three years preceding death. The assets of single survivors with characteristics similar to those of the deceased, on the other hand, are flat over the same period.

Our main finding is that death is often preceded by a costly illness. Out-of-pocket medical expenditures related to drug costs, doctor visits, and hospital and nursing home stays go up by about 200 percent in the few years before death. The increase in medical spending before death, combined with burial expenses, can explain about 24 percent of the decline in assets of the soon-to-be deceased and about 37 percent of the decline in assets in the last year of life. In short, out-of-pocket medical expenses right before death can deplete the assets of many elderly households and constitute an important reason to keep assets in old age. Our findings also suggest that even if government insurance and transfer programs, such as Medicare, Medicaid, or Supplemental Security Income, are shielding the elderly from medical expenses, the coverage is far from complete.

Related literature and contributions of our article

The principal model that economists use to understand saving and to project how policy changes—such as changes in social security or Medicare—will affect saving is the life-cycle model.[2] The life-cycle model assumes that people are forward-looking and base their consumption and saving decisions on their preferences for consumption and knowledge of their future income. In the simplest version of the model, individuals know with perfect certainty the age at which they will die, and they place no value on leaving bequests to their children. Under this framework, individuals choose to die with no wealth: An individual in the last period of life, knowing that he is going to die and receiving no utility from bequests, forgoes utility if he does not consume all of his remaining wealth.

Eric French and Mariacristina De Nardi are senior economists at the Federal Reserve Bank of Chicago. John Bailey Jones is an associate professor of economics at the University at Albany, State University of New York. Olesya Baker and Phil Doctor are associate economists at the Federal Reserve Bank of Chicago. *The authors thank Kristin Butcher, Gene Amromin, Craig Furfine, and Wojciech Kopczuk for helpful comments.*

This simple version of the life-cycle model, however, is at odds with a large body of empirical research showing that many households retain large amounts of assets even in very old age (see Hurd, 1990, for a review). There are several reasons why a simple life-cycle model might underpredict saving by the elderly. First, people are usually uncertain about the exact time of their death and might need to save against the possibility of a long life. Second, people might save in order to bequeath assets to their children. Third, people might retain assets in case they need to pay for large out-of-pocket medical expenses.[3]

In this article, we consider another possibility, namely, that the data are incomplete. Although the age profile of assets has been carefully documented, we are among the first to document the changes in assets and medical expenditures that occur immediately before death. This is extremely important because many studies do not include people that are just about to die. For example, people in nursing homes are not included in the core sample of the *Survey of Consumer Finances* (SCF), although they are included in the high wealth oversample. If high medical expenses incurred from nursing home visits cause assets to decline before death, this will be completely missed in any study that uses SCF data. The predictions of the life-cycle model refer to what happens immediately before death: We are constructing a more precise test of the theory.

In addition, by studying the medical expenditures that are incurred right before death, we can develop a better measure of how much medical expense risk households face. This helps us better understand the importance of precautionary motives. Finally, our analysis helps us evaluate the risk of poverty immediately before death, an issue of major policy relevance.

Our work in this article is arguably most related to work by Hurd and Smith (2001), Yun (2003), and Hoover et al. (2004). Hurd and Smith find that assets decline very little at death, but base their comparison on the 1993 wave of the AHEAD, for which assets appear to be underreported much more severely than in other waves. Yun uses AHEAD data to consider how "expecting death" affects asset growth and, like us, finds that people run down their assets as they approach death. Using the *Medicare Current Beneficiary Survey*, Hoover et al. find that medical expenditures rise dramatically in the final year of life.

Data

We use data from the *Asset and Health Dynamics Among the Oldest Old* data set. The AHEAD is a survey of individuals who in 1993 were both noninstitutionalized and aged 70 or older. A total of 8,222 individuals in 6,047 households were interviewed for the first wave in 1993. These individuals were interviewed again in 1995, 1998, 2000, and 2002. The AHEAD data include a nationally representative core sample, as well as additional samples of blacks, Hispanics, and Florida residents.

The AHEAD has comprehensive asset measures. It has information on the value of housing and real estate, autos, liquid assets (which include money market accounts, savings accounts, and Treasury bills), individual retirement accounts (IRAs), Keogh plans, stocks, the value of farms or businesses, mutual funds, bonds, and "other" assets and investment trusts. Our measure of wealth is the sum of all these assets, less mortgages and other debts. Following common practice (for example, Hurd, 1989; and Attanasio and Hoynes, 2000), we exclude pension and social security wealth. Because assets appear to be significantly underreported in the first wave (see Rohwedder, Haider, and Hurd, 2004), we begin our analysis with data from wave two (the 1995 wave).

Given their age, many members of the sample die over the sample period. A novel feature of the AHEAD data is that survivors of the deceased (usually either a surviving spouse or a child) are interviewed. Survivors are asked about the value of the estate, insurance payments, medical expenses immediately preceding death, and costs associated with death, such as burial expenses. The key variable is the value of the estate. The exact question asked of the survivors of the deceased varied year to year, but in most years they were asked: "Altogether, what was the value of (his/her) total estate?" How the survivors of the deceased interpret this question is unclear. However, in appendix A, we show that, at least for the children of the deceased, the most likely interpretation of the question is the total value of the estate, inclusive of all possessions, such as the house and other valuables. If one member of the household dies but the other survives, we do not use the response to the estate question, but instead use the survivor's responses to the usual asset questionnaire.

There are several problems with our asset data. The first is that the wealthy tend to underreport their wealth in virtually every household survey (Davies and Shorrocks, 2000). This leads us to understate asset levels at all ages. However, Juster, Smith, and Stafford (1999) show that the wealth distribution of the AHEAD matches up well with aggregate values for all but the richest 1 percent of households. A second problem with our data is that it spans the years 1995 to 2002, a period in which there was a rapid rise in asset prices. This makes it difficult for us to distinguish between intended asset growth due to active

saving and unintended asset growth due to unexpectedly high returns.

Our data also suffer from attrition—people leaving the sample over time—a problem common to all panel data sets. In the AHEAD, attrition is largely due to death: Reported deaths are confirmed using the National Death Index. However, in some cases, interviewers are unable to track down sample members as they move from house to house, and some individuals refuse to give follow-up interviews. If the people who are difficult to contact differ systematically from those we are able to keep track of, "nondeath" attrition could distort the composition of our sample. If, for example, it is more difficult to track down poor individuals, poor households will be dropped from the sample at greater rates than the rich ones.

Two additional problems arise from the fact that assets are a household-level rather than an individual-level variable. First, some of the households in our sample consist of two unmarried individuals. Because it is not clear how these respondents might answer the asset questions, we drop these households. Second, many sample members get married or divorced over the sample period. Therefore, changes in wealth over time reflect not only savings decisions (the object of interest in this study), but also household formation decisions. To counter this problem, we drop individuals who get married or divorced during the sample period. To sum up, we keep only those households that were either married or single living alone in wave one and that changed household structure only because of death.[4]

Table 1 presents some descriptive statistics of our sample and reports average asset holdings by wave. (Assets are measured in 1998 dollars and do not include the value of any estates.) Our analysis begins with 3,880 households in 1995, of which 2,312 have at least one surviving member in 2002. Housing is the largest component of our households' portfolios, but liquid assets (such as bonds) and stocks are also important.

Tables 2 and 3 summarize the demographic transitions in our sample by showing how household composition changes between 1995 and 2002. Table 2 shows that of the 501 single men alive in 1995, 62 percent (309) had died by 2002. Of the single women, 48 percent were dead in 2002. Table 3 shows that of the 1,165 married couples that were alive in 1995, 32 percent had just the male die, 11 percent had just the female die, and 18 percent had both members die during the sample period.

Table 4 shows starting and ending wealth, by 1995 and 2002 household structure, for people who were initially single. The leftmost column in table 4 shows 1995 and 2002 wealth for men who did not die during the sample period. The second column of table 4 shows wealth in 1995 and wealth at the time of death for men who did die between 1995 and 2002. Table 4 displays two measures of wealth at the time of death. The first measure ("excluding estates") is

TABLE 1

Household wealth, by asset type and year

	1995	1998	2000	2002
Housing	75,391	77,051	84,256	84,183
Liquid assets	52,078	47,236	48,081	61,021
Stocks	49,946	52,675	50,387	43,619
Automobiles	4,778	5,148	4,804	4,550
Businesses	12,057	8,916	8,480	17,183
Individual retirement accounts	7,558	9,013	11,312	7,963
Other assets	3,278	5,553	4,606	3,525
Debt	2,456	2,649	2,501	3,052
Total assets	202,630	202,943	209,425	218,992
Observations	3,880	3,303	2,777	2,312

Notes: Table does not include the value of estates. All values are in 1998 dollars.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

TABLE 2

Survival probabilities of singles

	Single male	Single female
Alive in 1995	501	2,214
Alive in 2002	192 (38)	1,161 (52)
Dead in 2002	309 (62)	1,053 (48)

Note: Percentage of sample is in parentheses.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

TABLE 3

Survival probabilities of married couples

Married couples in 1995	1,165
Both alive in 2002	457 (39)
Wife alive in 2002	376 (32)
Husband alive in 2002	128 (11)
Both dead in 2002	204 (18)

Note: Percentage of sample is in parentheses.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

TABLE 4

Singles' characteristics, by demographic status in 1995 and 2002

	Single male in 1995		Single female in 1995		All singles in 1995	
2002 household structure	**Single male**	**Dead**	**Single female**	**Dead**	**Single**	**Dead**
1995 assets	202,035	219,675	159,197	126,454	165,276	147,603
Final assets, excluding estates	226,228	210,740	156,258	107,634	166,187	131,026
Final assets, including estates	226,228	110,380	156,258	98,732	166,187	101,374
Death expenses		5,779		4,590		4,860
Death insurance payouts		422		528		504
Life insurance payouts		2,857		1,312		1,663
Observations	192	309	1,161	1,053	1,353	1,362

Notes: Assets in the final period refer to assets in 2002 if the household survived to 2002; otherwise, excluding estate value means excluding assets in the final period before death. Assets are in 1998 dollars. Observations refer to the number of observations that made the demographic transition.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

the wealth reported for the last year the person is alive. The second measure ("including estates") replaces, when possible, the previous measure of wealth with the value of the estate reported by the person's survivors. The third and fourth columns show the corresponding statistics for single women. Both sets of columns indicate that assets decline much more quickly for the deceased than for the survivors. Table 5 shows the same data for households that initially consisted of married couples. Table 5 also suggests that death is associated with a faster rate of asset decline.

Empirical methodology

In considering how assets behave immediately before death, it is useful to work with the asset accumulation equation:

$$1)\quad A(it+1) = (1+r)\,A(it) + y(it) - m(it) - e(it) - c(it),$$

where $A(it)$ denotes assets of individual i at time t, r denotes the interest rate, $y(it)$ denotes income (from social security, pensions, and so on), $m(it)$ denotes medical expenses, $e(it)$ denotes end-of-life expenses (burial fees, less insurance payouts), and $c(it)$ denotes consumption. In short, assets can fall at the end of life because consumption rises, medical expenditures rise, income falls, or end-of-life expenses are high.[5]

While we have good measures of $A(it)$, r, $y(it)$, $m(it)$, and $e(it)$, we do not have a good measure of consumption, $c(it)$. The consumption measures contained in the AHEAD are very poor, and moreover, it is always difficult to measure the service flow from housing and durables. Fortunately, we can use equation 1 to infer consumption:

$$2)\quad c(it) = [(1+r)\,A(it) - A(it+1)] + y(it) - m(it) - e(it).$$

Appendix B contains a detailed description of our consumption inference procedure. In interpreting this measure of consumption, it is important to note that we do not measure some key variables, such as *inter vivos* (nonbequest) transfers between parents and children. The measure of consumption given by equation 2 should thus be interpreted as the sum of consumption and any expenditure not assigned to medical or end-of-life expenses.

TABLE 5

Married couples' characteristics, by demographic status in 1995 and 2002

	Married in 1995			
2002 household structure	**Married**	**Single male**	**Single female**	**Dead**
1995 assets	345,878	252,879	220,426	268,196
Final assets, excluding estates	341,733	232,266	194,218	212,435
Final assets, including estates	341,733	232,266	194,218	224,801
Death expenses		4,897	4,403	8,031
Death insurance payouts		369	359	731
Life insurance payouts		2,773	5,418	2,999
Observations	457	128	376	204

Notes: Assets in the final period refer to assets in 2002 if the household survived to 2002; otherwise, excluding estate value means excluding assets in the final period before death. Assets are in 1998 dollars. Observations refer to the number of observations that made the demographic transition.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

In practice, all of the variables in equations 1 and 2 are measured with error. Therefore, even if the asset accumulation equation holds, its measured counterpart will not. As a matter of notation, we will use an asterisk to denote measured values, so $A(it)^*$ is the measured value of the asset level $A(it)$.

Our goal is to identify how assets and expenditures change immediately before death. The principal econometric problem we face in making this comparison is that people who die earlier might differ systematically from people who die later along a large number of (nondeath) dimensions. For example, poor people tend to die at younger ages than rich people (Shorrocks, 1975; and Attanasio and Emmerson, 2003). This means that people who have died in our sample might have lower assets than people who lived, not because they have run down their wealth in their final years of life, but simply because they were poorer all along.[6]

To deal with this problem while keeping our methodology simple and intuitive, we use a two-step approach. In step one, we collect the people in the AHEAD who die either between 1995 and 1998 or between 1998 and 2000. We then estimate fixed effects regressions on this group, using the methodology explained next.

In step two, we construct an artificial sample of people who have similar characteristics to those in step one described previously, but did not die between 1995 and 2000. For every household that did die during the sample period, and thus belongs in the sample described in step one, we find a household in the AHEAD that has the same age, the same 1995 composition (that is, for every single female who died, we find a single female who did not die), and a similar 1995 asset level. Next, we pretend that each of these matched survivors "died" at the same date as their counterparts who actually did not survive. If an individual in step one dies in 1998, we assign the comparison individual who did not die a fictitious time of death in 1998. We then repeat the fixed effects regressions described next with the sample of comparison individuals.

We then compare the various profiles of the people in these two groups by plotting the average fixed effect of each group and using the estimated coefficients from the fixed effects regressions. Since the households in the comparison group (who did not die) had the same age, sex, and 1995 wealth level as the household heads who did die, and since they faced the same aggregate environment, the wealth trajectories of the two groups arguably should differ only because of the event of death.

Our fixed effects regressions are computed as follows. Consider, for example, the regression for the assets of a single-person household:

$$3)\quad A(it)^* = f(i) + a0 \times 1\{\text{died between time } t-1 \text{ and } t\} + a1 \times 1\{\text{died between time } t \text{ and } t+1\} + e(it),$$

where $f(i)$ is a constant that varies across households but not across time and $a0$ and $a1$ are the parameters we wish to estimate. The $1\{.\}$ function is the 0–1 indicator function that returns 1 when the argument is true, so that $1\{\text{died between time } t-1 \text{ and } t\}$ equals 1 if the individual died between time $t-1$ and t, and equals 0 otherwise. (Since everybody in these exercises has a real or fictitious death, the omitted category is whether the individual dies between time $t+1$ and time $t+2$.) Thus, if the individual dies between 1998 and 2000,

4) $1\{\text{died between time } t-1 \text{ and } t\} = 1$ in 2000 and 0 in all other periods, and
$1\{\text{died between time } t \text{ and } t+1\} = 1$ in 1998 and 0 in all other periods.

Because the fixed effect $f(i)$ varies by household, it will capture systematic differences between people who die earlier and people who die later; put differently, the terms $a0$ and $a1$ capture the way in which dying causes assets to differ from their household-level average. The fixed effect term $f(i)$ is thus potentially correlated with the death indicators, although, by assumption, $e(it)$ is not.

Even though assets are measured with error—equation 3 uses $A(it)^*$ rather than $A(it)$—as long as the measurement error has a zero mean, and is uncorrelated with the fixed effects and the death indicators, the coefficients $a0$ and $a1$ will consistently estimate the way in which assets change immediately before death.

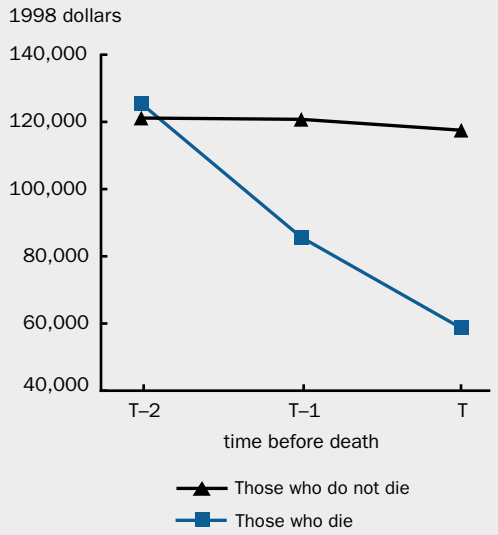
Asset rundown right before death: How important is it and why does it happen?

In this section, we document asset rundown right before death. We also document some of the possible reasons for this, such as high medical and death expenses.

Figure 1 presents average assets for our two groups of people. The first group consists of single people who died either between the 1995 and 1998 waves or between the 1998 and 2000 waves. (We consider only single-person households in order to simplify the interpretation of our results.) For the members of this group, we plot the average estate

FIGURE 1

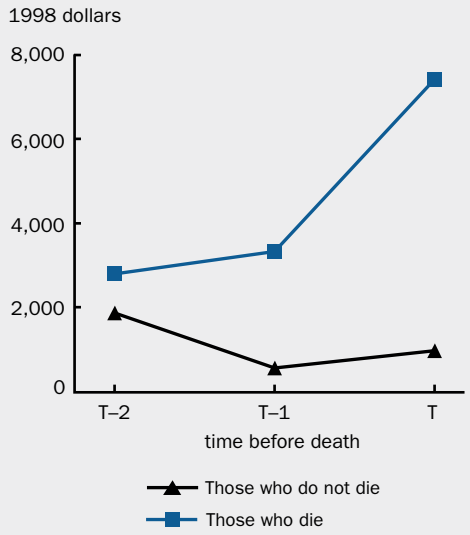
Assets, those who die vs. those who do not



Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

FIGURE 2

Annual medical costs, those who die vs. those who do not



Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old.*

and average assets one and two waves before they die. We derive these averages from the regression described previously, which contains a person-specific fixed effect and indicators for periods before death. We construct average assets, using the regression estimates and the average fixed effect for that group. On the vertical axis are asset levels. On the horizontal axis are waves before death, so that T is the time of death, $T-1$ is one wave before the time of death, and $T-2$ is two waves before death. Recall that waves in the AHEAD data are two or three years apart. Thus, the time difference between $T-1$ and T can be anywhere from zero to three years, and the time difference between $T-2$ and T can be anywhere from two to five years. The second group consists of comparable single people who did not die during the sample period.

Figure 1 shows that average assets of the soon-to-be deceased are $126,000 two periods before death, $86,000 one period before death, and $59,000 at the time of death. Thus, in the few years before death, average assets decline over 50 percent, and in the period just prior to death, they decline about 30 percent. These declines are statistically significant, with t-statistics of 2.3 for the difference in wealth between time $T-2$ and time $T-1$ and 3.8 for the difference in wealth between time $T-2$ and T. In contrast, individuals in the comparison group, who are similar in age, sex, and 1995 wealth, show relatively small asset declines. For this group, assets decline from $121,000 at time $T-2$ to $118,000 at time T.

The end-of-life wealth declines shown in figure 1 are much larger than those reported in Hurd and Smith (2001), who use 1993 AHEAD asset data and 1995 estate data for those who die between 1993 and 1995. Hurd and Smith find that assets only decline from $82,000 to $81,600. We find similar declines when comparing 1993 asset data with 1995 estates. As we noted previously, however, the 1993 asset measures are likely understated. This means that the asset declines between 1993 and 1995 are likely understated as well.

There are several reasons why assets might decline in the period preceding death. Perhaps the most obvious explanation is that medical expenses are high right before death. Figure 2 shows medical expenses before death for the same two groups of people, calculated with the same methodology used for assets. The vertical axis shows total out-of-pocket medical expenditures—the sum of insurance premiums and payments for drugs, doctor visits, and hospital and nursing home stays that were not covered by insurance. Figure 2 shows that for those who die, medical expenses rise rapidly before death. At time $T-2$, medical expenses are $2,800 per year; at time $T-1$, they rise to $3,300; and at time T (that is, the year before death), they are $7,400. The difference in medical costs between time $T-2$ and time $T-1$ is not statistically significant, but

the difference between time $T-2$ and time T is highly significant, with a t-statistic of 5.9. For the comparison group (with the same 1995 age, sex, and asset level), medical expenses remain roughly constant, and average $1,900 per year at time $T-2$, $500 at time $T-1$, and $1,000 at time T. The end-of-life increase in medical expenses shown in figure 2 is slightly higher than that found by Hoover et al. (2004), who estimate medical expenditure in the year before death using the *Medicare Current Beneficiary Survey*.

Note that in addition to affecting wealth directly, medical expenses can also affect wealth indirectly over the entire life cycle if people save to pay for future medical expenses. This preemptive effect might be especially important if people save to insure themselves against catastrophic medical expenses, as suggested by Hubbard, Skinner, and Zeldes (1994), Palumbo (1999), and De Nardi, French, and Jones (2006).

Figure 3 shows income, both for those who die and those who do not. Although we do not have good income data for time T, we do have good data for the two preceding waves ($T-1$ and $T-2$). Figure 3 shows that income for those who die and for those who do not is similar across both of these waves. Reported income falls modestly from about $11,500 to around $10,800 between time $T-2$ and time $T-1$ for those who do not die, and it remains roughly constant at approximately $10,900 for those who do die.

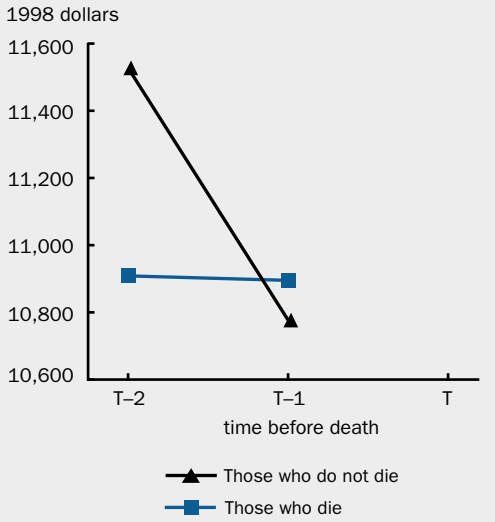
There are also several expenses associated with death, such as burial expenses. Average burial expenses for our sample are about $4,900. Part of these expenses are covered by "death insurance," small insurance policies designed to pay for burial expenses. These death insurance payments average a mere $500.

The measure of estates used in our analysis does not include the value of life insurance payments. The asset trajectories shown in figure 1 thus understate the estate actually received by the household's heirs. As it turns out, however, life insurance payouts are small, with an average value of $2,700.

Our findings suggest that much of the rundown in assets just before death can be attributed to medical costs and other end-of-life expenses. To give a sense of magnitude, recall that figure 1 shows that assets decline by $27,000 in the period just before death, and $67,000 in the two periods preceding death. Figure 2 shows that relative to the comparison group, the annual medical expenses of people who die are about $6,000 higher in the last year of life (our estimated gap between $T-1$ and T) and about $3,000 higher in the preceding two years (our estimated gap between $T-2$ and $T-1$). If death expenses (net of burial insurance) are about $4,000, it is then the case that

FIGURE 3

Annual household income, those who die vs. those who do not



Note: Data for time *T* are not yet available.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old*.

about 37 percent (10/27) of the asset decline in the last year of life and 24 percent (16/67) of the total end-of-life asset rundown are due to these expenses.

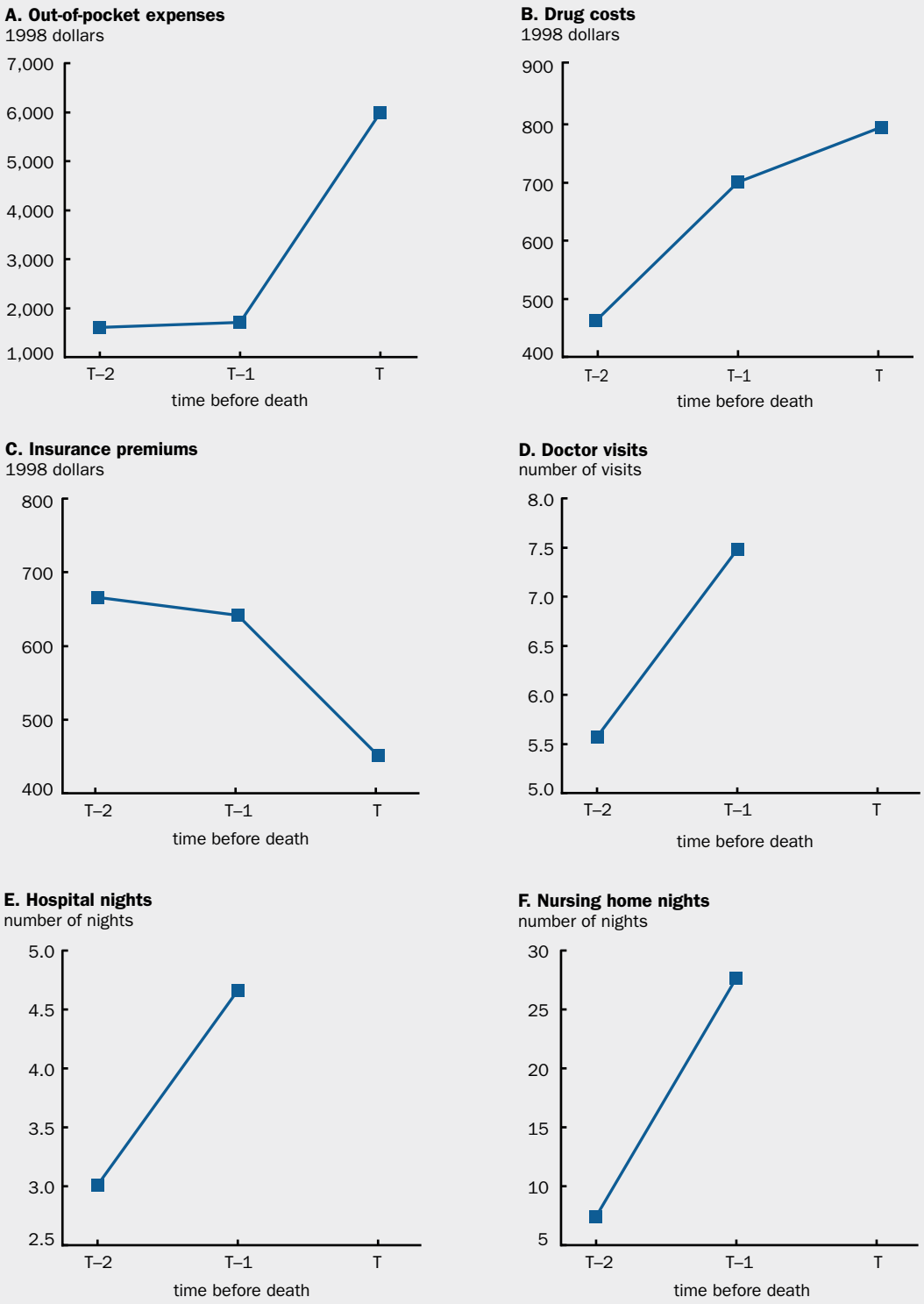
Understanding the rise in medical costs before death

Perhaps the most striking result we have shown is the sharp rise in medical expenses immediately preceding death. Although virtually all elderly individuals are covered by Medicare, there are gaping holes in Medicare coverage. Until the start of 2006, Medicare did not pay for prescription drugs. Medicare enrollees pay a 20 percent co-pay for doctor visits. Perhaps most importantly, Medicare puts caps on the number of hospital and nursing home nights that it covers per year. Medicare covers 100 percent of nursing home costs for only 20 days per year and only pays part of the cost of the next 80 days. Therefore, an individual who is in a nursing home for 365 days in a year will have to pay the full cost of the nursing home for 265 days out of the year, unless he or she has long-term care insurance or is financially destitute and eligible for Medicaid. French and Kamboj (2002) and French and Jones (2004) provide additional details about the medical expense data.

