Federal Reserve Bank of Chicago

Call for Papers

2001 Conference on Bank Structure and Competition

Fourth Quarter 2000

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.

Single-copy subscriptions are available free of charge. Please send requests for single- and multiple-copy subscriptions, back issues, and address changes to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, telephone 312-322-5111 or fax 312-322-5515.

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

ISSN 0164-0682

WWW.FRBCHI.ORG

Contents

Fourth Quarter 2000, Volume XXIV, Issue 4

A record current account deficit: Causes and implications

Jack L. Hervey and Loula S. Merkel

Introduction and summary

The U.S. deficit in international trade soared to new heights in 1998, again in 1999, and in all likelihood, will increase even further this year. Mirroring these deficits have been huge foreign capital inflows. In 1999, the U.S. current account deficit—that is, the difference between exports and imports of goods, services, receipts and payments of income from and to foreigners, and unilateral transfers—totaled $331 billion or 3.6 percent of nominal gross domestic product (GDP). This record deficit compares with the previous record of $217 billion (2.5 percent of GDP) in 1998 and $141 billion (1.7 percent of GDP) in 1997. The magnitude of the recent year-to-year increases in this deficit, as well as its absolute dollar size, has raised considerable concern among many public and private observers of the U.S. economy. Not since 1987, when the current account deficit peaked at a then record $161 billion, has the condition of the U.S. international accounts so captured the attention of economists, policymakers, and the popular press.

Further compounding uneasiness about the current situation is the expectation by many economists that the magnitude of the trade deficit will show a further increase this year and that only a modest reduction, if any, is likely in 2001. Indeed, trade developments thus far in 2000 indicate that at least the first half of that expectation (that is, an increase in the year-to-year size of the deficit during 2000) will be borne out. There are also fears surrounding an eventual economic adjustment—"the current account gap … is the single biggest threat to the current expansion of the economy."[1]

There is nothing inherently "bad" (or "good") about a current account deficit—or for that matter, a current account surplus. However, the concern about the deficit that has drawn the attention of reasonable observers centers on a specific issue: Does the deficit in the U.S. international accounts represent a risk to our economic well-being in the near term or in the longer term? To answer this question, we need to identify the underlying cause of the deficit. What developments during the past two or three years—in the domestic economy and in the rest of the world—have led the U.S. to purchase dramatically more goods and services from abroad than it sold abroad? Furthermore, can the U.S. economy maintain a deficit of this magnitude? And, if not, what are the likely implications of an adjustment for the U.S. economy?

Three rationales are commonly used to explain the sudden and dramatic increase in the U.S. current account deficit. The first rationale contends that U.S. consumers have shifted their preferences from saving for the future—witness the near zero personal savings rate—toward purchasing more consumption goods in the present.[2] This surge in demand for domestic consumption goods translates into a corresponding increase in imported consumption goods. We call this the *consumption boom* hypothesis. Certainly, trade in consumer-type goods has increased in recent years. Indeed, more than 60 percent ($52 billion) of the year-to-year increase in the goods trade deficit between 1998 and 1999 was accounted for by the year-to-year increase in consumer goods, foods and beverages, and automotive

Jack L. Hervey is a senior economist at the Federal Reserve Bank of Chicago. Loula S. Merkel, formerly an associate economist at the Federal Reserve Bank of Chicago, is currently with McKinsey & Company, Inc., United States/Chicago. *The authors especially wish to express their appreciation to Michael Kouparitsas and David Marshall for their helpful comments and suggestions and to the seminar participants at the Federal Reserve Bank of Chicago for their interest and suggestions.*

imports (most of which are broadly classed as consumer goods). If the consumption boom story is true, it implies that there has been excessive borrowing from abroad to finance a domestic consumption binge. And according to this argument, since this borrowing has not gone toward enhancing productivity, the economy will be forced to suffer a decline in consumption in the future as resources are diverted away from production for domestic use toward production to service the foreign debt.

A second hypothesis suggests that the financial/exchange rate crises in Asia, Russia, and Brazil from mid-1997 through early 1999 contributed to a "safe haven" inflow of short-term foreign capital into U.S. markets.[3] Briefly, the idea here is that the flight of capital from the foreign economies takes away from the productive and consuming capacity of those economies; it not only detracts from the capacity of their domestic economies to perform, but it also reduces their capacity to import from foreign markets, namely, the U.S. From the U.S. perspective, this flight of foreign capital into the economy does two things—it makes it more difficult for the U.S. to export goods and services to these now poorer performing foreign markets and it facilitates (makes cheaper, in terms of dollars) the U.S. importation of goods and services from these countries. Thus, other things remaining the same, the U.S. current account deficit increases. We call this the *safe haven* hypothesis. The concern implicit within this explanation for the capital inflow is that economic recovery and increased stability abroad might result in an abrupt and substantial outflow of short-term capital, with resulting disruption in U.S. financial markets.

A third potential explanation for the recent rapid increase in the current account deficit is associated with the technological restructuring of the U.S. economy. This hypothesis implies that a technology shift in the economy (largely related to the assimilation of advances in computer and communication technology) has increased the level of productivity, and returns on investment, in the economy. Demand for investment has increased in response to this technology shift, which in turn has stimulated the inflow (supply) of foreign capital in support of this new type of investment. We call this the *technological change* hypothesis. There is less concern about an eventual adverse adjustment in the economy in this case, because this hypothesis implies that productivity-enhancing investment will result in increased output in the economy, thereby facilitating the servicing and eventual repayment of the increased level of borrowing from abroad (the larger trade/current account deficit).[4]

Before we can examine the relationship between the international accounts and the domestic economy, we need to understand how these international transactions work. In the next section, we set out a simple framework for understanding these relationships, based on national income accounting identities. We then review the three hypotheses outlined above, which seek to explain the recent rapid increase in the current account deficit/capital inflow, and analyze how well they match the available evidence. Finally, we consider whether the deficit is sustainable and, if not, what the implications of each hypothesis might be for an eventual adjustment in the U.S. economy.

We find little support for the consumption boom explanation in the data. While consumption has increased, its share of total expenditures has declined. We find some evidence to support the safe haven rationale for the increase in capital inflows. However, because much of the capital inflow appears to represent long-term investment rather than a short-term flight to safety, we do not find the implications of this story to be particularly worrisome for the health of the U.S. economy. In other words, our view is that an unwinding of such capital inflows is unlikely to be overly disruptive to domestic financial markets. Finally, we find the technological change argument to have some merit. Much of the recent increase in goods imports has been in the "investment" goods categories—capital equipment, intermediate capital equipment components, and industrial supplies used in the production of capital goods. Recent gains in productivity measures and continuing structural changes across the spectrum of U.S. industry suggest that the economy may be shifting to a new and higher level of potential output. An economy in the process of such a shift has an incentive to increase borrowing from abroad to fulfill the increased demand for investment. We believe that the available data on the current U.S. economic environment fit well with this scenario.

International trade—the current account and the capital account

One can think of trade in the context of individual decision-making. Trade is a result of conscious and voluntary decisions made by individuals, firms, and public institutions. Any individual faces a budget in which current expenditures are constrained by current income and the ability to borrow. More specifically, allowing subscript t to represent the current period, the individual's flow budget constraint is:

1) $y_t + r_t a_t = c_t + i_t + (a_{t+1} - a_t)$,

where y is income, c is expenditure on the consumption of goods and services, i is investment expenditure, r is the interest rate, and a is net assets, which could be positive or negative. If a_t is greater than zero, then the individual is a net creditor. If a_t is less than zero, then the individual is a net debtor. The term $r_t a_t$ represents income or debt payments, depending on the sign of a_t. What this accounting relationship says is that one's current income is distributed over one's current consumption and savings with any shortfall resulting in an increased net liability in the next period $(a_{t+1} - a_t)$.

At the national level, the individual budget equation we presented above still holds, but some changes in notation are useful to see how the collective individual decisions—the sum of which are the national private budget decision—are related to the international economy. Allowing uppercase letters to represent the sum of the individuals' variables, the national budget constraint, or equivalently, gross national product, is:

2) $Y_t + r_t A_t = C_t + I_t + (A_{t+1} - A_t)$.

In a closed economy, that is, a country not open to foreign trade, the national debt must be zero, that is, $A_t = 0$. This means that the sum of all borrowers' debt must exactly offset the sum of all lenders' funds. In other words, the current expenditures of a country as a whole are constrained by its current income—it cannot borrow. In contrast, an economy that is open to international trade has the option of financing its aggregate demand for consumption and investment by borrowing abroad, that is, $A_t < 0$. Similarly, an open economy can lend or invest abroad, taking advantage of a wider market and enhanced risk/return choices for its assets.

The nation's expenditures $(C_t + I_t)$ can either be spent on domestic goods consumed at home $(Y_t - X_t)$ or on imports (M_t):

3) $C_t + I_t = (Y_t - X_t) + M_t$.

This is the familiar national accounting identity:

4) $Y_t = C_t + I_t + X_t - M_t$.

Rearranging terms in equation 4, we can express the *trade balance* as:

5) $X_t - M_t = Y_t - C_t - I_t$.

By substituting the trade balance equation 5 into the budget equation 2 and putting the $r_t A_t$ term on the left side, we get

6) $(X_t - M_t) + r_t A_t = (A_{t+1} - A_t)$.

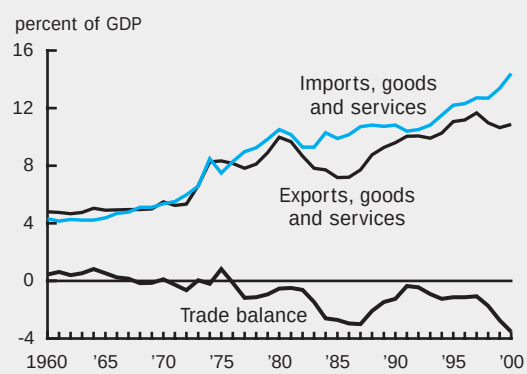
This equation represents a country's balance of international payments. The sum of the trade balance $(X_t - M_t)$ and net income receipts on net foreign assets (or net payments on liabilities) $(r_t A_t)$ is the *current account balance*. The change in the stock of foreign assets held by domestic individuals or firms is a *capital outflow,* and the change in the stock of domestic assets held by foreigners is a *capital inflow*, the net of which $(A_{t+1} - A_t)$ is the *capital account balance*.

Equation 6 shows that movements in the current account are matched by identical movements in the capital account. It also implies that if a country runs a current account deficit, there will be an increase in the stock of foreign liabilities in the next period to finance the difference (that is, foreign borrowing or a capital inflow). The interest rate is positive, so as A becomes a larger negative, the income balance in subsequent periods must be a larger negative, which leads to a larger current account deficit, all else equal. From here we can see that the increase in the stock of foreign debt, also referred to as the net international investment position (NIIP), can lead to an ever-increasing income deficit.

The recent behavior of these components of the U.S. balance of payments is the cause of the current concern. Data presented in figure 1 identify recent trends that have rekindled the debate about the "deterioration" in the U.S. international trade position. From 1992 through 1997, U.S. trade grew impressively, with the nominal value of exports of goods and services up 8.7 percent per year, on average, and imports up an even more robust 9.7 percent per year. The associated net export deficits for 1992 and 1997 rose from $36 billion to $106 billion, respectively, or from 0.4

FIGURE 1

U.S. trade balance



Source: Bureau of Economic Analysis.

percent to 1.1 percent of GDP. During the most recent two and a half years, however, the differential in the growth rates between exports and imports increased markedly, resulting in a dramatic increase in the trade deficit. Export growth slowed to an average of 1.3 percent per year, in the face of weak economic conditions in foreign markets, while average import growth held at a relatively strong 8.6 percent rate—stimulated by the robust domestic expansion. The trade deficit for 1999 stood at $254 billion and its share of GDP rose sharply, to 2.7 percent—approaching the record 3.0 percent reached in 1987. In the first half of 2000, the deficit rose further to $348 billion—at an annual rate, a level equivalent to 3.5 percent of GDP.

Figure 2 shows the behavior of the current account balance. The trade balance is the largest component of the current account; however, it is not the only item causing concern. Since 1998, the income balance has been in deficit; in 1999 it showed a deficit of 0.2 percent of GDP. This figure means that the income paid to foreigners on their holdings of U.S. assets exceeded the income received on U.S. assets held abroad. It is interesting to note that the nation's net international investment position (A_t) has been negative since 1986; however, only since 1998 has the income account been in deficit. This is because the rate of return on U.S. assets held abroad has historically exceeded the rate of return paid to foreign investors holding U.S. assets.[5]

The final category of international transactions included in the current account, which we neglected in our previous description, is unilateral transfers—government and private. These transactions are mostly in the form of U.S. government grants and aid to foreign countries and international institutions. On net, they have almost always constituted an outflow of funds from the U.S. The U.S. has recorded a net unilateral transfers inflow only in one year since World War II; this was in 1991 as a result of foreign governments' contributions to the U.S. for the Persian Gulf War.

As shown above, the capital account reflects the net acquisition or sale of assets by U.S. and foreign parties. Asset changes include changes in official assets (international reserves), the net outflow of funds from U.S. financial institutions, and direct investment abroad from U.S. firms and individuals. They also include the net inflow of funds from foreign financial institutions and the inflow of direct investment funds into the U.S. from foreign firms and individuals. Apart from "statistical discrepancy," the current and capital/financial account balances are always equal. Figure 3 shows the components of the capital account. The sum of all historical capital/financial account transactions equals the nation's NIIP. As mentioned earlier, the U.S. is currently a net foreign debtor. In 1999, foreign debt represented 15.9 percent of GDP.

Savings and investment imbalance

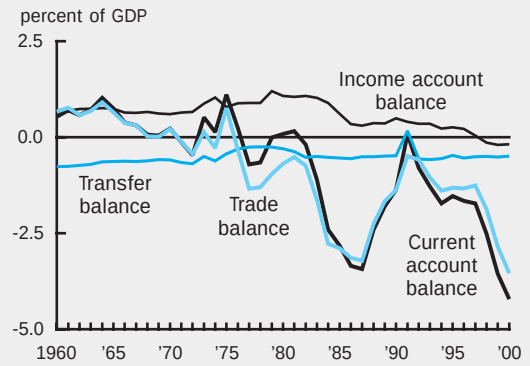
Looking again at the structural framework of the national accounting identities, we can examine how the equations presented above relate to the linkages between a country's domestic and international transactions. Equations 2 and 6 can be rearranged as:

$$7)\quad [Y_t + r_t(A_t) - C_t] - I_t = (A_{t+1} - A_t) = (X_t - M_t) + r_t A_t.$$

A country's inflow of foreign capital (debt) can be thought of as an import of foreign savings. A country's current income, less its current expenditures, equals its savings (S_t). It is easy to show that national savings less national investment is equivalent to the capital

FIGURE 2

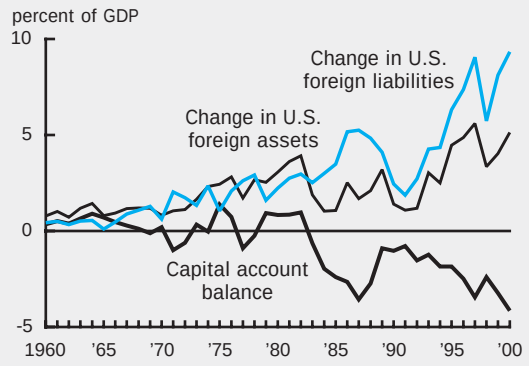
Components of current account



Source: Bureau of Economic Analysis.

FIGURE 3

Components of capital account



Source: Bureau of Economic Analysis.

account balance, which in turn equals the current account balance:

8) $S_t = [Y_t + r_t (A_t) - C_t]$.

9) $S_t - I_t = (A_{t+1} - A_t) = (X_t - M_t) + r_t (A_t)$.

This means that if a country's citizens in the aggregate decide to invest more than their available savings, then the country will run a current account deficit. This is matched by an increase in the stock of foreign debt. Thus, the current account deficit, representing the shortfall of domestic savings, will be financed by the net importation of foreign savings.[6]

To better understand this point, consider that when a country's aggregate demand for goods and services exceeds the aggregate domestic supply, it runs a trade deficit. Similarly, when aggregate supply outstrips aggregate demand, the country posts a trade surplus. This is what some economists term the "safety valve" quality of international trade. In the absence of the ability to trade with other countries, an excess aggregate demand situation would tend to bid up domestic prices. The trade balance, therefore, whether in deficit or surplus, is simply the residual of a country's aggregate demand and supply.

If a country has a trade deficit, it must finance it through foreign borrowing.[7] This description of market transactions, however, may give the misleading impression that debt is a sole consequence of trade flows. In reality, a country's net debt position is also a function of the relative risk and return preferences of investors. Foreign investors may want to invest in U.S. assets because they expect a higher risk-adjusted return than they might get at home or in a different country. As risk/return profiles around the world change, so do relative capital flows.

The relationship between a country's investment demand and the supply of domestic and foreign savings is shown in figure 4. In this graph of the market for loanable funds, the downward sloping demand curve for investment funds and the supply schedule of funds available for investment, or savings, are equilibrated by the interest rate. We see that the excess of domestic investment over savings is made up by foreign savings, or the current account deficit.

The size of an economy relative to the world economy is critical in determining the impact of changes in its investment demand and supply of funds. For example, a small economy in a world characterized by a high degree of capital mobility can have only a minuscule impact on the world interest rate. It essentially faces a perfectly elastic supply curve for funds; that is, any change in its domestic investment demand has little, if any, effect on the price it must pay for funds. By contrast, a large economy like the U.S., whose liabilities account for around one-third of the assets of the major OECD (Organization for Economic Cooperation and Development) countries,[8] has the power to influence the world interest rate. If, for example, there was an increase in demand for investment in the U.S., all else equal, the world interest rate would rise. Similarly, if for some exogenous reason, U.S. residents chose to save more (for example, due to a change in the demographic distribution over time), the world supply of loanable funds would increase and the world interest rate would fall.

Which hypothesis best fits the data?

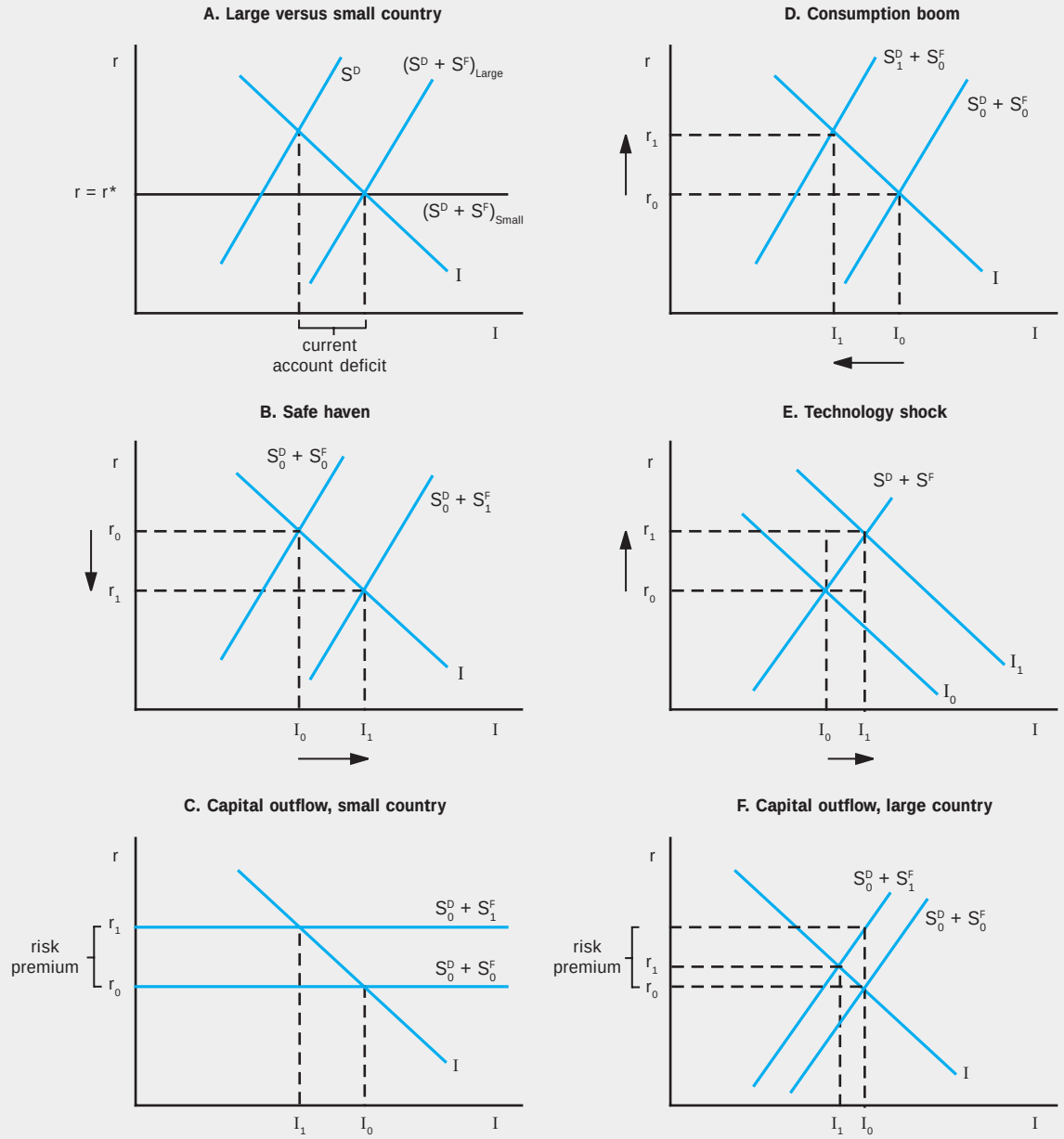
Using the framework we developed in the previous section, we briefly review our three hypotheses and examine to what extent they are supported by the data. While these hypotheses are not mutually exclusive, they do imply very different effects on interest rates and investment.

To many observers, the current boom in the U.S. economy is clearly manifested in a surge in personal consumption activity. U.S. consumers are purchasing more goods, including imports. At the same time, the personal saving rate has fallen to nearly zero (indeed, in July 2000 it was estimated at –0.2 percent).[9] These indicators lead to the view that increased consumption (and decreased saving) has led to the record trade deficit and capital inflows. This view translates into a shock to U.S. consumer tastes and preferences in favor of current over future consumption. This is very different from an increase in consumption associated with an increase in wealth. In the present situation, consumption is thought to be increasing at the expense of savings, regardless of income. In the context of the loanable funds market in a large economy, this would be reflected in an inward shift in the supply curve for funds, all else equal. From figure 4, we would expect this scenario to lead to a decline in investment and a rise in the interest rate.

The second potential explanation for the current account deficit and record capital inflows contends that the financial crises experienced by Asia, Russia, and Brazil from mid-1997 led to capital, domestic and foreign, fleeing these markets for the "safer" U.S market. This equates to the U.S. experiencing a beneficial shock to the perceived risk of its assets. Foreign capital (savings) flowed into the U.S. so more funds were available at any given interest rate. Again, all else equal, this would result in an outward shift in the supply curve for loanable funds, which would cause

FIGURE 4

Market for loanable funds



Notes: r* = world interest rate; S^D = domestic savings; S^F = foreign savings; I = investment.

interest rates to decline and investment to increase, as shown in figure 4. Note that this explanation is not inconsistent with an increase in the level of consumption, attributable to the increase in the quantity of investment leading to higher income.

The final explanation we consider argues that the U.S. is experiencing a positive technology shock, which has increased the economy's productivity and long-run level of potential output. The productivity of capital is higher and, hence, the incentives for investment are higher. As shown in figure 4, all else equal, this would correspond to an outward shift in the demand curve for investment and an increase in both interest rates and investment. Again, this explanation also justifies an increase in the level of consumption, since there is a wealth increase associated with the productivity shock.

Which of these explanations fits best with the recent behavior of the U.S. economy? Figure 5 shows that both gross domestic saving and investment as shares of GNP have been increasing since 1991. Investment however, has been increasing at a faster rate than national saving, resulting in the current account deficit.[10]

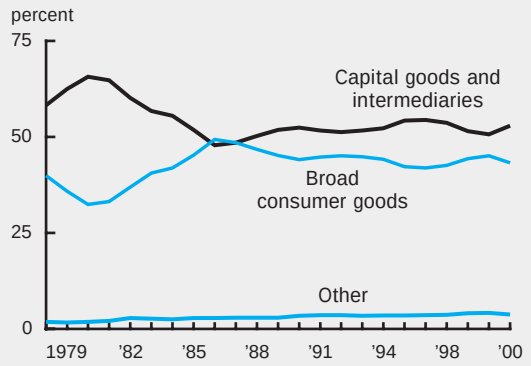
Immediately, we see that the behavior of investment does not support the consumption boom argument. We also find other evidence that refutes the consumption boom story. Imports of consumption goods, broadly defined to include automotive products and food, did represent more than half of the increase in the goods–trade deficit between 1998 and 1999.[11] However, looking at this figure alone may be misleading. Consumer goods historically have been the largest component of the goods–trade balance. Looking at the behavior and composition of imports, we see that capital goods, including non-oil industrial supplies and materials, actually comprise a larger share of our total imports than consumer goods (figure 6). Furthermore, there has been no apparent increase in consumer goods' share of imports. Note that we still run a trade surplus overall in capital goods, although it has been declining since 1970. Furthermore, to the degree that our capital goods imports constitute inputs into the production of intermediate and final goods, these figures may be reflecting re-exports. Figure 7 plots the components of our trade balance in capital goods, and indicates that since 1991 the U.S. has been a net importer of computers and related equipment. These capital goods are generally associated with productivity-enhancing investment.[12]

Perhaps a clearer view of whether the widening of the current account deficit is a result of a consumption boom can be seen from consumption and investment shares of gross domestic purchases (which include imports and exclude exports). This measure is equivalent to the resource constraint described in equation 3. Figure 8 plots these ratios (with government spending removed). An increase in investment since 1990 (above its historical trend) is evident, while total consumption expenditures have actually been declining in relative terms. The ratios are in real (price-adjusted) values, so they capture the volume effects of increased investment. Based on these statistics, one could argue that while the U.S. continues to have a growing trade deficit in consumption goods, these imports, as well as imports of capital goods, are allowing the economy to reallocate highly employed, scarce domestic resources toward productive investment.

While the investment behavior data refute the consumption boom theory, they support both the safe

FIGURE 6

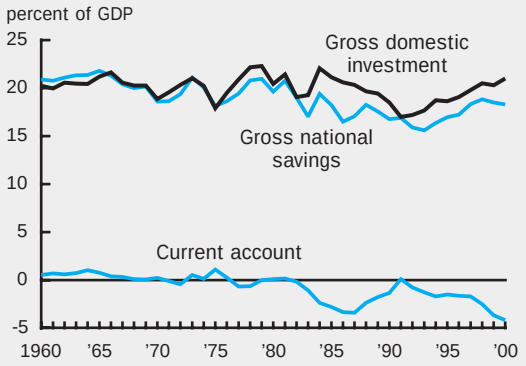
Share of total U.S. imports



Note: Broad consumer goods comprises consumer goods, foods, feeds and beverages, and automotive.
Source: Bureau of Economic Analysis.

FIGURE 5

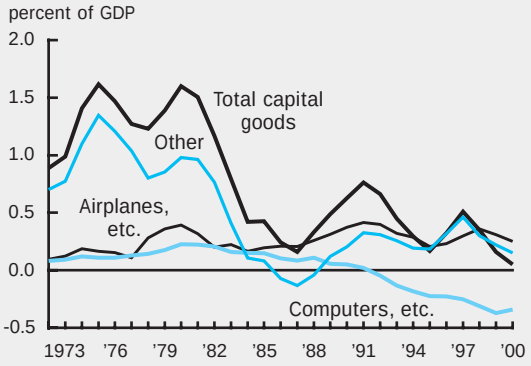
Savings and investment



Source: Bureau of Economic Analysis.

FIGURE 7

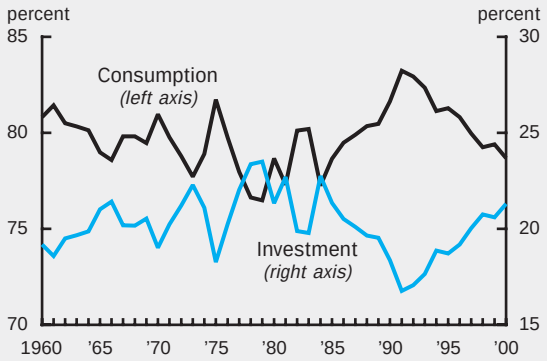
Capital goods trade balance



Source: Bureau of Economic Analysis.

FIGURE 8

Share of gross domestic purchases
(less government spending)



Source: Bureau of Economic Analysis.

FIGURE 9

Ten-year Treasury bond rate

Source: Federal Reserve Board of Governors.

haven and technology hypotheses. These two shocks would have opposite effects on interest rates, however. So how have U.S. interest rates behaved during the past two years? Figure 9 plots interest rates on the ten-year Treasury bond since 1990. From this figure we see that long-term interest rates declined from the middle of 1997 until late in 1998. From then until recently, U.S. long rates trended upwards. Obviously, these two effects did not operate in isolation—interest rates are affected by numerous developments. However, the "safe haven" story is supported by the decline in interest rates through 1998. During 1999, rates began to rise and capital flows and investment continued strong. This would indicate that increased domestic investment demand has been the dominating effect lately. So, while a combination of both the safe haven and technological change stories may have led to the record net capital inflows and current account deficit, it appears that technological change—increased U.S. demand for investment associated with the enhanced productivity of the economy—has dominated more recently.[13]

What would an adjustment mean?

With what we've learned about the likely sources of the current account and trade deficits (along with the corresponding capital flows), we now address whether these deficits are sustainable in the long term and, if not, what sort of adjustment the U.S. economy might ultimately undergo. If we consider the intertemporal qualities of the current account and capital account relationship, we can show that in the steady state, a trade deficit can be sustained as long as the growth rate of national income exceeds the rate of return paid on the nation's liabilities. Box 1 presents the details. How does the U.S. NIIP measure up against this sustainability requirement? Historically, the nominal growth rate of the U.S. economy has averaged around 7.4 percent per year.[14] The average nominal rate of return paid on U.S. foreign liabilities over the 1976 to 1999 period was around 5.5 percent. The 2 percentage-point difference between these rates implies that a trade deficit could be sustained.

The relationship also implies that for a given sustainable trade deficit, the U.S. can accumulate foreign liabilities up to some maximum level. Has this level been reached? If we consider that the average net exports share of GDP over the 1976 to 1999 period was –1.4 percent, and if we assume these averages to be the long-run values of these variables, then the corresponding long-run NIIP to GDP (or net foreign debt) position would be 70 percent. This level of indebtedness would likely not be desirable politically, given that the current NIIP to GDP ratio of 15.9 percent for 1999 is considered by some to be too large. Similarly, there is likely a foreign debt to GDP ratio above which foreign creditors would consider additional lending to the U.S. imprudent and, therefore, be less willing to purchase additional U.S. assets. Conversely, if one believes that there is some appropriate or desirable long-run level of foreign debt to income, say, for example, 20 percent, the corresponding long-run trade deficit to GDP ratio would be only around 0.4 percent.[15] If one were to subscribe to this second idea, then the current size of the deficit is "too large" and will adjust. What might an adjustment entail under our three different scenarios?

If the consumption boom argument held merit, then we would expect that, in the future, consumption would have to decline along with output as the

BOX 1

Sustainability of the current account deficit

The current account is the sum of the trade balance ($X_t - M_t$), income receipts on foreign assets held by U.S. citizens, income payments on U.S. assets held by foreigners, and unilateral transfers, which are typically foreign aid. The income receipts and payments net out to simply the rate of return (r) paid on the U.S.'s net international investment position (A), or stock of net foreign debt, multiplied by the current stock of NIIP. The capital account is equal to the change in the NIIP, and always equals the current account. Ignoring the foreign aid component, and allowing subscript t to represent the current period, we have

$$(X_t - M_t) + (1 + r_t)A_t = A_{t+1}.$$

As discussed in the text, the current account is equal to the capital account. Taking this relationship and dividing through by GDP (Y), we derive

$$((X_t - M_t)/Y_t) + (1 + r_t) \times (A_t/Y_t) = (Y_{t+1}/Y_t) \times (A_{t+1}/Y_{t+1}).$$

Now, allowing an S prefix to indicate share of GDP and g to be the growth rate of GDP, we can rewrite the equation as follows:

$$(SX_t - SM_t) + (1 + r_t)\, SA_t = (1 + g_t) \times SA_{t+1}.$$

In the long-run steady state, this equation can be represented as

$$(SX - SM) = (g - r) \times SA.$$

From this equation it can be seen that in the long-run steady state, the sustainability of a trade deficit depends on the relationship between the growth rate of the economy and the interest rate paid on U.S. liabilities. Specifically, a negative trade balance could be sustained as long as the rate of growth of income is greater than the interest rate ($g > r$).

U.S. begins running trade surpluses and repaying the foreign debt. However, as we noted earlier, we find this to be the least compelling hypothesis of the three we consider. So, in our view, this type of adjustment is unlikely.

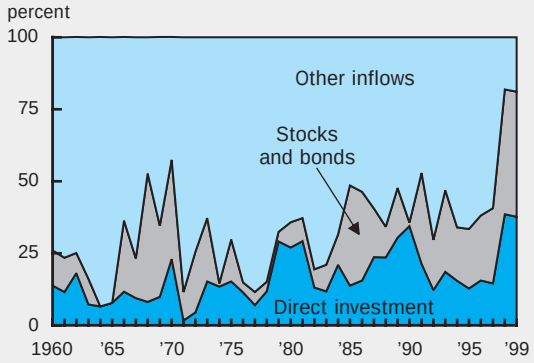
Under the safe haven scenario, for which we do find some support in the data, the adjustment would entail a reversal of the beneficial shock to the U.S. risk factor and a corresponding outflow of foreign capital from the U.S. It is not surprising that there is concern about the economic consequences of this type of adjustment in the U.S. current account deficit, given the financial crises we have seen in other parts of the world in recent years. The Mexican peso crisis in 1994–95 was followed by the Asian financial crisis beginning in mid-1997, the Russian bond default crisis in August 1998, and the Brazilian currency devaluation in early 1999. All of these developments displayed similar characteristics in that they involved a discord between countries' internal and external balances, capital flight, and sharp currency depreciations, followed by much slower or negative GDP growth during the adjustment phase.

However, as we noted above, the link between a country's internal and external balance is the interest rate, and a country's ability to influence the world interest rate is critical to the adjustment process. To see why this is important, let's examine the sequence of events if a country is confronted with an adverse shock to its perceived risk factor. This is the opposite of the shock described in the safe haven argument. In this case, foreigners demand a risk premium in excess of the world interest rate in order to supply the current level of funds. This is equivalent to a reversal or slowdown in capital inflows and would lead to a rise in national interest rates and a depreciation of the national exchange rate. Higher interest rates would lead to lower investment and consumption, while the exchange rate depreciation would lead to higher exports and lower imports. The fall in consumption and investment should outweigh the rise in exports, which means that output should fall. As shown in figure 4, however, these effects are less severe in a large country than in a small one.

Another important factor in determining the likely severity of an adjustment is the extent to which the capital inflows to the U.S. represent short-term, *hot money*. Some observers argue that speculative foreign capital has gone predominately into the U.S. stock market. However, a review of the structure of the U.S. capital account shows that this is not the case. Figure 10 shows that the largest shares of capital inflows since 1997 have been in the form of direct investment, defined as the purchase of a greater than 10 percent equity stake in a U.S. firm, and in long-term, non-Treasury securities, which include stocks and corporate

FIGURE 10

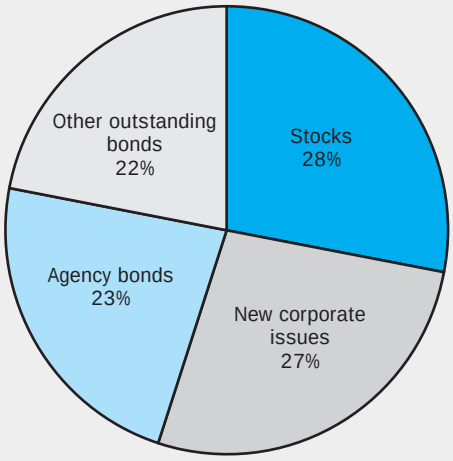
Share of capital inflows



Source: Bureau of Economic Analysis.

FIGURE 11

Foreign purchases of non-Treasury securities (1997–99 average)



Source: Bureau of Economic Analysis.

and municipal bonds with a maturity of at least one year. Such debt has contractual characteristics that make it harder to dissolve than short-term securities and bank loans. This is not to say that these assets are not liquid, but simply that they are not maturing and revolving at very short intervals.[16] Within the non-Treasury securities category, the share of equities was higher on average in the 1997 to 1999 period (28.6 percent) than in the entire 1991 to 1999 period (16.8 percent). However, figure 11 shows that the bulk of the inflows have been in the bonds, rather than the stocks subcategory.

Another important aspect of the capital account is that most of our existing debt, as well as the capital inflow, is dollar denominated. This means that in a capital outflow situation, the U.S. would not face the difficulty that some countries have of exchanging a depreciated domestic currency for the more expensive currencies in which payments need to be made.[17]

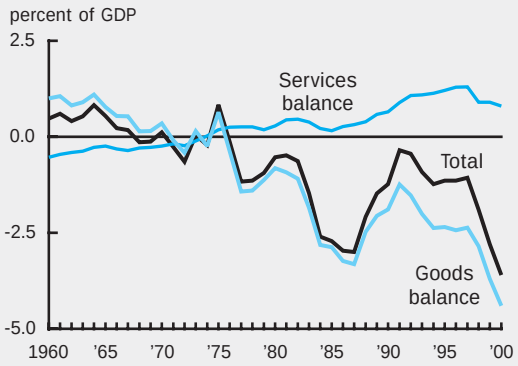
Finally, under the technological change scenario, for which we also find support in the data, the investment in productivity-enhancing capital stock leads to an increase in the output of the economy. As output increases over consumption, the country exports the excess and begins to run trade surpluses, thus beginning repayment of the foreign debt.[18]

An additional and related aspect of the technology shift argument derives from the composition of U.S. trade in goods versus services. Figure 12 shows that while the U.S. has run a persistent deficit in goods, the trade surplus in services is increasing. Trade in services represented 30 percent of U.S. goods and services exports in 1999. As the U.S. economy continues its transition toward a service economy, and as foreign countries continue to demand more services, we expect this trend will increasingly offset the goods deficit.

The scope of international trade in services, and the U.S.'s relative position in this market, is seldom given attention in policy discussions and the popular press. In fact, economists have traditionally regarded services as nontradable. Advances in technology, however, have allowed many services to transcend their historically local nature. While haircuts are still difficult to export, technological advances in communications are increasingly allowing U.S. companies to export business services, travel services, and financial services around the world. Deregulation of service industries within the U.S. and in other countries, international trade liberalization in services, and improvements in technology for service distribution

FIGURE 12

Trade balance and components



Source: Bureau of Economic Analysis.

channels will continue to expand the international market for services. With services already representing over half of real GDP, this is an area in which the U.S. will likely continue to be a competitive force in the expanding world market.[19]

If the increase in investment demand is truly due to an increase in productivity associated with new technology, and if this technological transformation continues to positively affect service industries, the U.S. can expect to reap substantial income gains in the future.[20] This would mean that the current U.S. international deficit position simply reflects the current and expected future prosperity of the U.S. and, importantly, that the adjustment process should be automatic and painless.