Figure 4 shows some of the subcomponents of health-related expenses, as well as the time spent using various health care services. In this figure, we

FIGURE 4

Annual health-related expenses and usage, those who die



Notes: Out-of-pocket expenses include those for doctor visits and nights spent in a hospital and nursing home. For panels D, E, and F, data for time *T* are not yet available.

Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old*.

divided the medical expenses into prescription drug costs, insurance premiums, and other out-of-pocket expenses, such as co-pays and deductibles for doctor visits and hospital and nursing home stays. Figure 4, panel A shows average out-of-pocket expenses for doctor visits and hospital and nursing home stays, which rise from $1,600 at time $T-2$ to $1,700 at time $T-1$ and $6,000 at time T. Figure 4, panel B shows drug expenses, which also rise before death, although not nearly as much as out-of-pocket expenses. Drug costs rise from about $500 at time $T-2$ to $700 at time $T-1$ and $800 at time T. Figure 4, panel C shows premiums for private health insurance (for example, "Medigap" coverage) and long-term care insurance, along with co-pays for Medicare part B insurance. Insurance premiums are approximately $700 at time $T-2$ and $600 at time $T-1$ and then fall to around $500 at time T. In short, the main expenditure change immediately before death is a sharp rise in out-of-pocket expenses associated with doctor visits and hospital and nursing home stays.

Further, figure 4 shows the rise in doctor visits, as well as nights spent in a hospital and in a nursing home. Unfortunately, we have not yet coded these variables at time T. However, panels D, E, and F of figure 4 show that all these utilization measures rise as individuals near death.

Figure 5 shows consumption before death, as well as consumption by those who do not die. As described in the empirical methodology section (and in appendix B), we infer consumption by using the asset accumulation equation. In order to infer consumption at a point in time, we need assets at two time periods. Therefore, because we only have assets over three periods, we can only infer consumption for two points in time. Figure 5 shows that consumption of those who die is much larger than for those who do not. Consumption rises from $25,000 at time $T-2$ to $33,000 at time $T-1$. For those who do not die, consumption rises from $13,000 to $15,000.

There are several potential explanations for the high level of inferred consumption right before death.

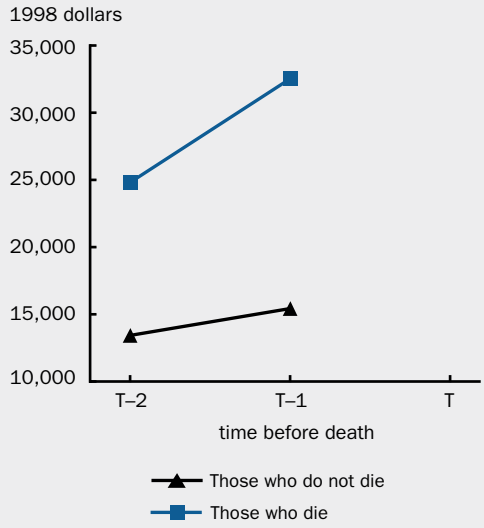
a) Individuals foresee the time of their death and run down their assets to enjoy higher consumption before death.

b) There is imprecise measurement. For example, there are many health-related expenditures (for instance, easy lifts that carry people up stairs) that are not included in our measure of medical expenditures.

c) There are large *inter vivos* transfers right before death. For example, Kopczuk (2005) compares individuals who die after short illnesses against those who die after lengthy illnesses. He finds that for those who die after lengthy illnesses, there are sizable declines in estates and sizable increases in gift giving immediately preceding their deaths. As we noted in our discussion of equation 2, these transfers would be included in our measure of consumption.

FIGURE 5

Annual consumption, those who die vs. those who do not



Note: Data for time *T* are not yet available.
Source: Authors' calculations based on data from the *Asset and Health Dynamics of the Oldest Old*.

Conclusion

A key implication of the basic life-cycle model is that assets are run down as individuals near death. Using data from the *Asset and Health Dynamics of the Oldest Old*, we present new evidence on asset rundown immediately before death.

We find that the assets of people who die decline much more quickly than those of people who survive. In single-person households, average wealth declines by 30 percent in the (roughly) one year preceding death and by 50 percent in the (roughly) three years preceding death. In contrast, the assets of comparable survivors are essentially flat over the same period.

We also find that death is often preceded by a costly illness. Out-of-pocket medical expenditures related to increased drug costs, doctor visits, and hospital and nursing home stays go up by about 200 percent in the few years before death. The increase in medical

spending before death, combined with burial expenses, can explain about 24 percent of the decline in assets of the soon-to-be deceased, and about 37 percent of the decline in assets in the last year of life. Our results thus suggest that end-of-life expenditures, medical and otherwise, provide an important reason for elderly households to retain their assets into very old age.

NOTES

[1]See table 11 of Wolff (2004). Estimates are based on the *2001 Survey of Consumer Finances* published by the Employee Benefit Research Institute.

[2]Recent surveys include Browning and Crossley (2001) and Carroll (2001). Altig et al. (2001) provide a good example of how the life-cycle model can be used for policy experiments.

[3]For example, in a recent survey of millionaires by the Northern Trust Corporation (2006), "nine out of ten … households are concerned that spiraling health care costs might affect their ability to enjoy retirement."

[4]Of 6,047 households in the AHEAD, we drop 362 households because of these criteria. We also drop 718 households whose heads were not retired, so our measure of income is clearer. We also drop 560 households who left the sample for reasons other than death. This leaves us with 4,407 households. As noted previously, we also drop wave-one data, as they are suspect. Because 527 households have all members die by 1995, we are left with 3,880 households alive in 1995 for the main analysis.

[5]We ignore taxes and assume that they are fairly minor. This is not too unreasonable, given that social security benefits are untaxed (so long as total income is below a certain threshold).

[6]A similar econometric problem is that in a cross-sectional or short panel data set, we observe individuals who were born at different times: Older people were born in earlier years than younger people. Households from older cohorts have, on average, lower real lifetime earnings than households from younger cohorts. Thus, we would expect the asset levels of households in older cohorts to be lower than those of younger cohorts in any given year. Therefore, comparing older households with younger households leads the econometrician to overstate assets when young and to understate assets when old when looking at a particular year. In other words, this will potentially lead one to infer that individuals run down their assets near the end of their lives when this is not actually the case. See Shorrocks (1975).

APPENDIX A: THE ESTATE DATA

We use estate data from 1998 and 2000 in the analysis. In 1995 and 1998, respondents (usually spouses or children of the deceased) were asked: "Altogether, what was the value of (his/her) total estate?" It is not totally clear whether respondents included the value of the house. However, in 2000, about half of respondents were asked the same question, and then in a later question were asked whether their previous answer had included the house. Most (but not all) responded that they were including the house. The remaining respondents in the year 2000 interview were asked: "Excluding (his/her) home and any life insurance, altogether, what was the value of (his/her) estate?" We make no attempt to add in the value of the house in the estate. Thus, we may be overstating asset declines at the time of death. We also assume that the value of the estate is net of all expenses related to death (the death expenses themselves are net of death insurance), but does not include life insurance.

For married couples, Hurd and Smith (2001) show that if one member of a couple dies, individuals are most likely interpreting "estate" to mean net of housing. To address this issue, we use reported assets when one member of a couple dies, as the surviving spouse usually gives both asset and estate information. We assume that when one member of a couple dies, no money goes to children or other individuals.

A second important issue is nonresponse. This is a particularly important issue for estates. About 55 percent of all estate values are actual reports. The other 45 percent are imputations made by the AHEAD. Many of the imputations use "unfolding brackets," where the respondents (usually children or spouses of the deceased) state that the estate of the deceased was worth more than some amount and less than another (for example, between $100,000 and $500,000). However, many of the estates are merely calculated using a hot-deck procedure that uses very little information. Thus, estates are likely measured with considerable error. Moreover, the 2002 data do not include any imputations at all and are thus excluded from the main analysis.

A third issue is that it is not clear whether the reported estate values are net of any estate taxes.

A final issue is that for some variables, such as death expenses, many people do not report a value. We set these values to zero, and thus, we likely understate death expenses of these households.

APPENDIX B: CONSUMPTION INFERENCE

As mentioned in the main text, although the AHEAD's measure of consumption is poor, we can use equation 2 in the text to estimate consumption. Consider an individual who died between survey years $t-1$ and t. For this individual, anywhere from zero to three years may have elapsed between the survey interview at time $t-1$ and the time of death. For the purpose of this exercise, we assume that exactly one year passes between survey year $t-1$ and the time of death, which allows us to use equation 2 exactly as it is formulated in the main text.

Imputing the consumption that occurred between one period and two periods prior to death is a bit trickier. Anywhere between two years and three years may have elapsed between these two survey waves. In this instance, we assume that exactly three years pass between interviews. This introduces compounded returns into our equation, yielding:

$$\frac{C_{it-3} = y(it-3) + (1+r)^3 A(it-3) - A(it) - m(it-3)}{(1+r)^2 + (1+r) + 1}.$$

Death expenses and insurance payouts are omitted, as the person has not yet died.

REFERENCES

Altig, David, Alan J. Auerbach, Laurence J. Kotlikoff, Kent A. Smetters, and Jan Walliser, 2001, "Simulating fundamental tax reform in the United States," *American Economic Review*, Vol. 91, No. 3, June, pp. 574–595.

Attanasio, Orazio P., and Carl Emmerson, 2003, "Mortality, health status, and wealth," *Journal of the European Economic Association*, Vol. 1, No. 4, June, pp. 821–850.

Attanasio, Orazio P., and Hilary Williamson Hoynes, 2000, "Differential mortality and wealth accumulation," *Journal of Human Resources*, Vol. 35, No. 1, Winter, pp. 1–29.

Browning, Martin, and Thomas F. Crossley, 2001, "The life-cycle model of consumption and saving," *Journal of Economic Perspectives*, Vol. 15, No. 3, Summer, pp. 3–22.

Carroll, Christopher D., 2001 "A theory of the consumption function with and without liquidity constraints," *Journal of Economic Perspectives*, Vol. 15, No. 3, Summer, pp. 23–45.

Davies, J. B., and A. F. Shorrocks, 2000, "The distribution of wealth," in *Handbook of Income Distribution*, Vol. 1, A. B. Atkinson and F. Bourguignon (eds.), Amsterdam: Elsevier Science, chapter 11.

De Nardi, Mariacristina, Eric French, and John Bailey Jones, 2006, "Differential mortality, uncertain medical expenses, and the saving of elderly singles," Federal Reserve Bank of Chicago, working paper, No. WP-2005-13, revised March 2006.

French, Eric, and John Bailey Jones, 2004, "On the distribution and dynamics of health care costs," *Journal of Applied Econometrics*, Vol. 19, No. 6, pp. 705–721.

French, Eric, and Kirti Kamboj, 2002, "Analyzing the relationship between health insurance, health costs, and health care utilization," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 26, No. 3, Third Quarter, pp. 60–72.

Hoover, Donald R., Stephen Crystal, Rizie Kumar, Usha Sambamoorthi, and Joel C. Cantor, 2004, "Medical expenditures during the last year on life: Findings from the 1992–96 *Medicare Current Beneficiary Survey*," Center for Discrete Mathematics and Theoretical Computer Science, technical report, No. 2004-27, May.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes, 1994, "The importance of precautionary motives in explaining individual and aggregate saving," *Carnegie-Rochester Series on Public Policy*, Vol. 40, June, pp. 59–125.

Hurd, Michael D., 1990, "Research on the elderly: Economic status, retirement, and consumption and saving," *Journal of Economic Literature*, Vol. 28, No. 2, June, pp. 565–637.

_____, 1989, "Mortality risk and bequests," *Econometrica*, Vol. 57, No. 4. July, pp. 779–813.

Hurd, Michael D., and James P. Smith, 2001, "Anticipated and actual bequests," in *Themes in the Economics of Aging*, David A. Wise (ed.), Chicago: University of Chicago Press, pp. 357–389.

Juster, F. Thomas, James P. Smith, and Frank Stafford, 1999, "The measurement and structure of household wealth," University of Michigan, manuscript.

Kopczuk, Wojciech, 2005, "Bequest and tax planning: Evidence from estate tax returns," Columbia University, manuscript, November.

Northern Trust Corporation, 2006, "Northern Trust study reports soaring healthcare costs may threaten retirement," press release, Chicago, April 17, available at www.northerntrust.com/pws/jsp/display2.jsp?TYPE=interior&XML=http://web-xp2a-pws/content//primary/pressrelease/1145292555550_431.xml.

Palumbo, Michael G., 1999, "Uncertain medical expenses and precautionary saving near the end of the life cycle," *Review of Economic Studies*, Vol. 66, No. 2, April, pp. 395–421.

Rohwedder, Susann, and Steven J. Haider, and Michael D. Hurd, 2004, "Increases in wealth among the elderly in the early 1990s: How much is due to survey design?," National Bureau of Economic Research, working paper, No. 10862.

Shorrocks, A. F., 1975, "The age–wealth relationship: A cross-section and cohort analysis," *Review of Economics and Statistics*, Vol. 57, No. 2, May, pp. 155–163.

White House and Council of Economic Advisers, 2006, *Economic Report of the President*, Washington, DC: U.S. Government Printing Office, February.

Wolff, Edward, 2004, "Changes in household wealth in the 1980s and 1990s in the U.S.," Levy Economics Institute, working paper, No. 407.

Yun, Heesuk, 2003, "Empirical investigation of dissaving near the end of life," Columbia University, dissertation.

The self-employment duration of younger men over the business cycle

Ellen R. Rissman

Introduction and summary

There are two competing views of self-employment that appear to be at odds with one another. On the one hand, entrepreneurs are credited with stimulating job growth and encouraging innovation. President George W. Bush clearly expressed this sentiment in a recent speech in which he stated that cutting taxes is important because "70 percent of the new jobs in America are created by small businesses."[1] Policies that encourage the creation of small businesses are based upon just such a perspective.

A somewhat contrarian viewpoint argues that although the ranks of the self-employed may include future successful entrepreneurs, the majority of workers choose self-employment because of limited opportunities in the wage sector.[2] Put another way, workers are pushed into self-employment as a stopgap measure, rather than being drawn to self-employment because of the opportunities self-employment itself creates. Wage sector prospects for these workers may be limited either because they do not have the necessary skills to be successful in the wage sector or because of weak labor demand, such as that experienced during a recession. If workers select self-employment as a second-best alternative to wage sector work, then an economy that is *growing* should actually feature *reductions* in self-employment. In this case, well-intentioned policies aimed at buttressing self-employment may be misguided. Rather than encouraging workers to start small businesses, policymakers may prefer to support investment in human capital or subsidize job search in the wage sector.

Being able to distinguish those who are entrepreneurs from those who are discouraged wage workers would be useful in understanding the role each plays in economic growth. Certainly, identifying entrepreneurs earlier would be useful for channeling resources. In a recent study, Davis et al. (2006) calculate that in 2000, about 74 percent of businesses had no employees, apparently being operated entirely by the business owner. These nonemployer businesses are quite small, accounting for about 4 percent of aggregate U.S. business revenue. Yet small nonemployer firms do sometimes grow into employers. Davis et al. (2006) find that about 16 percent of young employers came from the nonemployer ranks.

However, distinguishing one from the other is not an easy task. The U.S. Bureau of Labor Statistics collects data in its monthly household survey on "class of worker." The survey asks employed respondents, "Last month, were you employed by government, by a private company, a nonprofit organization, or were you self-employed?" In 2005, approximately 6.8 percent of all nonagricultural workers aged 16 and older were self-employed.[3]

Although self-employment appears to represent a relatively small share of total employment, the incidence of self-employment is relatively high. In fact, almost 27 percent of all younger males aged 21 or older experience at least some period of self-employment. Furthermore, of those young men who do experience self-employment, approximately 26 percent of their time is spent in self-employment.[4] These numbers suggest that self-employment is a fluid state with a great deal of turnover within the ranks of the self-employed.

Many models of self-employment and entrepreneurship focus on the decision to become self-employed. Some of these emphasize the importance of liquidity constraints in starting a business. Evans and Jovanovic (1989), Holtz-Eakin, Joulfaian, and Rosen

Ellen R. Rissman is an economist in the Economic Research Department at the Federal Reserve Bank of Chicago. *The author is indebted to Olesya Baker for her excellent research assistance, and she thanks Anna Paulson, Craig Furfine, and seminar participants at the Federal Reserve Bank of Chicago for providing many helpful comments.*

(1994), and, more recently, Cagetti and De Nardi (2003) are among these.[5] In Rissman (2003), on the other hand, I modeled the decision to become self-employed as an alternative to unemployment and analyzed the cyclical implications of the model. Aaronson, Rissman, and Sullivan (2004) provide some evidence from the *Current Population Survey* (CPS) about the cyclicality of self-employment. Other researchers have examined the intergenerational linkages in self-employment and its role in intergenerational mobility. Dunn and Holtz-Eakin (2000) find that children of self-employed parents are more likely to be self-employed themselves. Hipple (2004) and Fairlie (2005a) provide an interesting overview of self-employment. Hipple (2004) focuses on characteristics of the self-employed. Fairlie (2005a) surveys the work of other researchers who have used the *National Longitudinal Survey of Youth 1979* (NLSY79) to analyze questions related to self-employment.

Although we are far from a complete understanding of the decision to enter self-employment, there is even less of a body of research on exits from self-employment. The purpose of this article is to examine the determinants of turnover in self-employment, with particular emphasis on the role economic activity plays. It is hoped that a fuller understanding of which factors influence the duration of self-employment spells may help identify relevant models. In the first section, I introduce the U.S. Bureau of Labor Statistics' *National Longitudinal Survey of Youth 1979* data set that I use to examine self-employment and provide some summary statistics. In the next section, I examine exits from self-employment with a particular emphasis on the role of aggregate and local economic conditions.

I find that durations of self-employment tend to be short, with many first spells of self-employment terminating within a year. But the longer a worker has been self-employed, the less likely he is to leave self-employment. Spell duration does not seem to be influenced by educational levels and, after controlling for aggregate and local labor market conditions, does not seem to be influenced by race (white and nonwhite), marital status, or region. Aggregate and local labor market conditions play an important role in determining the duration of self-employment spells. A growing economy appears to encourage people who are self-employed to exit self-employment, suggesting that small business owners take the opportunities provided by growth to enter the wage sector.

Self-employment and its duration in the National Longitudinal Survey of Youth 1979

Ideally, one would want to continuously observe a large number of people over an extended period and to know their entire employment histories, including information about self-employment experiences. One would want detailed information about their businesses while self-employed, including such things as corporate structure, capitalization, industry, number of employees, time spent working, and profitability. This is in addition to other individual characteristics such as wealth, income, consumption, family background, age, race, geographical region, education, marital status, sex, health, and prior work history. Unfortunately, the ideal data set does not yet exist.[6]

Although the U.S. Bureau of Labor Statistics collects detailed information about a large number of individuals through its *Current Population Survey*, the survey has a severe limitation for studying the problem at hand: The CPS is a short panel. Data on individuals are collected for four consecutive months, followed by a hiatus of eight months, and then followed again for four consecutive months. Accordingly, only short spells of self-employment are observed directly. This data set is, thus, not particularly well-suited for studying the determinants of self-employment duration and its dynamics. A longer panel is better suited for this purpose.

The *National Longitudinal Survey of Youth 1979* follows a group of individuals over time, first surveying them in 1979. Annual interviews were conducted through 1994, after which the survey was conducted biannually. The most recent data contained in the analysis presented here are for 2002. There were 12,686 original participants, ranging in age at the time of the initial survey from 14 to 22. In 2002, these same individuals ranged in age from 37 to 45, enabling us to follow them for an important formative part of their working lives. Males accounted for 6,486 of the initial respondents. The empirical work presented here focuses exclusively on males aged 21 and older. The self-employment decisions of males are less complicated than those of females, who during their younger years must also make childbearing decisions that may complicate their employment choices. The reason for focusing on those aged 21 and older is to reduce the effects of school attendance on the employment decision.[7]

At the time of each interview, respondents were asked a number of questions about their current employment or most recent job. Class of worker data indicate whether a worker 1) works for a private company or individual for wages, salary, or commission; 2) is a government employee; 3) is self-employed in his own business, professional practice, or farm; or 4) is working without pay in a family business or farm. Similar to the CPS, respondents are also asked whether their business is incorporated or unincorporated.[8] For the

TABLE 1

Transitions between wage work and self-employment, males aged 21 and older, 1979–2002

Status at time t–1	Status at time t: Wage work	Self-employment	Total
Wage work	52,138	1,832	53,970
Percent	96.61	3.39	100.00
Self-employment	1,506	2,686	4,192
Percent	35.93	64.07	100.00
Total	53,644	4,518	58,162
Percent	92.23	7.77	100.00

Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

1994 survey, class of worker data are not comparable to earlier years.[9]

Self-employment is transitory

Transitions from self-employment to wage work are quite common. Table 1 shows transition rates between self-employment and wage work for consecutive observations from 1979 to 2002.[10] These transition rates are reported for males aged 21 and older. Over the sample period, once a person is employed in the wage sector, on average, he tends to stay in the wage sector. Only 3.4 percent of males move from wage work to self-employment from one observation to the next. Self-employment, however, is a far more transient state, with 35.9 percent of self-employed males moving from self-employment to wage work from one reporting period to the next.[11]

Self-employment is common

Table 2 illustrates how widespread self-employment is. The columns headed "overall" give the incidence of self-employment in terms of people-years. These are annual observations on an individual. Self-employment occurred in about 7.5 percent of the observations on men.[12] The "between" calculations repeat the analysis, but this time in terms of individuals rather than people-years. The incidence of self-employment is more pervasive than the simple percentage would lead us to believe. In fact, over a quarter of all men in the sample who ever worked in the wage sector (25.9 percent) also experienced self-employment at some point during their observed young work lives.[13] Finally, the last column gives the fraction of time a person has the specified employment status, conditional on ever having that status. Given that a male is ever self-employed, he spends about 26 percent of his time self-employed.[14]

Self-employment increases with age for young males

The incidence of self-employment has risen over time within the NLSY79.[15] Figure 1, panel A shows the percentage of workers aged 21 years and older at the interview date who are self-employed. For young males in the NLSY79, the fraction of the self-employed increased steadily until 1992, cresting temporarily at 9.6 percent before increasing to 10.9 percent in 2002.[16]

It is misleading to think that self-employment has become more prevalent over time. In fact, this upward trend reflects the maturing of the sample population. In 1979, only 1,079 of those males interviewed were over the age of 20. Of these young men, only 520 were working, and only 3.5 percent of these were engaged in self-employment. Ten years later, 5,196 of those sampled were aged 21 and older, and 4,719 were working, of whom 7.6 percent were self-employed.[17]

Figure 1, panel B examines this age effect in more detail. This panel shows the percentage of workers self-employed by age. Young workers just starting out are, on average, less likely to be self-employed.

TABLE 2

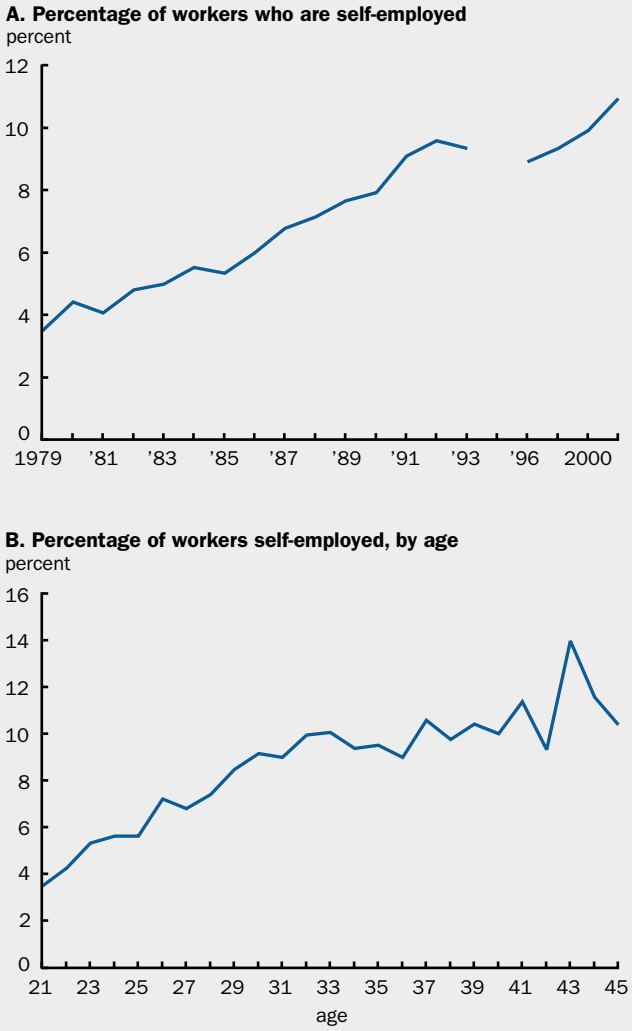
Incidence of wage work and self-employment, males aged 21 and older, 1979–2002

	Overall: Frequency	Overall: Percent	Between: Frequency	Between: Percent	Within: Percent
Wage work	60,695	92.53	6,034	99.47	92.82
Self-employment	4,903	7.47	1,572	25.91	25.87
Total	65,598	100.00	7,606	125.38	78.98

Notes: The sample consists of 6,175 individuals, with 6,034 ever having wage work and 1,572 ever being self-employed. Numbers in the columns marked "between" and "within" need not total 100 percent, since some people experience both labor market outcomes over the time they are observed.
Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

FIGURE 1

Incidence of self-employment, males aged 21 and older



Note: In panel A, the data point for 1994 is missing because of noncomparability with earlier years.

Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

The percentage self-employed by age rises steadily, reaching between 9 percent and 10 percent of workers by the time they are 30 years old. The incidence of self-employment remains fairly stable after this, with some volatility due to the dwindling number of older people in the survey: Those who are the oldest in the sample in 2002 were the first to enter the labor force in 1979.

Other factors in addition to age influence the decision to enter self-employment and the duration of self-employment once entered. Some simple statistics on a variety of factors shown in table 3 help clarify which features may be important to consider. The sample focuses on males aged 21 and older and is divided into two groups—those who are never self-employed (4,494) and those who experience at least one spell of self-employment (1,572). I used standard t-tests to test the equality of the statistic between the two groups. The self-employed were slightly more likely to be white (68.7 percent versus 63.4 percent), more likely to have been married (79.0 percent versus 75.9 percent), and are about half a year older, on average, than those who are never self-employed. They are also more likely to have lived at some point in an urban area and have higher average wages and salaries. The self-employed also tend to have higher percentile scores on the Armed Services Vocational Aptitude Battery (ASVAB) exam.[18] Despite their having slightly lower average local unemployment rates, they have at times tended to live in high unemployment rate areas, experiencing fewer weeks of unemployment, on average, than those who have never been self-employed. In addition, the self-employed are more likely than those who have never been self-employed to claim that their health has prevented them from working.

Self-employment duration

The evidence presented in the previous section suggests that many workers explore self-employment for a relatively short period, but most of these workers eventually return to wage work. Closer examination of self-employment duration may help illuminate what factors are important in explaining the longevity and entrepreneurial nature of self-employment.

In earlier work (Rissman, 2003), I explicitly modeled the decision to enter self-employment when wage work and unemployment are alternatives to self-employment. In this model, people self-select into self-employment because of limited wage sector opportunities. As unemployment insurance benefits expire, some choose self-employment as a way to make ends meet until a better wage sector offer is obtained.

Accordingly, the business cycle affects the wage offer distribution and frequency of offers. As the economy expands, discouraged wage workers who

TABLE 3

Some characteristics of wage workers and the self-employed, males aged 21 and older, 1979–2002

	Never self-employed	Ever self-employed
	(percent, unless stated otherwise)	
Number	4,494	1,572
White***	63.4	68.7
Foreign born	6.9	7.2
Average local unemployment rate***	2.93	2.88
Ever high local unemployment rate***	4.12	4.19
Ever married***	75.9	79.0
Age (years)***	27.97	28.53
Ever has more than four years of college	8.8	9.0
Weeks unemployed***	5.06	4.53
Ever urban**	93.6	95.7
Real wages and salary (dollars)**	13,659	14,998
Ever claimed health prevents working***	12.3	15.6
Mean ASVAB percentile***	40.7	42.2

**Significant at the 5 percent level.
***Significant at the 2 percent level.
Notes: ASVAB is the Armed Services Vocational Aptitude Battery. Tests of differences in the means were conducted.
Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

have entered self-employment as a stopgap will return to the wage sector. This suggests that economic conditions at the time of entry into self-employment and over the course of self-employment are important determinants of self-employment duration. Discouraged wage workers are more likely to enter self-employment during economic downturns and are more likely to exit during economic expansions.[19]

Focusing on the dichotomy between discouraged wage workers and entrepreneurs may be misleading. Instead, it may be more useful to think about the pool of self-employed in terms of their degree of attachment to self-employment. Those workers who have greater attachment to self-employment are likely to have longer durations. Those who have less attachment to self-employment are more likely to enter self-employment during economic downturns and to exit self-employment during economic expansions.

People can be attached to self-employment for a number of reasons. One possibility is that self-employment pays well for the individual. In this case, it would take a large wage offer to induce the worker to leave self-employment for work in the wage sector. Similarly, the individual may value being his own boss, so again a high wage offer is necessary to induce him to exit self-employment. Yet another possibility is that those who anticipate being self-employed for longer periods are more likely to make investments in the business. These sunk costs effectively reduce the option value of wage work and increase the "attachment" of the worker to self-employment. Whatever the source of the attachment, the business cycle will have a differential impact on entry to and exit from self-employment, depending upon the degree of attachment.

Nonparametric analysis of self-employment duration

For the analysis presented here, the focus is on exits from self-employment. A person is deemed to be self-employed if he responds that he was "self-employed in his current or most recent job." Unfortunately, we cannot be sure that the person is currently self-employed because the wording of the question leaves open the possibility that the person is no longer working. This could happen if, for example, the respondent had been self-employed but either has left the labor force or is currently unemployed.

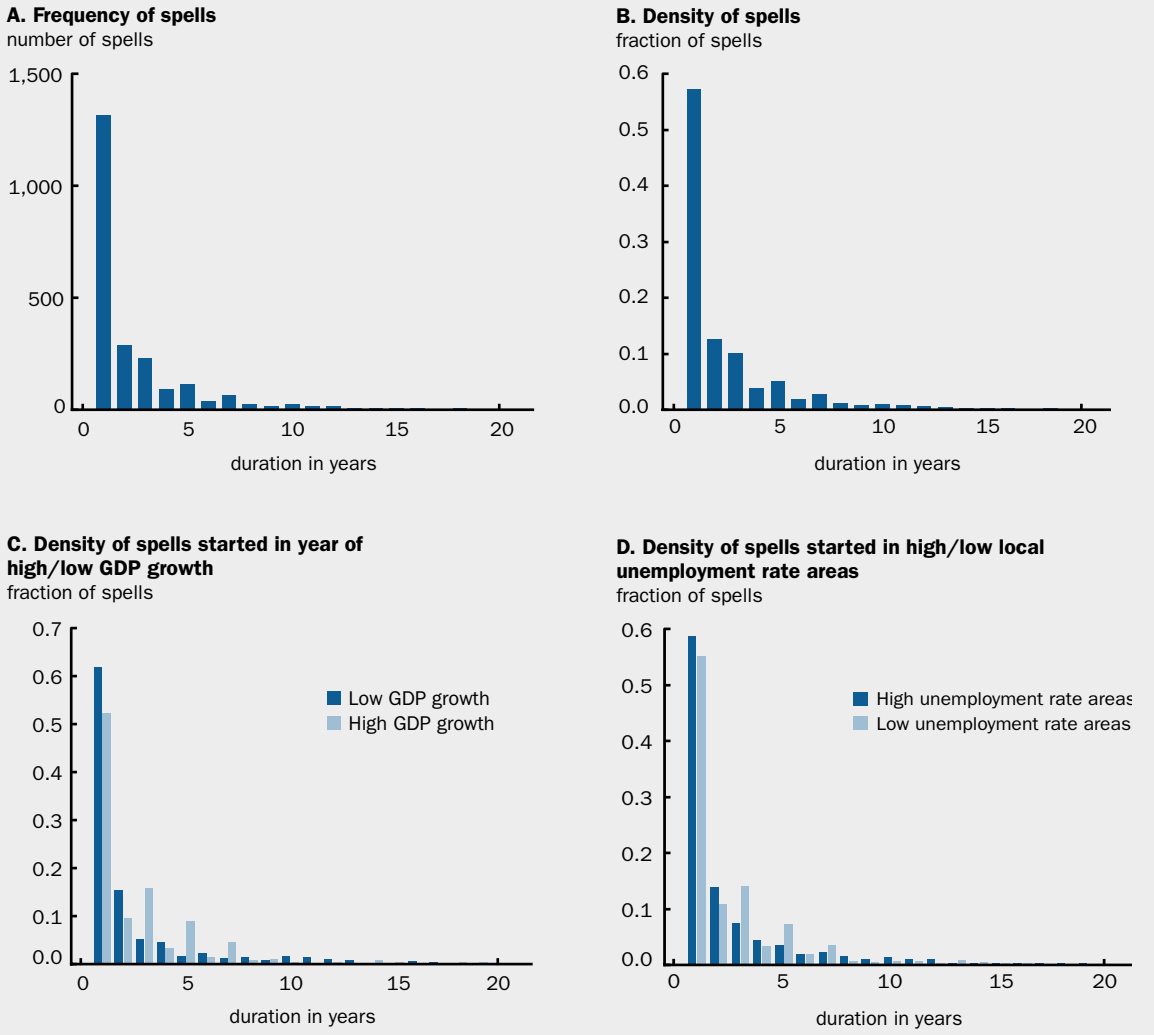
Changes over time in the response to the class of worker question can help identify shifts in self-employment status. Typically, the data set provides annual observations on the individual respondent. Ignore for a moment the complication that jobs with short tenure may not be observed. Tenure in self-employment may be overstated. To see this, imagine a person who is currently self-employed. There are various potential labor market outcomes that can subsequently occur. First, he can remain self-employed. Second, he can move to a job in the wage sector. Third, he can become unemployed. And fourth, he can drop out of the labor force altogether. The person who moves to the wage sector from self-employment will respond to the class of worker question that he is employed by a private company, the government, or employed without pay.[20]

Complications arise in interpreting how someone who becomes unemployed or drops out of the labor force answers the question. If there is no intervening job, the worker who exits self-employment to become unemployed will likely answer that he was self-employed in his current or *most recent* job. Even though he exits self-employment, by focusing on changes in class of worker, we do not observe the transition. The same issue arises for the transition to nonparticipation.

In the sample of young men aged 21 and older in the NLSY79, there were 1,479 who experienced self-employment. Of these, 938 had only one spell of self-employment and the rest went on to have

FIGURE 2

Duration of self-employment spells



Note: Gross domestic product (GDP) growth is measured as real GDP growth less average real GDP growth from 1979 to 2002.
Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

multiple spells. Figure 2, panel A shows the number of spells by duration. Approximately 1,350 spells lasted for only one year and around 300 lasted for only two years. The frequency drops off quickly. Figure 2, panel B exhibits the same data, but in terms of density. A little more than half of all spells terminate after one year. Less than 15 percent last for two years, and a little more than 10 percent last for three years.

Theory suggests that those who are less attached to self-employment generally have shorter spells of self-employment. Figure 2, panel C provides some evidence on this point. The spells have been divided into two groups: those that started during years when real gross domestic product (GDP) growth was below average and those that started when real GDP growth was above average.[21] Spells that started when the economy was doing poorly tended to have shorter durations than spells that started when the economy was doing well.

The NLSY79 provides information on the unemployment rate for the labor market of the respondent's current residence at the time of the interview. Because of privacy concerns, the public use data files do not include the exact local unemployment rate, only a range for the individual. For example, if the local unemployment rate is between 6.0 percent and

BOX 1

Estimated hazard, survivor, and integrated hazard functions

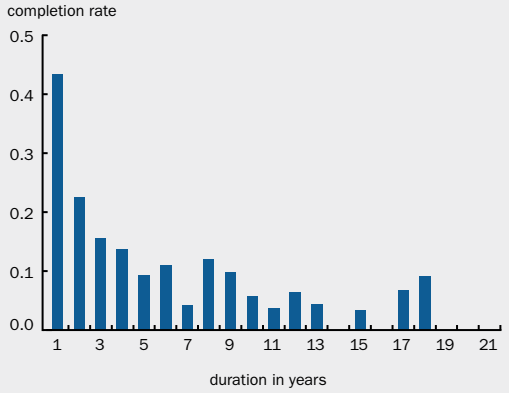
The panels in figures 2 are somewhat misleading because they do not distinguish between spells that result in exits to wage work after *x* years and those that are censored after *x* years. Censoring occurs when no further observations are available for the respondent or when observations are missing. For example, all workers who are self-employed in 2002 are by definition censored, since no subsequent observations are available. Similarly, workers who drop out of the sample are said to be censored. The appendix provides additional information on censoring.

Figure B1 exhibits the estimated hazard function for males at least 21 years of age from the NLSY79. Roughly, the hazard is the rate at which spells will be completed at duration *t*, given that they last until *t*. This hazard is calculated as the number of exits from self-employment at duration *t* divided by the number "at risk of failure" at duration *t*. The appendix provides additional information about the hazard rate. About 43 percent of observations last for only one year; that is, the respondent identifies himself as self-employed at the time of one interview and as employed in the wage sector in the subsequent interview occurring about a year later. The hazard function exhibits negative duration dependence, since the hazard rate declines with duration.

The sample survivor function for the data is shown in figure B2, and the sample integrated hazard function is shown in figure B3. Both of these functions are described more fully in the appendix. The survivor function is the probability that the duration will last at least *t* periods.[1] The integrated hazard has no precise interpretation. It is, however, particularly helpful in assessing departures from a constant hazard model. If the data exhibited no duration dependence, then the integrated hazard would be linear in duration, rising by a constant amount with each increase in duration. From figure B3, it is apparent that the integrated hazard has a more concave shape, increasing by less for each unit increase in duration. This suggests that negative duration dependence is a fixture of the data and should be accounted for in the estimation.

FIGURE B1

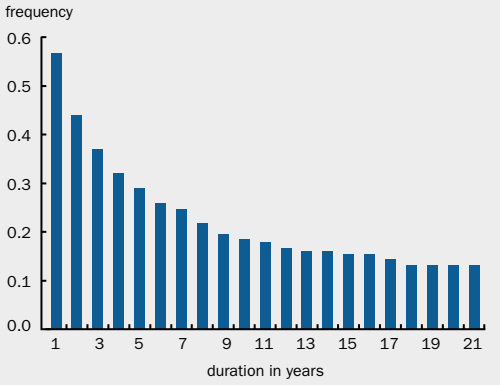
Estimated hazard function



Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

FIGURE B2

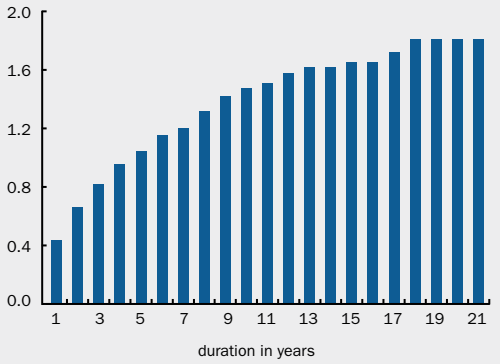
Estimated survivor function



Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

FIGURE B3

Estimated integrated hazard function



Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

[1]The sample survivor function shown in figure B2 is the Kaplan-Meier estimate. This estimator is obtained by setting the estimated conditional probability of completing a spell at duration *t* equal to the observed relative frequency of completion in the data. It expressly considers censored observations.

8.9 percent, the NLSY79 assigns the unemployment rate variable a value of 3. Clearly, the range for each value of the local unemployment rate is quite large at 3.0 percent. However, changes do occur from one survey year to the next.

Figure 2, panel D shows the density of self-employment spell duration for those spells initiated in areas with a high unemployment rate and a low unemployment rate. Durations in low unemployment rate areas tend to be longer than in high unemployment rate areas (see box 1 for a more rigorous analysis of the data).

Logit analysis

It is assumed that a person is either self-employed or is not self-employed.[22] The probability of an individual *i* exiting self-employment at time *t* is modeled as a function of individual-specific characteristics and the amount of time the person has been self-employed. Specifically,

$$\Pr(y_{it} = 1) = \frac{\exp(\beta' x_{it})}{1 + \exp(\beta' x_{it})}.$$

The dummy variable y_{it} takes on the value 0 if the self-employed worker remains self-employed and 1 if he leaves self-employment at time *t* to go to the wage sector. The probability of exiting self-employment depends upon a vector of variables x_{it}. Individual characteristics that could potentially be included as explanatory variables in this vector are age, human capital, marital status, the number of jobs held, and race. Recalling the previous discussion, other variables that reflect the performance of the aggregate and local economies may also be included as explanatory variables. β is a vector of constant parameters to be estimated that measures the effect of the independent variables x_{it} on the probability of exiting self-employment.

The data set is constructed as follows. Only people who are self-employed are included in the estimation, and the individual must have at least one period of self-employment. He exits the data set if he either transitions to wage work or is censored, meaning no current and possibly future observations are available for the individual. Many people exhibit multiple spells of self-employment. In this case, the individual enters the data set multiple times. The standard error estimates explicitly consider that observations from the same individual are not independent. Roughly 67 percent of all individuals who are ever self-employed exhibit only one spell of self-employment, and around 23 percent have two. Three spells are far less common, with only about 8 percent of all self-employed exhibiting a maximum of three spells. The maximum number of spells is five (recorded for seven workers).

Table 4 provides results for logit estimates. The dependent variable is the hazard of self-employment, taking on the value of 0 while the individual is self-employed and 1 if the individual exits self-employment. Explanatory variables include a dummy variable taking on the value 1 if the respondent scored above the 50th percentile on the 1989 equivalent of the Armed Forces Qualifications Test, age and its square, a dummy variable for race[23] (white and nonwhite), a dummy variable indicating if the respondent is married, the number of jobs the respondent had in the last calendar year, educational attainment variables described later, the index of the current spell of self-employment, and the natural logarithm of the duration of the current spell of self-employment and its square.[24] In addition, I included variables capturing local and aggregate economic conditions.

I assessed a number of different educational variables. These included years of education and stratifying the level of education at various cutoffs. Generally, after including the other variables listed previously, education variables were not significantly different from 0 in the estimation results. The results reported in table 4 include stratified educational variables. Categories include those having a high school degree or less, those having some college, those having four years of post-high-school education, and those having more than four years of post-high-school education.

Spells are numbered from 1 to 5. A spell can be thought of as a continuation of self-employment from

TABLE 4

Probability of leaving self-employment, logit results

	1	2	3
GDP growth less mean GDP growth	4.6675** (2.1347)	—	4.9373** (2.1375)
Local unemployment rate	—	0.1076*** (0.0402)	0.1120*** (0.0403)
Number of observations	4,804	4,804	4,804
Number of people	1,551	1,551	1,551

**Significant at the 5 percent level.
***Significant at the 2 percent level.
Notes: GDP is gross domestic product. The dependent variable is the hazard of self-employment. Standard errors are in parentheses. Statistically significant variables include the test score dummy (–), age (+) and its square (–), the number of jobs in the past calendar year (+), race (+), spell (–), log duration (–) and its square (+), and a constant term (–).
Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

the time of entry into self-employment to either the time of exit or the time of censoring. The reason for incorporating such a variable is that it is reasonable to assume that those workers who have multiple episodes of self-employment are different from those who have only one or two spells of self-employment. Those who enter multiple times are either very loosely attached to the wage sector or have a greater affinity for self-employment.

Theory suggests that economic conditions play a role in both entrance into self-employment and exit from self-employment. According to Rissman (2003) and Paulson and Townsend (2005), people should enter self-employment when the economy is doing relatively poorly. During contractions, opportunities in the wage sector dry up and self-employment becomes relatively more attractive. However, these workers are less attached to self-employment and are likely to be the first to exit self-employment as the economy improves. Thus, the likelihood of exiting self-employment depends upon two factors—the state of the economy at the time of entry and how the economy changes during the time the worker is self-employed.

I employ two different ways of assessing the economy in my analysis. One way is to consider aggregate economic conditions over the duration of self-employment. The other way is to focus on the role of local economic conditions and the local unemployment rate in particular. Aggregate economic conditions are measured as deviations of real annual GDP growth from its mean from 1979 to 2002. There is one observation for each year. If the economy is growing more quickly in real terms than average, we are in an expansion. Contractions are characterized by real GDP growth below average. People entering and exiting self-employment at the same time face the same aggregate economic conditions.[25]

Examination of the local unemployment rate provides a source of variation across individuals. As noted earlier, this measure of local unemployment rates is from the public use data files of the NLSY79 and is highly stratified. As a result, changes in local unemployment can occur that are not captured in the data. Only changes that result in a change in the stratification level are observable.

Column 1 of table 4 shows the logit results for regressions that include the explanatory variables discussed previously and the measure of aggregate economic conditions over the spell of self-employment.[26] Positive GDP growth relative to the average significantly increases the probability of exit from self-employment. This effect is significant at the 5 percent confidence level. Thus, an expanding economy appears to encourage self-employed workers to exit self-employment and return to wage work. Why this occurs is somewhat ambiguous. It may be that they are discouraged wage workers who find more and better opportunities in a growing economy. It might also be the case that the small businesses they own continue on in some fashion, but that their businesses are no longer their primary jobs. Or these businesses may be merged or acquired by other firms, so that, technically, the respondent is no longer self-employed.

The effect of local economic conditions on exits from self-employment is shown in column 2 of table 4. As the local unemployment rate rises, exits from self-employment increase. The reason for this may be that conditional on starting a business, a worsening local economy has a larger effect on small businesses than on larger businesses. Thus, failure rates of small businesses would be higher during times when the local economy is poor, inducing business owners to seek work elsewhere.

Both the measure of aggregate economic activity and the local unemployment rate are included in the findings in column 3 of table 4. It presents a picture similar to the results reported in columns 1 and 2. Specifically, real GDP growth relative to the average over the period of self-employment tends to encourage people to exit self-employment, as does a rising local unemployment rate.

To summarize the results of the other variables included in the logits, but not reported in table 4, nonwhite respondents are more likely than white ones to exit self-employment *ceteris paribus* and those scoring above the 50th percentile on the 1989 equivalent of the Armed Forces Qualification Test tend to have longer spells of self-employment. The number of spells significantly influences the duration of self-employment, with duration increasing with the number of spells for the individual. As the number of jobs in the past calendar year rises, the likelihood of exiting self-employment rises significantly. The natural logarithm of duration enters with a negative and significant coefficient estimate, indicating that as the length of time in self-employment rises, the likelihood of exiting declines. Exit probabilities rise with age, but at a decreasing rate. Marital status does not significantly affect the likelihood of exiting self-employment.

Heterogeneity

As noted previously, duration dependence is captured by the natural logarithm of duration and its square. The results suggest that people who have longer self-employment tenure are *less* likely to leave self-employment. However, this effect is nonlinear. After about

seven years, the likelihood of leaving self-employment starts to rise. The hazard for self-employment is said to exhibit negative duration dependence. The proper interpretation of this result hinges upon how well the model captures heterogeneity in the sample of self-employed.[27]

In general, heterogeneity leads to a downward-biased estimate of duration dependence. The intuition is straightforward. Suppose that there are two types of self-employed individuals—those that are good at self-employment and those that do better in the wage sector. Let's assume that both groups have a constant probability of exiting self-employment, regardless of the time spent self-employed to date. Furthermore, I assume that this constant exit probability is lower for those who are good at self-employment than for those who are not. If I could identify which people belonged to which group, I would estimate the logit regressions for each separately and find that duration does not significantly influence the likelihood of exiting self-employment. Unfortunately, I am unable to distinguish between the two types of individuals, so my sample contains both. The combined sample looks as if those who have been self-employed longer are less likely to exit and, therefore, exhibit negative duration dependence, despite there being no duration dependence for the individual. This merely reflects the fact that the composition shifts, with those who have a lower exit probability making up an increasing proportion of the self-employed over time.

To examine this issue of worker heterogeneity a little more carefully, I consider how multiple job holders may differ from those who have had only a single job in the past calendar year. Those who have had only one job are likely to be different from those who have had many. Those with many jobs may be using self-employment to moonlight and make a little extra money on the side. Or they may be poor-quality wage workers, being repeatedly hired and fired. In the NLSY79, 539 young men aged 21 and older reported having had a single job in the past calendar year at the time they entered self-employment, while 1,183 respondents reported having had multiple jobs.

I provide logit results for these two samples in table 5. I find that those having a single job in the previous calendar year are far more sensitive to aggregate economic conditions than the full sample. Furthermore, these workers apparently are not affected by local economic conditions, as evidenced by the insignificant coefficient on the local unemployment rate variable. In comparison, those having multiple jobs in the past calendar year are affected less by changes in aggregate economic conditions, but more by the local economy. These multiple job holders are the ones more likely to exit self-employment when the local economy worsens. One possible interpretation is that these small business owners are moonlighting and when a contraction occurs, their businesses are more likely to suffer. They then turn to other wage sector opportunities.

Conclusion

Self-employment is a fluid labor market state, exhibiting a great deal of turnover. Because it is believed that self-employment is a significant determinant of economic growth, it is important for policymakers to understand both what factors influence people to enter self-employment and what encourages them to exit self-employment. This article has focused on the latter.

Data from the NLSY79 for young males suggest that durations of self-employment tend to be short, with many first spells of self-employment terminating within a year. However, the longer a worker has been self-employed, the less likely he is to leave self-employment. Spell duration does not seem to be influenced by educational levels and, after controlling for aggregate and local labor market conditions, appears to be the same for whites and nonwhites. Marital status and region are also not significant.

TABLE 5

Probability of leaving self-employment for single and multiple job holders, logit results

	Single job at time of entry	Multiple jobs at time of entry
GDP growth less mean GDP growth	11.2989** (5.1145)	2.2091 (2.3061)
Local unemployment rate	0.0759 (0.0837)	0.1200*** (0.0439)
Number of observations	1,163	3,484
Number of people	539	1,183

**Significant at the 5 percent level.
***Significant at the 2 percent level.
Notes: GDP is gross domestic product. The dependent variable is the hazard of self-employment. Standard errors are in parentheses. For column 1, statistically significant variables include the test score dummy (–), race (+), having some college (–), and spell (–). For column 2, statistically significant variables include age (+) and its square (–), number of jobs in the past calendar year (+), spell (–), log duration (–) and its square (+), and a constant term (–).
Source: Author's calculations based on data from the *National Longitudinal Survey of Youth 1979*.

Aggregate and local labor market conditions play an important role in determining the duration of self-employment spells. A growing economy appears to encourage people who are self-employed to exit self-employment. This fact suggests that small business owners take the opportunities provided by growth to enter the wage sector. Local labor market conditions are also important. As the local economy worsens, small business owners are likely to suffer disproportionately more. These aggregate and local effects should be evaluated more comprehensively in a framework that incorporates exits from self-employment to the wage sector, unemployment, and nonparticipation.

NOTES

[1]For a transcript of President Bush's January 6, 2006, speech to the Economic Club of Chicago, see www.whitehouse.gov/news/releases/2006/01/20060106-7.html.

[2]Rissman (2003) provides a theoretical framework for this hypothesis.

[3]This number includes both males and females. In contrast, Fairlie (2005b) computes a much higher 9.8 percent self-employed in 2003, with male self-employment accounting for 12.4 percent of total male wage and salary workers. The discrepancy appears to be the result of how incorporated and unincorporated businesses are treated.

[4]Rissman (2003).

[5]In contrast to others, Hurst and Lusardi (2004) find no evidence that financial constraints affect entrepreneurial activity.