Conclusion

The U.S. deficit in international trade soared to new heights in 1998, again in 1999, and in all likelihood, will increase even further this year. Mirroring these deficits have been huge capital inflows from foreign investors. Is the condition of the U.S. international accounts placing the domestic economy in jeopardy? Can the U.S. continue to run such large trade deficits—continue to borrow abroad to finance the deficits—without facing an adjustment that will severely disrupt the domestic economy, along the lines of what some other countries have experienced in recent years?

In reviewing three commonly cited explanations for the source of the current account deficit—that it is a result of a boom in consumer spending at the expense of savings; that it is a result of short-term capital inflows fleeing disruptive economic conditions abroad; and/or that it is a result of a transition of the economy toward a higher level of productivity—we find that the evidence supports the notion that the current account deficit reflects a technological shift that has led to an increase in the relative prosperity of the U.S. economy. Furthermore, to the extent that an adjustment in the international sector may take place in the future, we would expect it to be relatively smooth and gradual—an adjustment that can be accomplished without serious adverse consequences to the U.S. economy. We also find some support for the safe haven story, but we do not believe that the U.S. economy faces the kind of sudden reversal in capital inflows that would be highly disruptive. This view is based on the large relative size of the U.S. economy, and its consequent ability to influence world interest rates, and the fact that much of the capital inflows of recent years have been in the form of longer-term investments.

NOTES

[1]Fred Bergsten, as quoted in Stevenson (2000).

[2]Note that this reason implies that the increase in consumption was a result of a shock to consumer tastes and preferences and is very different from an increase associated with an increase in wealth.

[3]This explanation was put forth in Hervey and Kouparitsas (2000).

[4]See, for example, Pakko (1999) and Hervey (1986).

[5]The March 2000 issue of *Survey of Current Business* reviews the behavior of returns to foreign direct investment in the U.S. The author finds that the return on assets (ROA) of U.S. owned companies in the U.S. exceeded the ROA of foreign owned companies in the U.S. by between 1 percent and 2 percent over the 1988–97 period.

[6]In this description of a national budget constraint we ignore the public sector. Government spending and saving would, of course, also affect a country's aggregate demand, and so its external balance. In particular, a government budget deficit lowers national savings and so lowers the current account balance. This is the basis of the twin deficits argument that was popular in explaining current account behavior during the 1980s.

[7]This is true if A_t is less than or equal to zero. If the country has a net stock of foreign assets, it need not borrow but can simply draw down its foreign asset stock.

[8]Humpage (1998).

[9]The personal savings rate may not be an appropriate measure, however. See Velde (1999).

[10]Another interesting aspect of figure 5 is that gross savings and gross investment in the U.S. are highly positively correlated, while investment and the current account deficit have a weaker, negative correlation. A number of studies have shown that investment booms are highly negatively correlated with increases in current account deficits, especially in smaller countries. See discussion in Baxter (1995). The difference between the excess investment and the current account deficit presented in figure 5 is termed the statistical discrepancy. Some argue that the increase in this component of the national accounts since 1997 has led to the current account deficit being overstated. See Koretz (2000).

[11]Cars and trucks account for about 85 percent of the automotive category. The food category here includes feed products.

[12]The U.S. is running a surplus in "advanced technology products," although it declined from $32.3 billion in 1997 to $19.1 billion in 1999.

[13]What the dominant force has been on the way to the current situation in the U.S. current account may not be as important as whether these shocks are permanent or temporary. Here, we assume that the shock is permanent or at least very persistent. See Baxter (1995).

[14]Over the 1955–99 period.

[15]Our choice of 20 percent in this example is arbitrary.

[16]The maturity structure of countries' foreign liabilities becomes very important in a capital outflow situation. In comparison, of Mexico's foreign liabilities in 1994, 55 percent were short term. When foreign sentiment towards Mexican assets changed, Mexico faced difficulty in rolling this debt over.

[17]Mann (1999), chapter 9.

[18]The technology shock would eventually be transmitted to the rest of the world, making investment in foreign countries relatively more attractive and encouraging a current account adjustment in the U.S.

[19]In 1999, services represented 52 percent of real GDP. See Mann (1999), chapter 6, for a further discussion of trade in services.

[20]The jury is still out on whether the U.S. is truly experiencing a "new economy" technology shock. See Kouparitsas (1999).

REFERENCES

Baxter, Marianne, 1995, "International trade and business cycles," in *Handbook of International Economics*, Vol. 3, G. Grossman and K. Rogoff (eds.), New York: Elsevier, North Holland, chapter 35, pp. 1801–1864.

Glick, Reuven, and Kenneth Rogoff, 1995, "Global versus country specific shocks and the current account," *Journal of Monetary Economics*, Vol. 35, No. 1, February, pp. 159–192.

Hervey, Jack L., 1986, "The internationalization of Uncle Sam," *Economic Perspectives*, Federal Reserve Bank of Chicago, May/June, pp. 3–14.

Hervey, Jack L., and Michael A. Kouparitsas, 2000, "Should we be concerned about the current account deficit?," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, April.

Humpage, Owen, 1998, "Is the current account deficit sustainable?," *Economic Commentary*, Federal Reserve Bank of Cleveland, October 15.

Koretz, Gene, 2000, "The trade gap may be inflated," *Business Week*, July 3, p. 32.

Kouparitsas, Michael A., 1999, "Is there evidence of the new economy in the data?," Federal Reserve Bank of Chicago, working paper, No. 99-21, December.

Krugman, Paul R., and Maurice Obstfeld, 1994, *International Economics—Theory and Policy*, third edition, New York: Harper Collins.

Mann, Catherine L., 1999, *Is the U.S. trade deficit sustainable?,* Washington, DC: Institute for International Economics, September.

Mataloni, Jr., Raymond J., 2000, "An examination of the low rates of return of foreign-owned U.S. companies," *Survey of Current Business*, Bureau of Economic Analysis, March.

McDonough, William, 2000, "McDonough—3: Higher growth in Japan, Europe needed," quoted in *Dow Jones Newswire*, March 30.

Obstfeld, Maurice, and Kenneth Rogoff, 1995, "The intertemporal approach to the current account," in *Handbook of International Economics*, Vol. 3, G. Grossman and K. Rogoff (eds.), New York: Elsevier, North Holland, chapter 34, pp. 1731–1799.

Pakko, Michael R., 1999, "The U.S. trade deficit and the 'new economy,'" *Review*, Federal Reserve Bank of St. Louis, September/October.

Stevenson, Richard W., 2000, "Economy may have a soft spot," *New York Times*, June 10, p. B-1.

Velde, François R., 1999, "Americans are not saving: Should we worry?," *Chicago Fed Letter*, Federal Reserve Bank of Chicago, May.

Effect of auto plant openings on net migration in the auto corridor, 1980–97

Thomas H. Klier and Kenneth M. Johnson

Introduction and summary

How do newly opened auto plants influence the patterns of demographic change in an area? An answer to this question has important policy implications. Competition among communities to attract new manufacturing jobs is substantial. Local governments often provide significant economic incentives to firms to induce them to locate a new plant in a given community. Such financial commitments are often justified, in part, by the argument that new jobs will aid the community in retaining its population, particularly its young adult population. The young population is viewed as critical to the future of communities because it represents a significant amount of human capital. In that context, we believe it is important to document the impact that such industrial development has on the local demographic structure. We would expect such demographic change to be most evident in the patterns of migration to and from the respective counties.

Our research links demographic trends of the last two decades to the geographic dispersion of the auto industry. The analysis focuses on the nonmetropolitan areas of seven states that make up the core of the auto industry. This "auto corridor" includes 66 percent of the employees and 70 percent of the plants engaged in the production of cars and light trucks in the U.S. Our primary interest is in estimating the impact that the presence of auto plants has on the pattern of migration in the immediate and proximate counties. We accomplish this by combining county-level migration data with data on the spatial and longitudinal distribution of auto industry plants. Our auto industry dataset is unique, consisting of plant-level information for auto assembly plants plus data on the notoriously hard-to-track auto supplier plants. It encompasses over 2,200 individual plants, representing just under 900,000 employees for the seven auto corridor states. Such comprehensive coverage of this industry represents a significant contribution to the literature.

Our models estimate the impact of auto plants on county-level net migration during the 1980s and 1990s. Explanatory variables include measures of economic, locational, and demographic characteristics and several variables measuring the presence and structure of auto plants within and proximate to the county.

Consistent with previous empirical work, we find that a set of background variables widely used in demographic research accounts for the bulk of the variation in county-level migration. However, including variables measuring the presence and addition of auto plants does add to the explanatory power of the model. The addition of a large plant to a county appears to have a significant positive influence on migration. This effect is evident not only in the county where the plant locates, but also in the contiguous counties. The effect of smaller plants is much more limited, but it is in the expected direction.

Review of the literature

Our work draws on several strands of literature. First, we examine demographic trends between 1980 and 1997. Review of such timely information is important, because metropolitan and nonmetropolitan demographic trends have been extremely fluid during the past 30 years in the nation as a whole (Long and DeAre, 1988; and Frey and Johnson, 1998).

Thomas H. Klier is a senior economist at the Federal Reserve Bank of Chicago. Kenneth M. Johnson is a demographer and professor of sociology in the Department of Sociology and Anthropology at Loyola University, Chicago. *The authors would like to thank Paul Huck and David Marshall for helpful comments and Mike Rorke and Tommy Scheiding for excellent research assistance.*

Historically, nonmetropolitan demographic change both in the auto corridor and elsewhere in the U.S. has been dominated by an excess of births over deaths sufficient to offset the net outmigration of population to metropolitan areas. This pattern of slow nonmetropolitan population gain through an excess of natural increase over migration loss was so consistent that it came to be taken for granted (Fuguitt et al., 1989). However, the pattern changed abruptly in the 1970s with the onset of what came to be called the nonmetropolitan population turnaround. Nonmetropolitan areas experienced widespread and substantial population gains and net inmigration during the turnaround (Beale, 1975; Johnson and Purdy, 1980; and Fuguitt, 1985). Nonmetropolitan population redistribution patterns shifted again in the 1980s. Most nonmetropolitan counties lost population during the decade because they had a modest net outflow of population combined with low levels of natural increase (Johnson, 1993). Many researchers regarded the diminished nonmetropolitan growth of the 1980s as evidence that U.S. population redistribution trends had reverted to historical form, with the turnaround of the 1970s being just a short-term fluctuation. Yet, there is now evidence of another upturn in population growth rates in nonmetropolitan areas during the late 1980s and 1990s (Johnson and Beale, 1994; and Johnson, 1998). Our purpose here is to examine the linkages between such demographic change and trends in the spatial structure of the auto industry.

The U.S. auto industry has undergone major changes during the last 30 years (see McAlinden and Smith, 1993; Rubenstein, 1992; and Harbour, 1990). Three developments have shaped the spatial pattern of the industry during this period—the reconcentration of auto assembly facilities in the heart of the country, the southward expansion of the traditional Midwest Auto Belt, and the arrival of Japanese auto assembly and parts plants.

The spatial changes affecting the auto industry have been reported in Rubenstein (1992). Most notable is the reconcentration of auto assembly facilities in the heart of the country. In the early days of the industry, assembly plants were built close to population centers, since it was cheaper to ship parts to assembly plants than to ship finished vehicles across the country. This approach worked well as long as consumer demand for specific models was sufficient to support production at multiple assembly plant locations. However, since the 1960s, the proliferation of different models of cars has far outstripped the growth in overall production of vehicles. As a result, particular models must now be produced and shipped from only one or two assembly plants if they are to be profitable. This development has led many companies to concentrate their assembly plants in the heart of the U.S. to minimize the costs of distributing the final product to a national market. It also allows the assembly plants to be located near the plants that produce engines, transmissions, drive trains, and a host of other components. The net result of this trend has been the closing of many coastal assembly plants and the reconcentration of light vehicle assembly plants in the auto corridor (see table 1).

As the auto industry reconcentrated in the nation's heartland, much of the growth occurred in the historical Auto Belt around Detroit. However, the industry simultaneously expanded southward, forming an auto corridor that includes not only the traditional auto states of Michigan, Illinois, Indiana, Wisconsin, and Ohio, but also Kentucky and Tennessee. This southward extension of the auto corridor started in the 1970s with efforts by General Motors to lower procurement costs by building component plants south of the traditional auto region. Facilitating the southward expansion of the auto corridor was the arrival of Japanese-owned assembly and supplier facilities during the 1980s (see table 2) (see Kenney and Florida, 1992; McAlinden and Smith, 1993; Smith and Florida, 1994; and Head et al., 1995).[1] Plant location decisions by Japanese car companies reflect a preference for greenfield locations on the southern periphery of the traditional Auto Belt. Hence Kentucky and Tennessee more than tripled their share of light vehicle assembly plants, from 4 percent to 13 percent, while the other five states of the auto corridor increased their overall share from 43 percent to 50 percent between 1970 and 1997 (table 1).

TABLE 1

Reconcentration of light vehicle assembly plants

	Plants operational in			
Auto corridor	**1970**		**1997**	
	Number	(%)	Number	(%)
Illinois	2	(4)	3	(5)
Indiana	0	(0)	2	(4)
Kentucky	2	(4)	5	(9)
Michigan	15	(29)	15	(27)
Ohio	3	(9)	7	(13)
Tennessee	0	(0)	2	(4)
Wisconsin	2	(4)	1	(2)
Total	24	(47)	35	(63)
U.S. total	**51**	**(100)**	**56**	**(100)**

Source: *Ward's Automotive Yearbook*, various years.

TABLE 2

Newly opened light vehicle assembly plants, 1980–97

Company	State	Start-up year
GM	Kentucky	1981
GM	Louisiana[a]	1981
GM	Ohio	1981
Honda	Ohio	1982
Nissan	Tennessee	1983
NUMMI[b]	California[a]	1984
GM	Michigan	1984
GM	Missouri[a]	1984
GM	Michigan	1985
AutoAlliance[b]	Michigan	1987
DiamondStar	Illinois	1988
Toyota	Kentucky	1988
Honda	Ohio	1989
Subaru-Isuzu	Indiana	1989
Saturn	Tennessee	1990
Chrysler	Michigan	1991
Ford	Ohio	1991
BMW	So. Carolina[a]	1994
Toyota	Kentucky	1994
Mercedes-Benz	Alabama[a]	1997

[a]Indicates state not in auto corridor.
[b]Reopened previously closed facility.
Source: ELM Guide database, 1997.

Several papers examine the economic impact of locating an auto assembly facility on the proximate areas (see Haywood, 1998; Center for Business and Economic Research, 1992; Fournier and Isserman, 1993; and Marvel and Shkurti, 1993). These studies examine the balance between the incentives used to attract a plant and the resulting development of the region as measured by income and employment. Their findings suggest that adding an assembly plant can have spatially disparate effects on growth. For example, the host county for a Honda assembly plant that opened in Ohio in 1982 experienced much stronger employment and income growth than the contiguous counties (Fournier and Issermann, 1993). At the state level, the impact of attracting an assembly plant depends on the timing of a particular plant relative to others in a region. Murray et al. (1999) suggest that spinoff effects derived from the subsequent location of supplier facilities near new assembly plants are strongest in areas that were first to attract an automobile assembly plant. The economic development literature, however, provides very little evidence on possible linkages between plant location and demographic trends. Two studies of the impact of new assembly plants suggest that approximately 80 percent of those who migrated to obtain work in the newly opened assembly plants came from within the same state (Elhance and Chapman, 1992; and Marvel and Shkurti, 1993).

In sum, the literature suggests there have been significant shifts in the demographic trends in the auto corridor during the past three decades. During the same period, the auto industry has experienced a reconcentration of activity in the auto corridor and a simultaneous southward expansion of this corridor. The literature provides some evidence that the opening of new auto plants has an impact on the economic and, perhaps, the demographic character of the proximate area. Our purpose here is to more clearly delineate the linkages between recent spatial shifts in the location of the auto industry and demographic change in the seven auto corridor states, using new data on the distribution of auto industry plants.

Data and procedures

We use data on demographic change since 1990 from the *Federal–State Cooperative Population Estimates* series, developed jointly by the U.S. Bureau of the Census and the states. Additional data are from the U.S. decennial censuses of population for 1970, 1980, and 1990. Births and deaths for 1980 to 1990 are from special tabulations of the *Federal–State Cooperative* series. The typology used to classify counties by economic function was developed by the Economic Research Service of the U.S. Department of Agriculture (Cook and Mizer, 1994). The recreational specialty variable is from Beale and Johnson (1998). We calculate net migration by subtracting natural increase from the population change during the appropriate period.

Counties are the unit of analysis and are appropriate for this purpose because they have historically stable boundaries and are a basic unit for reporting fertility, mortality, and census data. This article focuses on the auto corridor, which is defined as the following seven states: Illinois, Indiana, Michigan, Ohio, Kentucky, Tennessee, and Wisconsin. There are 652 counties in the auto corridor, with a total population of 53.1 million people in 1997. This region encompasses about two-thirds of the total number of light vehicle assembly and supplier plants in the U.S. (see table 3).[2]

Metropolitan reclassification complicates our efforts to compare the trends of various periods. We use the latest (1993) metropolitan definition to classify counties as metropolitan or nonmetropolitan. According to this definition, there were 455 nonmetropolitan

TABLE 3

Auto corridor share of auto industry, 1997

	Plants		Employment	
	Number	(%)	Number	(%)
Major plants	156	(72)	353,392	(70)
Independent suppliers	2,043	(68)	533,808	(60)
Total	**2,199**		**887,200**	

Notes: Major plants are light vehicle assembly plants and captive supplier plants. Numbers in parentheses indicate percent of U.S. total.
Sources: ELM Guide database, 1997; various state manufacturing directories, 1997

counties in the auto corridor and 197 metropolitan counties in the auto corridor. Because counties are reclassified from time to time as new metropolitan areas are formed or territory is added to existing areas, the demographic implications of using one definition of metropolitan areas in preference to another are far from trivial (Johnson, 1989). There is no simple resolution to the problem of metropolitan reclassification nor is any one approach clearly superior to all others (Fuguitt et al., 1988). Using the 1993 definition results in greater nonmetropolitan losses during the 1980s and slower nonmetropolitan gains during the early 1990s than would have been the case had we used the earlier metropolitan definition.[3]

We use auto industry data from the ELM Guide database, a set of plant-level data developed by a private company in Michigan. This database includes information on auto assembly facilities, supplier plants owned by assembly companies (so-called captive suppliers), and independent supplier plants (the database focuses on suppliers that deal directly with assembly companies). The data represent the year 1997 and identify, among other variables, for each plant the address, a list of the plant's products, the production processes used, and employment. We obtained information on the plants' start-up year from various state manufacturing directories and the plants themselves. The data represent over 2,200 individual plants and approximately 900,000 employees.[4] While the data are very comprehensive for the year 1997, due to their cross-sectional nature, they do not include information on plant deaths during the period analyzed. In other words, all information on plant opening years is conditional on the plant surviving through 1997, leading to *survivor bias* in the data. (See box 1 for an explanation of the implications of this data problem.)