[6]Davis et al. (2006) are making progress in merging the employer and nonemployer universes in the Integrated Longitudinal Business Database (ILBD). Efforts are also under way at the U.S. Census Bureau to integrate business and household data in the Longitudinal Employer–Household Dynamics (LEHD) Program. Abowd, Haltiwanger, and Lane (2004) discuss the data set. The University of Michigan's *Panel Study of Entrepreneurial Dynamics* (PSED) follows a group of individuals who are considering starting a business and tracks them over time to determine the steps and outcomes of their decisions.

[7]Analysis of data for those aged 24 and older yields no difference in the results.

[8]Data are collected for the "CPS job." The CPS job is the respondent's current or most recent job at the interview date. If more than one job is held at that time, the CPS job/employer is the one at which the respondent works the most hours. If the respondent is not working, the CPS job is the job most recently held since the date of the last interview.

[9]The class of worker questions asked in the 1994 wave of the NLSY79 inquired about a number of jobs. The counts of self-employed workers are very different for 1994 than 1993 or in the next survey year 1996. Estimation results assume that if the class of worker was the same in 1993 as recorded in 1996, then the class of worker had the same value in 1994.

[10]Technically, the transition is from self-employment to other employment. This other employment can, in practice, involve a transition to nonpaid employment in a family business or farm. There are few instances of this recorded in the data.

[11]For females, the transition rate from wage work to self-employment was similar at 2.1 percent, and the transition rate from self-employment to wage work was 41.6 percent. The sample covers younger workers, so women are followed during their childbearing years. Family and work choices are complicated decisions with women weighing if and when to have children, and whether and how much to work after children are born. For women, self-employment may offer young mothers flexibility in work scheduling, suggesting that women would be more likely than men to engage in self-employment. Alternatively, it may be that self-employment is too time-consuming and the income too unstable relative to wage work so that young self-employed mothers would have even less flexibility. The results suggest that once women opt to work, they have about the same transition rates between wage work and self-employment as do men.

[12]For women, this figure was lower at 4.8 percent.

[13]More than one in five women aged 21 and older (22.8 percent) experience self-employment at some point in their young working lives. Similar to males, a female who has ever been self-employed spends about 19.7 percent of her time in self-employment.

[14]Numbers in the columns marked "between" and "within" in table 2 need not total 100 percent, since some people experience both labor market outcomes over the time they are observed.

[15]Fairlie (2005a) makes the same point.

[16]In figure 1, panel A, note that the data point for 1994 is missing because of noncomparability with earlier years.

[17]Unfortunately, although an effort is made to track everyone in the original sample, not everyone is interviewed each time. Sometimes respondents cannot be tracked either because they cannot be located or because they have died or are otherwise incapacitated.

[18]The ASVAB is a standardized test that was administered to all NLSY79 respondents in 1980. Subject areas include word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge. The results of these subtests were used to create a percentile ranking that is comparable to another standardized test, the Armed Forces Qualification Test (AFQT), which was reweighted in 1989. The percentiles reflected in table 3 are for the 1989 weighting.

[19]In related work, Paulson and Townsend (2005) find that in Thailand more new businesses were started around the time of the currency crisis in the late 1990s, suggesting that economic activity is an important factor in the decision to start a new business.

[20]It is difficult to know what employment without pay entails. Fortunately, such incidences are rare, with less than 0.3 percent of both men and women in this classification.

[21]The GDP variable is constructed as $y_t - Y$ where y_t is real annual GDP growth at time t and Y is average annual GDP growth from 1979 to 2002.

[22]This simple model does not differentiate between unemployment, wage sector employment, or nonparticipation. A richer model

would distinguish among these alternatives. Because of the nature of the question asked, the transition is almost certainly from self-employment to wage work.

[23]Fairlie and Meyer (2000) document the convergence of black/white self-employment rates over a long period.

[24]In 1980 and again in 1989, the ASVAB subtests were used to create a composite percentile ranking that facilitated comparisons between the ASVAB and the AFQT. The reason for the re-creation of the composite score in 1989 was to ensure comparability with the U.S. Department of Defense's revised AFQT. The AFQT variable employed in the logit analysis is a dummy variable taking on the value 1 if the respondent scored above the 50th percentile in the AFQT equivalent ASVAB score and 0 otherwise. Thus, it is a crude measure of aptitude and intelligence.

[25]A time variable indicating the year of entry and years since entry could capture the same information.

[26]Aggregate conditions are measured as real GDP growth less average GDP growth over the period from 1979 to 2002.

[27]Ham and LaLonde (1996) examine biases introduced by sample selection and initial conditions.

APPENDIX: AN INTRODUCTION TO SURVIVAL ANALYSIS

The discussion given here draws heavily from the discussion found in Kiefer (1988). Kalbfleisch and Prentice (1980) is a useful text for the interested reader. The focus of this article is on the dynamics of self-employment, in particular on the duration of self-employment. Figure A1 is helpful in understanding key concepts. This figure shows hypothetical self-employment durations for four individuals. The individuals are assumed to be surveyed at time A and then again at time B.

Short spells can be overlooked in the data. This is the case for persons 1 and 2 who are self-employed for periods of time that fall between A and B. The survey never records them as having been self-employed. Estimates of average self-employment duration are likely to be overstated as a result.

Surveys typically have a beginning and an end.[1] This can lead to censoring problems. For example, person 3 was not self-employed at time A, but was self-employed at time B. However, there are no subsequent observations for him because the survey was terminated. This individual was "right-censored," meaning that at the time of the last observation, he was self-employed and therefore still at risk of leaving self-employment. We know the duration of self-employment up to the time of his final observation but do not know what happened thereafter. Right-censoring may lead to an underestimation of average durations. The last individual, person 4, was self-employed even before he was observed at time A. This individual is said to be "left-censored." Left-censoring is not a large problem for studying self-employment in the NLSY79 because most respondents were younger than 21 at the time of the initial survey, and we are interested in their self-employment experiences only after they become attached to the labor force, which is typically at an age 21 and older.

To deal with the unique problems that arise from duration data, statisticians have modeled the hazard function. To start, the probability distribution of durations is given by the distribution function:

$$F(t) = \Pr(T < t).$$

This is the probability that a spell of self-employment T will be less than some value t. The corresponding density function is given by:

$$f(t) = dF(t)/dt.$$

Other functions are also now easily defined. The survivor function $S(t)$ is given by:

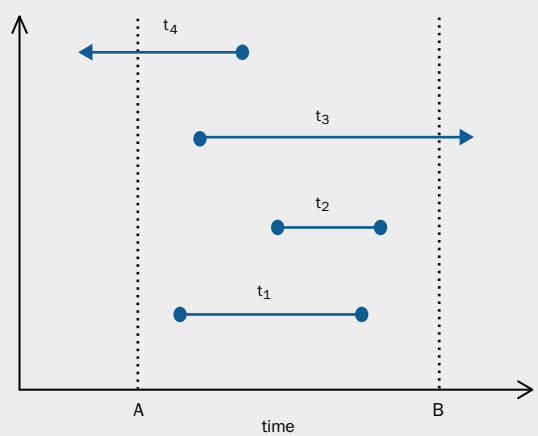
$$S(t) = 1 - F(t) = \Pr(T \geq t).$$

This is the probability that the duration of self-employment is greater than or equal to t and gives the upper tail of the distribution. The hazard function can now be defined. It is the rate at which spells are completed at duration t, given that they last until t. The hazard function is given by:

$$\lambda(t) = f(t)/S(t).$$

FIGURE A1

Hypothetical self-employment duration data



Source: Kiefer (1988).

Lastly, the integrated hazard is a useful function for empirical testing. It is defined as:

$$\Lambda(t) = \int_0^t \lambda(u)du.$$

Its relation to the survivor function is given by:

$$S(t) = \exp[-\Lambda(t)].$$

Note that knowing the hazard function uniquely characterizes the other functions.

A number of different distributions have been used successfully in modeling duration data. An exponential distribution is perhaps the most widespread because it is easy to work with. This distribution has the interesting property that its hazard function is constant: The rate at which people exit is not dependent on how much time they have already spent in the state. Because of this "memoryless" property, the exponential distribution is not a good description of data that contains both very short and very long durations. Other distributions that have been successfully employed include the Weibull, log-logistic, Gompertz, lognormal, and generalized gamma. The Weibull distribution permits the exponential distribution as a special case. These other distributions have the benefit of capturing a time-varying hazard. For example, if the statistician believes that workers are quick to leave self-employment at short durations but that they are less likely to leave once they have been around for awhile, then a log-logistic or lognormal distribution may be a better description than the exponential.

[1]In the NLSY79, the respondents were interviewed in 1979 and followed periodically through 2002. Although the survey is ongoing, this is the most recent observation.

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, Second Quarter, pp. 2–20.

Abowd, John M., John C. Haltiwanger, and Julia I. Lane, 2004, "Integrated longitudinal employee–employer data for the United States," U.S. Census Bureau, Longitudinal Employer–Household Dynamics Program, technical paper, No. TP-2004-02, May, available at http://lehd.dsd.census.gov/led/library/techpapers/tp-2004-02.pdf.

Cagetti, Marco, and Mariacristina De Nardi, 2003, "Entrepreneurship, frictions, and wealth," Federal Reserve Bank of Minneapolis, staff report, No. 322, September.

Davis, Steven J., John Haltiwanger, Ron S. Jarmin, C. J. Krizan, Javier Miranda, Alfred Nucci, and Kristin Sandusky, 2006, "Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes," U.S. Census Bureau, Center for Economic Studies, discussion paper, No. CES 06-04, February, available at http://webserver01.ces.census.gov/index.php/ces/1.00/cespapers?down_key=101744.

Dunn, Thomas, and Douglas Holtz-Eakin, 2000, "Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links," *Journal of Labor Economics*, Vol. 18, No. 2, April, pp. 282–305.

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.

Fairlie, Robert W., 2005a, "Self-employment, entrepreneurship, and the NLSY79," *Monthly Labor Review,* Vol. 128, No. 2, February, pp. 40–47.

_____, 2005b, "Self-employed business owners—Nonagricultural industries: Calculations from the *Current Population Survey*, Outgoing Rotation Group files (1979–2003)," University of California, Santa Cruz, table, available at http://econ.ucsc.edu/faculty/fairlie/serates/sesex7903data.pdf.

Fairlie, Robert W., and Bruce D. Meyer, 2000, "Trends in self-employment among white and black men during the twentieth century," *Journal of Human Resources*, Vol. 35, No. 4, Fall, pp. 643–669.

Ham, John C., and Robert J. LaLonde, 1996, "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training," *Econometrica*, Vol. 64, No. 1, January, pp. 175–205.

Hipple, Steven, 2004, "Self-employment in the United States: An update," *Monthly Labor Review*, Vol. 127, No. 7, July, pp. 13–23.

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.

Kalbfleisch, J. D., and R. L. Prentice, 1980, *The Statistical Analysis of Failure Time Data*, New York: John Wiley and Sons.

Kiefer, Nicholas M., 1988, "Economic duration data and hazard functions," *Journal of Economic Literature,* Vol. 26, No. 2, June, pp. 646–679.

Paulson, Anna L., and Robert M. Townsend, 2005, "Financial constraints and entrepreneurship: Evidence from the Thai financial crisis," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 29, No. 3, Third Quarter, pp. 34–48.

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

The Ninth International Banking Conference

October 5–6, 2006
International Financial Instability: Cross-Border Banking and National Regulation

The Federal Reserve Bank of Chicago will host its ninth annual International Banking Conference in Chicago on October 5 and 6, 2006, in conjunction with the International Association of Deposit Insurers. This year's theme will be "International Financial Instability: Cross-Border Banking and National Regulation."

Financial stability depends, in large part, on appropriate and prudential supervision and regulation. While banking has increasingly become international and operates across borders, prudential regulation has largely remained national. Its jurisdiction has been primarily limited to financial institutions chartered within national borders. This conference will identify the implications of this mismatch for international financial stability and examine recommendations for mitigating any adverse effects without reducing gains from cross-border banking activities.

The two-day conference will feature keynote presentations by **Sheila Bair,** Chairman of the U.S. Federal Deposit Insurance Corporation; **Stefan Ingves**, Chairman of the Executive Board, Bank of Sweden/Riksbank; **Raghuram G. Rajan**, Economic Counselor and Director of Research, International Monetary Fund; and **J. P. Sabourin**, Chief Executive Officer, Malaysia Deposit Insurance Corporation, and Chair of the Executive Council and President, International Association of Deposit Insurers. There will also be sessions on the cross-border transmission of financial instability, home–host country supervisory conflicts, the role and effectiveness of the government safety net across borders, resolution problems associated with cross-border failures, and public policy alternatives to prevent cross-border crises.

As with past conferences, the theme was selected because it is an important and current concern in a broad array of countries. The targeted audience is industry practitioners, financial consultants, financial regulators, and researchers interested in public policy toward global financial markets. The conference will be held in the Federal Reserve Bank of Chicago's Conference Center, 230 South LaSalle Street, Chicago, IL 60604.

Additional information, including the conference agenda and conference and hotel registration information can be found at:

www.chicagofed.org/InternationalBanking Conference.

For additional questions, please contact:

Ms. Pam Suarez
Phone: (312) 322-8103
Email: psuarez@frbchi.org

or

Ms. Regina Langston
Phone: (312) 322-5641
Email: rlangston@frbchi.org

The great turn-of-the-century housing boom

Jonas D. M. Fisher and Saad Quayyum

Introduction and summary

In the last ten years, residential investment as a share of gross domestic product (GDP) has reached levels not seen since the 1950s. At the same time, the homeownership rate has climbed to levels never before achieved. This article discusses the forces underlying these developments and argues that they are connected.

Figure 1 shows the ratio of nominal residential investment to GDP from 1947 to 2005, with shaded regions indicating years in which the economy was in recession. The spending share of residential investment is clearly highly cyclical, but in the last ten years it has seemed relatively immune to macroeconomic disturbances. Indeed, from a near historic low below 3.5 percent in 1991, residential investment spending has grown rapidly, passing 6 percent of nominal GDP in 2005.

Figure 2 shows the history of the homeownership rate from 1890 to 2004. The homeownership rate equals the number of owner-occupied housing units divided by the number of occupied housing units. Between 1890 and 1940, the homeownership rate varied between 43 percent and 48 percent. After World War II, the homeownership rate rose rapidly, and by the mid-1960s it had surpassed 64 percent. Upward progress in homeownership stalled in the 1970s and even fell in the 1980s. It began growing again in the mid-1990s and by 2005 had reached a new high of 69 percent.

Understanding why residential investment and homeownership have reached such unusually high levels is useful from a policymaking standpoint. For example, monetary policy has been traditionally viewed as having a strong influence over new home construction. Have the high levels of residential investment been driven by unusually loose monetary policy? Another concern of policymakers is that the unusually high level of spending on new housing reflects speculation and is not driven by underlying fundamentals. The increase of rates of homeownership has long been an announced goal of policymakers. Indeed, both Presidents Clinton and George W. Bush have touted the rising levels of homeownership as accomplishments of their administrations. So, understanding why homeownership rates have risen should help in the development of policies directed at establishing socially and economically desirable levels of homeownership.

Much has been said in the press about high levels of house prices and the possibility of a house price bubble. Figure 3 displays the median sales price of existing single family homes, converted into real terms by dividing total sales price by the Consumer Price Index (CPI) for all urban consumers. This figure shows that indeed the real price of single-family homes, after being roughly stable from the early 1980s to the mid-1990s has grown considerably since. This article does not address house prices directly. Rather, it seeks to understand recent developments by focusing on quantities. To the extent that the quantities can be understood by considering the underlying economic fundamentals, such as productivity growth and the evolution of the mortgage market, then the recent growth in house prices is probably not due to excessive speculation in the housing market, such as occurs in a bubble. We argue that our findings point toward the high prices being driven by fundamentals.

The article begins by describing the evolution of key variables that should influence residential investment. While informative, this discussion has the drawback that it is difficult to distinguish the truly exogenous factors driving the spending. For example, showing that real interest rates have been relatively

Jonas D. M. Fisher is an economic advisor and team leader of the macroeconomics group and Saad Quayyum is an associate economist in the Economic Research Department at the Federal Reserve Bank of Chicago.

low as residential investment has surged beyond its trend level does not establish that unusually loose monetary policy is to blame. Consequently, the next phase of the analysis involves an econometric study of the effects of identified exogenous shocks to the economy. This study focuses on the roles of technological change and monetary policy.

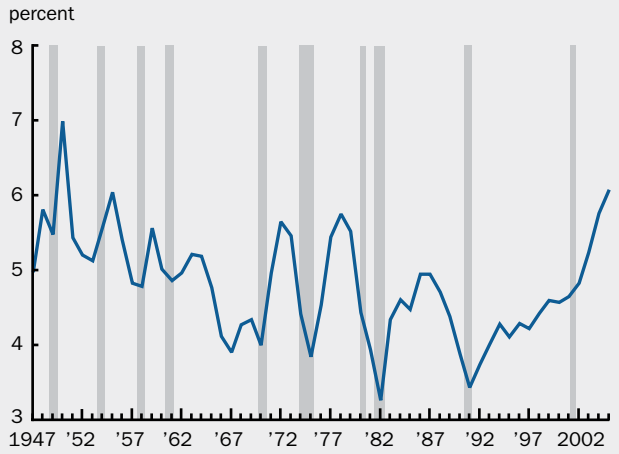
We then turn to the homeownership issue. We begin by describing how homeownership rates have changed across various racial, generational, educational, and income categories. Then we address the question of how much of the increase in homeownership can be explained by changes in the distribution of the population across these categories. For example, older people have higher homeownership rates than younger people, so, all else being equal, an aging population tends to increase the homeownership rate. By accounting for all easily measurable factors, this analysis provides a bound on what needs to be explained by other, more difficult to measure factors, such as the increased use of new mortgage products.

The final part of the article connects the overall increase in residential investment with the increase in homeownership. This analysis focuses on the impact of the rapid growth of the subprime mortgage market.

Our main findings are as follows. First, it appears that the housing boom has not been driven by unusually loose monetary policy. This is not to say that monetary policy has not been unusually loose, but that to the extent it has been loose, this is not what has been driving spending on housing. Second, the current levels of spending on housing are largely explained by the wealth created by dramatic technological progress over the previous decade. Third, changes in the demographic, income, educational, and regional structure of the population account for only one-half of the increase in homeownership. That is, without any other developments, the homeownership rate is likely to have gone up anyway, but not by nearly as much as it has done. The last finding is that substitution away from rental housing made possible by technology-driven developments in the mortgage market, such as subprime lending, could account for a significant fraction of the increase in residential investment and homeownership. The current spending boom thus may be a temporary transition toward an era with higher homeownership rates and a share of spending on housing that is nearer historical norms.

FIGURE 1

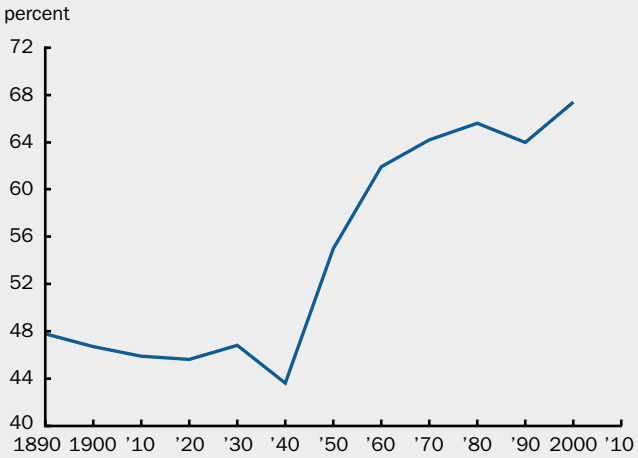
Nominal GDP share of residential investment, 1947–2005



Note: Shaded areas indicate recessions as determined by the National Bureau of Economic Research.
Source: U.S. Bureau of Economic Analysis, national income and product accounts data, from Haver Analytics.

FIGURE 2

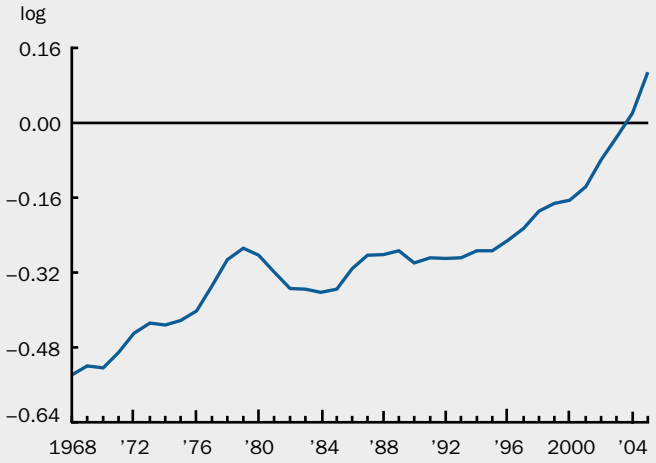
U.S. homeownership rate, 1890–2004



Sources: U.S. Census Bureau, *Decennial Census*, and the U.S. Bureau of Labor Statistics, *Current Population Survey/Housing Vacancy Survey*, Series H-111, reports, Washington, DC, from Haver Analytics.

FIGURE 3

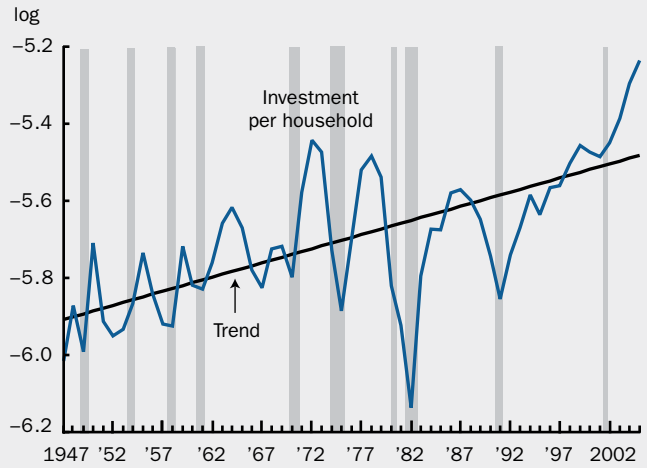
Log real median sales price of existing single family homes



Sources: Authors' calculations based on data from the National Association of Realtors, Economics and Research Division, "Home sales," report, and the U.S. Bureau of Labor Statistics, Consumer Price Index, report, from Haver Analytics.

FIGURE 4

Log residential investment per household



Note: Shaded areas indicate recessions as determined by the National Bureau of Economic Research.

Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis, national income and product accounts data, and the U.S. Bureau of Labor Statistics, *Current Population Survey*, report, from Haver Analytics.

Factors affecting residential investment

Figure 4 is helpful as a starting point for gauging whether residential investment is currently at unusual levels. The figure displays the log of real residential investment per household from 1947 to 2005 along with a trend line. While certainly subject to large cyclical variations, residential investment seems to follow a linear trend quite closely. So while investment grew rapidly after 1991 until the latter part of the decade, this was largely a return to trend. Only after the 2001 recession has spending grown substantially above trend. By 2005, it was roughly as far above trend as occurred during the boom part of the boom–bust cycle of the 1970s and the early 1980s. The dramatic swings in residential investment in the 1970s and early 1980s contributed significantly to three recessions. From this perspective, the current levels of residential investment may seem alarming. In the remainder of this section, we consider some of the factors that may underlie the current high levels of residential investment.

Household formation

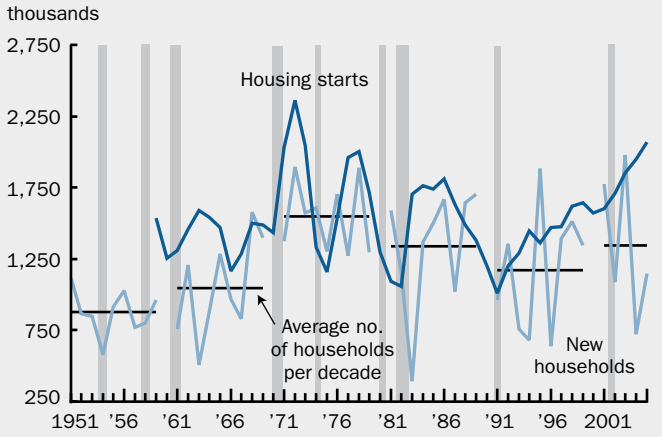
Household formation, to the extent that it is governed primarily by long-term social and demographic developments, is the most basic determinant of home building and residential investment.[1] Indeed, if vacancy rates were constant and houses were never torn down to be replaced by new residential structures, new home building would be exactly proportional to new household formation.[2] Figure 5 shows the evolution of new households and housing starts since 1960. The light blue lines indicate the number of new households in a given year, the black lines indicate the average number of new households per year for each decade, and the dark blue line indicates the level of housing starts. In the 1970s, 1980s, and 1990s, there seems to be a close association between home building and household formation. The increases in household formation and home building over the 1970s are an example of the strong influence of demographic factors—this is when the baby boom generation moved out on its own. The close association of home building with household formation is less true of the 1960s, when home building is near its 1980 levels, but new household formation is much lower. Since the 2001 recession, housing starts have also risen to levels that do not seem closely tied to new household formation.

Migration

Another factor determining new home building is migration. With migration, the number of households can stay fixed while there is still a demand for new homes. In the region where households migrate from, vacancy rates go up, while in the region where households migrate to, houses need to be built. To assess the possible impact of migration, consider figure 6. Panel A shows the shares of national population increases attributable to the four census regions from 1982 to 2005. Panel B displays the corresponding shares of all housing starts. In each case, a rising or falling share indicates that either population or housing starts are increasing faster in the region than for the nation as a whole. This figure shows that the relative shares of housing starts generally correspond to the relative population shares. Some of the trends seem to correspond as well. For example, the dip in the share of housing starts for the South in the 1980s is associated with a downward trend in the population share as well. Of particular interest is the uptick in the population share of the South from 2000 to 2005. Consistent with a role for migration in the current housing boom, the share of housing starts also picked up, although with a delay.

FIGURE 5

New households and housing starts



Note: Shaded areas indicate recessions as determined by the National Bureau of Economic Research.

Sources: U.S. Census Bureau, *Decennial Census*, report, and the U.S. Bureau of Labor Statistics, *Current Population Survey*, report, from Haver Analytics.

Interest rates

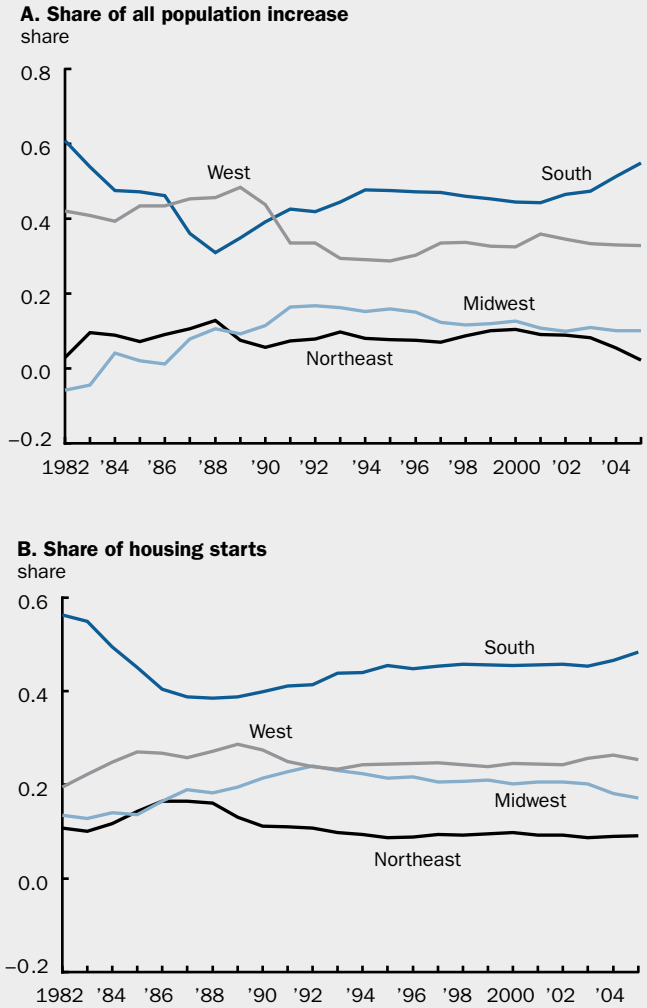
The discussion until now has focused on relatively long-term developments. As is evident from figures 1 and 2, home building historically has been highly cyclical. The conventional wisdom on why this is so is that the demand for housing is sensitive to movements in mortgage rates. If mortgage rates are unusually low, then this could fuel unusually high rates of residential investment. Figure 7 displays a measure of the nominal effective mortgage rate along with an estimate of the corresponding real rate. The term "effective" means that the mortgage rate incorporates the various points, fees, and other closing costs associated with a mortgage. These have generally been declining since the early 1980s. The real rate is equal to the nominal rate less an estimate of the expected rate of inflation. The inflation rate used for this figure is equal to the inflation rate in the national income and product accounts (NIPA) personal consumption expenditure deflator over the previous year. This probably overstates actual changes in the expectations relevant for determining housing demand.[3] Figure 7 shows that both real and nominal mortgage rates were low in the 1990s compared with the 1980s. This presumably contributed to the return to trend of residential investment over this period. Interestingly, the period when home building accelerated beyond its trend level was also a time when nominal and real rates were falling even further. While over the last two years nominal mortgage rates have started creeping up, real rates have continued to fall because inflation expectations have been rising. These considerations suggest that sustained low interest rates, possibly driven by unusually loose monetary policy, could be fueling the housing boom.

Wealth

The final factor affecting home building that we consider is household wealth. All else being equal, the richer households are, the more housing they demand. The latter half of the 1990s was a period of rapid wealth accumulation. For example, according to the *Survey of Consumer Finances*, average family net worth increased by 72 percent between 1995 and 2001. These increases in wealth were primarily due to the large increases in stock values over this period. As a consequence, housing and other nonfinancial assets' average share of total assets, fell from 63.3 percent in 1995 to 58 percent in 2001. If the share in 1995 was "normal" or close to the "desired" household allocation of nonfinancial assets in households' portfolios, then it is to be expected that the share

FIGURE 6

New households and housing starts, by region



Sources: U.S. Census Bureau, *Decennial Census*, report, and the U.S. Bureau of Labor Statistics, *Current Population Survey*, report, from Haver Analytics.

would eventually start rising again. Indeed, by 2004, nonfinancial assets' share of total assets had risen to 64.3 percent, near its 1995 level. While the share of housing in total assets, in the form of primary residences, rose faster over this period, the behavior of nonfinancial assets as a whole suggests that much of the acceleration in residential investment after the 2001 recession might be due to households rebalancing their portfolios. That is, it may be a natural consequence of the stock market boom of the 1990s.[4]

The macroeconomic shocks driving residential investment

The foregoing discussion suggests various factors that may be influencing the high levels of residential investment, but the underlying causes remain unclear. Determining the causes of macroeconomic fluctuations is notoriously difficult because essentially all the variables of interest are endogenous—no single variable moves independently and drives movements in other variables. The traditional approach to assessing the causes of fluctuations is to posit that the economy is subject to exogenous random disturbances, which are called shocks. Macroeconomists have formulated methods for identifying three kinds of shocks—two kinds of shocks to technological possibilities and one kind of monetary policy shock.[5] The procedure for identifying these shocks involves specifying a statistical model of the variables of interest and making a series of identification assumptions that make it possible to extract the exogenous shocks from the statistical model. Once the shocks have been identified, it is possible to determine how much of the growth in residential investment from 1995 to 2005 can be attributed to these shocks using what is called a *historical decomposition*. The strategy for identifying the technology shocks builds on Fisher (2006) and the monetary policy shock identification builds on Christiano, Eichenbaum, and Evans (2005).

Identifying the exogenous shocks

We begin by supposing that the economy evolves according to the following vector autoregression (VAR):

$$1) \qquad Aq_t = \Gamma(L)q_{t-1} + \varepsilon_t,$$

where q_t is a vector of variables of interest to be specified in a moment, ε_t is a vector of fundamental shocks, and A is a matrix of coefficients conformable with q_t. The term $\Gamma(L)$ is called a *lag polynomial*. It specifies how many lags of q_t appear in equation 1 and is defined as

$$\Gamma(L) = \Gamma_0 + \Gamma_1 L + \Gamma_1 L^2 + \cdots + \Gamma_M L^M,$$

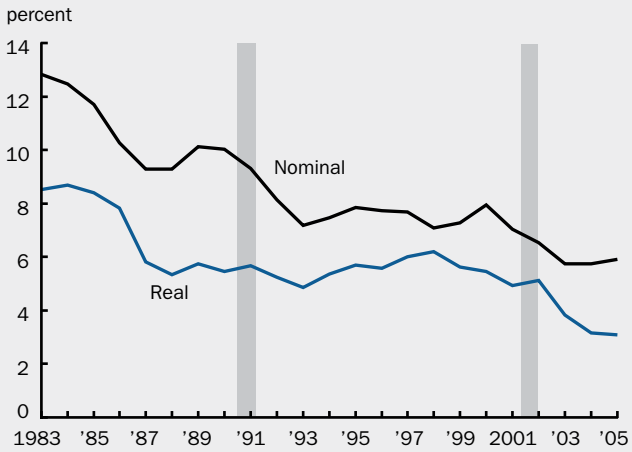
where the Γ_i, $i = 1,2,\ldots, M$ are matrices of the same dimensions as A, and L is a *lag operator*. Lag operators have the property that $L^n x_t = x_{t-n}$, for any variable x_t. Equation 1 specifies the exogenous shocks that drive fluctuations in the variables in q_t and how these variables interact contemporaneously and dynamically. We suppose that

$$q_t = [\Delta p_t, \Delta y_t, \pi_t, R_t, h_t]',$$

where p_t is the log real price of capital equipment, y_t is the log of real per capita GDP, π_t is inflation, R_t is the federal funds rate, h_t is the log ratio of nominal residential investment to nominal GDP, and Δx_t for some variable x_t, such as p_t or y_t, is shorthand for writing $x_t - x_{t-1}$.[6] We use a series of instrumental variables regressions to estimate A, $\Gamma(L)$, and the fundamental shocks, and simulate equation 1 with the identified shocks to determine how these shocks have influenced residential investment. The equations were estimated using data from the third quarter of 1982 to the fourth quarter of 2005 with four lags.

FIGURE 7

Nominal and real effective mortgage rates



Note: Shaded areas indicate recessions as determined by the National Bureau of Economic Research.
Sources: Federal Housing Finance Board data and U.S. Bureau of Economic Analysis, national income and product accounts data, from Haver Analytics.

The assumptions used to identify the two technology shocks are derived from growth theory. In particular, we assume that all growth and wealth accumulation derive from two kinds of exogenous random technological change. One source of technological change increases the quality and efficiency of capital equipment, that is, it is embodied in new capital equipment. Shocks to the rate of accumulation of this kind of technology are called investment-specific technology shocks. For simplicity, we will call them "I-shocks." To identify these shocks, we assume that they are the only disturbances which have a long-run impact on the real price of capital equipment. The other source of technological change improves the economy's ability to produce all kinds of goods. Shocks to the rate of accumulation of this kind of technology are called neutral technology shocks, "N-shocks" for short. To identify these shocks, we assume that, along with I-shocks, they are the only shocks that affect output in the long run.

Our strategy for identifying the monetary policy shock, an "R-shock," has been widely applied in the empirical macroeconomic literature. The basis of this strategy is the assumption that the Federal Reserve follows a simple rule for setting the federal funds rate. This rule is subject to shocks, which can be thought of as randomness in the deliberations of the Federal Open Market Committee or factors uncorrelated with the variables in the rule that affect decisions on how to set the federal funds rate. Examples of the latter might be a hurricane or a terrorist attack. We assume that the Fed looks at contemporaneous values of the equipment price, output and inflation, as well as lags of q_t when setting the federal funds rate.

The three sets of identifying assumptions can be translated into assumptions about the structure of equation 1. First, consider the assumption used to identify the I-shock. To apply this assumption we use the first regression of equation 1, which can be written

$$2)\quad \Delta p_t = \Gamma_{pp}(L)\Delta p_{t-1} + \Gamma_{py}(L)\Delta y_t + \Gamma_{p\pi}(L)\pi_t + \Gamma_{pR}(L)R_t + \Gamma_{ph}(L)h_t + \varepsilon_{it},$$

where the $\Gamma_{xy}(L)$ values here and below are the relevant lag polynomials plucked from $\Gamma(L)$. From this equation we can see that the contemporaneous effects of all non-ε_{it} shocks influence Δp_t through Δy_t, π_t, R_t, and h_t. Our assumption for identifying the I-shocks implies that the long-run multipliers from these variables to p_t are zero. The long-run multiplier associated with a variable in equation 2 is given by the sum of the lag coefficients for that variable. This sum can be calculated by evaluating the lag polynomial associated with that variable at $L = 1$. So the identifying assumption for the I-shock is equivalent to assuming

$$\Gamma_{py}(1) = \Gamma_{p\pi}(1) = \Gamma_{pR}(1) = \Gamma_{ph}(1) = 0.$$

This means that each $\Gamma_{pj}(L), j = y, \pi, R, h$ can be written, $\Gamma_{pj}(L) = \tilde{\Gamma}_{pj}(L)(1-L)$. So, by imposing the I-shock identifying assumption, equation 2 becomes

$$3)\quad \Delta p_t = \Gamma_{pp}(L)\Delta p_{t-1} + \tilde{\Gamma}_{py}(L)\Delta^2 y_t + \tilde{\Gamma}_{p\pi}(L)\Delta\pi_t + \tilde{\Gamma}_{pR}(L)\Delta R_t + \tilde{\Gamma}_{ph}(L)\Delta h_t + \varepsilon_{it}.$$

In general, disturbances to Δp_t affect the contemporaneous values of Δy_t, π_t, R_t, and h_t. That is, ε_{vt} is correlated with the other right-hand side variables in equation 3. Consequently, equation 3 cannot be estimated by ordinary least squares. However, assuming ε_{vt} is exogenous means this shock is independent of all variables dated $t-1$ and earlier. So equation 3 is estimated by instrumental variables, using M lags of q_t as instruments. The coefficients of the first regression of equation 1 are found by unraveling the resulting regression coefficients from the instrumental variables estimation. The residuals from equation 3 are our estimates of $\varepsilon_{vt}, \hat{\varepsilon}_{it}$.

By a similar argument used with the first regression, the assumptions used to identify the N-shocks imply the long-run multipliers from π_t, R_t, and h_t are zero in the second regression of equation 1. It follows that this second regression can be written

$$4)\quad \Delta y_t = \Gamma_{yp}(L)\Delta p_t + \Gamma_{yy}(L)\Delta y_{t-1} + \tilde{\Gamma}_{y\pi}(L)\Delta\pi_t + \tilde{\Gamma}_{yR}(L)\Delta R_t + \tilde{\Gamma}_{yh}(L)\Delta h_t + \varepsilon_{nt},$$

where the $\tilde{\Gamma}_{yj}(L), j = h, q$ are defined in the same way as the similar terms in equation 3. As before, this equation is estimated by instrumental variables, and the resulting coefficient estimates are used to assign values to the second row of coefficients in equation 1. The instruments are $\hat{\varepsilon}_{it}$ and M lags of q_t. The residuals from equation 4, $\hat{\varepsilon}_{nt}$, are our estimates of ε_{nt}. Including $\hat{\varepsilon}_{it}$ as an instrument ensures $\hat{\varepsilon}_{nt}$ is uncorrelated with the I-shock within the sample period.

We estimate the remaining regressions of equation 1 sequentially by instrumental variables, using the residuals from the previously estimated regressions and M lags of q_t as instruments. We do not formally identify the residuals of the third and fifth regressions with any particular shock, since, unlike the other residuals, we do not have a theory to justify doing so. Without loss of generality, in the third regression we assume that inflation does not respond to contemporaneous movements in R_t. The fourth regression of equation 1 can be written

$$R_t = \Gamma_{Rp}(L)\Delta p_t + \Gamma_{Ry}(L)\Delta y_t + \Gamma_{R\pi}(L)\pi_t + \Gamma_{RR}(L)R_{t-1} + \Gamma_{Rh}(L)h_{t-1} + \varepsilon_{rt}.$$

This is our hypothesized monetary policy rule, which incorporates our assumption that it depends on contemporaneous values of p_t, y_t, and π_t. The residuals from this equation are our estimates of the monetary policy shocks, $\hat{\varepsilon}_{rt}$.

The effects of the identified shocks

To build intuition for our historical decomposition of the path of residential investment after 1995, it is helpful to study some impulse response functions. An impulse response function describes how a variable, for example, residential investment, responds to a hypothetical exogenous shock with all other shocks set to zero. To the extent that the impulse response functions make sense, we can have confidence in the quality of the historical decompositions. Responses of output (per capita GDP in consumption units), the real equipment price, and per capita residential investment in consumption units to positive one standard deviation I-, N-, and R-shocks are displayed in figure 8. The units of the responses are percentage deviations from the path that would have been followed absent the shock. The magnitude of the response of a variable to a given shock indicates that shock's importance to the variable's fluctuations around trend.

The responses to the I-shock show the equipment price falling to its new long-run level, output rising to its new long-run level, and residential investment initially falling and then rising to the same long-run level as output.[7] The long-run responses of these variables are predicted by theory. Cheaper equipment encourages capital accumulation, which, in turn, raises labor productivity and output. The short-run response of residential investment makes sense because, until capital and wealth accumulate and increase the demand for housing, a fall in the equipment price should induce substitution away from other capital goods, including housing.

The increases in output, the equipment price, and residential investment after an N-shock are, at least qualitatively, also consistent with theory. An N-shock raises productivity, thereby encouraging capital accumulation, higher employment, and an increase in output. Since an N-shock directly affects output, while an I-shock affects output indirectly through capital accumulation, the N-shock has a faster impact on output. The equipment price responds by only a small amount. Qualitatively, it rises after the shock and then returns to its pre-shock level. This makes sense if the N-shock

encourages a short-run increase in the demand for investment goods relative to consumption goods, which it does in many theoretical models. Output rises toward its new long-run level. Residential investment initially surges, then falls back toward zero, before rising toward its long-run level. This kind of response is more difficult to reconcile with existing theories. However, the initial surge in residential investment is consistent with the rise in output, if housing demand is increasing in income.

The R-shock responses show that output and the equipment price respond by very little, but residential investment responds strongly, falling by about 1 percent. That residential investment is particularly sensitive to a monetary policy shock is consistent with traditional views about the monetary transmission mechanism.

The historical decomposition of the path of residential investment from 1995 to 2005 is displayed in figure 9. The historical decomposition for a given shock is based on simulating equation 1 with the estimated values of the coefficients, assuming all other shocks are equal to zero and that the given shock is equal to its estimate for each period of the decomposition. The contributions of the individual shocks, $\hat{\varepsilon}_{It}, \hat{\varepsilon}_{Nt}, \hat{\varepsilon}_{Rt}$, plus the two unidentified residuals add up to the observed path of residential investment. Consequently, figure 9 can be used for assessing which shocks contributed the most to the dynamics of residential investment. In the figure, the dark blue line represents the trend path of residential investment implied by the VAR when all the shocks are set to zero, the black line is the empirical path, and the light blue line is the path corresponding to the shock(s) indicated in the header of the individual panels of the figure.

Figure 9, panel A reveals that the investment-specific technological change of the late 1990s acted as a drag on residential investment. However, by 2004, the capital accumulation generated by this technological change meant that households were wealthier than otherwise.

FIGURE 8

Responses to shock

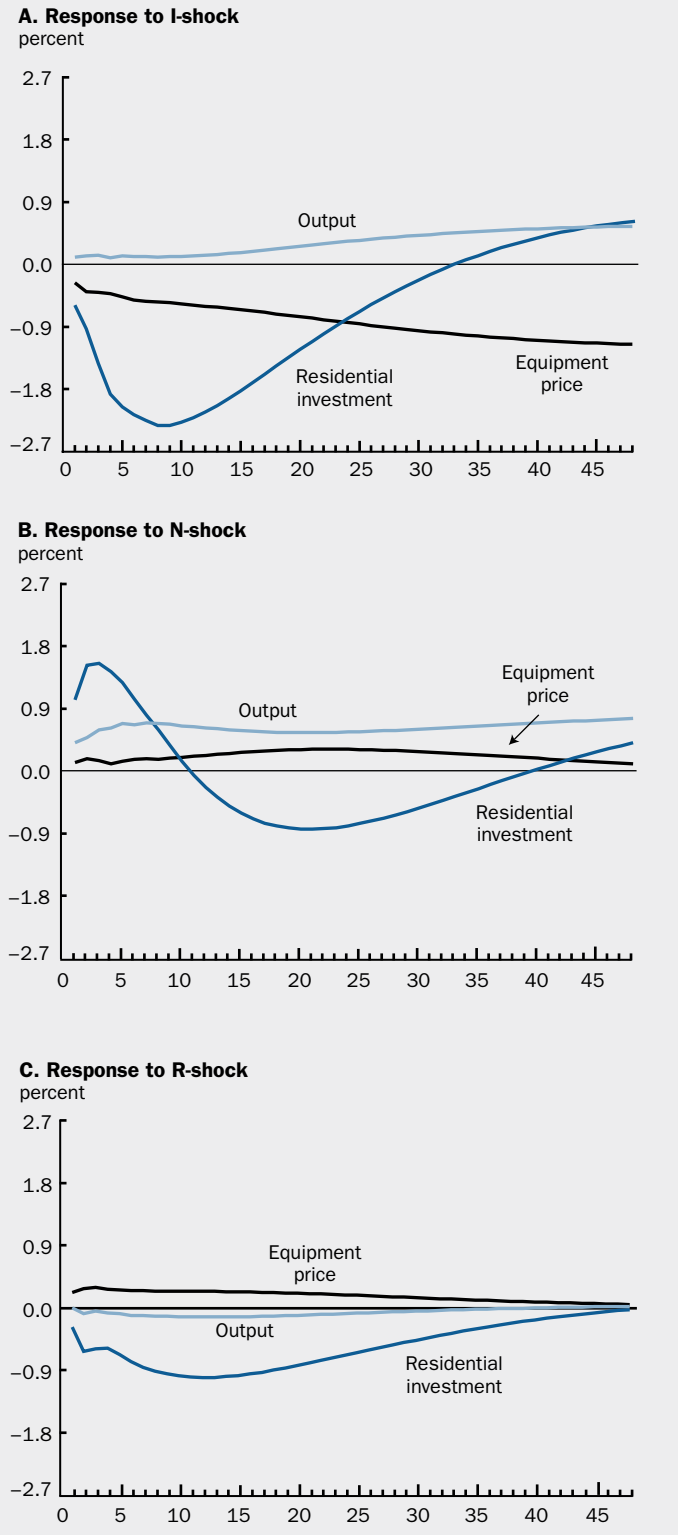
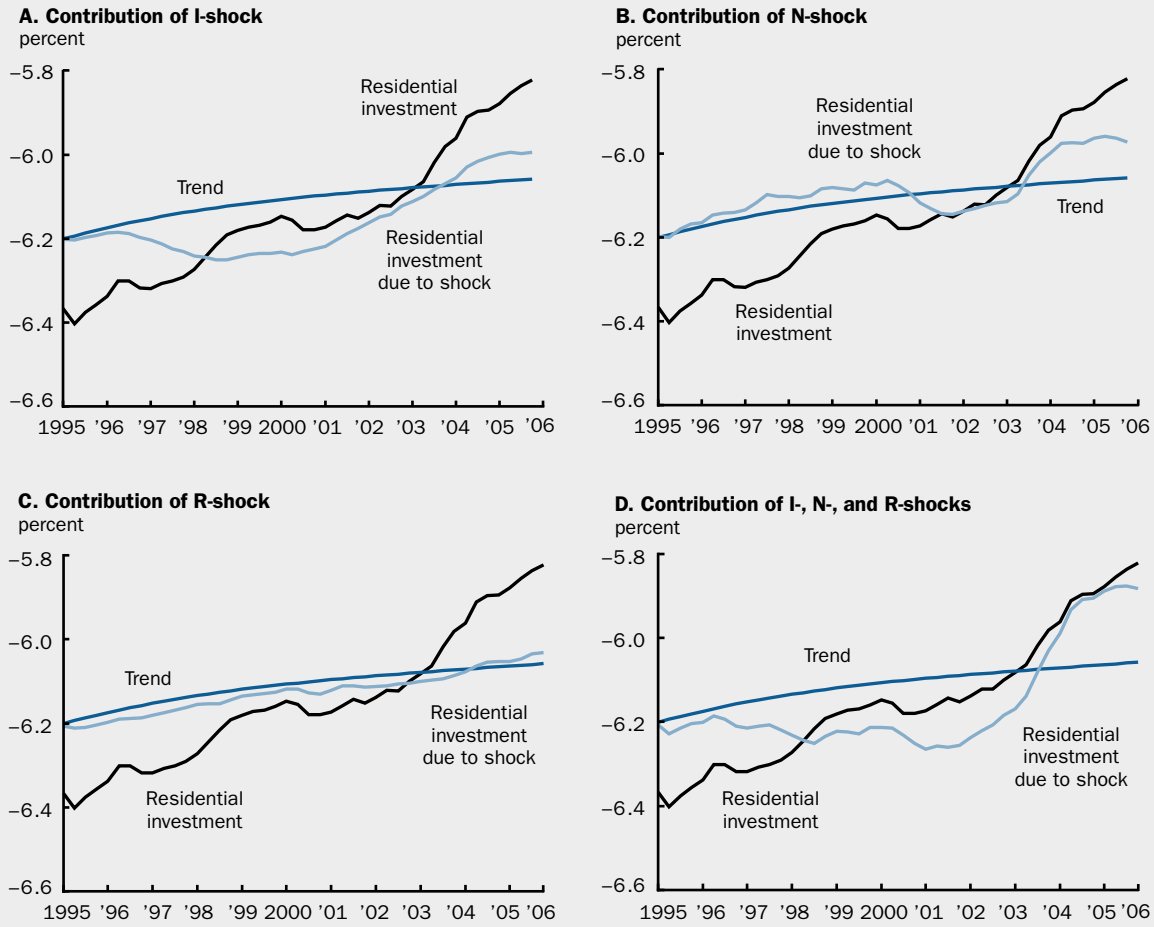


FIGURE 9

Residential investment, 1995:Q1–2005:Q4



Consequently, residential investment is eventually driven above its trend by the end of the sample. Panel B shows that N-shocks have had relatively little impact on residential investment, except during the 2001 recession and that recession's slow recovery. However, these shocks began exerting a strong positive influence toward the end of 2003. The mechanism for this is similar to that for I-shocks—technological change increases capital accumulation and wealth. From panel C, we see that monetary policy, through the R-shocks, seems to have had very little impact. Toward the end of the sample, these shocks exert a small positive influence, however. Panel D shows that the three identified shocks together account for almost all the surge of residential investment above its trend toward the end of the sample. Given the small contribution of the R-shocks around this time, this result suggests that the unusually high levels of residential investment in recent years may just be the direct result of the wealth accumulation from previously high rates of technological progress. In other words, according to the econometric analysis, the recent high rates of residential investment appear to have been driven mostly by fundamentals and not unusually loose monetary policy or speculative building.