Descriptive statistics

Demographics in the auto corridor

In a reversal of the trend of the 1980s, there was widespread population growth in nonmetropolitan areas of the auto corridor during the 1990s. More than 87 percent of the 455 counties in the auto corridor classified as nonmetropolitan in 1993 gained population between 1990 and 1997 (table 4). In all, 192 more nonmetropolitan counties gained population than in the 1980s. The estimated nonmetropolitan population gain in the auto corridor between April 1990 and July 1997 was 693,000. In contrast, nonmetropolitan areas lost nearly 20,000 in population during the 1980s. Although the nonmetropolitan population gain of 1.3 million in the auto corridor during the 1970s was greater than the gain of the 1990s, this recent gain is substantial compared with any other in recent decades.

The nonmetropolitan population gains are even more surprising given that, historically, the metropolitan areas of the auto corridor have been the major growth centers of the region. Yet, in two of the last three decades, nonmetropolitan growth rates have actually exceeded those in the region's metropolitan areas. The nonmetropolitan population grew at a faster pace (5.7 percent) than the metropolitan population

BOX 1

Survivor bias

The data represents information from 1997 and includes the opening year for individual plants. However, it does not represent time-series information as it only includes plants that were operational in 1997. In other words, the data include survivor bias, because all information on the history of individual plants is conditional on their surviving until 1997. Given that constraint, what assumptions do we make in interpreting the empirical results?

In interpreting the data on spatial distribution of plant location, we assume that plants located in nonmetropolitan counties do not show higher exit rates than plants located in metropolitan counties. If they did, the dispersion of the industry into nonmetropolitan counties during the 1980s and 1990s, as measured by plant openings, would be overstated. To our knowledge, there is no empirical work that could back up this assumption. However, it seems a reasonable assumption to make. In fact, it might be rather conservative in light of the fact that older manufacturing plants tend to be concentrated in urban areas, which might lead to *higher* exit rates for metropolitan county plants.

TABLE 4

Auto corridor population by metro status

	Number of cases	Initial population	Population change: Absolute change	Population change: Percent change	Population change: Percent growing	Net migration: Absolute change	Net migration: Percent change	Net migration: Percent growing	Natural increase: Absolute change	Natural increase: Percent change	Natural increase: Percent growing
1970 to 1980											
Nonmetropolitan	455	10,799,742	1,322,914	12.2	94.3	718,423	6.7	81.8	604,491	5.6	93.0
Metropolitan	197	36,609,737	1,201,872	3.3	87.8	–1,465,839	–4.0	63.5	2,667,886	7.0	100.0
Total	652	47,409,479	2,524,786	5.3	92.3	–747,416	–1.6	76.2	3,272,377	6.9	95.1
1980 to 1990											
Nonmetropolitan	455	12,122,650	–19,966	–0.2	44.8	–544,481	–4.5	23.3	524,515	4.3	91.6
Metropolitan	197	37,811,609	657,130	1.7	70.1	–1,979,479	–5.2	35.0	2,639,609	7.0	99.5
Total	652	49,934,259	637,164	1.3	52.5	–2,523,960	–5.1	26.8	3,164,124	6.3	94.0
1990 to 1997											
Nonmetropolitan	455	12,104,092	693,026	5.7	87.0	441,416	3.6	78.2	251,610	2.1	78.2
Metropolitan	197	38,468,739	1,900,129	4.9	88.3	71	0.0	72.1	1,900,058	4.9	98.0
Total	652	50,572,831	2,593,155	5.1	87.4	441,487	0.9	76.4	2,151,668	4.3	84.2

Note: 1993 metropolitan status used for all periods.
Sources: 1970–90 Census and *Federal–State Cooperative Population Estimates*.

(4.9 percent) between 1990 and 1997. Metropolitan growth did exceed that in nonmetropolitan areas of the auto corridor during the 1980s. However, during the turnaround of the 1970s, nonmetropolitan gains (12.2 percent) exceeded metropolitan gains (3.3 percent) by a substantial margin. Geographically, population gains were widespread in nonmetropolitan areas of the corridor. Population losses were common in the core counties of the older industrial areas of the region.

The renewed nonmetropolitan population growth in the 1990s, as well as the earlier growth during the 1970s, is due, in large part, to migration gains. For example, such migration gains accounted for 64 percent of the total estimated population increase between April 1990 and July 1997. Nonmetropolitan areas had an estimated net inflow of 441,000 people during the period. In contrast, metropolitan areas of the auto corridor experienced no migration gain during the 1990s. This is a sharp contrast to the pattern during the 1980s when both metropolitan and nonmetropolitan areas had net outmigration. The auto corridor's metropolitan areas were particularly hard hit by outmigration during the 1970s and 1980s, losing nearly 3,445,000 people between 1970 and 1990. In comparison, nonmetropolitan areas enjoyed substantial migration gains during the population turnaround of the 1970s. The complex pattern of migration change over the past three decades in the auto corridor is of particular interest here because it coincides with a period of change in the auto industry.

The differential impact of net migration on metropolitan and nonmetropolitan areas is clearly evident when we look at spatial patterns (see figures 1 and 2 where migration patterns are shown). We see migration from both metropolitan and nonmetropolitan areas during the 1980s. Nonmetropolitan counties with net inmigration are located primarily in recreational and high amenity areas in the northern and southern periphery of the seven-state area. Migration from metropolitan counties was also evident, particularly in many of the traditional Auto Belt cities of southern Michigan and northern Ohio. The few metropolitan counties that were growing were in suburban rings around older cities.

There are dramatic changes in the spatial patterns of migration in the 1990s. Nonmetropolitan migration gains are extremely widespread except in agricultural areas near the center of the corridor. We also see a migration recovery in the region's metropolitan counties, though migration losses continued in many cities traditionally associated with car production.

Natural increase accounted for 36 percent of the nonmetropolitan population increase in the auto corridor between April 1990 and July 1997. In all, births exceeded deaths by 252,000 in nonmetropolitan areas. The annualized gain through natural increase in nonmetropolitan areas was somewhat lower between 1990 and 1997 than it had been during the 1980s. In contrast, the annualized rate of natural increase remained constant in the auto corridor's metropolitan regions, and natural increase accounted for all of the metropolitan population increase in the 1990s.[5]

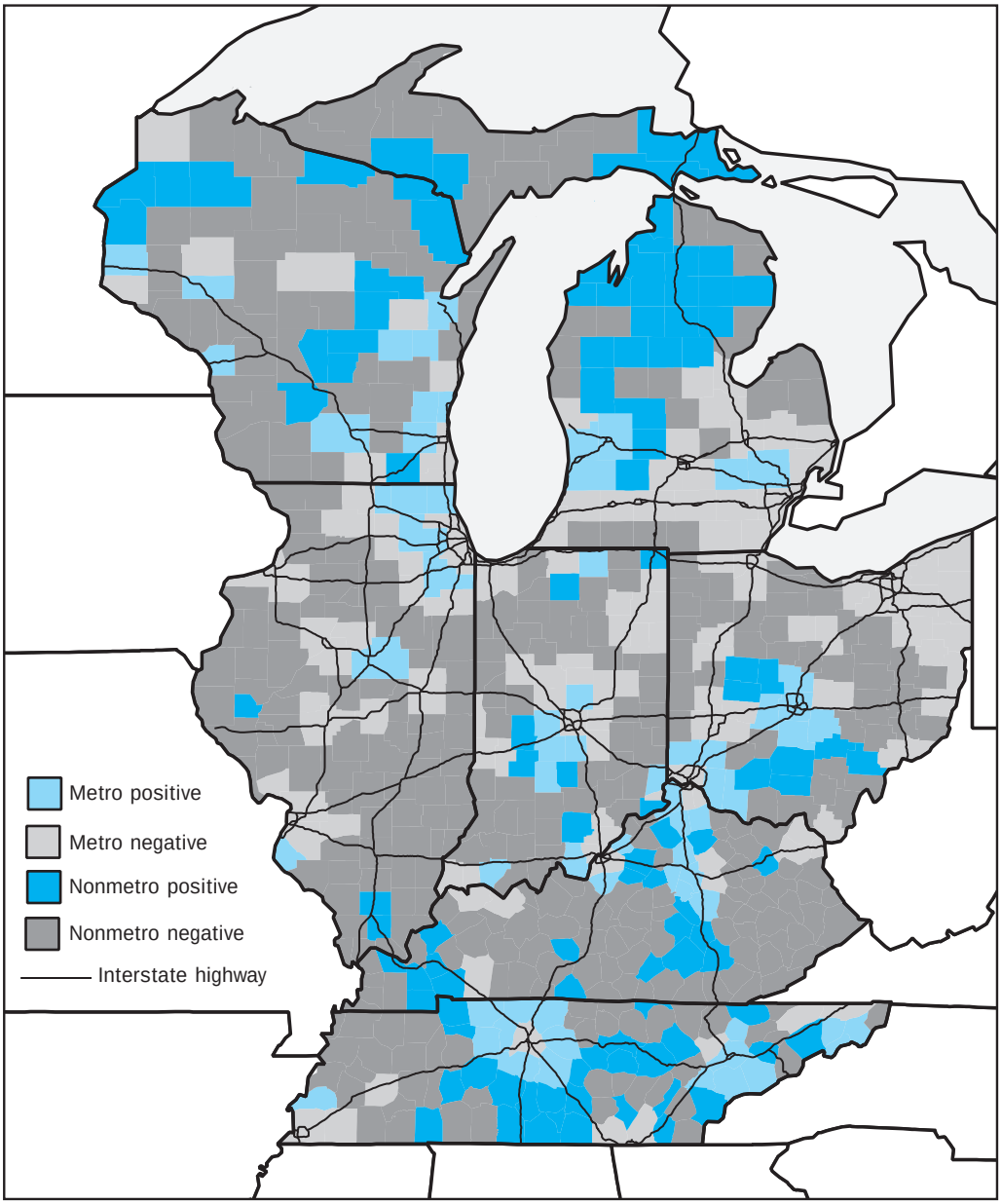
Nonmetropolitan population gains were more likely in counties near metropolitan centers. Nearly 92 percent of these adjacent counties gained population in the 1990s, and 80 percent had net inmigration (table 5). Even among more remote nonmetropolitan counties, recent population gains have been significantly greater than during the 1980s. Growth occurred

in 82 percent of counties not adjacent to metropolitan areas in the 1990s. Such nonadjacent counties had net inmigration (3.7 percent) during the 1990s.

Nonmetropolitan counties that were destinations for retirees or centers of recreation were the fastest growing counties during the early 1990s. All of the 24 nonmetropolitan retirement destination counties in the auto corridor gained population and had net inmigration between 1990 and 1997. These areas, common in the Upper Great Lakes and Appalachians (Cook and Mizer, 1994), are attracting retirees while retaining their existing population (Fuguitt and Heaton, 1993). Population gains also occurred in 95 percent of the 41 nonmetropolitan recreational counties during the 1990s, with a large majority (93 percent) receiving net inmigration. Such counties had been prominent growth nodes during the 1970s and 1980s and the trend persisted in the 1990s. There is

FIGURE 1

1980s net migration for metro and nonmetro counties



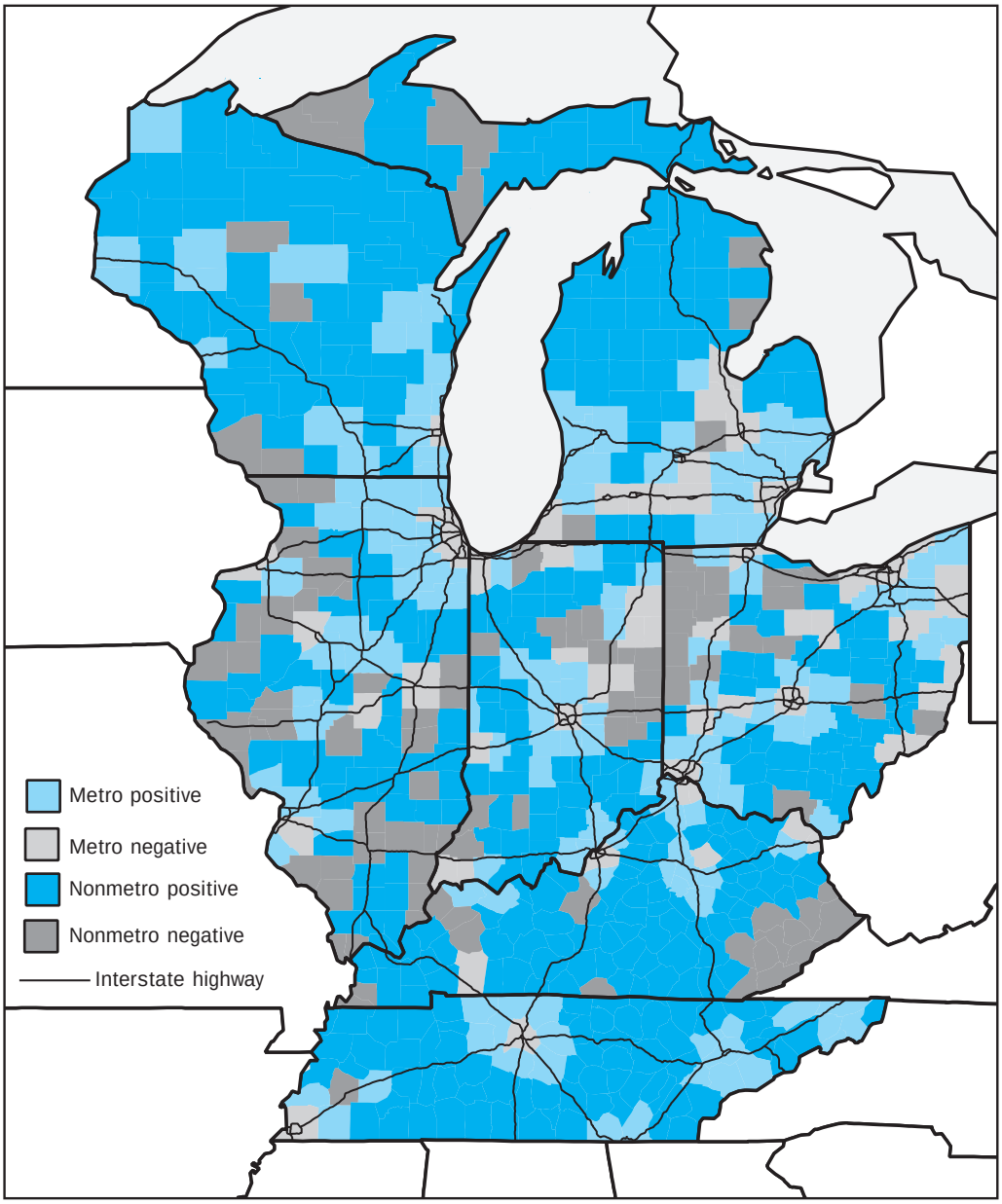
Note: Interstate highways are shown only for the seven auto corridor states.
Sources: 1970–90 Census and *Federal–State Cooperative Population Estimates.*

significant overlap between the recreational and retirement destination counties, because the amenities and scenic advantages that attract vacationers and seasonal residents also appeal to retirees.[6]

Nonmetropolitan population gains were also widespread in manufacturing and commuting counties in the auto corridor, though the gains were smaller than those in recreational and retirement counties. Growth in such counties was more evenly balanced between natural increase and net migration. The proximity of the lower Great Lakes manufacturing belt and the emergence of new industrial areas in the southern part of the region in recent years accounts for the large number (178) of rural manufacturing counties. The expansion of the auto industry during the past several decades has certainly been a factor in this. Both migration gains and natural increase were common in manufacturing counties. A large number of auto corridor nonmetropolitan counties have a substantial share of their labor force commuting to

FIGURE 2

1990s net migration for metro and nonmetro counties



Note: Interstate highways are shown only for the seven auto corridor states.
Sources: 1970–90 Census and *Federal–State Cooperative Population Estimates*.

TABLE 5

Auto corridor population in nonmetropolitan counties, selected variables

	Number of cases	Population change: Percent change	Population change: Percent growing	Net migration: Percent change	Net migration: Percent growing	Natural increase: Percent change	Natural increase: Percent growing
Adjacent	241	6.1	92	3.6	80	2.5	87
Nonadjacent	214	5.2	82	3.7	77	1.5	69
Retirement	24	14.7	100	14.0	100	0.7	42
Recreational	41	9.8	95	8.5	93	1.2	63
Manufacturing	178	6.5	93	3.9	80	2.6	88
Commuting	117	7.7	92	5.7	87	2.0	75
Mining	33	0.9	52	–0.7	45	1.6	73
Farming	28	5.0	93	4.2	86	0.9	68
Total nonmetropolitan	455	5.7	87	3.6	78	2.1	78

Notes: 1993 metropolitan definition. Percent change is aggregate change for all cases in category.
Recreational counties defined by Beale and Johnson (1998). All other types defined as in Cook and Mizer (1994).
Nonmetropolitan counties divided into those adjacent to a metropolitan county and those not adjacent.
Sources: 1970–90 Census and *Federal–State Cooperative Population Estimates*.

jobs in other counties, often in proximate metropolitan counties. This allows rural workers to access the urban labor market, while retaining their rural place of residence and lifestyle. The substantial migration gains in such counties reflect their significant appeal.

The 33 counties dependent on mining in the auto corridor were the least likely to gain population during the 1990s. Only 52 percent of these counties gained population and only 45 percent had net inmigration. Population gains were considerably more widespread in farming counties, but the magnitude of the gains was relatively small. The smaller than average population gains for mining and farming dependent counties in the 1990s represents a continuation of the trends of the 1980s. However, even among these counties the population and migration trends moderated in the 1990s compared with the 1980s, when population decline and migration losses were much more prevalent.

Evolving spatial distribution in the auto corridor

Table 6 shows the distribution of auto plant openings for plants surviving through 1997 across time and by county type. It distinguishes independent supplier plants from captive suppliers and light vehicle assembly plants (referred to as "major" plants). Because we do not know the employment history for the plants, we use this distinction to approximate small and large plants. The average independent supplier plant employed 258 workers in 1997, compared with 2,265 for an assembly or captive parts plant.

Table 6 shows the industry's growth during the last three decades as measured by the growth in newly opened independent supplier plants. Their number more than doubled since 1970, with the largest absolute increase occurring during the 1980s. Since 1970, the industry has also spread out within the auto corridor, indicated by the increase in the share of supplier plants located in nonmetropolitan counties from 28 percent in 1970 to 33 percent in 1997. Plant openings among major plants largely reflect the opening of new assembly plants.