The increase in homeownership

We now turn to the homeownership rate. Recall from figure 3 that the homeownership rate rose from about 64 percent in 1995 to about 69 percent in 2005. In this section, we describe how homeownership has changed along various demographic, income, educational, and regional lines. It then addresses the question of how much of the overall increase in homeownership can be attributed to changes in the distribution of households among these different categories. This is

TABLE 1

Homeownership rates and percent of population for various household characteristics

	Percent of population			Homeownership rates		
	1993	2003	Change 1993–2003	1993	2003	Change 1993–2003
Overall				64.68	68.39	3.71
Race of household head						
White	84.50	82.69	–1.81	68.62	72.32	3.70
Black	11.77	12.31	0.60	42.96	47.58	4.62
Other	3.73	4.99	1.26	44.04	54.53	10.49
Age of household head						
18–24	4.90	4.81	–0.09	12.43	15.91	3.48
25–29	8.55	7.20	–1.35	34.63	39.64	5.01
30–34	11.49	9.88	–1.61	50.98	54.75	3.77
35–39	12.08	10.06	–2.02	61.55	64.59	3.04
40–44	11.03	11.37	0.34	69.05	70.96	1.91
45–54	17.35	20.79	3.44	75.42	76.02	0.60
55–74	25.01	25.35	0.34	80.63	81.71	1.08
>74	9.59	10.54	0.95	72.54	78.48	5.94
Gender of household head						
Male	70.68	70.50	–0.18	70.95	74.31	3.36
Female	29.32	29.50	0.18	49.51	54.21	4.70
Marital status of household head						
Married, spouse present	54.02	51.51	–2.51	79.41	83.37	3.96
Unmarried, or spouse absent	45.98	48.49	2.51	47.38	52.47	5.09
Children in the household						
None	61.28	62.77	1.49	64.77	68.17	3.40
One	15.73	15.48	–0.25	63.36	68.04	4.68
Two	14.59	13.76	–0.83	67.59	71.26	3.67
Three	5.84	5.65	–0.19	64.98	68.00	3.02
Four or more	2.57	2.34	–0.23	53.50	60.42	6.92
Adults in the household						
One	30.87	32.54	1.67	45.65	52.27	6.62
Two	54.34	52.47	–1.87	71.88	75.47	3.59
Three	10.61	10.64	0.03	77.32	78.55	1.23
Four or more	4.18	4.35	0.17	79.40	78.67	–0.73
Region						
North East	19.85	18.66	–1.19	62.16	64.64	2.48
North Central	23.89	23.12	–0.77	67.85	73.14	5.29
South	35.54	37.00	1.46	66.33	70.08	3.75
West	20.72	21.22	0.50	60.69	63.79	3.10
Education of household head						
Less than high school	19.33	17.44	–1.89	58.77	57.75	–1.02
High school graduate	35.50	28.35	–7.15	64.67	68.52	3.85
Some college	19.88	27.26	7.38	63.20	67.71	4.51
College graduate	14.08	17.30	3.22	67.25	73.25	6.00
Postgraduate	11.20	9.64	–1.56	74.35	80.40	6.05

Source: American Housing Survey.

useful for assessing the extent to which other factors, including the introduction of new mortgage products, are needed to account for the increase in homeownership. The analysis in this section borrows from Segal and Sullivan (1998). The data underlying the analysis are from the American Housing Survey (AHS).[8]

Cross-sectional changes in homeownership

Table 1 displays how homeownership has changed between 1993 and 2003 along various demographic, educational, and regional lines, and table 2 does the same for income deciles. When referring to an individual characteristic, the unit of analysis is the household head. Also contained in these tables is the change

TABLE 2

Homeownership rate, by income deciles

1993 income decile	Percent of population: Change 1993–2003	Homeownership rates: 1993	2003	Change 1993–2003
1	–0.33	0.39	0.43	0.04
2	–1.67	0.45	0.49	0.04
3	–0.81	0.52	0.52	0.00
4	0.75	0.54	0.58	0.04
5	–1.29	0.63	0.61	–0.02
6	–1.25	0.64	0.67	0.03
7	0.18	0.72	0.75	0.03
8	–0.55	0.79	0.82	0.03
9	0.90	0.86	0.88	0.02
10	4.06	0.92	0.93	0.01

Source: American Housing Survey.

in the proportion of household heads that belong to each category. We do not discuss all the entries in these tables, but instead focus on two features of the tables that play roles in the analysis to follow.

The first and most important observation is that all but two categories in both tables display an increase in homeownership rates between 1993 and 2003. The two categories with declining homeownership rates are households with four or more adults, and household heads with less than a high school education. That the increase in homeownership cuts across so many different categorizations suggests that the overall homeownership rate is not merely reflecting changes in the distribution of the population among the categories. Something fundamental about the homeownership process has changed.

The second key observation is that among the different age groups, younger household heads experienced larger increases than middle-aged household heads. Consistent with the large increases in homeownership rates of younger household heads, single household heads have larger increases in homeownership rates than married household heads and other households with more than one adult. The pattern of homeownership among the young is a striking reversal of a trend seen between 1978 and 1993. Between 1978 and 1993, homeownership rates dropped for household heads under 40. This drop in homeownership rates coincided with a fall in marriage rates and the fraction of households with children for household heads under 40. This is consistent with the fact that starting a family has traditionally been one of the main instigators of homeownership. From this perspective, the increase in homeownership among younger household heads from 1993 and 2003 suggests that there might have been a reversal in rates of family formation among the young. Yet, marriage rates and the likelihood of a household having children by age of household head are about the same for 2003 and 1993. Clearly some other factor is driving homeownership among younger household heads.

The impact of changes in the distribution of households

Tables 1 and 2 show that there were noticeable changes between 1993 and 2003 in the distribution of household heads among the various categories. For example, the share of younger household heads fell, while that of older household heads has risen. Since older household heads tend to have higher homeownership rates than their younger counterparts, this change in distribution raises the economy-wide homeownership rate. Next, we consider, a decomposition of the overall homeownership rate into parts due to changes in the household head distribution and changes in the homeownership rate.

The basis for the decomposition is a simple linear probability model. This relates the probability that a household head owns her house as a linear function of household characteristics. Specifically,

$$h(x_i, t) = x_i'\beta_t, \qquad t = 1993, 2003,$$

where x_i is a column vector of dummy variables for household i corresponding to each of the characteristics in tables 1 and 2, and β is a column vector of coefficients. For example, if the household head is

TABLE 3

Contributions to 1993–2003 change in homeownership rate

Included variables	Base year 1993: Δ*h*	Δ*w*	Base year 2003: Δ*h*	Δ*w*
Age only	2.56	1.14	2.45	1.25
Demographic + regional	3.51	0.20	3.51	0.19
Demographic + regional + ed. + inc.	1.86	1.84	1.94	1.76

Notes: The figures reflect actual 1993–2003 data. The change in the homeownership rate is 3.70.

between the ages of 18 and 24, then the dummy corresponding to this characteristic is set equal to one and the dummies corresponding to the other age categories are equal to zero. We estimate two linear probability models, one using data from the 1993 AHS and one using data from the 2003 AHS.

The linear probability model is related to the overall homeownership rate in year t, $\bar{h}_t$, as follows:

$$\bar{h}_t = \sum_{i \in N_t} w_i h(x_i, t), \qquad t = 1993, 2003,$$

where w_i is the sample weight (the number of households in the population that each individual household in the sample represents divided by the total number of households in the population) for household i, and N_t is the year t sample. Given this relationship, the change in the overall homeownership rate can be decomposed into two parts. In addition, there are two ways to construct this decomposition, depending on the choice of the base year.

The decomposition is

$$\begin{aligned}\bar{h}_{2003} - \bar{h}_{1993} &= \left[\sum_{i \in N_B} w_i \left(h(x_i, 2003) - h(x_i, 1993)\right)\right] \\ &\quad + \left[\sum_{i \in N_{2003}} w_i h(x_i, \sim B) - \sum_{i \in N_{1993}} w_i h(x_i, \sim B)\right] \\ &= \Delta h + \Delta w.\end{aligned}$$

Here, B denotes the base year for the decomposition, and $\sim B$ means "not the base year." So if B is 1993, then $\sim B$ is 2003, and vice versa. The term Δh corresponds to the first term in square brackets and Δw corresponds to the second term. These variables correspond to the two channels through which the homeownership rate changes. The term Δh captures the part due to changes in the household-level homeownership rates, holding fixed the sample weights. The term Δw captures the part due to holding fixed the individual-level homeownership rates, but allowing the sample weights, that is, the household distribution, to change. Since the sample weights and their changes depend on the choice of base year, the decomposition also depends on which base year is chosen. Without a compelling reason to choose one base year over the other, we consider decompositions based on each base year.

The decomposition for the two possible choices of base year is shown in table 3. Each row of the table corresponds to different sets of variables in the linear probability model. The table shows that when the income and education variables are excluded from the model (the first two rows), then the Δh term accounts for most of the overall change in homeownership. If true, this would mean that factors influencing individual homeownership rates, and not changes in the characteristics of the population, are by far the most important factor underlying the overall change in homeownership. However, once the education and income variables are included in the model (the third row), then the share attributable to the change in weights rises to about one-half. This result is driven primarily by the income variables due to the fact that the shares of the higher income household heads have risen, and higher levels of income are associated with higher levels of homeownership.

Table 4 decomposes the effects of the changes in weights (the Δw terms) into the effects of the main groups of characteristics. This verifies that age and income are the main factors driving this component of the change in homeownership. An older and higher income population will tend to have a higher overall homeownership rate without any other changes in the economy. Still, table 3 indicates that part of the increase in homeownership remains unexplained. In the next section, we consider this unexplained portion of the rise in homeownership and relate it to the high levels of residential investment.

Connecting the booms in residential investment and homeownership

Here, we connect the boom in homeownership with the boom in residential investment. We argue that developments in the mortgage market have led to a large expansion of the pool of potential homeowners by lowering borrowing constraints.[9] Suppose a householder in 1995 would have preferred to buy at terms available in 2005 but, because these terms were

TABLE 4

Effects of changes in distribution of household characteristics

	Base year	
	1993	2003
Demographic and regional	0.48	0.47
Age	1.15	1.29
Sex	0.01	0.01
Marital status	–0.35	–0.36
Household size and composition	–0.13	–0.21
Race	–0.22	–0.27
Region	0.02	0.01
Education and income	1.36	1.29
Education	0.15	0.06
Income	1.21	1.24

not available in 1995, chose to rent. If rental housing is not perfectly substitutable with owned housing in the short run, as is likely the case because of moral hazard considerations, then the availability of 2005 mortgage terms increases the demand for owned housing. All else being equal, this should increase the quantity of housing supplied and raise the share of residential investment in GDP. In the long run, rental and owned housing are essentially perfect substitutes. As the homeownership rate reaches its new, less borrowing-constrained, equilibrium level, we would expect residential investment as a share of GDP to return to "normal" levels. The remainder of this section considers the possible connection between mortgage innovation and homeownership and residential investment in more detail.

The dramatic rise in residential investment and homeownership has coincided with equally dramatic developments in the mortgage market. Over the past ten years to 15 years, the mortgage market has developed substantially in four areas. First, technological progress has reduced the cost of approving a mortgage under a standardized set of lending guidelines, in part by allowing more precise measurement of a borrower's credit risk. Second, mirroring developments in financial markets more generally, many new kinds of mortgages have become available. Third, the secondary mortgage market has grown and matured so that many kinds of mortgages can now be packaged and sold as mortgage-backed securities. Fourth, the mortgage market has become more specialized, as firms concentrate on different pieces of the market, including origination, servicing, and securitization.

Of these developments, the second is most important for our argument. We think that the main impact of the other three developments is to increase competition and lower transactions costs. Also, the development of the secondary mortgage market probably improved the risk–return tradeoff between mortgage-backed securities and other financial instruments. This would have the effect of increasing the supply of capital to mortgage markets. All these developments drive down mortgage rates. Historically, we have seen large swings in mortgage rates without large changes in the homeownership rate. So we conclude that the cost reductions and increases in the supply of capital to the mortgage market are likely to have had a relatively small impact on homeownership.

In contrast, the development and dissemination of many new mortgage products have made it possible for large numbers of people to acquire mortgages who would have been unable to previously. Before the 1990s, the standard kind of mortgage required the potential home buyer to satisfy a relatively rigid set of criteria on loan-to-value ratios, income, and other measures of creditworthiness. This rigidity was necessary for the development of the secondary market for mortgages. While beneficial in this respect, it effectively shut many potential homeowners out of the market. By reducing the remaining rigidities, the new mortgage products have expanded the pool of potential homeowners.

Two developments seem to have had a particularly large impact along this dimension. One development involves the so-called combo loan. These mortgages reduce or even eliminate entirely the need for a down payment. The second involves mortgages aimed at the "subprime" market. Subprime borrowers are individuals with low credit ratings. Subprime lending allows borrowers who, in the past, would not have qualified for a mortgage to qualify by paying higher interest rates and offering more equity or lower loan-to-value ratios. Before the development of the combo loan, potential buyers had to accumulate sufficient savings to afford the necessary down payment. Before the development of the subprime market, most borrowers with poor credit would not have been able to get a mortgage at all. It is easy to see how these developments have lowered borrowing constraints and made it possible for potential buyers to buy earlier or buy at all.

The information on mortgages collected by the AHS which is consistent with the above interpretation of how new kinds of mortgages have affected the choices available to households. For instance, among first-time home buyers with a mortgage, 7.9 percent report that no down payment was required in 1993, and 12.1 percent report this in 2003. It is not possible to determine from the AHS data whether a borrower has a subprime mortgage. However, the survey does report the interest rate paid on the primary mortgage. We use this to compute the coefficient of skewness of interest rates in 1993 and 2003. A large positive skewness coefficient indicates that the distribution of interest rates includes a larger fraction of relatively high interest rates. In 1993, the skewness coefficient is 0.59. In 2003, it is 1.84. This increase in skewness is consistent with a greater fraction of mortgages being high interest subprime mortgages, although it could also arise with an increased usage of adjustable rate mortgages.[10]

What impact have these changes had on home buying? Since the most constrained home buyers are first-time buyers and first-time buyers are typically young, we can use table 1 to assess the possible impact of the changes in the structure of mortgages. Table 1 shows that homeownership rates for young

buyers have risen by much more than for older buyers, except for the oldest buyers (>74). So, for buyers under 40 the changes in homeownership rates are all greater than 3 percent, while for buyers from age 40 to age 74 the changes are all less than 2 percent.

The evidence so far is consistent with the mortgage developments increasing the pool of buyers. The last question we address is whether this increase in the pool is large compared to the recently high levels of residential investment and homeownership. Given the paucity of data we have to work with at this time, our calculations are rough and tentative. We first make a rough estimate of the additional homeowners due to the subprime lending. Our hypothesis is that these homeowners come from the ranks of renters, so we compare this magnitude with changes in the rental vacancy rate. To assess the potential impact on residential investment and homeownership, we also compare our calculated increase in the number of new homeowners due to subprime lending with changes in the number of housing completions and the unexplained portion of the increase in homeownership displayed in table 3.

Using data on the volume of home purchase mortgage originations in 2002 (from the Mortgage Bankers Association), the average loan amount in 2002 (Federal Housing Finance Board), and the fact that 10 percent of such originations were subprime mortgages (Gramlich, 2005), we calculate that about 673,000 subprime mortgages were issued for home purchases in 2002. In 1994, 76 percent of new originations were for home purchases (Federal Housing Finance Board). Combining this information with numbers on the volume of the subprime mortgage market and the average loan amount in 1994 (Gramlich, 2005), we calculate that about 242,000 subprime mortgages were issued for home purchases in 1994. So between 1994 and 2002, there was an increase of 431,000 subprime home purchases.

Assuming that the home buyers using these subprime mortgages would previously have been excluded from the mortgage market, this increase in subprime purchases must be accompanied by a reduction of similar scale in renting households (assuming no substitution between rental and owned housing units and no impact on the formation of new households). In this case, we should have seen an increase in the rental vacancy rate. In fact, between 1994 and 2002, there was an increase of 1.5 percentage points in the rental vacancy rate (U.S. Census Bureau). This increase in the vacancy rate translates into 570,000 additional vacant rental units in 2002.[11] By this calculation, the additional subprime lending accounts for 76 percent (431,000/570,000) of the increase in the rental vacancy rate. Additionally, the magnitude of the new subprime lending accounts for about 72 percent of the roughly 600,000 additional housing completions (U.S. Census Bureau) in 2005 compared with 1995.

According to table 3, about 1.75 percent of the increase in homeownership between 1995 and 2005 remains unexplained by changes in the cross-sectional characteristics of the population. With 113 million households in 2005, this translates into two million additional homeowners in 2005, or an average of 200,000 additional homeowners each year between 1995 and 2005. The volume of subprime lending we calculated for 1994 is much higher in the years afterward (Gramlich, 2005). Consequently, by our calculations, the subprime market can easily account for the additional homeowners unaccounted for by changes in the cross-sectional characteristics of the population.

We conclude that substitution away from rental housing made possible by developments in the mortgage market, such as subprime lending, could account for a significant fraction of the increase in residential investment and homeownership. The current spending boom thus may be a temporary transition toward an era with higher homeownership rates and spending on housing, which will ultimately move nearer to historical norms.

It may appear that the role of mortgage markets in accounting for the increase in residential investment contradicts our finding earlier in this article that the current levels of spending on housing are largely explained by technology-driven factors. We do not view this as a contradiction, because the kind of technological change that made the mortgage market developments possible affected many parts of the economy. In particular, much of the technological change underlying our previous finding can be attributed to firms finally working out how to take advantage of innovations in information technology. The advances we have in mind have found uses in all sectors of the economy, including the financial services industry, and so can be viewed as neutral technological change. In addition, much of the decline in the equipment price underlying our estimates of the investment-specific shocks can be attributed to information technology. So investment-specific technological change has also contributed to the evolution of the mortgage market.

Conclusion

This article has attempted to explain two features of the turn of the twenty-first century U.S. economy: high levels of residential investment and homeownership rates. Our main findings are as follows. First, it appears that the housing boom has not been driven by unusually loose monetary policy. This is not to say the monetary policy has not been unusually loose, but that to the extent it has been loose, this is not what has been driving spending on housing. Second, the current levels of spending on new housing are largely explained by technology-driven wealth creation over the previous decade. Third, changes in the demographic, income, educational, and regional structure of the population account for about one-half of the increase in homeownership. That is, without any other developments, the homeownership rate is likely to have gone up anyway, but not by as much as it has done. The last finding is that substitution away from rental housing made possible by developments in the mortgage market, such as subprime lending, could account for a significant fraction of the increase in residential investment and homeownership.

We view our findings as supporting the view that the current housing boom may be a temporary transition toward an era with higher homeownership rates in which spending is temporarily higher than historical norms but will eventually return to such norms. While we have so far mostly avoided discussing housing prices, our findings do suggest that to the extent that house prices have grown considerably in recent years, this is not due to unusually excessive speculation in the housing market, such as would occur in a bubble. Instead, our findings point toward the high prices being driven by fundamentals.

NOTES

[1]Home building and residential investment are not quite the same thing, since average home size and quality as well as construction costs vary over time. However, they are closely related, and this article will use the terms interchangeably.

[2]Household formation is also affected by conditions in the housing market. For example, high house prices or rental rates may induce singles to remain at home rather than find a place of their own. The discussion here implicitly assumes that these factors are overwhelmed by social developments, such as declines in marriage rates among the young, and demographic developments, such as the baby boom.

[3]If households typically stay in a house five years before they move, then the relevant expected inflation rate is over five years. Expectations of inflation over five years should be slower to change than expectations over one year.

[4]Median wealth has risen less than average wealth, indicating that wealth has become more unevenly distributed. Suppose the demand for housing is an inferior good, so that demand for it grows less than in proportion to growth in wealth. Think Bill Gates. In this case, greater dispersion in wealth should lead to a decline in the share of housing in the aggregate portfolio.

[5]Progress has been made in identifying fiscal shocks. See, for example, Burnside, Eichenbaum, and Fisher (2004). However, there is much less consensus on the viability of the available identification strategies.

[6]The real price of capital equipment is measured as the NIPA deflator for equipment and software divided by a consumption deflator derived from the NIPA deflators for consumer nondurables and services. Real GDP is measured in consumption units by dividing nominal GDP by the consumption deflator. Time t inflation is the difference between the date t and $t - 1$ values of the log of the consumption deflator.

[7]The response of output to a shock is the accumulated response of Δy_t. The response of residential investment is the response of h_t plus the response of output.

[8]The AHS is a survey that asks questions about the quality of housing in the United States. In gathering information, the U.S. Census Bureau interviewers visit or telephone the household occupying each housing unit in the sample. For unoccupied units, they obtain information from landlords, rental agents, or neighbors. The data used for this article is taken from the national survey (there is also a metropolitan area survey). The national survey is conducted during a three-month to seven-month period during which interviewers gather information on housing throughout the country. The survey covers about 55,000 housing units every two years, in odd-numbered years, and is available only through 2003.

[9]Chambers, Garriga, and Schlagenhauf (2005) make a similar argument in the context of a formal model of the life cycle.

[10]Note that looking at average down payments as a fraction of price is not very informative about the relaxation of borrowing constraints, since combo loans involve smaller down payments and subprime loans involve larger down payments, compared to conventional mortgages.

[11]In 2002, 32.1 percent of households rented and the vacancy rate was 8.9 percent. With 108.2 million occupied housing units, the number of vacant units equals 0.321 × 108.2 million / 0.911 = 38 million. Thus, 1.5 percent of 38 million is 0.57 million.

REFERENCES

Burnside, Craig, Martin Eichenbaum, and Jonas D. M. Fisher, 2004, "Fiscal shocks and their consequences," *Journal of Economic Theory*, Vol. 115, No. 1, March, pp. 89–117.

Chambers, Matthew, Carlos Garriga, and Don E. Schlagenhauf, 2005, "Accounting for changes in the homeownership rate," Florida State University, manuscript.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, 2005, "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, Vol. 113, No. 1, February, pp. 1–45.

Fisher, Jonas D. M., 2006, "The dynamic effects of neutral and investment-specific technology shocks," *Journal of Political Economy*, Vol. 114, No. 3, June, pp. 413–451.

Gramlich, Edward M., 2005, "Subprime mortgage lending: Benefits, costs, and challenges," remarks at the Financial Services Roundtable Annual Housing Policy Meeting, Chicago, IL, May 21.

Segal, Lewis M., and Daniel G. Sullivan, 1998, "Trends in homeownership: Race, demographics, and income," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 22, No. 2, Second Quarter, pp. 53–72.

Are U.S. and Seventh District business cycles alike?

Michael A. Kouparitsas and Daisuke J. Nakajima

Introduction and summary

When academic economists talk about business cycles, they have something more general in mind than persistent fluctuations of gross domestic product (GDP) about its trend, which is the definition typically used by business economists. For an academic economist, the business cycle describes the way that cyclical fluctuations of GDP typically relate to cyclical fluctuations of other economic time series (such as consumption and investment) from the same economy. One of the most striking findings of the vast academic business cycle literature is that irrespective of the time period or particular country, business cycles are all alike. This means that the typical relationship between cyclical fluctuations of GDP and cyclical fluctuations of other economic time series of the U.S. economy is similar to the typical relationship between cyclical fluctuations of the same time series in all other market-based economies. As Lucas (1977) notes, this finding is both important and challenging for the study of business cycles, since it suggests the possibility of a unified explanation of business cycles based on general laws governing market economies.

So, we know that national business cycles are alike in important ways. What do we know about subnational business cycles? Given that subnational economies, such as those of U.S. states, are as large as some national economies, one would expect their business cycles to have been well studied. In contrast to national business cycles, little is known about subnational business cycles. The goal of this article is to expand our knowledge of subnational business cycles by testing whether the proposition that all business cycles are alike extends to U.S. states. We limit our analysis to the business cycles of the U.S. and the five states (Iowa, Illinois, Indiana, Michigan, and Wisconsin) of the U.S. Federal Reserve's Seventh District, home of the Federal Reserve Bank of Chicago.[1]

Our approach follows that of international business cycle studies, such as Backus and Kehoe (1992), by conducting a detailed analysis of the way in which activities within a regional economy relate to the region's aggregate business cycle and the way in which regional aggregate business cycles relate to one another. The main limitation on subnational business cycle research stems from a deficiency of state-level data analogous to the national income and product accounts data that are typically used in the analysis of national business cycles. We overcome this problem by finding suitable state-level proxies for national account aggregates (please see the appendix for further details). Consumption expenditure is proxied by real retail sales taxes, investment is proxied by the number of residential home sales and housing permits, while output is proxied by real personal income, various measures of labor input (including nonfarm payroll employment and average manufacturing hours) and capital utilization. One of the byproducts of this analysis is that we uncover new leading variables that could serve as useful indicators of the future direction of District state business cycles.

We find that District state business cycles are like the national business cycle along a number of dimensions. Turning to the within-region analysis, which explores the way activities within a regional economy relate to the region's aggregate business cycle, we find that state-level analogues of consumption and

Michael A. Kouparitsas is an economist and Daisuke J. Nakajima is an associate economist in the Economic Research Department at the Federal Reserve Bank of Chicago. *The authors thank Richard Kaglic for the suggestion of this topic and many useful discussions in the early stages of this article; Craig Furfine, François Velde, and Jonas Fisher for their comments and suggestions on an earlier draft; and, above all, Alexei Zelenev for his insightful comments and perseverance in assembling the data used in this article.*

residential investment have similar business cycle characteristics to their national counterparts. Moreover, they tend to be strong leading indicators of the national and state business cycles. Our labor market analysis yields much stronger results. Here, we find that the business cycle characteristics of national and state-level measures of labor market activity are virtually identical, especially with regard to whether they are procyclical or countercyclical, as well as leading or lagging.

A similar conclusion emerges from the across-region analysis, which explores the way in which regional aggregate business cycles relate to one another. We find that irrespective of the data source for measuring aggregate business cycles (we consider real personal income and nonfarm payroll employment), national and state-level business cycles are highly correlated at zero leads and lags. For Michigan and Indiana, there is weak evidence that the business cycles of these states lead the national business cycle by one quarter. It is fair to say based on these statistics that the District state business cycles are well described by the national business cycle. We conjecture that the high correlations are the result of important common shocks that affect both the national and regional business cycles.

These results are important not only for the study of business cycles, but also for the conduct of monetary policy. Mundell (1961) and others have argued that under certain conditions two or more economies may be better off if they abandon their individual monetary policies and pursue a common monetary policy with a common currency. The key provision governing which economies should form a single so-called optimal currency area (OCA) is the extent to which they have similar business cycles. If two or more economies have similar business cycles, then a common monetary policy for them is optimal. Our results suggest that the U.S. and the Seventh District clearly fit Mundell's notion of a single OCA, which suggests that optimal monetary policy for the U.S. is also the optimal monetary policy for the Seventh District and the District states.

Brief literature review

Researchers have pursued a number of different avenues to overcome the apparent lack of an aggregate monthly or quarterly measure of state-level economic activity. A popular approach constructs composite indexes of regional activity from a set of disaggregate measures of state economies released on a monthly or quarterly basis. Crone and Clayton-Matthews (2005) use Stock and Watson's (1989) coincident indicator method to derive consistent coincident indexes for all 50 U.S. states. Their indexes rely exclusively on state-level labor market data, including total nonfarm payroll employment, average weekly hours in manufacturing, the unemployment rate, and real labor income. Crone and Clayton-Matthews do not explore the business cycle aspects of their coincident indicators. However, their approach follows earlier work by Orr, Rich, and Rosen (1999), which focuses on New Jersey and New York. One of the findings of this study is that while the amplitudes of the coincident index and nonfarm payroll employment differ, they have identical business cycle turning points. This suggests that one could rely solely on the nonfarm payroll employment series to identify the start and end dates of New Jersey and New York state business cycles. A similar picture emerges from the closely related study by Clayton-Matthews and Stock (1998–99), which constructs similar coincident indicators for Massachusetts. In this more ambitious study, the authors estimate a coincident index using data on state taxes and labor market indicators. As in the other study on New Jersey and New York state business cycles, dating the Massachusetts business cycle from the estimated coincident index yields the same turning points as using only the nonfarm payroll employment of the state. Based on these studies, we conclude that there are reliable high frequency indicators of aggregate state activity, such as nonfarm payroll employment, which can be used to identify state business cycles.