TABLE 6

Plant openings across time and county type

	Prior to 1970	Plants opened: 1970s	Plants opened: 1980s	Plants opened: 1990s	Plants open in 1997
Independent supplier plants					
Nonmetropolitan	265	110	215	75	665
Metropolitan	653	210	303	212	1,378
Total	918	320	518	287	2,043
Major plants					
Nonmetropolitan	9	0	4	3	16
Metropolitan	104	11	17	8	140
Total	113	11	21	11	156

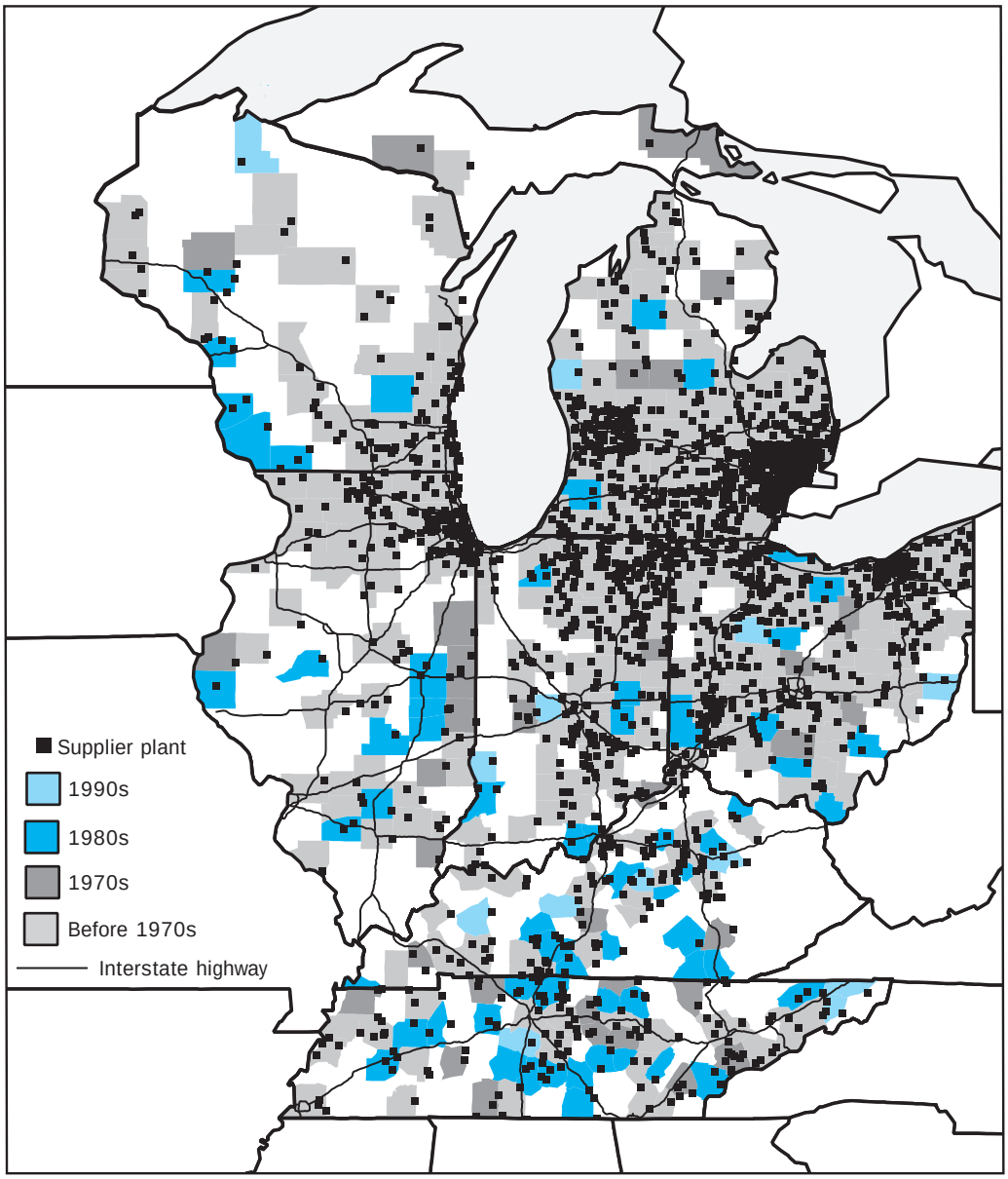
Note: Major plants are light vehicle assembly plant and captive supplier plants.
Sources: ELM Guide database, 1997; various state manufacturing directories, 1997.

As mentioned above, the location choices of auto plants in the corridor states during the last 30 years can be characterized by dispersion as well as southward expansion. Figure 3 shows this development based on our database, using information on start-up years for plants that were in business in 1997. The counties are color-coded to indicate the decade during which the first independent auto supplier plant opened. Finally, figure 3 also shows interstate highways and the density of auto supplier plants in 1997.

Figure 3 shows the core of the industry to be located in southern Michigan, as well as northern Indiana, northern Ohio, and the Chicago area.[7] From there, plants dispersed to the west and north, but mostly to the south. Such dispersion peaked during the 1980s, when 64 counties that did not previously have auto supplier plants gained at least one (versus 34 in the 1970s, and 14 in the 1990s). Most of the newly occupied auto corridor counties were in Kentucky and Tennessee. During the last three decades, the two southern states'

FIGURE 3

Dispersion and plant density of auto suppliers



Note: Interstate highways are shown only for the seven auto corridor states.
Sources: ELM Guide database, 1997; various state manufacturing directories, 1997.

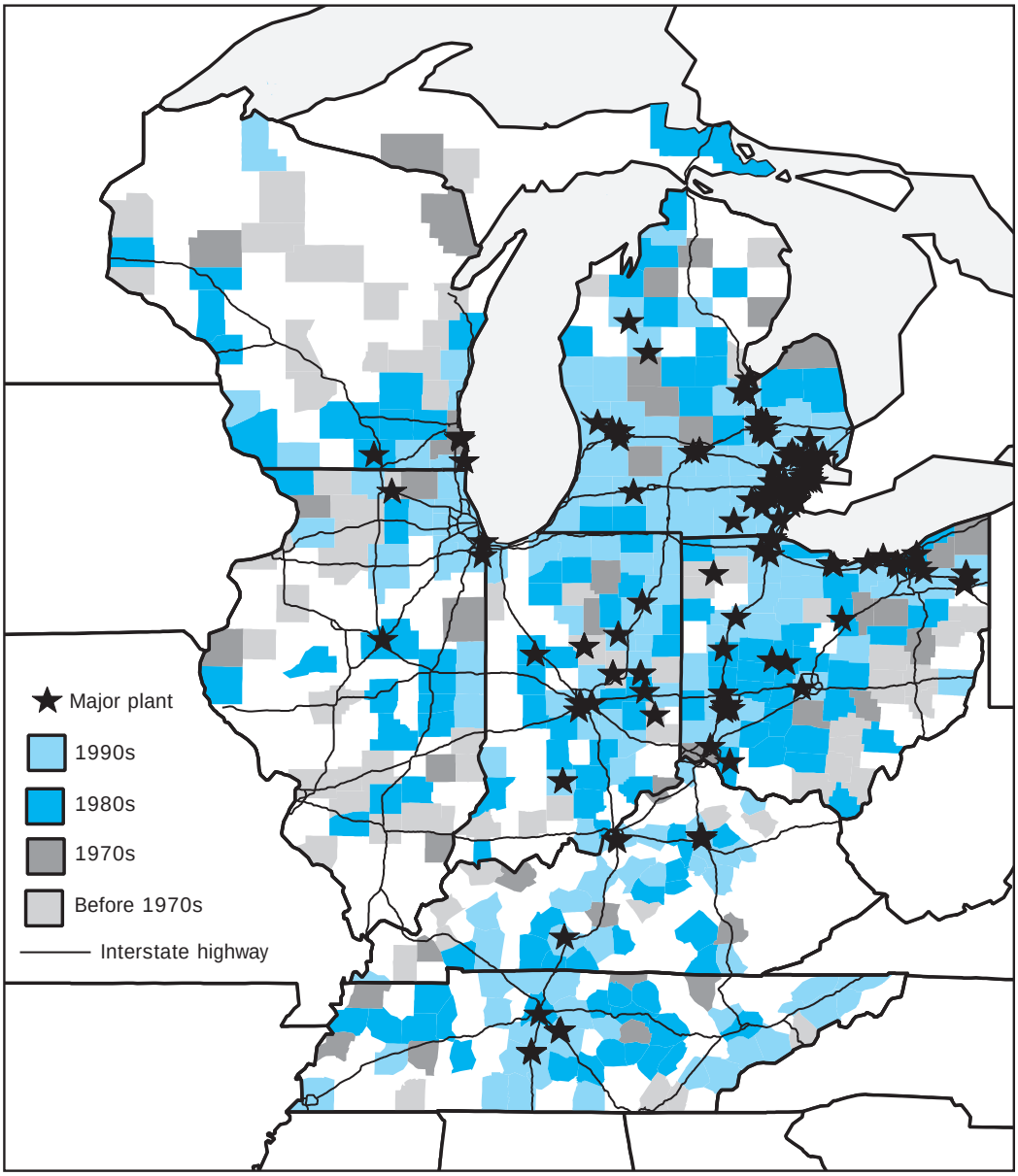
share of newly opened supplier plants has steadily increased from 41 percent in the 1970s, to 53 percent in the 1980s and 57 percent in the 1990s. This trend also holds for new assembly plants (see table 2).

The importance of highway transportation is also evident, especially in the southern half of the auto corridor.[8] Nearly every county with an auto supplier plant is on or near an interstate highway, and supplier facilities cluster around transportation hubs such as Indianapolis and Nashville.

Figure 4 adds a longitudinal perspective to the analysis by showing the year during which the last independent supplier plant was added to a county. It complements figure 3 and demonstrates that the core of the auto corridor continued to be the preferred location choice for plant openings by auto supplier companies in the 1990s. Correspondingly, the share of counties last occupied within the two southern states increased only slightly from 24 percent in the 1970s to 31 percent in the 1990s. Further, this figure

FIGURE 4

Auto corridor counties by last year of entry



Note: Interstate highways are shown only for the seven auto corridor states.
Sources: ELM Guide database, 1997; various state manufacturing directories, 1997.

underscores the continuing importance of highway transportation.

Model and results

The net migration evident in the auto corridor during the 1980s and 1990s is the product of a myriad of economic, demographic, locational, and historical factors. To estimate the combined influence of these factors, we need to perform a *multivariate analysis*. Here, we examine the impact of these background factors and the influence of the auto industry using *ordinary least squares regression*. We estimate a separate cross-sectional model for each of the two decades. The dependent variable in each model is the net migration during the decade (defined as population change net of natural increase) relative to the level of population at the beginning of the period. The analysis covers the 455 nonmetropolitan counties in the seven auto corridor states.

We group the independent variables into two major categories. The first represents economic, locational, and demographic variables recognized as important in previous work (Johnson, 1998; and Goetz and Rupasingha, 1999). We include measures of labor force structure, commuting, metropolitan adjacency, and whether the county is a retirement or recreational node (Beale and Johnson, 1998; and Cook and Mizer, 1994). Because there has been considerable regional variability in nonmetropolitan demographic trends recently, we include a dummy variable to differentiate the two southern states from the five midwestern states. Demographic change may also be influenced by the size of the local population; therefore, we include a county's population at the beginning of each period. Table 7 provides a detailed description of the variables included in the models.

We supplement these standard economic, locational, and demographic variables with a block of

TABLE 7

Variable key

Dependent variable	
Net migration	Population change minus natural increase relative to population at beginning of period.
Independent variables	
Control variables	
Metro adjacency	1 if nonmetropolitan county is adjacent to metropolitan county, 0 otherwise.
Recreational county	High proportion of spending and employment in recreational industries, large concentration of second homes, high per capita spending on hotels and motels, contextual data indicating presence of major tourist activity.
Retirement county	Net migration gain for those over the age of 60 by 15 percent or more between 1980 and 1990.
Percent employed in agriculture	Ratio of employment in agriculture to total employment at beginning of decade.
Percent employed in manufacturing	Ratio of employment in manufacturing to total employment at beginning of decade.
Percent work outside the county	Ratio of employees who had jobs outside the county of residence to total employment at beginning of decade.
Population	Population at beginning of period.
South	1 if county is in Kentucky or Tennessee, 0 otherwise.
Unemployment rate	Annual average rate at beginning of decade.
Auto variables	
Supplier base	Number of independent supplier plants at the beginning of decade.
Supplier addition	Number of independent supplier plants added during decade.
Major auto plant base	Number of assembly and captive supplier plants at beginning of decade.
Major auto plant addition	Number of assembly and captive supplier plants added during decade.
Contig. major auto plants base	Number of assembly and captive supplier plants in contiguous counties at beginning of decade.
Addition of contiguous major plants	Number of assembly and captive supplier plants added in contiguous counties during decade.

variables measuring the presence of the auto industry in a county. This characterization of the auto industry distinguishes between the assembly and parts plants owned by major foreign or domestic automakers (labeled "major") and independent supplier plants (labeled "supplier").[9] The independent supplier plants tend to be smaller, more numerous, and more widely distributed throughout the nonmetropolitan areas of the auto corridor. The company-owned plants are considerably larger and tend to be located in metropolitan areas (see table 6). For each of these two major plant types we measure the number of plants in operation at the beginning of the modeling period and the number of plants added during the decade. Finally, it is possible that the impact on migration of locating a plant spills over into surrounding counties. We model this so-called contiguity effect only for assembly and captive parts plants, as these plants employ substantial numbers of workers. We use variables measuring the number of major plants in contiguous counties at the beginning of the period as well as the number of new plants added during the period.[10]

The explanatory power of the estimated model for migration is similar in each period (see table 8 and the appendix for the estimated coefficients). It accounts for 37 percent of the variation in net migration between 1990 and 1997, compared with 43 percent between 1980 and 1990. There is also considerable consistency in the contribution of specific variables during both periods. Among the control variables, greater migration gains were likely in counties that were centers of recreation and retirement, had a higher share of commuters to neighboring counties, and were located in Kentucky or Tennessee. Other things being equal, counties with employment concentrations in agriculture and those with a larger population tended to gain less or lose more from migration than other counties. For each of these variables, results for both periods are statistically significant.

The block of six variables representing the auto industry provides a statistically significant improvement in explanatory power during the 1990s.[11] The incremental improvement during the 1980s does not quite reach statistical significance. The directional impact of the individual variables is also quite consistent for the two decades. The addition of assembly and captive supplier plants (major) in either the county of interest or a contiguous county has a positive impact on migration. This effect is statistically significant for the 1990s. In the immediate county it increased net migration by 7.68 percent. The effect on net migration spills into the contiguous counties, albeit at a reduced level (3.29 percent in the 1990s). The size of this spillover effect is similar in magnitude to the effect of being a retirement county.

In contrast, counties containing assembly and captive supplier plants at the beginning of a given decade were likely to be adversely affected with respect to migration, though the impact was statistically significant only for the 1980s. This may reflect the cutbacks experienced in the auto industry during the 1980s and early 1990s. A county containing such plants at the beginning of the 1980s experienced an additional net migration of –2.5 percent whereas the mean value for nonmetropolitan counties during that decade was –3.9 percent. That result suggests that during the 1980s the presence of auto plants worsened the negative migration experience of nonmetropolitan counties. Once again, this effect spills over into contiguous counties. For these counties we estimate that the presence of assembly and captive supplier plants

TABLE 8

Summary of results for nonmetropolitan counties

	1990s	1980s
Intercept	+	+
Control variables		
Metropolitan adjacency	+	+
Recreational county	+***	+**
Retirement county	+***	+***
Percent employed in agriculture	–**	–***
Percent employed in manufacturing	+*	–
Percent work outside the county	+***	+*
Population	–***	**
South	+***	+***
Unemployment rate	–***	+
Auto variables		
Supplier base	+	+
Supplier addition	–	+*
Major auto plant base	–	–*
Major auto plant addition	+***	+*
Contiguous major plants base	–***	–***
Contiguous major plants addition	+***	+
R-squared	0.37	0.43
Number of observations	455	455

*Indicates significance level of 90 percent; **indicates 95 percent; and ***indicates 99 percent.
Note: See table 7 for variable definitions.

TABLE 9

Results for nonmetropolitan counties

	1990s	1980s
Intercept	2.348	–1.88
	(1.26)	(–1.06)
Control variables		
Metropolitan adjacency	0.854	0.615
	(1.58)	(1.22)
Recreational county	3.321	2.06
	(3.33)	(2.31)
Retirement county	10.970	13.590
	(9.45)	(11.10)
Percent employed in agriculture	–0.145	–0.315
	(–2.07)	(–6.48)
Percent employed in manufacturing	0.061	–0.040
	(1.83)	(–1.06)
Percent work outside the county	0.107	0.058
	(4.15)	(1.50)
Population	–0.00005	–0.00003
	(–2.76)	(–2.34)
South	2.865	2.310
	(4.13)	(4.17)
Unemployment rate	–0.257	0.023
	(–2.55)	(0.28)
Auto variables		
Supplier base	0.051	0.191
	(0.41)	(1.55)
Supplier addition	–0.002	0.347
	(–0.006)	(1.85)
Major auto plant base	–0.971	–2.470
	(–0.70)	(–1.74)
Major auto plant addition	7.683	2.560
	(6.41)	(1.80)
Contiguous major plants base	–0.429	–0.300
	(–4.04)	(–3.86)
Contiguous major plants addition	3.293	0.531
	(4.34)	(1.00)
R-squared	0.37	0.43
Number of observations	455	455

Notes: See table 7 for variable definitions. Numbers in parentheses are t-stats. The error terms are White-corrected for heteroskedasticity.

at the beginning of the decade lowered net migration by –0.4 percent in the 1990s and –0.3 percent in the 1980s. The estimated effects of the presence and addition of independent supplier plants, which generally are much smaller plants, tend not to be statistically significant. However, for the 1980s, the decade that saw the largest number of independent supplier plants start up during the time period analyzed (see table 6), adding a supplier plant increases net migration by 0.3 percent.[12]

In order to address the effect of plant size more directly, we reestimate the model for the 1990s, distinguishing auto plants by their employment level. Consequently, we redefine all the auto industry variables to represent either large (1,000 employees or more) or small plants. The estimates we obtain are virtually identical to the ones reported in table 9, which suggests that locating a large plant in a nonmetropolitan county raises net inmigration into that county by 7.68 percent and by 3.29 percent in the surrounding nonmetropolitan counties.[13]

In sum, we find that accounting for the presence of auto plants adds to the explanatory power of a model of county-level net migration. Our results reproduce earlier findings for a fairly standard set of control variables. Furthermore, we find that adding a large plant to a nonmetropolitan county triggers a sizeable positive net migration response, both in the county where the plant locates and in the surrounding counties.

Conclusion

This article addresses possible linkages between the recent spatial shifts in the auto industry and demographic change at the county level, a question that had previously received very little attention. We perform the analysis for the seven states that represent the core of the U.S. auto industry. We use a standard set of control variables measuring economic, locational, and demographic characteristics together with a comprehensive set of data on the distribution of auto plants across space and time.

Consistent with previous empirical work, we find that the background variables widely used in demographic research account for a substantial proportion of the variation in county-level migration. However, adding variables measuring both presence and addition of two types of auto plants adds to the explanatory power of the model. As a group, the auto industry variables provide a modest incremental improvement in explanatory power for net migration. Most prominent among the industry variables is the addition of a large plant (that is, 1,000 employees or more) to a county. This has a significant positive influence on migration. This effect is evident both in the county that receives the plant and in those contiguous to it.

Our finding regarding the importance of auto industry variables has significant policy implications. It

suggests that development efforts aimed at retaining or attracting population will have greater immediate success if they focus on attracting larger plants. Furthermore, this result underscores the importance of cooperative efforts to obtain such plants, given that they positively affect population in a multicounty area. Future research will look more specifically at the effect on migration of the young adult population.[14]

APPENDIX

Means and standard deviations (455 nonmetropolitan counties)

	Mean	Standard deviation
Net migration 1990s	5.135	6.809
Net migration 1980s	–3.872	6.615
Control variables		
Metro adjacency	0.530	0.500
Recreational county	0.090	0.287
Retirement county	0.053	0.224
Percent employed in agriculture, 1990s	6.457	4.214
Percent employed in agriculture, 1980s	11.260	7.170
Percent employed in manufacturing, 1990s	25.570	9.729
Percent employed in manufacturing, 1980s	27.420	10.120
Percent work outside the county, 1990s	31.112	13.808
Percent work outside the county, 1980s	25.880	12.860
Population, 1990	26,602	19,014
Population, 1980	26,643	19,068
South	0.367	0.483
Unemployment rate, 1991	9.040	2.960
Unemployment rate, 1981	11.010	3.630
Auto variables		
Supplier base, 1990	1.295	2.169
Supplier addition, 1990s	0.165	0.515
Supplier base, 1980	0.822	1.526
Supplier addition, 1980s	0.473	1.108
Major auto plant base, 1990	0.029	0.191
Major auto plant addition, 1990s	0.007	0.105
Major auto plant base, 1980	0.020	0.168
Major auto plant addition, 1980s	0.009	0.093
Contiguous major plants base, 1990	0.532	1.923
Contiguous major plants addition, 1990s	0.059	0.279
Contiguous major plants base, 1980	0.413	1.792
Contiguous major plants addition, 1980s	0.119	0.419

Note: See table 7 for variable definitions.

NOTES

[1]The number of plant openings by Japanese auto parts suppliers in the U.S. peaked in the late 1980s (Klier, 1994).

[2]Except for Illinois, all the states in the auto corridor have a higher than average motor vehicle and equipment (Standard Industrial Classification 371) share of gross state product (GSP). The data are averaged over 1995, 1996, and 1997. The specific industry shares of GSP are: Illinois, 0.79 percent; Indiana, 4.91 percent; Kentucky, 5.27 percent; Michigan, 8.52 percent; Ohio, 3.66 percent; Tennessee, 2.74 percent; and Wisconsin, 1.46 percent. The U.S. average for that period is 1.09 percent (data from U.S. Bureau of Economic Analysis).

[3]Between 1970 and 1993, 43 formerly nonmetropolitan counties were redefined as metropolitan, and 13 formerly metropolitan counties were reclassified as nonmetropolitan.

[4]In all, 8.4 percent of the original database entries could not be confirmed by review of state directories nor could they be reached by phone. However, we are confident that the coverage afforded by the database is high. The employment estimates for Michigan assembly and supplier plants are only slightly below those reported by McAliden and Smith (1999) using ES 202 data.

[5]The demographic trends in the auto corridor during the past 30 years have been generally consistent with those in the nation. The only exception to this general consistency between population growth patterns in the auto corridor and the nation is that metropolitan areas of the auto corridor lost a significant amount of population during the 1970s and 1980s, whereas metropolitan areas in the U.S. as a whole generally gained population.

[6]Fourteen counties of the nonmetropolitan counties in the auto corridor are both recreational and retirement counties.

[7]It slightly overstates the concentration of independent supplier plants in Michigan by showing all plants, regardless of their age. If we were to present information on plants opened since 1980 only, Michigan's share would fall from 39 percent to 36 percent.

[8]Within a just-in-time production environment, inventories at assembly and supplier plants are being minimized, which puts a premium on being able to deliver parts on time to the assembly line. Consequently, the majority of parts shipments to assembly facilities is delivered by truck, whereas the final product is distributed across the country by a combination of truck and rail.

[9]We chose this categorization, which indirectly distinguishes plant size, because plant-level employment data are available only for 1997.

[10]In defining this variable, we take account of contiguous plants in metropolitan and nonmetropolitan counties.

[11]An F-test shows it to be significant at the 95 percent level.

[12]The difference in the estimated effect of adding a major plant and adding a supplier plant for the 1980s is approximately commensurate to the factor by which an average major plant is larger, in terms of employment, than an average supplier plant.

[13]That corresponds to an estimated net inmigration of 12.7 percent for the immediate county. The actual net migration rates for the two nonmetropolitan counties in which large plants opened during the 1990s, two Saturn facilities in Spring Hill, Tennessee, and one large independent supplier plant in central Michigan, are 22.1 percent and 6.1 percent, respectively.

[14]Data will not become available until after the release of the 2000 Census.

REFERENCES

Beale, C. L., 1975, "The revival of population growth in nonmetropolitan America," U.S. Department of Agriculture, Washington: U.S. Government Printing Office, report, No. ERS-605.

Beale, C. L., and K. M. Johnson, 1998, "The identification of recreational counties in nonmetropolitan areas of the United States," *Population Research and Policy Review*, Vol. 17, pp. 37–55.

Byerly, E. R., 1994, "Population estimates for counties and metropolitan areas: July 1, 1991," U.S. Bureau of the Census, Washington: U.S. Government Printing Office, report, No. P25-1108.

Center for Business and Economic Research, 1992, "The economic significance of Toyota Motor Manufacturing, U.S.A. Inc., in Kentucky," *Review and Perspective*, University of Kentucky, December.

Cook, P. J., and K. L. Mizer, 1994, "The revised ERS county typology: An overview," Washington: U.S. Department of Agriculture, Economic Research Service, report, No. RDRR-89.

Elhance, Arun P., and Margaret Chapman, 1992, "Labor market of a U.S.–Japanese automobile joint venture," *Growth and Change*, Vol. 23, No. 2, pp. 160–182.

ELM International, Inc., 1997, "The ELM GUIDE supplier database," East Lansing, MI, database file.

Fournier, Stephen F., and Andrew M. Isserman, 1993, "Putting it all together: The effects of the Honda plant on its host county and the rural hinterland," West Virginia University, Regional Research Institute, research paper, No. 9305.

Frey, W. H., and K. M. Johnson, 1998, "Concentrated immigration, restructuring, and the 'selective deconcentration' of the United States population," in *Migration into Rural Areas*, P. Boyle and K. Halfacree (eds.), London: Wiley, pp. 79–106.

Fuguitt, G. V., 1985, "The nonmetropolitan turnaround," *Annual Review of Sociology*, Vol. 11, pp. 259–280.

Fuguitt, G. V., D. L. Brown, and C. L. Beale, 1989, *Rural and Small Town America*, New York: Russell Sage Foundation.

Fuguitt, G. V., and T. B. Heaton, 1993, "The impact of migration on the nonmetropolitan population age structure, 1960–1990," University of Wisconsin, Department of Sociology, Madison, WI, unpublished manuscript.

Fuguitt, G. V., T. B. Heaton, and D. L. Lichter, 1988, "Monitoring the metropolitan process," *Demography*, Vol. 25, pp. 115–128.

Goetz, Stephan J., and Anil Rupasingha, 1999, "Determinants and impacts of net migration at the county-level," paper presented at the North American Regional Science Association meetings, Montreal, November 11–14.

Harbour & Associates, 1990, *The Harbour report—A decade later. Competitive Assessment of the North American Auto Industry 1979–1989*, Troy, MI.

Haywood, Charles F., 1998, "A report on the significance of Toyota Motor Manufacturing Kentucky, Inc. to the Kentucky economy," University of Kentucky, paper.

Head, Keith, John Ries, and Deborah Swenson, 1995, "Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States," *Journal of International Economics*, Vol. 38, pp. 223–247.

Johnson, Kenneth M., 1998, "Renewed population growth in rural America," *Research in Rural Sociology and Development*, Vol. 7, pp. 23–45.

_____, 1993, "Demographic change in nonmetropolitan America, 1980 to 1990," *Rural Sociology*, Vol. 58, pp. 347–365.

_____, 1989, "Recent population redistribution trends in nonmetropolitan America," *Rural Sociology*, Vol. 54, No. 3, pp. 301–326.

Johnson, K. M., and C. L. Beale, 1994, "The recent revival of widespread population growth in nonmetropolitan areas of the United States," *Rural Sociology*, Vol. 59, pp. 655–667.

Johnson, K. M., and R. L. Purdy, 1980, "Recent nonmetropolitan population change in fifty year perspective," *Demography*, Vol. 17, pp. 57–70.

Kenney, Martin, and Richard Florida, 1992, "The Japanese transplants—Production organization and regional development," *Journal of the American Planning Organization*, Vol. 58, No. 1, pp. 21–38.

Klier, Thomas H., 2000, "Does 'just-in-time' mean right next door? Evidence from the auto industry on the spatial concentration of supplier networks," *Journal of Regional Analysis & Policy*, forthcoming.

_____, 1994, "The impact of lean manufacturing on sourcing relationships," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 18, No. 4, pp. 8–19.

Long, L., and D. DeAre, 1988, "U.S. population redistribution: A perspective on the nonmetropolitan turnaround," *Population and Development Review*, Vol. 14, pp. 433–450.

Marvel, Mary K., and William J. Shkurti, 1993, "The economic impact of development: Honda in Ohio," *Economic Development Quarterly*, Vol. 7, No. 1, pp. 50–62.

McAlinden, Sean P., and Brett C. Smith, 1999, "The Michigan Automotive Policy Survey," University of Michigan, Transportation Research Institute, Office for the Study of Automotive Transportation, paper, No. UMTRI-99-1.

_____, 1993, "The changing structure of the U.S. automobile parts industry," University of Michigan, Transportation Research Institute, Office for the Study of Automotive Transportation, February.

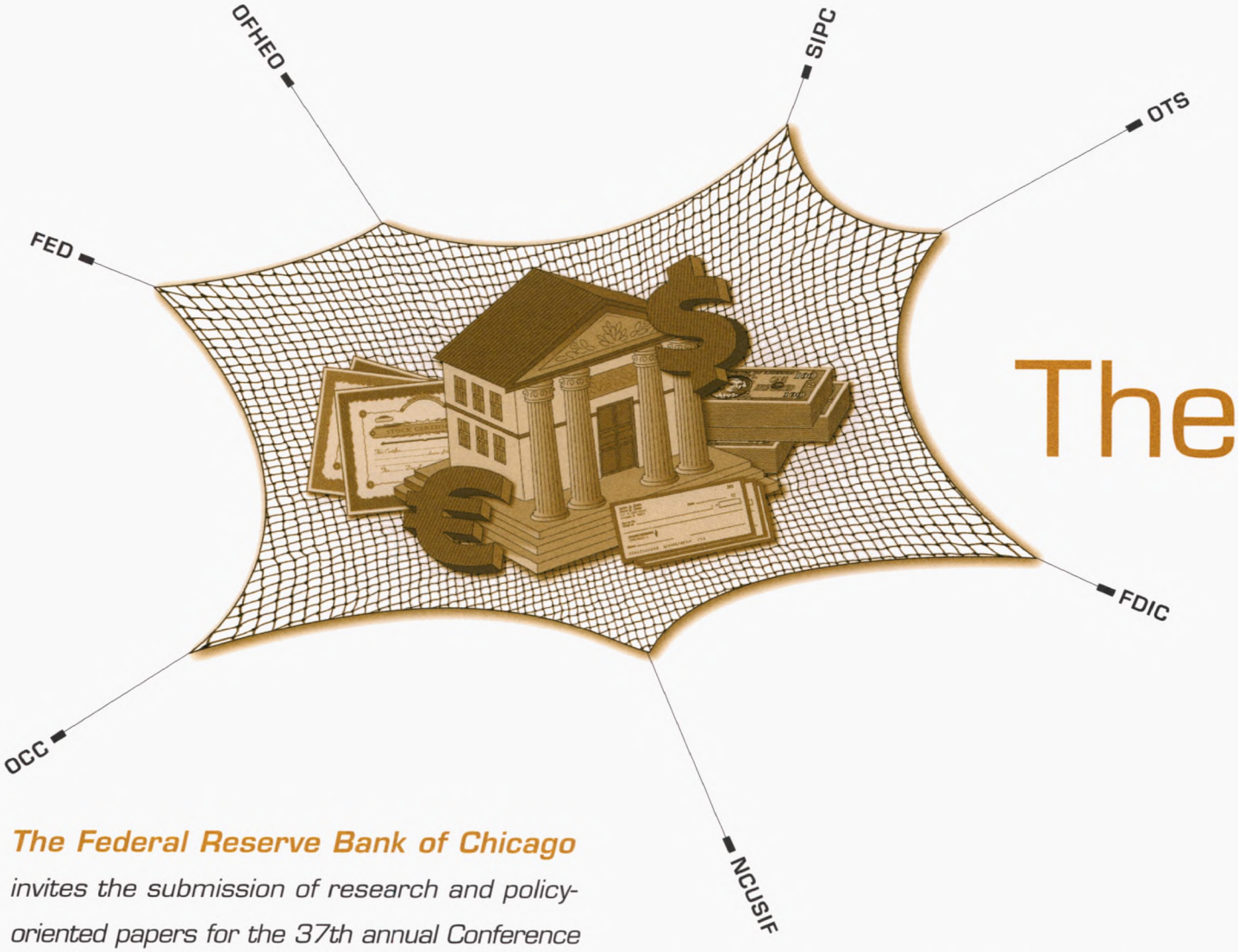
Murray, N. Matthew, Paula Dowell, and David T. Mayes, 1999, "The location decision of automotive suppliers in Tennessee and the Southeast," University of Tennessee, Center for Business and Economic Research, report.

Rubenstein, James M., 1992, *The changing US auto industry*, London: Routledge.

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

The 37th Annual Conference on Bank Structure and Competition

MAY 9–11 * 2001



The

The Federal Reserve Bank of Chicago *invites the submission of research and policy-oriented papers for the 37th annual Conference on Bank Structure and Competition to be held May 9–11, 2001, at the Fairmont Hotel in Chicago. Since its inception, the conference has aimed to encourage an ongoing dialogue on current public policy issues affecting the financial services industry. Although we are interested in papers related to the conference theme, we are most interested in high quality research addressing public policy and the financial services industry.*

The theme of the 2001 conference will be the purpose, role, and implications of the current structure of the financial industry safety net. The most obvious component of the safety net, and the one that receives most public attention, is deposit insurance. The reach of the safety net, however, is more extensive than the stated coverage extended by the deposit insurance funds. It includes both implicit and explicit guarantees, the means by which these guarantees are delivered, and the resulting behavioral changes induced by their presence.

Public discussion of these safety net issues has grown louder in recent years. There has been significant criticism of the current structure of the federal deposit insurance system and numerous recommendations for its reform. Recently the FDIC initiated what was intended to be a comprehensive review of the deposit insurance system in three major areas: pricing, fund maintenance, and extent of coverage. Input from banks, consumer groups, and trade associations has been requested and will be used in determining what adjustments, if any, should be made.

There has recently been discussion in the U.S. about reforming the lender of last resort function: the Federal Reserve's discount window. Numerous concerns have also been expressed about the appropriate role of government-sponsored enterprises (GSEs). For example, Fannie Mae and Freddie Mac have established lines of credit with the U.S. Treasury, and it is widely perceived that their debt is guaranteed by the U.S. government. This gives them a funding advantage in debt markets and in their capital requirements compared with other purely private mortgage market participants. The competitive advantages and potential taxpayer liability implied by this arrangement have brought these institutions under close scrutiny by both Congress and the industry. Similar concerns have been raised about the Federal Home Loan Banks (FHLB) as more and more banks join the FHLB System. Another area of concern is the liability guarantees available to nondepository institutions (for example, insurance companies have access to explicit state guarantee programs). These programs offer a different laboratory to examine the consequences of liability guarantee systems. Perhaps the most widely discussed feature of the safety net is the expectation that large banks are "too-big-to-fail," a perception that endures despite legislative and regulatory efforts to change it. Such implicit guarantees may encourage risk-taking by large banks, and may increase the relative funding cost of small banks.

These safety net issues have raised a number of public policy questions. Has the financial safety net expanded in recent years? Does the Gramm-Leach-Bliley Act, by allowing for broader product expansion of financial holding companies, open the door for "safety net leakages" to a host of nonbanks? If so, what are the implications of this expansion and how can it best be contained? After years of criticism, are regulators now ready to introduce true reform through an incentive-compatible deposit insurance scheme? What are the implications of such reform for small banks, which apparently have been having problems attracting the deposits that are fundamental to their success? For large banks? Has the modern-day role of deposit insurance changed? Concerning the discount window, is there a subsidy associated with the current means by which the Federal Reserve operates its credit facility? Would movement toward use of an above-market or "penalty" rate be beneficial? Do these directed liquidity injections by the central bank have advantages over general liquidity injections which can then be allocated in private markets?

Concerning implicit guarantees, how can GSE oversight best be handled? Is bank-like oversight appropriate? What is the extent of the subsidy associated with the current government guarantees? Should the activities of GSEs be constrained? And finally, do the markets believe a too-big-to-fail policy is still in effect? What are the resulting market distortions?

The 2001 conference will focus on these and related questions. There will also be a number of additional sessions on industry structure and regulation concerning topics such as:

- Financial modernization or the implications of the Gramm-Leach-Bliley Act;
- Bank capital standards;
- Fair lending issues and predatory pricing issues;
- Measuring and managing risk, particularly for transnational/global financial services companies;
- Alternative approaches to dealing with financial crises;
- Reforming the international financial institutions; and
- The implications of technology on bank delivery systems (for example, Internet banking) and payment innovations.

Financial Safety Net:
Its Role, Benefits, and Costs

Continuing the format of recent years, the final session of the conference will feature a panel of industry experts who will discuss the purpose, structure, problems, and proposed changes associated with an important and topical banking regulation. A record of the panel presentations will be included in the *Proceedings* of the conference. Past topics discussed at this session include bank antitrust analysis, bank capital regulation, optimal regulatory structures, and the appropriate role of the lender-of-last-resort. Proposals for this session are also welcome.

If you would like to present a paper at the conference, please submit four copies of the completed paper or a detailed abstract (the more complete the paper the better) with your name, address, affiliation, telephone number, and e-mail address, and those of any co-authors, by December 18, 2000. Correspondence should be addressed to:

Conference on Bank Structure and Competition
Research Department
Federal Reserve Bank of Chicago
230 South LaSalle Street
Chicago, Illinois 60604-1413

For additional information contact:

Douglas Evanoff
(312) 322–5814
devanoff@frbchi.org

or

Portia Jackson
(312) 322–5775
portia.jackson@chi.frb.org

Why do consumers pay bills electronically? An empirical analysis

Brian Mantel

Introduction and summary

Although the checkless society has been predicted for decades, checks remain the most frequently used noncash payment method in the U.S., contrary to trends in a number of other countries. Despite the debate over why consumers do or do not adopt new payments technology, little is known about the subject. Given unsuccessful efforts to induce a shift away from checks, some industry observers have even suggested that consumers are "irrationally" wedded to their checks. As a result, the financial services industry faces significant uncertainty regarding potential investments in electronic bill payment technologies as well as in debit cards, smart cards, stored value, e-cash, check imaging, and check conversion technologies. The goal of this article is to provide some insight into the consumer's decision to use electronic payments technology—What factors influence this decision and what might financial industry leaders do to encourage greater numbers of consumers to make the transition to electronic payments?

The study of payment methods is of interest for several reasons. First, technology is enabling new payment methods to be introduced more easily and frequently. As a result, the very characteristics of what constitutes a payment instrument are changing over time. Second, recent research highlights the importance of payment-related revenues to financial institutions.[1] Consequently, payment providers will continue to look for ways to increase the value of payment products to customers, thereby enhancing potential revenue streams. Likewise, companies will continue to look for ways to reduce the costs of payments (for example, by reducing the fees they pay to payment providers). For instance, checks are being converted from paper into electronic items and cleared via the automated clearinghouse (ACH) at the point of sale.[2] Firms are also considering new ways to leverage current electronic payment networks to make payments electronically, for instance, experimenting with the ACH network to make debit transactions at the point of sale[3] or using automated teller machine (ATM) networks to make debit transactions for Internet payments.[4]

Ultimately, some combination of consumers, corporations, and financial service providers will determine the success of various payment instruments. These innovations will put increasing pressure on the structure of the rights, warranties, and incentives associated with different payment instruments. Therefore, in order to make better forecasts for business planning and enhance public policy decision-making, we need to better understand the factors influencing consumer choice among alternative payment options.