Other researchers have arrived at a similar conclusion about regional measures of economic activity and have gone on to explore the relationship between subnational and national business cycles in a variety of ways. Some examples of these business cycle analyses include Owyang, Piger, and Wall (2005), Kouparitsas (2002), and Hess and Shin (1998). Owyang, Piger, and Wall (2005) explore the synchronicity of business cycle phases for all 50 states. Their approach to identifying business cycles is fundamentally different to the method used in this article. They adopt a statistical method that identifies cyclical components of the state data, which is consistent with the approach used by the National Bureau of Economic Research (NBER) to date business cycles. The NBER rule-based approach to identifying expansions and contractions often yields different business cycle characteristics to the cyclical estimates identified using the statistical methods that we apply in this article. Despite these methodological differences, Owyang, Piger, and Wall find, as we do, that the business cycles of the five Seventh District states (Illinois, Indiana, Iowa, Michigan, and

Wisconsin) are closely related to the U.S. national business cycle.[2] Kouparitsas (2002) also explores the synchronicity of U.S. regional business cycles. The analysis is limited to the eight U.S. Bureau of Economic Analysis regions. Using an observed component technique, the study finds that while economic fluctuations of these economies are similar at business cycle frequencies, they are quite different at lower frequencies due to the different industry mix of each region. Hess and Shin (1998) is the most closely related study. They use a similar approach to the one used in this article for their study of the business cycles of nine U.S. Census regions and 13 U.S. states. Their results, however, are not directly comparable to ours, since they focus on annual data and report mean statistics across the 22 regions used in their study.

Methodology

Our approach follows the macroeconomic business cycle literature by assuming that the data can be broken down into three distinct parts: a trend component, which captures permanent changes in the data series; a cyclical component, which captures persistent temporary deviations from the trend; and a high frequency component, typically referred to as noise, which is uninformative about the cyclical or trend components. Although it is possible that innovations to the trend and cyclical components are common, business cycle studies typically assume that these components are driven by independent innovations, which allow them to ignore the trend properties of the data.

A common misconception is that the cyclical component of any aggregate time series is a measure of the aggregate business cycle. This is only true for broad indicators of economic activity, such as GDP and gross national product (GNP). To see this we need to review Burns and Mitchell's (1946) classic definition of business cycles, which describes them as "expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals."[3] Although this definition is somewhat vague, Burns and Mitchell were quite specific about the data that should be used to measure the aggregate business cycle.[4] They recommended that the aggregate business cycle be measured as the cyclical component of the broadest measure of activity, in particular GNP. However, they were quick to add that estimates of GNP on a quarterly and monthly basis were at that time in the experimental stage, which led to their recommendation that aggregate business cycles be measured by isolating common cyclical components from a wide set of variables measured at monthly or quarterly intervals. We now have timely quarterly measures of aggregate economic activity, especially at the national level, so we can measure Burns and Mitchell's definition of the aggregate business cycle by extracting the cyclical component of a broad measure of activity, such as quarterly GNP.[5]

As discussed earlier, District states have a limited number of broad indicators of aggregate state-level activity. The most obvious choice is the state equivalent of GDP: gross state product (GSP). However, GSP arrives with a considerable lag of up to two years and is only available at an annual frequency. The alternatives are total personal income and total nonfarm payroll employment, which are published at quarterly and monthly intervals, respectively. Both measures have been used in the study of regional economic fluctuations. For brevity's sake and in keeping with the recommended approach of Burns and Mitchell (1946)—that is, to use the broadest measure of regional income to measure business cycles—we limit our discussion in this article to aggregate business cycles described by total personal income. However, we did explore the behavior of District state business cycles using both these measures of aggregate regional activity. After controlling for the fact that employment is a lagging indicator, we found that there was very little difference between aggregate business cycles described by personal income and nonfarm employment.[6]

Summarizing business cycles

We follow the macroeconomic business cycle literature by summarizing regional business cycles along three dimensions: comovement, persistence, and volatility.

Comovement reflects the extent to which the cyclical component of an activity within a region, comprising several states, is correlated with the region's aggregate business cycle. If the correlation of a cyclical component of an activity with the region's aggregate business cycle is close to one, the activity is procyclical, which implies the within-region activity has similar cyclical peaks and troughs to the region's aggregate business cycle. It also means that the idiosyncratic component of the cyclical component of the within-state activity is small. Alternatively, if the correlation is close to minus one, this also implies the idiosyncratic component of the state activity is small, but now the activity is countercyclical, suggesting that a peak for one series is synchronized with a trough for the other.

Correlations can also be used to identify whether a within-region activity leads or lags the region's business cycle. We do this by plotting the cross-correlation function of the region's business cycle and the cycle

of the within-region activity, which is the sequence of correlations of the region's aggregate business cycle at time t with the cycle of the within-region activity at time $t - j$ over a range of j. We limit our analysis to leads and lags of up to two years. If the maximum absolute correlation occurs at time $k > 0$, this implies the within-region activity is a leading indicator of the region's aggregate business cycle. If the maximum occurs at $j = 0$, the activity is called a coincident indicator. Otherwise the activity is a lagging indicator. Most cases we explore have a U-shape or an inverted U-shape cross-correlation function, where the single trough or peak identifies the lead or lag. In some instances, the cross-correlation functions take on a horizontal S-shape or a horizontal Z-shape. We identify the lead or lag in those cases as the maximum of the absolute value of the peak and trough.

The persistence of the cyclical component of an activity within a region reflects the size of the correlation of the cyclical component with itself at time t and time $t + 1$. If this first-order autocorrelation is close to one, the activity has a highly persistent cycle.

The volatility of the cyclical component of a time series is captured by its standard deviation. Researchers also study relative volatility, which is simply the ratio of the standard deviations of the cyclical component of a within-region activity and the region's aggregate business cycle. A relative volatility of one implies that the within-region activity has the same cyclical amplitude as the region's aggregate business cycle, while a relative volatility above one implies that the within-region activity has a larger cyclical amplitude than the region's aggregate business cycle.

Estimating cyclical components of time series

There are several competing methods for performing univariate trend/cycle/noise decompositions. We use the most widely used approach, which is based on spectral analysis. This technique takes advantage of the fact that every time series can be thought of as the sum of components spanning different frequencies of oscillation. Business cycle studies typically assume that fluctuations with low frequency oscillations, lasting more than eight years, capture the trend, while high frequency oscillations, lasting less than 18 months, capture noise. Fluctuations that occur within the range of 18 months to eight years are referred to as business cycle frequencies. Some researchers have adopted slightly wider business cycle ranges by assuming the trend is captured by oscillations of ten or more years. We found that widening the interval affected the volatility and persistence of business cycles, but had little effect on business cycle comovement.

The easiest way to extract these business cycle frequencies is to use a band-pass filter, which essentially zeros out the frequencies of the data that the researcher is not interested in. Due to data limitations, researchers need to use approximate band-pass filters. The results reported in this article are based on data filtered by the widely used approximate band-pass filter developed by Christiano and Fitzgerald (2003), although we find that our main conclusions are not affected by the choice of approximate band-pass filter.

Are subnational and national business cycles synchronized?

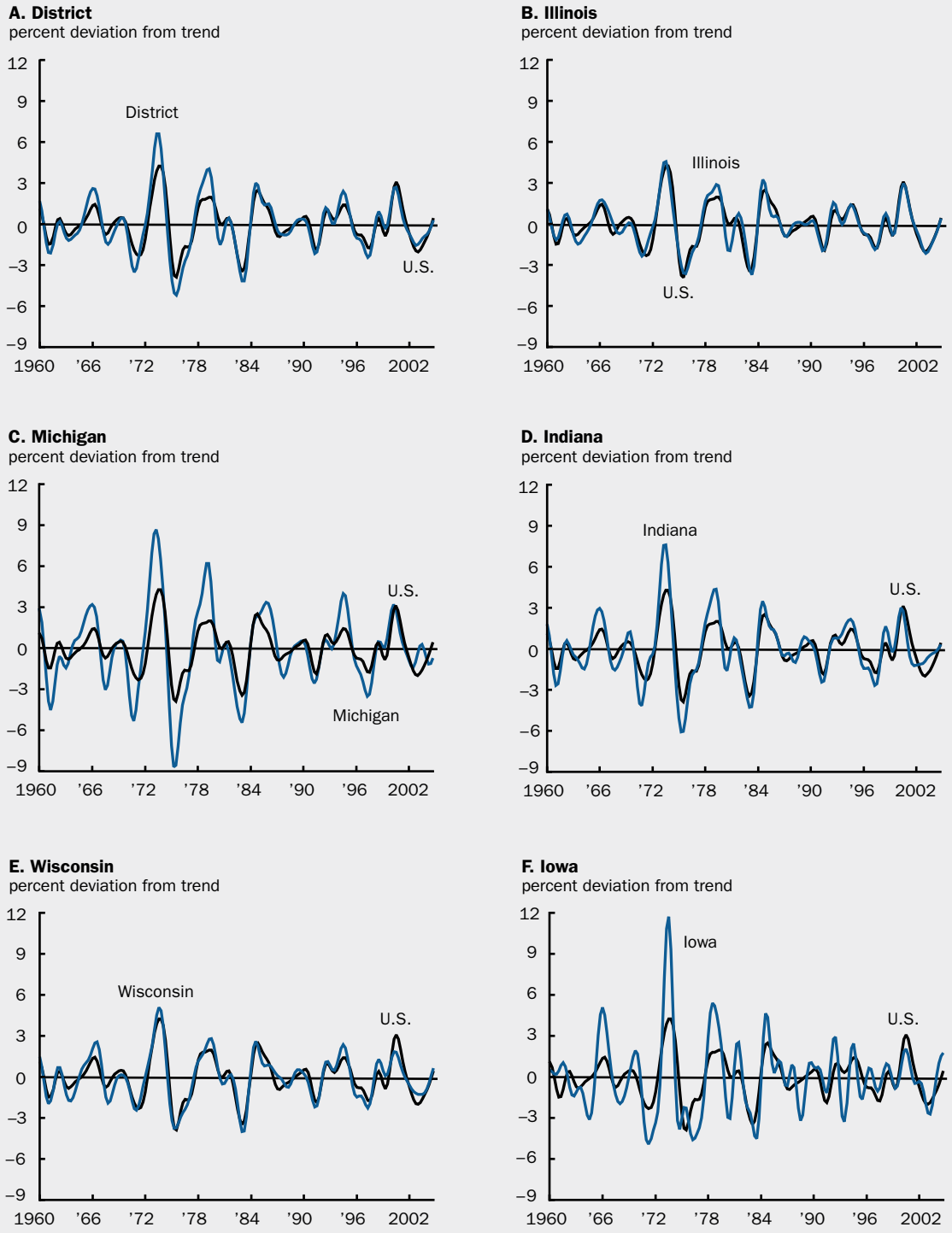
We begin our analysis by exploring the synchronicity of regional and national aggregate business cycles. Figure 1 plots the business cycles of U.S. national income (black line) with the business cycles of aggregate income for the Seventh District and aggregate income for the individual states (blue lines). It is obvious from this figure that there is a high degree of business cycle comovement. This observation is confirmed in figure 2, panel A, which plots the cross-correlation function of the national business cycle at time t and the District and state cycles from time $t - 8$ to time $t + 8$. In most cases the peak correlation (ranging from 0.76 for Iowa to 0.94 for Illinois) occurs at a zero lead/lag. The exceptions are Michigan and Indiana, which lead the nation by one quarter. However, the point estimates of these correlations are marginally higher than those for their contemporaneous counterparts. These results suggest that state business cycles have relatively small idiosyncratic components, so to a very large extent, they reflect the national business cycle.

Figure 1 also implies that the persistence of the District state business cycles, with one exception, is roughly the same as the persistence of the national business cycle. Iowa appears to have a less persistent cycle. These observations are confirmed in table 1, panel A, which reports the first-order autocorrelations of the District state business cycles and the national business cycle. According to this table, while Iowa has a less persistent cycle with a first-order autocorrelation of 0.89, the persistence is roughly 0.93 for the business cycles of the other District states and the nation.

Variations in the volatility of the regional business cycles are also revealed in figure 1. The business cycles of Michigan, Indiana, and Iowa appear to be more volatile than the national business cycle, whereas the business cycles of Illinois and Wisconsin appear to be roughly similar to the national business cycle in terms of volatility. The extent of these differences is quantified in table 1, panel B, which reports

FIGURE 1

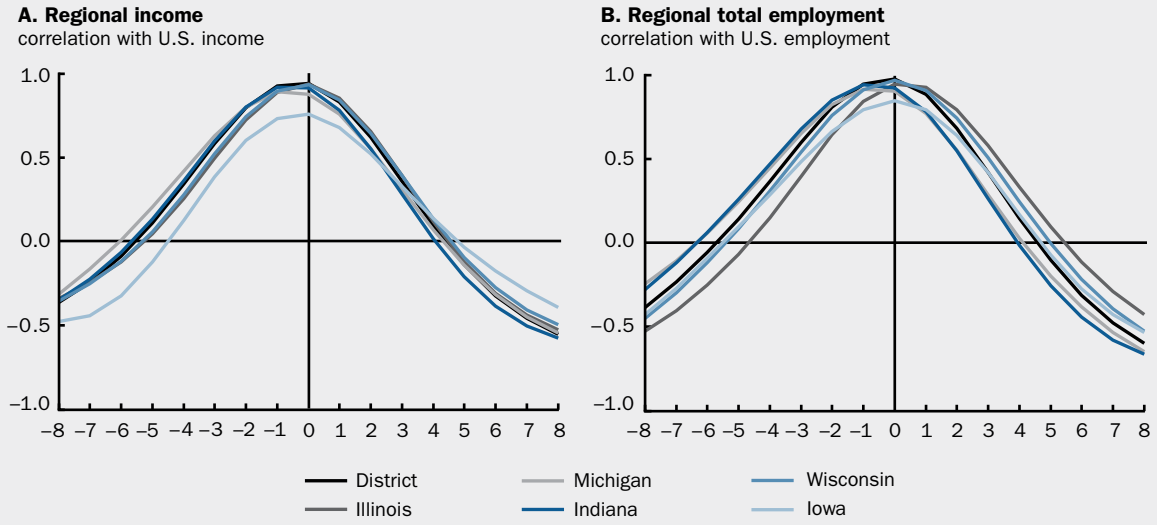
Aggregate business cycles



Source: Authors' calculations based on data from the U.S. Bureau of Economic Analysis from Haver Analytics.

FIGURE 2

Across region: Income and total employment



Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis and the U.S. Bureau of Labor Statistics from Haver Analytics.

the absolute volatility for the national business cycle in the first column and the volatility of the state business cycles relative to the national business cycle in the remaining columns. According to these estimates, Michigan and Iowa have business cycles with amplitudes that are twice as large as those of the national business cycle.

The most obvious explanation for differences in state business cycles is that each state has a different mix of industries. Table 2 explores this by reporting the share of gross state product accounted for by major industries in each of the state economies; it also shows this breakdown by industry for the District and national economies. Based on these data, Iowa differs significantly from the other states in the District in that it has an agriculture share that is more than three times as large as the share for the District. This suggests that Iowa's business cycle is more volatile and less persistent because Iowa is more sensitive to fluctuations in commodity prices, which are considerably more volatile than other prices. In contrast, Illinois is an outlier in that it has a manufacturing share (and general industry mix) that looks more like the national average than the District average, which explains why the Illinois business cycle is virtually identical to the national business cycle.

Because of the widespread use of state-level employment data for business cycle analysis, we repeated

TABLE 1

Across region

	U.S.	District	Illinois	Michigan	Indiana	Wisconsin	Iowa
A. Persistence							
Income	0.93	0.93	0.92	0.94	0.92	0.93	0.89
Nonfarm payroll employment	0.94	0.93	0.92	0.93	0.93	0.94	0.93
B. Volatility							
		Relative volatility					
Income	1.49	1.35	1.04	1.95	1.48	1.10	1.79
Nonfarm payroll employment	1.17	1.20	0.96	1.75	1.48	1.12	1.04

Note: In panel B, income and nonfarm payroll employment are measured in absolute volatility for the U.S.
Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis and the U.S. Bureau of Labor Statistics from Haver Analytics.

TABLE 2

Percent of regional gross state product accounted for by major industry

Region	Agriculture	Mining	Construction	Manufacturing	Transportation & public utilities	Trade	FIRE	Services	Government
Illinois	0.45	0.23	4.51	14.86	5.95	13.05	22.15	29.32	9.48
Michigan	0.47	0.19	4.71	22.64	4.39	12.66	16.46	28.50	9.98
Indiana	0.62	0.32	4.62	29.73	5.69	11.76	15.56	21.95	9.74
Wisconsin	1.34	0.15	4.42	24.55	4.87	12.33	18.02	23.43	10.89
Iowa	2.70	0.20	3.90	21.93	5.67	13.03	19.18	21.62	11.77
District	0.77	0.22	4.53	21.09	5.32	12.65	18.82	26.59	10.02
U.S.	1.01	1.24	4.47	14.63	5.04	12.86	19.81	29.30	11.64

Note: FIRE is finance, insurance, and real estate.
Source: Authors' calculations based on data from the U.S. Bureau of Economic Analysis from Haver Analytics.

this exercise using total nonfarm payroll employment. Panel B of figure 2 reports the cross-correlation functions for national and state employment business cycles. These results are virtually identical to our findings for personal income. Just as in the income case, most states have peak correlations at a zero lead/lag, while Michigan and Indiana weakly lead the nation by one quarter. With the exception of Iowa, there is little difference in the relative persistence and volatility of District state business cycles measured using income and employment (see table 1 for details). In contrast, Iowa's employment business cycle is more persistent and less volatile than its income business cycle. This result reinforces our argument that differences in Iowa's income business cycle largely reflect differences in income due to price fluctuations, since fluctuations in employment reflect fluctuations in the volume rather than the value of production.

Are national, District, and state business cycles alike?

The previous section established that there is a relatively high synchronization of national and District state aggregate business cycles. Does this mean that national and state business cycles are alike? The answer to this question is no. When Lucas (1977) and other business cycle researchers talk about business cycles being alike, they are referring to similarities in the way that cyclical fluctuations of activities within a region relate to the aggregate cycle of the region. The goal of this section is to compare national and subnational business cycles along as many within-region dimensions as we can.

Considerable effort has gone into understanding the dynamics of national income and various expenditure aggregates at the national and international levels. Therefore, any analysis of regional business cycles would be lacking if it did not provide analogues to these widely known facts about national business cycles. We begin with investment activity.

Investment

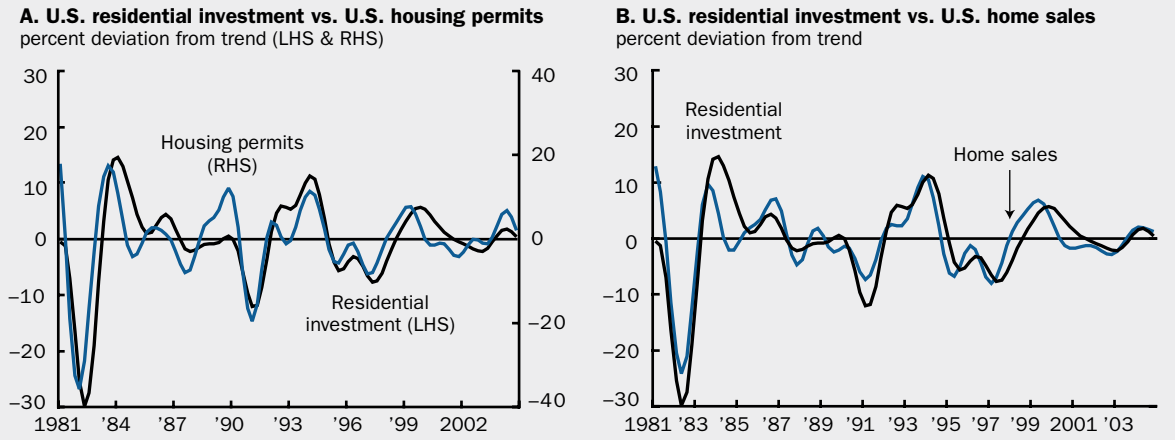
Investment expenditure is not available at the state level, so we must explore other indicators of investment activity. These data are limited to residential investment, since there are no state measures of business investment. Our residential investment indicators are the number of construction permits for housing and the number of home sales (new and existing). Figure 3 reveals that these indicators have similar cyclical properties to residential investment.

It is well known from the work of Fisher (2001) and others that residential investment tends to lead business investment over the business cycle. Turning to panel A of figure 4, we see the cross-correlation function of U.S. business investment from time $t - 8$ to $t + 8$ and national income has a peak correlation of 0.93 with the national income cycle at a zero lead/lag. The cross-correlation function for national residential investment and income, on the other hand, leads national income by three quarters with a peak correlation of 0.83. Panel B of figure 4 reveals that the cross-correlation functions for both of our indicators of residential investment with income have a similar shape to that of residential investment, with a peak correlation of 0.58 occurring at a lead of four quarters for housing permits and a peak correlation of 0.63 occurring at a lead of three quarters for home sales.

In figure 5, cross-correlation functions for District-level housing permits and income, as well as District home sales and income, are shown to have the same shape as the national level data, including peak correlations of 0.47 and 0.69, respectively, at a lead of three

FIGURE 3

U.S. residential investment



Notes: LHS means left-hand scale. RHS means right-hand scale.

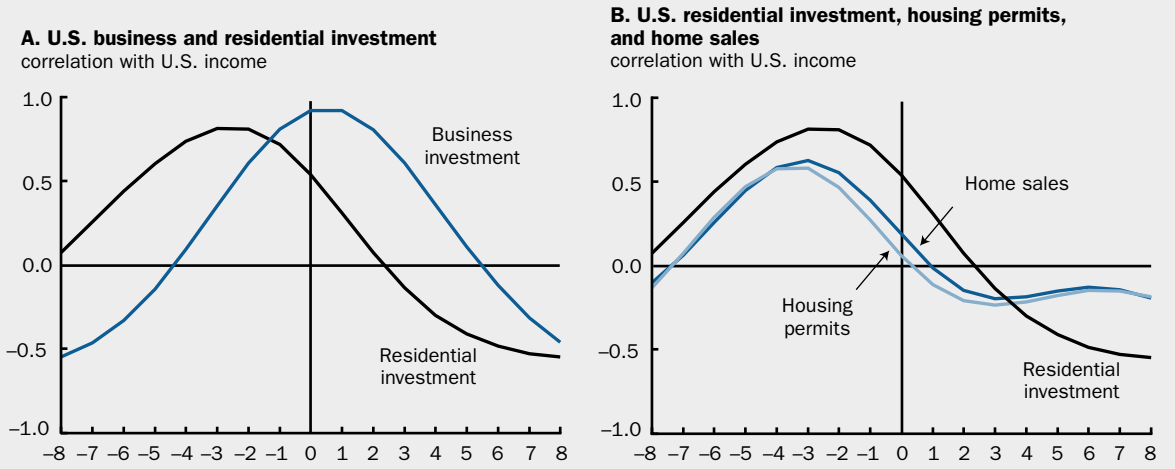
Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis, U.S. Census Bureau, and National Association of Realtors from Haver Analytics.

quarters. Results for state-level data echo the national and District-level data, with statistically significant positive correlations for leading housing permits and home sales and insignificant correlations lagged of housing permits and home sales. The peak correlations between state-level income and housing permits and state-level income and home sales are similar in magnitude to the District-level results and vary from leads of two to four quarters. These findings suggest that housing permits and home sales are reliable leading indicators of regional cyclical income fluctuations.

The persistence of housing permits and home sales cycles is reported in table 3, panel A. These estimates suggest that there is very little variation across the region in terms of persistence. Panel B of table 3 reports the volatility of housing permits and home sales. The top row of this panel reports the absolute volatility of regional income, or the standard deviation of the

FIGURE 4

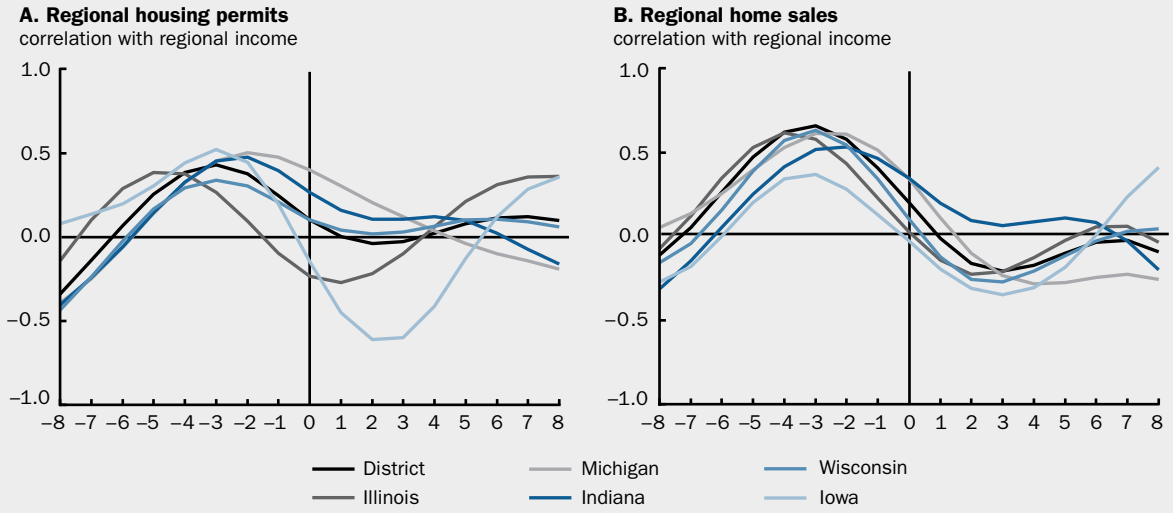
Within region: Investment



Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis, U.S. Census Bureau, and National Association of Realtors from Haver Analytics.

FIGURE 5

Within region: Housing permits and home sales



Sources: Authors' calculations based on data from the U.S. Census Bureau and the National Association of Realtors from Haver Analytics.

percentage deviation of the cycle from its trend. The following rows report the relative volatility of an activity, which is the ratio of the standard deviation of the cycle of the regional activity to the standard deviation of the region's income cycle. Values above one imply the activity is more volatile than the region's income cycle. The U.S. column reveals that housing permits have a relative volatility of 7.71, which is slightly higher than the relative volatility of residential investment, while home sales, with a relative volatility of 4.80, are less volatile than both residential investment and housing permits. Moving along the housing permits and home sales rows, we see that the District states have the same pattern.

Consumption

Countless business cycle studies have documented that national consumption expenditure measured by personal consumption expenditure exhibits strong positive comovement with national income. Personal consumption expenditure is not available at the state level, so we use a measure of retail sales derived from retail sales tax revenue, which is available at the state level, as a proxy for consumption spending. We plot the cyclical components of these national time series in figure 6. National retail sales tax revenue is roughly twice as volatile as national consumption expenditure (see table 3, panel B). Along all other cyclical dimensions, these series appear to be very similar.