This article analyzes the extent to which various factors influence consumers' willingness to use electronic bill payment. I review the economic, marketing, consumer decision-making, and payments literatures. Then, I analyze a unique 1,300-person survey to evaluate the factors associated with usage of electronic bill payment. I find that several broad factors influence the consumer's preference for electronic bill payment:

Brian Mantel is the Emerging Payment Studies Department program manager at the Federal Reserve Bank of Chicago. *The author would like to thank Dan Aaronson, David Allardice, Eric Berggren, Ed Green, Rick Kolsky, David Marshall, Kathy Paese, Ann Spiotto, Joanna Stavins, and Dan Sullivan for their comments and suggestions. The article benefited from comments from participants at the Financial Services Technology Consortium's Annual Spring Meeting, the University of Michigan's Electronic Payments Symposium, and a Federal Reserve Bank of Chicago research seminar. The author would also like to acknowledge the excellent research assistance of Patricia Rozaklis, Sonalee Shah, and Alpa Shah.*

1) wealth; 2) personal preferences for control, record keeping, convenience, incentives, personal involvement, and/or privacy; and 3) transaction-specific factors associated with different types of payments. I also find that certain demographic factors are significantly associated with the use of electronic bill payment services. My findings are consistent with new product adoption theories, supporting the idea that some consumer segments are natural "first adopters" of electronic bill payment services. However, while new product diffusion theories assume that consumers will begin to experiment and adopt innovations as they learn more about the product's features, my analysis suggests that fundamental consumer needs still must be addressed before a broader portion of consumers will adopt electronic bill payment services.

As a result, I find that an important portion of consumers do not perceive checks and some electronic bill payment services as substitutes at this time. Some analysts suggest that many consumers are likely to remain reluctant to adopt new payment technologies.[5] However, my analysis suggests that a larger fraction of consumers would adopt these new technologies if important product features such as error resolution, service level guarantees, customer service, the ability to make partial payments, and more convenient sign-up were bundled with electronic bill payment services. My results suggest that the next stage of migration towards electronic bill payment may depend more on firms' willingness to fund the development of these new product features than on overcoming consumer resistance to change. This article also highlights the need for policymakers to better understand the diversity of consumer preferences when considering public policy questions relating to consumer protection.

Overview of the payments marketplace

The payments mechanism, like the electricity power grid, is an important piece of the foundation that supports our economy. Today's payment instruments have evolved from barter to commodity-based, to currency and coin, to card-based and, more recently, to electronic network-based systems. The introduction of commodity money reduced the costs and risks associated with trade. Coins and paper currency brought greater standardization, broader acceptance, and lower transaction costs than previous commodity-money or barter-based economies. Card-based systems have extended the reach of one's wealth and creditworthiness, lowered costs, and improved access to customer information. Recent advances in technology now make further improvements possible when consumers value them and when providers have a clear business case to offer the improvement.

According to McKinsey and Company research (Stevenson, 1997), consumers initiate approximately 90 percent of all transactions. Table 1 provides an overview of the mix of different payment instruments across the U.S. economy. See MacKie-Mason and White (1996) for a detailed comparison of the different attributes associated with different payment instruments.

Two theories of how new products are adopted

There are two general, complementary theories of how new products are adopted. The first theory, the *new product diffusion model*, assumes that the primary determinant of new product adoption is the time it takes consumers to learn about a product, to experiment with it, and then ultimately to use it.[6] This theory assumes that consumers view a new product or service as a clear and valuable substitute

TABLE 1

Estimates of historical U.S. payment volumes
(Items in billions)

Payment	1995	1996	1997	1998	1999	CAGR
Currency	500	—	—	—	—	—
Postal money order	0.2	0.2	0.2	0.2	.2	1.5%
Check	63.0	64.7	66.0	67.5	68.8	2.3%
Credit card	14.9	16.1	16.9	17.5	NR	11.9%
Electronic funds transfer	10.5	11.8	12.6	13.2	13.3	6.3%
ATM	9.7	10.7	11.0	11.2	10.9	3.1%
Debit at point of sale	0.7	1.1	1.6	2.0	2.4	14.8%
ACH	3.4	3.9	4.5	5.3	6.2	16.0%

Notes: Volumes include payments initiated by business and government, in addition to those by consumers.
NR indicates not reported. Columns may not total due to rounding.
Sources: Hancock and Humphrey (1997); Federal Reserve Bank of St. Louis, *Annual Report*, Green Sheet, various years; Faulkner & Gray; National Automated Clearinghouse Association; and Bank for International Settlements.

for past products or services and that risks associated with trial can be managed by some combination of consumers, distributors, and producers. According to the new product diffusion model, if consumers perceive the new product to be a substitute for a product they currently use and understand, providers can more easily leverage existing distribution and communications channels to generate awareness and demand for the innovation.

The second theory, the *new market development model*, suggests that a new product by itself will have a limited market potential. In order to reach mass consumer markets, firms need to offer additional product features, services, and/or infrastructure over time, tailoring the product to new customer segments and/or to new uses, as well as making products interoperable.[7] Under this theory, new products are introduced and evolve, new features are added, and over time the product reaches a mass and mature stage of acceptance.

The first model suggests a heavy focus on building awareness and trial while the second theory suggests staged introduction of new product features to new and different customer segments. Thus, it is critical to assess whether consumers perceive new payment innovations to be substitutes for past products or whether new innovations are viewed as fundamentally new products, requiring significantly more resources to promote adoption.

Literature review

Consumer payments decision-making

In an extensive survey of the payments literature, Hancock and Humphrey (1997) provide an overview of the factors associated with electronic banking adoption, including incentives, the nature of a country's financial infrastructure, and the role of network economics in electronic banking adoption. Using a longitudinal Norwegian survey (1989–95), Humphrey, Kim, and Vale (1998) conclude that efficient payment instrument pricing would induce greater electronic payment use because of its lower cost relative to paper-based payments.

Using the Federal Reserve's 1995 *Survey of Consumer Finances*, Kennickell and Kwast (1997) analyze the influence of demographic characteristics on the likelihood of electronic payment instrument usage. As shown in table 2, higher levels of education and financial assets increase the likelihood of electronic payment usage. Carow and Staten (1999) investigate consumer preferences among debit cards, credit cards, and cash for gasoline purchases. Higher levels of education and income and having more than one credit card are associated with greater use of credit cards than cash. However, convenience, rather than borrowing capacity, was the greatest determinant of a credit card user. Lastly, Carow and Staten find that a consumer's ownership of credit cards and use of credit cards are related, because having a certain type of account reveals payment preferences. The American Bankers Association and Dove Associates (1999) analyze a survey of 1,400 consumers to investigate the factors motivating consumer payment instrument choice between online and offline debit. They find that consumers exhibit strong and distinct payment preferences, with different segments of consumers valuing different debit attributes.

Wells (1996) finds that check float does not explain the persistence of consumer check use; alternative explanations include the consumer perception of checks and ACH as dissimilar payment instruments, market failure, and measurement errors. Using 1997 data to investigate consumer responsiveness to changes in checking account costs, Stavins (1999) finds that the supply of bank deposits to checking accounts is sensitive to banks' per item fees and check return, teller, and foreign ATM restrictions. Using the Federal Reserve Board's *Terms of Credit Card Plans Survey* to investigate consumers' willingness to pay for credit card service, Stavins (1996) finds that consumers respond to product offerings that bundle other services. Research suggests that despite the fact that banks could earn higher revenues by lowering

TABLE 2

Factors in use of payment technology

	Income	Financial assets	Age	Education
In person	–	+	0	0
Mail	+	+	–	+
Telephone	+	+	0	+
Electronic transfer	0	+	–	+
ATM	0	+	–	+
Debit card	0	+	–	+
Automatic deposit/withdrawal	–	+	+	+
Direct deposit	–	+	+	+
Pre-authorized debit	0	+	–	+
Computer	0	+	–	+
Smart card	0	0	0	+

Note: Statistically significant positive/negative factor (+/–); not statistically significant factor (0).
Source: Kennickell and Kwast (1997).

the price of services, they would not necessarily maximize the profit from each account.

MacKie-Mason and White (1996) provide a detailed review of the characteristics that are important to consider when designing new payment innovations. Mantel (2000) surveys the literature on consumer payment decision-making and proposes a framework in which three factors explain consumer electronic banking usage: 1) wealth; 2) personal preferences, such as incentives, convenience, control, budgeting, privacy, security, and personal involvement; and 3) transaction-specific factors. Including this broad list of factors helps explain sometimes hard-to-explain or inconsistent behaviors. For instance, Mantel's (2000) framework helps explain why consumers are increasingly choosing to use debit cards, based on the changes in the attributes financial institutions have begun bundling with debit cards, although credit cards are well known for providing convenience and short-term, "interest-free" loans. Similarly, this framework helps explain why consumers in different countries have adopted smart cards at significantly different rates, again based on the nature and/or importance of the attributes bundled with these payment products.

Consumer awareness

New product diffusion theories point to the important role of consumer awareness in promoting adoption. There is relatively little public data on consumer awareness and perceptions of electronic bill payment. A 1998 Federal Reserve Bank of St. Louis study finds that 99 percent of consumers say they understand direct deposit and 97 percent of current users report satisfaction with the system. However, only 55 percent of consumers feel they understand electronic bill payment and ACH well, while 84 percent of electronic bill payment users report satisfaction with this type of payment instrument. A study for the New York Clearing House (1997) conducted by Wirthlin Worldwide measures direct deposit usage before and after a marketing campaign was employed from September 1996 to February 1997. Roughly half of all nonusers surveyed remembered the principal messages of the campaign, including the ideas that direct deposit is convenient (18 percent), easy to use (17 percent), and available (16 percent). However, the study does not find evidence that communication efforts increase usage.

While the fact that a significant fraction of consumers may not fully understand electronic bill payment services might indicate a problem to some, the new market development model might suggest that this is not a problem per se. After all, if the evidence continues to suggest that an important fraction of consumers do not yet perceive electronic bill payment and checks as clear substitutes, it may be that the electronic bill payment market is still developing. In this case, a significant fraction of consumers will likely continue to report a lack of familiarity, even as significant improvements are made over time, until the product's functionality is fully developed. When future studies find evidence that a larger portion of consumers see electronic bill payment and checks as clear substitutes, then firms then may be better able to target their communications campaigns to consumers' unique needs. Clearly, communications efforts in the early stages of a product life cycle will continue to be important; nonetheless, they will likely serve different purposes than communications efforts used for more mature products.

Analysis

Description of data and variables

I use a dataset collected by Vantis International on behalf of the Federal Reserve Bank of St. Louis and the Federal Reserve Bank of Atlanta that consists of responses to a national 1,300-person survey on consumer decision-making among competing bill payment instruments. The household's primary bill payer served as the survey respondent. The survey collected data pertaining to consumer demographic characteristics, payment behaviors, and self-reported payment preferences and evaluations of different payment options. The survey focused primarily on consumers' experiences with checks and electronic bill payment, but also considered other payment instruments such as debit cards, credit cards, and money orders. The appendix provides descriptions and summary statistics of the variables included in the analysis.[8] I expect these factors to influence the likelihood of electronic bill payment use generally and/or the likelihood of high usage. For ease of interpretation, I group them into the following broad categories: demographic, new product adoption, control and budgeting, convenience, incentives, privacy and security, and personal involvement.

Model

In this article, I analyze why consumers choose among alternative payment instruments for bill payment. This analysis uses a series of *binomial logistic regressions*, a statistical technique that allows one to examine the extent to which various factors influence the likelihood of direct bill payment usage. First, I explore the factors that affect the initial choice of whether to use electronic bill payment. Second, I investigate the factors that influence the extent or frequency of electronic bill payment use among users of these

types of services. For this part of the analysis, I classify consumers as "low users" if they pay fewer than 20 percent of their bills electronically and "high users" if they pay more than 30 percent of bills with direct bill payment technology.[9] Third, I examine the factors that influence the use of electronic bill payment for specific types of bills—mortgage, loan, lease, telephone, cable, credit card, and insurance.

The regression model analyzes consumer C_i's choice of payment instrument for bill payment P_i (for example, mortgage payment, credit card payment, or telephone bill).[10] The consumer has two payment options—1) paper-based payment instruments such as cash, checks, or money orders and 2) electronic-based payment instruments such as electronic bill payment and ACH.[11] I impose several simplifying assumptions on the analysis. First, I assume that all payments initiated electronically by the consumer are paid electronically by the payment intermediary. This assumption allows the analysis to focus only on consumers' willingness to choose electronic payments. Next, I assume that consumers have access to all payment options. Therefore, findings are applicable only to the 90 percent of consumers who have checking accounts. This article does not address the important issue of identifying the needs and preferences of unbanked consumers.

Limitations of the analysis

First, my analysis focuses solely on consumer decision-making pertaining to electronic bill payment and does not address the perspective of the payment provider or the entity receiving the payment. I am primarily interested here in whether consumer preferences are a significant limiting factor to the migration towards electronic payments. Second, because this model evaluates the attributes of desirable payment instruments using self-reported data, my results depend on the accuracy with which consumers recall and report actual behavior. Third, the survey focuses on the primary bill payer rather than households in general. As a result, one must exercise care in extending these results to the general population. Finally, this research does not consider how consumer behavior has changed over time nor does it provide insight into how and when a specific factor (such as price, product attributes, or promotions) induces changes in electronic bill payment use.

Empirical results

Table 3 provides the results of binomial logistic regressions comparing nonusers of electronic bill payment with users and low users with high users. Table 4 provides the results of binomial logistic regressions comparing nonusers and users by type of bill (that is, cable, credit card, insurance, loan, mortgage, telephone, and utility). Overall, four general findings emerge. First, there are important and significant differences between nonusers, low users, and high users of electronic bill payment. Second, consumer demographic and financial characteristics are important in influencing whether consumers use electronic bill payment. However, these factors do not strongly distinguish high users from low users, lending support for the idea that there are certain natural first adopters of electronic bill payment technology. Third, several consumer preferences, including the desire for control, convenience, incentives, privacy, and personal involvement, are also significant in whether consumers use electronic payments. Furthermore, these factors distinguish low users from high users. At a more detailed level, this analysis suggests that incentives may be a valuable tool to induce introductory usage of electronic bill payment, although they may not be needed to increase usage. Similarly, services that promote greater consumer recourse and control may be important in inducing some consumers to adopt electronic bill payment and others to expand the number of bills they pay electronically. Finally, payment-specific factors, such as the dollar size of a payment and whether a payment amount varies, are also important in explaining why certain consumers choose electronic bill payment or paper checks to pay for certain bills.

Comparison of users and nonusers

Referring to table 3, I find that several demographic factors influence the initial decision of whether to use direct payment methods. Holding other factors constant, older individuals are more likely to use direct electronic payments than younger individuals. For instance, a 40-year-old is 2.3 percent more likely to use electronic bill payment than a 39-year-old. Women are 49.2 percent more likely to use electronic bill payment. Neither college education nor market size variables are statistically significant factors in influencing the use of direct electronic payment. The absence of a statistically significant education variable differs from Kennickell and Kwast's (1997) findings, but is attributable to the inclusion of other nondemographic factors, such as personal preferences, income, and lifestage (for example, single, married with children, or retired), that are related to education level.[12]

New product adoption factors do play a role in influencing the likelihood of electronic bill payment usage across the proposed factors. Consumers who

TABLE 3

Binomial regression results, odds ratio

	Nonusers vs. users	Low vs. high users
Demographics		
Age	1.023*** (0.007)	1.031*** (0.008)
Female	1.492** (0.196)	0.623** (0.235)
Race	1.013 (0.259)	1.570 (0.331)
College	0.910 (0.193)	1.031 (0.219)
Market size: under 100,000[a]	1.441 (0.233)	0.741 (0.278)
Market size: 100,000–499,999	1.241 (0.231)	0.914 (0.271)
Market size: 500,000–1,999,999	1.121 (0.251)	0.864 (0.303)
New product adoption		
Understand direct payment	1.768*** (0.135)	0.767 (0.199)
Understand set-up of direct payment	1.122*** (0.033)	1.029 (0.036)
PC owner	2.106*** (0.195)	1.085 (0.217)
Cellular phone owner	1.159 (0.204)	0.762 (0.234)
Internet purchase	0.902 (0.285)	0.997 (0.340)
Consumer financial		
Household income: $20,000–$39,999[b]	2.144*** (0.239)	1.257 (0.312)
Household income: $40,000–$74,999	1.836** (0.257)	1.088 (0.324)
Household income: over $75,000	2.038** (0.352)	2.382** (0.432)
Homeowner	1.439* (0.208)	1.268 (0.276)
Savings account	1.264 (0.214)	1.492 (0.266)
Credit card	1.485** (0.179)	0.671* (0.222)
Regional/national bank account[c]	1.154 (0.186)	0.878 (0.216)
Credit union account	1.630*** (0.183)	0.708 (0.215)
Brokerage account	0.984 (0.269)	0.794 (0.302)
Savings and loan account	0.755 (0.220)	0.913 (0.262)
Control		
Control when bill is paid	0.898* (0.061)	0.873** (0.067)
Option to stop payment	1.093** (0.041)	1.001 (0.050)
Receipt for payment	0.891*** (0.036)	0.989 (0.041)
Balance checkbook once/month	1.050 (0.033)	0.929* (0.042)
Disciplined about finances	0.988 (0.034)	1.047 (0.043)
Use toll-free number to check balances	0.951* (0.027)	1.031 (0.032)
Person available	0.989 (0.046)	0.879** (0.056)
Convenience		
Bill paid when out of town	1.065* (0.037)	1.099* (0.051)
Saving time	1.061 (0.047)	0.980 (0.066)
Banks not open convenient hours	1.007 (0.031)	1.106*** (0.038)
Incentives		
Least expensive payment method	1.025 (0.046)	1.065 (0.064)
Use shopping coupons	1.071** (0.033)	1.042 (0.040)
Avoid penalties for late payments	0.977 (0.055)	1.098 (0.074)
5% discount	1.882*** (0.241)	1.060 (0.379)
Privacy/security		
Not comfortable giving account number to salesperson	0.984 (0.027)	1.039 (0.224)
Dislike automatic withdrawal	0.806*** (0.031)	0.914** (0.013)
Credit card on Internet is secure	0.965 (0.037)	0.900** (0.044)
Personal involvement		
Enjoy talking with bank teller	1.082** (0.031)	1.069* (0.036)
Number in sample	956	556
Likelihood ratio	387.6***	120.4***

[a]Market size reference variable baseline characteristic: under 100,000.
[b]Income reference variable baseline characteristic: $20,000 and under.
[c]Financial institution reference variable baseline characteristics: local bank.
*Indicates 0.10 statistical significance level; ** indicates 0.05 significance level; and *** indicates 0.01 significance level.
Notes: Standard errors are in parentheses. Of the 956 cases, 400 are nonusers. Of the 556 users of electronic payments, 236 are low users and 320 are high users. See appendix for description of variables.

TABLE 4

Statistically significant results by bill type

	Cable	Credit card	Insurance[c]	Loan[d]	Mortgage[e]	Telephone	Utilities[f]
Demographic factors							
Age		+**	+***		+***		+*
Race				+***		–*	
Lifestage				+**	–*		+**
Market size			+*				+*
New product adoption							
Understand direct payment	–**		+***				
Understand set-up of direct payment		+**	+*		+**	+**	+***
PC owner			+***				
Purchase on Internet		+***					
Consumer financial							
Household income: $20,000–$39,999[a]						+*	
Household income: $40,000–$74,999						+**	
Household income: over $75,000					+**		
Homeowner	+*			+**		+**	+**
Savings account				+***			
Regional/national bank account[b]				–**			+*
Credit union account	–*						
Control							
Control when bill is paid					–**	–***	–**
Option to stop payment			+*				
Receipt for payment			–***				
Balance checkbook once/month				+*			
Disciplined about finances						–*	+***
Use toll-free number to check balances						–*	
Person available				–**	–***		
Convenience							
Bill paid when out of town			+*		+*	+**	+***
Saving time							
Banks not open convenient hours		+*		+**			+**
Incentives							
Least expensive payment method							
Use shopping coupons							
5% discount	–**		+**	+*			
Privacy/security							
Not comfortable giving account number to salesperson	–***	–***				–**	
Dislike automatic withdrawal	–**		–***	–***		–***	–***
Credit card on Internet is secure			–**				
Personal involvement							
Enjoy talking to bank teller	+*	+***	+*	+*		+**	+**

[a]Income reference variable baseline characteristic: $20,000 and under.
[b]Account reference baseline characteristic: local/community bank.
[c]Market size is under 100,000 people.
[d]Lifestage is young couple, retired couple, or middle parent.
[e]Lifestage is middle-aged single.
[f]Lifestage is middle-aged single, older single, or retired couple. Market size is 100,000–499,999.
*Indicates 0.01 significance level; ** indicates 0.05 significance level; and *** indicates 0.10 significance level.
Notes: "+" Indicates variable increased odds of electronic payment use; "–" indicates variable decreased odds.
See appendix for description of variables. Lifestage denotes period in life. For example, young single, middle-aged married with children, or elderly with a roommate are lifestage categories.

own personal computers are twice as likely to use electronic bill payment. Consumers who report "understanding direct payment technologies" are 79 percent more likely to use electronic bill payment. Likewise, "understanding how to set up direct payment" is associated with a 12 percent greater likelihood of using electronic bill payment. However, these findings regarding consumer awareness are contrary to the results of the New York Clearing House (1997) study that finds no correlation between increased direct electronic payment use and consumers' understanding of key communication messages. Furthermore, I find that "cellular phone ownership" and a self-reported "perception that the Internet is secure for purchases" are not associated with greater usage of electronic bill payment. Clearly, we need to know more about how and why consumers choose to adopt new electronic payment technologies.