Panel A of figure 7 reveals that the cross-correlation function of U.S. retail sales tax revenue and income has an identical shape to that of U.S. consumption and income, including a peak correlation of 0.87, with a one quarter lead. This suggests that retail sales tax revenue is a reliable proxy for consumption.

Turning to the District-level data in panel B of figure 7, we find that although the relationship is weaker, there is a relatively strong positive relationship between retail sales tax revenue and District income, with the peak correlation of 0.66 occurring at a zero lag/lead. The cross-correlation functions of Michigan and Illinois display a similar pattern to the District. However, the peak correlation is somewhat lower for Illinois at 0.52. This suggests that retail sales tax data are a coincident indicator of business cycles of the larger District states. For the remaining states, Iowa, Indiana, and Wisconsin, the correlation between cyclical fluctuations in state income and state retail sales tax revenue is positive at a zero lead/lag, but it is close to zero and not statistically significant. Panel A of table 3 reveals that the persistence of the state retail sales tax revenue cycles is less than that of the national cycle, with a range of 0.81 for Indiana to 0.91 for Illinois. With the exception of Wisconsin, the state retail sales tax revenue cycles are roughly twice as volatile as the state income cycles, which are slightly more volatile than the national cycle (see table 3, panel B). Wisconsin's retail sales tax revenue cycle has a relative volatility of eight. Overall, these results suggest that over the business cycle, Michigan's and Illinois's consumption and income behave in a similar

TABLE 3

Within region

	U.S.	District	Illinois	Michigan	Indiana	Wisconsin	Iowa
A. Persistence							
Income	0.93	0.93	0.92	0.94	0.92	0.93	0.89
Business investment	0.94						
Residential investment	0.94						
Existing home sales	0.87	0.88	0.87	0.88	0.88	0.86	0.87
Housing permits	0.86	0.86	0.84	0.87	0.87	0.85	0.85
Personal consumption	0.94						
Retail sales tax revenue	0.92	0.87	0.91	0.89	0.81	0.83	0.86
Nonfarm payroll employment	0.94	0.93	0.92	0.93	0.93	0.94	0.93
Average hours in manufacturing	0.89	0.88	0.86	0.89	0.90	0.90	0.86
Initial unemployment claims	0.91	0.91	0.92	0.90	0.91	0.91	0.86
Unemployment rate	0.92	0.90	0.90	0.92	0.89	0.90	0.90
Real wage	0.87	0.89	0.87	0.90	0.82	0.84	0.84
Capital utilization (industrial sector)	0.87	0.87					
B. Volatility							
Income	1.49	2.00	1.54	2.90	2.20	1.64	2.67
	Relative volatility						
Business investment	3.21						
Residential investment	6.79						
Existing home sales	4.80	5.15	5.86	3.87	5.64	7.00	5.05
Housing permits	7.71	10.36	12.10	8.91	10.10	11.56	7.15
Personal consumption	0.79						
Retail sales tax revenue	1.87	1.64	2.31	2.11	2.70	8.12	2.31
Nonfarm payroll employment	0.79	0.70	0.72	0.70	0.78	0.80	0.45
Average hours in manufacturing	0.56	0.53	0.57	0.77	0.69	0.47	0.43
Initial unemployment claims	7.43	7.58	7.83	6.76	8.24	9.09	5.04
Unemployment rate	0.43	0.38	0.45	0.40	0.36	0.48	0.14
Real wage	0.74	0.69	0.93	0.65	0.59	0.71	0.36
Capital utilization (industrial sector)	1.56	1.82					

Note: In panel B, income is measured in absolute volatility for the U.S., the District, and District states.
Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis, U.S. Bureau of Labor Statistics, U.S. Department of Labor, U.S. Census Bureau, and National Association of Realtors, all from Haver Analytics, and the Board of Governors of the Federal Reserve System.

way to their national counterparts; this also applies to the District as a whole. No clear conclusions can be reached for the other states. In light of these results, we are working on improving the quality of our retail sales tax data for these states by doing a more complete correction for tax rate changes and harmonizing the accounting rules, so we consider the findings of this section to be incomplete at this time.

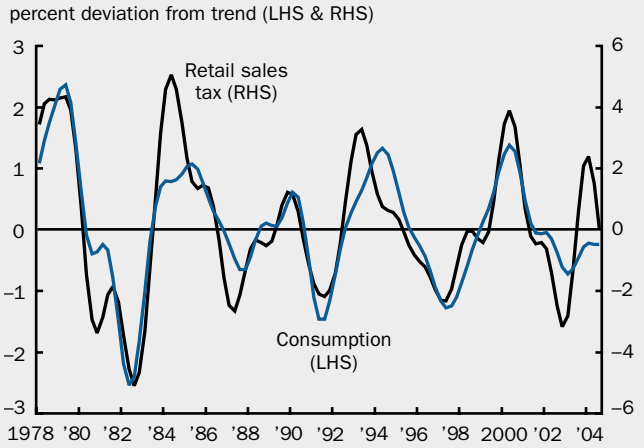
The remaining components of national expenditure are government spending and net exports. We are working on assembling data for state-level government spending, so we are unable to report those results at this time. Exports are the only component of net exports that are available at the state level, but our data samples are too small for any meaningful business cycle analysis, so we leave it to future research to explore the business cycle properties of exports at the state level.

Labor market

The cyclical behavior of labor market data has been the focus of many business cycle studies (see the surveys of Cooley, 1995; King and Rebelo, 1999). There is considerable overlap in labor market data available at the national and state levels. We explore five measures of labor market activity: total nonfarm

FIGURE 6

U.S. consumption vs. U.S. retail sales tax revenue



Notes: LHS means left-hand scale. RHS means right-hand scale.
Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis from Haver Analytics and the U.S. Census Bureau.

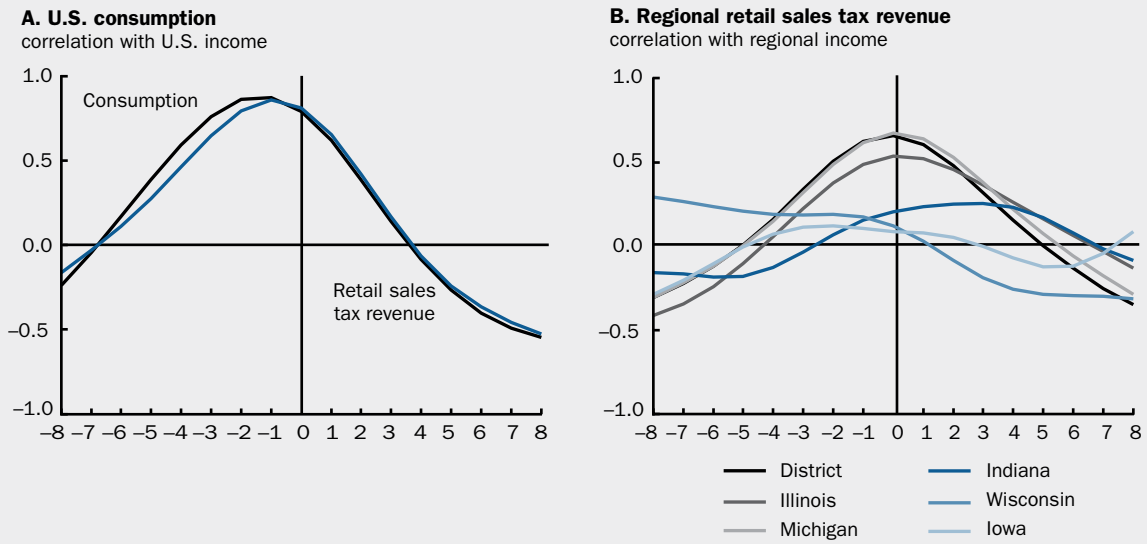
payroll employment, average weekly hours in manufacturing, initial unemployment insurance claims, the unemployment rate, and the real wage.

We find that both national measures of labor input—nonfarm employment and average hours—are positively correlated with national income over the business cycle (see figure 8). Total employment's peak correlation of 0.91 occurs at a lag of one quarter, while average hours has a peak correlation of 0.83 at a lead of two quarters. In contrast, both indicators of national unemployment—initial unemployment claims and the unemployment rate—have a negative correlation with national income over the business cycle. Initial unemployment claims is a leading indicator, with a minimum correlation of –0.86, which occurs at a lead of two quarters. The unemployment rate, which is well known to be highly countercyclical, has a minimum correlation with national income of –0.91, which occurs at a zero lead/lag, making it a coincident indicator. Finally, we find that the correlation of the real wage and income is not significantly different from zero at all leads and lags. All these findings are consistent with earlier business cycle studies of national labor market dynamics.

Figure 9, which features regional labor market data, reveals that the cross-correlation functions of District-level labor and income data have very similar shapes to their national counterparts in figure 8. There are some slight differences in the location of peak/minimum correlations. However, variables that were leading or lagging indicators at the national level continue to be leading or lagging indicators at the District level.

FIGURE 7

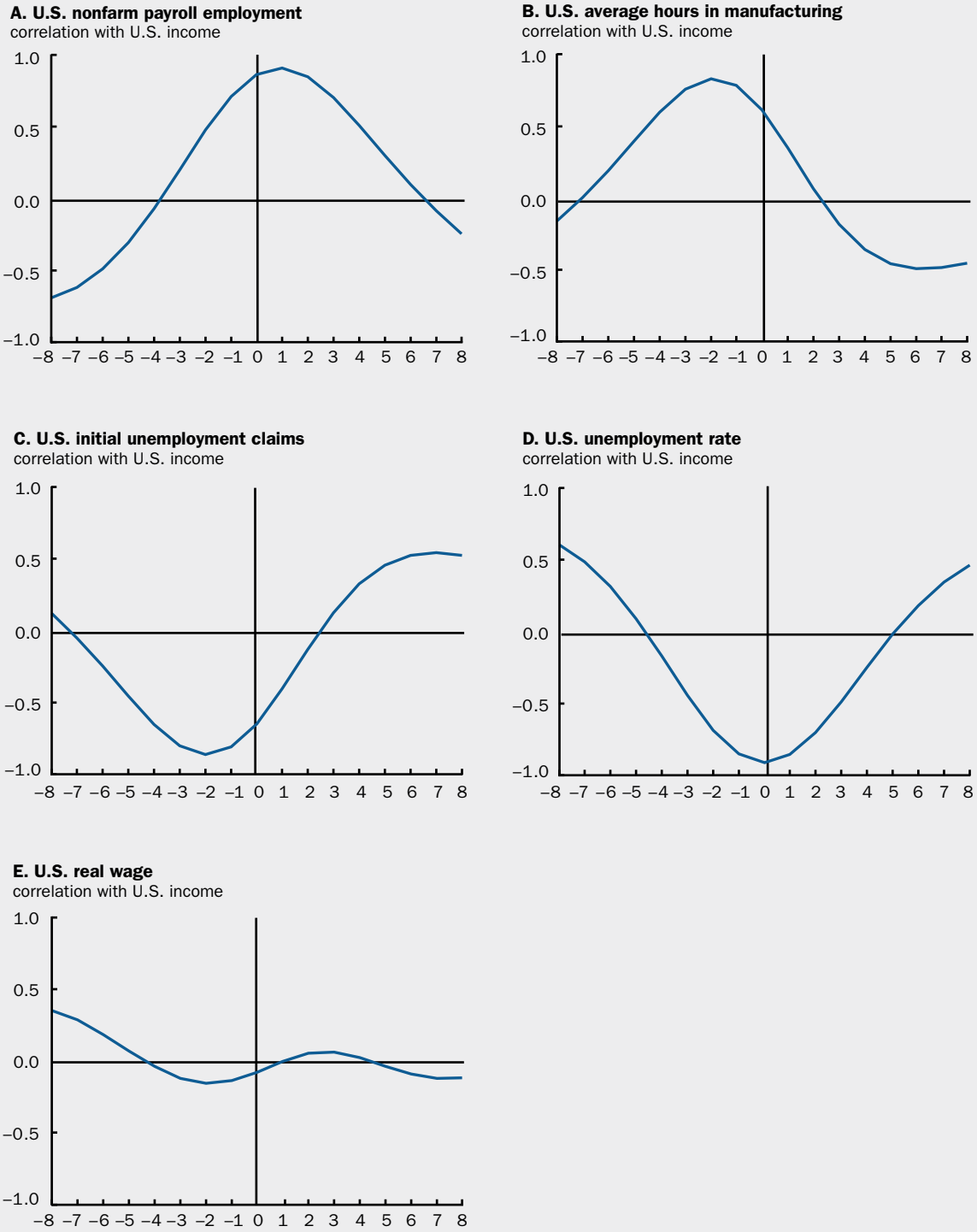
Within region: Consumption and retail sales tax revenue



Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis from Haver Analytics and the U.S. Census Bureau.

FIGURE 8

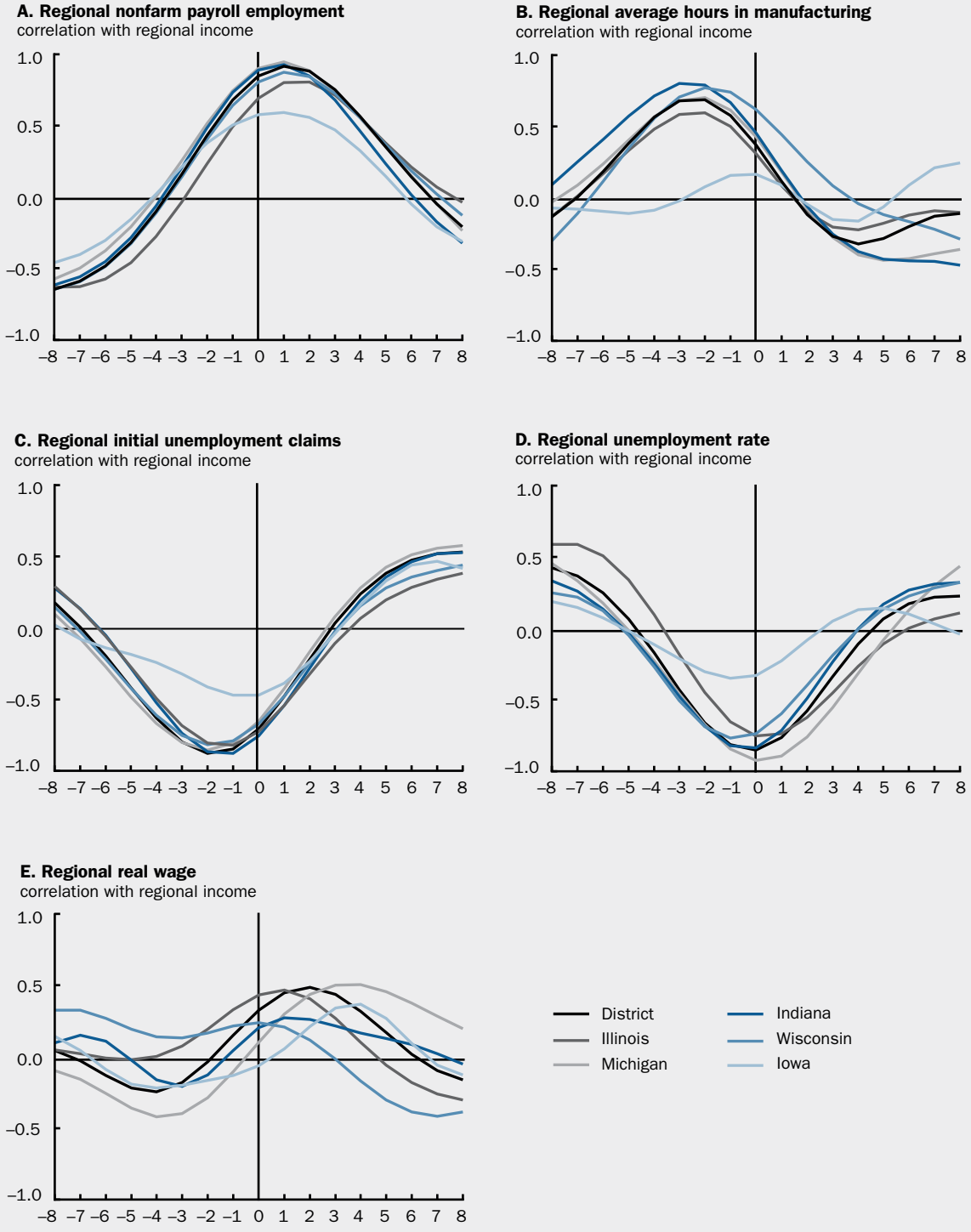
Within region: U.S. labor



Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis, U.S. Bureau of Labor Statistics, and U.S. Department of Labor from Haver Analytics.

FIGURE 9

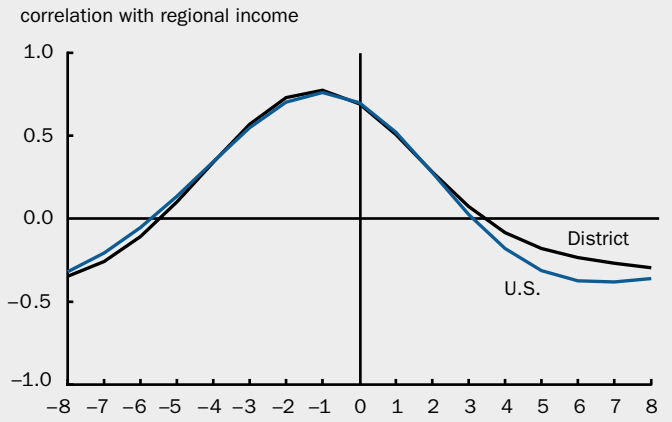
Within region: Regional labor



Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis, U.S. Bureau of Labor Statistics, and U.S. Department of Labor from Haver Analytics.

FIGURE 10

Within region: Industrial electricity consumption



Sources: Authors' calculations based on data from the U.S. Bureau of Economic Analysis from Haver Analytics and the Board of Governors of the Federal Reserve System.

Labor market data for Illinois, Indiana, Michigan, and Wisconsin also closely resemble the national labor market data. Total employment and average hours have significant positive correlations with income, while initial unemployment claims and the unemployment rate have significant negative correlations with income at similar leads and lags as those for the District and nation. We also find that the correlation of the real wage and income is close to zero at all leads and lags. In the case of Iowa, estimates of the correlation between labor market activity and income over the business cycle are less precise than for the other states. The most obvious example is the cross-correlation function of average hours and income for Iowa, which is insignificant at all leads and lags.[7]

Panel A of table 3 (p. 54) reveals that the state labor market fluctuations and national labor market fluctuations show similar persistence. All measures are highly persistent, with the real wage displaying the lowest persistence and employment the highest persistence.

With the exception of initial unemployment claims, cyclical fluctuations in labor market variables at both the national and state levels are less volatile than for income (see table 3, panel B). In most cases, employment is more volatile than average hours and unemployment. There is no clear ranking for the real wage. The initial unemployment claims category, on the other hand, is considerably more volatile than income, with a relative volatility of 7.43 at the national level and between 5.04 and 9.09 for the District states.

Overall, these findings suggest that the cyclical properties of District state labor markets are virtually identical to those of the national labor market.

Capital utilization

The final dimension of comparison is capital utilization. Measures of capital utilization, such as industrial electricity usage, are only available at the national and District levels. We find at the national level that electricity usage/capital utilization has a strong positive relationship with income (see figure 10). The peak correlation for these series (0.76) occurs with a lead of one quarter. The District cross-correlation function has an almost identical shape to the nation's, including a peak correlation of 0.77, which occurs at a lead of one quarter. These results suggest that electricity usage is a reliable leading indicator of District-level cyclical activity.

Conclusion

Our within-region results suggest that we can expand Lucas's (1977) comment about business cycles to read that irrespective of the time period or *size of the economy*, business cycles are all alike. This is welcome news for business cycle theorists, since it suggests that models that have been developed or are being developed to explain national business cycles are applicable to subnational business cycles. Our results are also of interest to those charged with the formulation of U.S. monetary policy, since they imply that the U.S. and the Seventh District fit Mundell's notion of a single optimal currency area, which suggests that the "best" monetary policy for the U.S. is also the "best" monetary policy for the Seventh District and the District states.

NOTES

[1]Iowa is the only state that is wholly within the Seventh Federal Reserve District's boundaries, while the District includes the southern portions of Wisconsin and Michigan and the northern portions of Illinois and Indiana. Since data at the state level are based on state boundaries, our study includes parts of Wisconsin, Michigan, Illinois, and Indiana that are in other Federal Reserve districts.

[2]See Owyang, Piger, and Wall (2005), table 6, p. 615.

[3]See Burns and Mitchell (1946), p. 3.

[4]For detailed information on the discussion that follows, please see Burns and Mitchell (1946), pp. 72–76.

[5]This article follows Harding and Pagan (2002). Broad indicators of activity, such as GDP, typically arrive with a considerable lag and are only available on a quarterly basis, so researchers have followed Burns and Mitchell's approach to identifying business cycles by using techniques that extract common fluctuations from a potentially large cross section of disaggregate data that arrives more frequently than broad indicators. This is the thinking that underlies the popular diffusion index approach of Stock and Watson (1998). The outcomes from these exercises, such as the Chicago Fed National Activity Index (CFNAI), yield measures of the business cycle that are consistent with business cycles estimated using a single broad indicator, such as GDP.

[6]Interested readers may contact the authors for copies of these additional figures and tables.

[7]Labor market data are widely used in studies of regional business cycles. For example, Crone and Clayton-Matthews (2005) use monthly total nonfarm payroll employment, average weekly hours in manufacturing, real wage, and the unemployment rate to estimate all 50 state-level business cycles using the Stock and Watson (1989) common factor approach. They find that these employment indicators are coincident indicators of the employment business cycle for four of the District states. The one exception is Illinois, for which average weekly hours in manufacturing is a leading indicator, with a lead of one month.

APPENDIX: DESCRIPTION OF DATA

The data on real U.S. gross domestic product, consumption, and investment come from the national accounts published by the U.S. Bureau of Economic Analysis (BEA).

For the regional level analysis, we use quarterly personal income data (published by the BEA). Personal income includes all sources of earnings, such as wages, interest and dividends, proprietor's income, and other miscellaneous labor income, by place of residence.

Since consumption is not directly available at the state level, we proxy for consumer expenditure by using retail sales tax revenue data. The data on retail sales tax revenue are published by the U.S. Census Bureau on a quarterly basis and provide the quarterly estimates of sales and gross receipts taxes on goods and services for individual states.

To proxy for residential investment at the state level, we use the data on housing construction permits and home sales. The data set on construction permits covers new privately owned housing units and is collected and published by the U.S. Census Bureau. We also use data on the number of home sales, including new and existing privately owned single-family houses, condominiums, and cooperative housing; these data are published by the National Association of Realtors.

Our measure of capital utilization is electricity power use, by industry, which is published by the Board of Governors of the Federal Reserve System as part of its industrial production and capacity utilization data release, Federal Reserve Statistical Release G.17. The data are limited to national and District measures of electricity power use, by industry.

Our labor market indicators include data on initial unemployment claims (that is, the number of people filing new claims for state unemployment insurance), average weekly hours in the manufacturing sector, and total nonfarm payroll employment. These data are recorded by the local state governments and published by the U.S. Department of Labor. Also included in our data set is the unemployment rate, which represents the fraction of the labor force that is unemployed. These unemployment rate series are published monthly by the U.S. Bureau of Labor Statistics. Our measure of nominal wages is total wage and salary disbursements, by place of work, from the BEA's personal income data divided by the product of total nonfarm payroll employment and average weekly hours in manufacturing.

All nominal data are deflated by the national Personal Consumption Expenditures Price Index to yield real consumption based measures of economic activity. We explored alterative approaches that deflated nominal data by either the Midwest Census Region Consumer Price Index or metropolitan Consumer Price Indexes. These approaches yield virtually identical results to those reported in this article.

REFERENCES

Backus, D. K., and P. J. Kehoe, 1992, "International evidence of the historical properties of business cycles," *American Economic Review*, Vol. 82, No. 4, September, pp. 864–888.

Burns, A. F., and W. C. Mitchell, 1946, *Measuring Business Cycles*, New York: National Bureau of Economic Research.

Christiano, L. J., and T. J. Fitzgerald, 2003, "The band pass filter," *International Economic Review*, Vol. 44, No. 2, May, pp. 435–465.

Clayton-Matthews, A., and J. H. Stock, 1998–99, "An application of the Stock/Watson index methodology to the Massachusetts economy," *Journal of Economic and Social Measurement*, Vol. 25, No. 3/4, pp. 183–233.

Cooley, T. F. (ed.), 1995, *Frontiers of Business Cycle Research*, Princeton, NJ: Princeton University Press.

Crone, T. M., and A. Clayton-Matthews, 2005, "Consistent economic indexes for 50 states," *Review of Economics and Statistics*, Vol. 87, No. 4, November, pp. 593–603.

Fisher, J. D. M., 2001, "A real explanation for heterogeneous investment dynamics," Federal Reserve Bank of Chicago, working paper, No. WP-2001-14.

Harding, D., and A. Pagan, 2002, "Dissecting the cycle: A methodological investigation," *Journal of Monetary Economics*, Vol. 49, No. 2, March, pp. 365–381.

Hess, G. D., and K. Shin, 1998, "Intranational business cycles in the United States," *Journal of International Economics*, Vol. 44, No. 2, April, pp. 289–313.

King, R. G., and S. T. Rebelo, 1999, "Resuscitating business cycles," in *Handbook of Macroeconomics*, Vol. 1B, M. Woodford and J. B. Taylor (eds.), Amsterdam: Elsevier.

Kouparitsas, M. A., 2002, "Understanding U.S. regional cyclical comovement: How important are spillovers and common shocks?," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 26, No. 4, Fourth Quarter, pp. 30–41.

Lucas, R. E., 1977, "Understanding business cycles," *Carnegie-Rochester Series on Public Policy*, Vol. 5, pp. 7–29.

Mundell, R. A., 1961, "A theory of optimum currency areas," *American Economic Review*, Vol. 51, No. 4, September, pp. 657–665.

Orr, J., R. Rich, and R. Rosen, 1999, "Two new indexes offer a broad view of economic activity in the NewYork–New Jersey region," *Current Issues in Economics and Finance*, Federal Reserve Bank of New York, Vol. 5, No. 14, October, pp. 1–6.

Owyang, M. T., J. Piger, and H. J. Wall, 2005, "Business cycle phases in U.S. states," *Review of Economics and Statistics*, Vol. 87, No. 4, November, pp. 604–616.

Stock, J. H., and M. W. Watson, 1998, "Diffusion indexes," National Bureau of Economic Research, working paper, No. 6702, August.

_____, 1989, "New indexes of coincident and leading economic indicators," in *NBER Macroeconomic Annual 1989*, O. J. Blanchard and S. Fischer (eds.), Cambridge, MA: MIT Press, pp. 351–394.