Consistent with Kennickell and Kwast (1997) and others, I find a significant relationship between consumer financial characteristics—income and home ownership as imperfect proxies for wealth—and increased electronic bill payment usage. Relative to households with incomes under $20,000, consumers at all other income levels are approximately twice as likely to use direct payment technologies. I also find that homeownership is positively related to the use of direct electronic payment methods, with homeowners being 44 percent more likely to use electronic bill payment. Consumers with credit cards are 49 percent more likely to be electronic bill payment users. Members of credit unions are 63 percent more likely to use direct electronic bill payment than other consumers. It may be worth exploring why this occurs. For instance, are there synergies between credit unions and their members' workplaces that promote better communication, make sign-up easier, or make customer service and error resolution more effective or credible? Are users of credit unions fundamentally different from other consumers?

I find that preferences for greater "control" over payments and household finances have a significant influence on the likelihood of direct electronic payment usage. As the importance placed on "having control over when a bill is paid" increases, the likelihood of electronic bill payment use decreases by 11 percent. Increased importance of "receiving a receipt of payment" is associated with an 11 percent lower likelihood of being a user of electronic bill payment methods. Consumers who report "using toll-free telephone numbers to check account balances" are 5 percent less likely to use electronic bill payment. These factors prevail over measures of financial behaviors such as "self-reported financial discipline" and the "frequency of checkbook balancing." I find one result to the contrary. The greater the importance consumers place on "the ability to stop a payment," the higher the likelihood of electronic payment use by 9 percent. While this is a surprising result, the survey instrument does not allow one to distinguish between consumers who pay bills electronically through automatic, preauthorized electronic debits and those who pay bills through electronic banking packages. The latter form of electronic bill payment gives consumers some improved ability to stop payments, which would be consistent with the above finding.

Contrary to a priori expectations, I do not find strong support for the importance of increased convenience in distinguishing users of electronic bill payment from nonusers. Only one hypothesis for this category is slightly significant (at the 10 percent level). Those who report a higher importance of "having bills paid when out of town" are 6.5 percent more likely to use electronic bill payment. These findings are not surprising, given that many bills offer grace periods of several weeks and convenience may be relevant only for intense electronic bill payment users. Indeed, preferences for convenience do influence the level of use of electronic bill payment, a topic I explore in the next section.

Some incentive factors influence the likelihood of electronic bill payment use. Consumer preference for "using the least expensive method of payment" and importance placed on "avoiding penalties due to late payments" do not influence the probability of electronic bill payment use. Consumers who report they "would use direct payment if offered a 5 percent discount on the monthly bill" are 88 percent more likely to use electronic bill payment. The heavier "use of coupons when shopping" is associated with a 7 percent greater likelihood of using electronic bill payment.

Self-reported preferences for privacy and security are generally not significant in determining the likelihood of choosing direct payments. Self-reported "dislike of the idea of someone automatically withdrawing money from one's bank account" is associated with a 20 percent decreased likelihood of electronic bill payment use. Consumer perceptions of "Internet transaction security" and self-reported "discomfort with giving account information to sales representatives" do not influence the decision to use electronic bill payment. Contrary to popular belief, other factors that reflect consumers' potential distrust of technology and security, such as "distrust of ATMs" and self-reported "preferences for account privacy," are not statistically significant determinants of direct payment

use. Nonetheless, I do find that privacy concerns influence the intensity of direct payment use, as discussed in the next section.

Preferences for personal involvement are associated with an increased probability of being a direct payment user. Contrary to prior expectations, consumers who report they "enjoy talking with bank tellers" are 8 percent more likely to be electronic bill payment users. However, this finding may be the result of segmentation across consumer choice of banking institutions. For instance, consumers who choose to use direct payment may have stronger relationships with their banks, making them more likely to report enjoying interactions with bank personnel. This finding warrants further investigation, given the importance of the related electronic banking product strategy, customer service, and branch infrastructure issues that financial services industry leaders face.

Comparison of low users and high users

Table 3 also highlights important differences between low and high users of electronic bill payment. As noted earlier, low users pay less than 20 percent of all bills using electronic bill payment and high users pay 30 percent or more of their bills electronically. Of the demographic variables included in the analysis, only age is statistically significant. A 38-year-old consumer is 3 percent more likely to be a high electronic bill payment user than a 37-year-old. While women are more likely to be direct payment users, they are 40 percent less likely to be high users of electronic bill payment. Race, educational attainment, and market size are not statistically significant factors.

The new product adoption factors proposed, a priori, to influence the likelihood of high electronic bill payment use are not statistically significant. This might suggest that new product adoption factors influence the likelihood one will adopt the technology in the first place, but do not affect how much one uses it upon adoption.

In terms of wealth and financial variables, households with incomes over $75,000 are three times more likely to be high users of electronic bill payment than households with incomes below $20,000. Households with incomes between $20,000 and $75,000 are not more likely than low-income consumers to be high users of electronic bill payment. This emphasizes the critical role that wealth and budgeting play in enabling electronic bill payment. Home ownership is not statistically significant in explaining the differences between low and high users. Account ownership at different types of financial institutions also does not influence the intensity of electronic bill payment use. While owning a credit card increases the likelihood that one uses electronic bill payment, perhaps as a proxy for financial activity, individuals who own at least one credit card are 33 percent less likely to be high users of electronic bill payment than low users. One explanation for this is that the analysis needs to account for the outstanding dollar value on credit cards as well. I tested the potential role of credit card debt levels, but found it to be insignificant. It is possible that consumers systematically misreport their level of credit card balances and/or that the survey instrument does not allow for adequate variation in responses. Clearly, we need to know more about this.

In terms of control-related factors, several variables are statistically significant in distinguishing low users from high users. The greater the importance placed on "control over when a bill is paid," the lower the likelihood that a consumer is a high user of electronic bill payment by 13 percent. A self-reported importance rating of "a person being available if a problem arises" is associated with a 13 percent lower likelihood of being a high user of electronic bill payment. Individuals who agree with the statement that they "balance their checkbooks at least once per month" are 9 percent less likely to be high users of electronic bill payment. While this finding points to the potential convenience, it may be important to better understand consumer budgeting practices in the context of using electronic banking services. To what extent do consumers rely on PC-based balancing of accounts? How much emphasis do consumers now place on managing traditional checking/transaction accounts?

Of the convenience factors in the survey, consumers who place higher ratings on the statement that "bank hours are inconvenient" are 11 percent more likely to be high users of electronic bill payment. Consumers who place importance on "having a bill paid when out of town" are 10 percent more likely to be high electronic bill payment users. Preferences for "saving time" do not influence the likelihood of high electronic bill payment use.

I find that incentive factors do not influence the likelihood of high electronic bill payment use. As with new product adoption factors, this might suggest that incentives influence the initial decision between traditional payment instruments and electronic bill payment, but do not affect the degree to which a consumer later uses electronic bill payment.

My results show that privacy concerns and preferences for personal involvement influence the likelihood of high electronic bill payment use. Self-reported "dislike of the idea of someone automatically withdrawing money from one's account" decreases the likelihood that a consumer is a high user of electronic

bill payment by 9 percent. The belief that the "Internet is secure for purchases" is associated with a 10 percent lower likelihood of high electronic bill payment use. This counterintuitive result is consistent with the findings that younger consumers, who are more likely to view the Internet as secure, are less likely to pay bills electronically because of lower incomes and other factors correlated with lifestage (for example, moving more often than older consumers and having to change payment relationships). Lastly, similar to the findings for the initial electronic bill payment choice, consumers who report that they "enjoy talking with bank tellers" are 7 percent more likely to be high electronic bill payment users.

Comparison of users and nonusers by bill type

Table 4 provides the statistically significant results of my binomial logistic regressions of nonusers versus users of electronic bill payment by bill type. The bill types I consider are cable, credit card, insurance, general consumer installment loan, mortgage, telephone, and utilities.[13] In terms of demographic and new product adoption factors, older consumers are more likely to use electronic bill payment than younger consumers for credit card, insurance, mortgage, and utility bills, though not for cable, loans, and telephone bills. Nonwhite consumers are significantly less likely to pay telephone bills electronically, but more likely to pay loan bills via electronic bill payment than white consumers. Lifestage factors prove to be significant in explaining the use of electronic bill payment for certain types of bills. Young and retired couples and middle-aged parents are more likely to pay loans electronically; middle-aged singles are less likely to pay mortgages electronically; and middle-aged and older singles and retired couples are more likely to pay utility bills electronically.

Consumers living in markets with fewer than 100,000 people are more likely to pay insurance bills electronically. Consumers in markets with population between 100,000 and 500,000 are more likely to pay utility bills electronically. Contrary to popular belief, in no case does living in a very large urban area increase the likelihood of electronic bill payment use. Future research will likely want to investigate such questions as whether institutions in smaller markets tend to promote electronic bill payment more frequently or whether consumers perceive that billing authorities in these markets provide better service. The extent to which a consumer understands how to set up direct payment positively influences the likelihood of paying credit card, insurance, mortgage, telephone, and utility bills electronically. Consumers who have made purchases over the Internet are also more likely to use electronic bill payment for credit card bills.

Relative to consumers with incomes below $20,000, higher income levels are associated with a greater likelihood of using electronic bill payment for mortgages and telephone bills. I find that homeowner status increases the probability of using direct payment for cable, loan, telephone, and utility bills. Savings account ownership increases the likelihood of making loan payments electronically.

Consumers' self-reported preference for "controlling when a bill is paid" decreases the likelihood of paying mortgage, telephone, and utility bills electronically. Control over when a bill is paid may be related to consumer budgeting concerns and a desire to minimize the risk of insufficient funds. This concern could also stem from a consumer's preference to review a variable bill and minimize the risk of errors. Having the "option to stop payment" increases the likelihood of electronic bill payment for insurance bills, while the importance placed on receiving a receipt of payment decreases the likelihood of electronic payment for insurance. This finding underscores the notion that consumers want receipts and control over bills they see as "critical," so as to avoid potentially larger bills in the future or the potential loss of insurance coverage. Higher self-reported scores for "financial discipline" and "use of toll-free telephone numbers to check account balances" decrease the likelihood of paying telephone bills electronically. This finding may be explained by the variable-dollar nature of some bills. Consumers who check account balances may be more likely to be financially constrained and less likely to use electronic bill payment if they worry about having sufficient funds to cover variable bills.

The greater the importance placed on "having a bill paid when out of town," the greater the likelihood of using electronic bill payment for insurance, mortgage, telephone, and utility bills. The extent to which a consumer believes that "bank hours are not convenient" increases the probability of paying credit card, loan, and utility bills electronically. Incentives such as a "5 percent bill discount if electronic payment is used" positively influence the likelihood of paying insurance and loan bills electronically. Higher levels of "discomfort associated with giving one's account number to a salesperson" and "dislike of someone automatically withdrawing funds from one's account" decrease the likelihood of paying cable, credit card, insurance, loan, telephone, and utility bills electronically. Contrary to prior expectations, the more a consumer reports he or she "enjoys talking with bank tellers," the greater the likelihood of paying cable, credit card, insurance, loan, telephone, and utility bills electronically. The results for cable bills across

the board indicate that consumers view them differently from all other bill types.

Summary of model results

The results of the model suggest that there are important differences between nonusers, low users, and high users of electronic bill payment services. Heavy users of electronic bill payment tend to be wealthier and to place a higher premium on convenience. Moderate users of electronic bill payment have at least modest levels of wealth and tend to value convenience. Nonetheless, these individuals do not use electronic bill payment for a broad segment of bills because of the potential risk of errors, which could result in overdrawn checking accounts or require significant time and energy following up with financial institutions and merchants. More importantly, these consumers are also subject to periodic swings in incomes or expenses that create budgeting challenges. For this broad group of consumers, the choice not to use electronic bill payment is akin to the purchase of a low-cost insurance contract that limits the potential risk of payments-related problems.[14]

Why do other consumers not use electronic bill payment? Clearly, some consumers expect some sort of incentive to change. These consumers shop for the best deal and may not change until they receive a benefit, particularly if they believe their institution is benefiting by moving to a more efficient form of payment. There are also consumers who do not pay electronically because it is not convenient enough or because some bills cannot be paid electronically. But more importantly, there is a significant fraction of consumers who do not use electronic bill payment because they lack the financial resources to even consider paying electronically. Some low-income consumers may use the ability to avoid paying a bill on time as a short-term funding vehicle, sometimes at low cost if there are limited penalties associated with late payment. Other consumers prefer a personal involvement in bill payment or seek to limit potential risks to their privacy by avoiding the use of electronic bill payment. These groups, some of which are of potentially significant size, do not yet perceive checks and electronic bill payment services as clear substitutes.

Conclusion

This analysis highlights the importance for policymakers to understand consumers' varying preferences and needs.[15] After all, consumers' desires for "control" vary significantly and encompass concerns about the ability to review bills, initiate payments, and have errors resolved. Public sector involvement in the rights, warranties, consumer protections, and incentives associated with different payment instruments may have significant implications for the adoption of electronic payments. To some extent, legally mandated business practices and consumer protection may motivate increased adoption of electronic payments.[16] An argument can be made that there is a positive externality in setting standards or in developing common rules for consumer protection, particularly in an industry with significant fragmentation and perhaps uneven bargaining power between consumers and financial institutions.[17]

Yet, we must recognize that setting these types of rules may in some cases bring costs as well as benefits. For instance, rules on what firms must do to resolve errors may have the effect of implementing a price floor, which may lead to the unintended result that it is uneconomical or unprofitable to serve some consumer segments.[18] One potential alternative is to have public entities work to coordinate the development of a reasonably small number of standards rather than one standard. The net effect would be a greater emphasis on transparency and disclosure and less on public determination of the final outcome.[19] Clearly, more needs to be known about the costs and benefits of potential public policy decisions. At a minimum, frameworks like the one presented in this article help identify where public policy decisions may be expected to have an effect, as well as where unintended consequences may arise.

This analysis suggests that, despite speculation to the contrary, consumers may not be as resistant to new payment innovations as has been proposed in the past. My results show that consumers' choices are consistent with their preferences. These preferences vary across bills and depend on the consumer's level of wealth, but include elements of preferences for control, convenience, incentives, privacy, and personal involvement. Consumers' financial positions and transaction-specific characteristics clearly have a significant impact on their decisions. The importance of these factors may help explain why consumers sometimes appear to exhibit "irrational" behavior, that is, behavior that is not consistent with self-reported preferences. This behavior may be driven by situational factors. My work suggests that the next stage of migration towards electronic bill payment may be more dependent on establishing the business cases to justify investment in new product features that address consumer preferences than on overcoming consumer resistance to change.

There is an important need to perform this type of research on data representing actual consumer behavior rather than on self-reported data. There are many

unanswered questions for future research to address, some of which I raised earlier in this article. Researchers may also want to assess the links between consumer income, expenditures, savings, borrowing, and payment methods. How are consumers' payment preferences changing over time and, more specifically, how are they responding to market stimuli, such as pricing, advertising, and changes in product attributes? Which payment instruments are substitutes and for which consumer segments? How do consumers perceive the relative merits of different payment instruments and different account structures? Finally, how do public policy decisions influence the migration to alternative payment methods?

APPENDIX: VARIABLE DESCRIPTIONS

Variable	Scale	Summary statistics
Demographics		
Age of respondent	continuous	Mean: 50.4
Female	0: male, 1: female	34% male 66% female
Race	0: white, 1: non-white	86% white 14% non-white
College	0: no, 1: yes	64% no 36% yes
Market size: under 100,000 people	0: no, 1: yes[a]	23% under 100,000
Market size: 100,000–499,999	0: no, 1: yes[a]	16% 100,000–499,999
Market size: 500,000–1.9 million	0: no, 1: yes[a]	21% 500,000–1 million
New product adoption		
Understand direct payment	1: no–4: yes	Mean: 3.5
Understand how to set up direct payment	1: disagree completely– 10: agree completely	Mean: 4.2
PC owner	0: no, 1: yes	71% no 29% yes
Cellular phone owner	0: no, 1: yes	75% no 25% yes
Purchase over the Internet	0: no, 1: yes	88.9% no 11.1% yes
Consumer financial		
Household income: $20,000–$39,999	0: no, 1: yes[b]	30% under $20,000 29% $20,000–$39,999
Household income: $40,000–$74,999	0: no, 1: yes[b]	30% $40,000–$74,999
Household income: over $75,000	0: no, 1: yes[b]	11% over $75,000
Homeowner	0: no, 1: yes	24% no 76% yes
Savings account	0: no, 1: yes	25% no 75% yes
Credit card	0: no, 1: yes	38% no 62% yes
Regional/national bank account	0: no, 1: yes[c]	36% regional/national bank 54% local bank
Credit union account	0: no, 1: yes[c]	41% credit union
Brokerage account	0: no, 1: yes[c]	12% brokerage
Savings and loan account	0: no, 1: yes[c]	20% savings and loan

Variable	Scale	Summary statistics
Control		
Control when bill is paid	1: not important– 10: extremely important	Mean: 8.9
Option to stop payment	1: not important– 10: extremely important	Mean: 8.1
Receipt for payment	1: not important– 10: extremely important	Mean: 7.8
Balance checkbook at least once/month	1: disagree completely– 10: agree completely	Mean: 8.3
Disciplined about finances	1: disagree completely– 10: agree completely	Mean: 7.4
Frequently use toll-free number to check account balances	1: disagree completely– 10: agree completely	Mean: 4.4
Person available to talk to if there's a problem	1: not important– 10: extremely important	Mean: 8.4
Convenience		
Bill paid even when I'm out of town	1: not important– 10: extremely important	Mean: 7.2
Saving time	1: not important– 10: extremely important	Mean: 7.7
Banks not open convenient hours	1: disagree completely– 10: agree completely	Mean: 4.6
Incentives		
Least expensive payment method	1: not important– 10: extremely important	Mean: 7.6
Frequently use shopping coupons	1: disagree completely– 10: agree completely	Mean: 7.8
Avoid penalties for late payments	1: not important– 10: extremely important	Mean: 8.9
Would use electronic payment if offered a 5% discount on monthly bill	0: no– 1: yes	21% no 79% yes
Privacy/security		
Not comfortable giving my account number to salesperson	1: disagree completely– 10: agree completely	Mean: 7.0
Dislike someone automatically withdrawing from my account	1: disagree completely– 10: agree completely	Mean: 6.6
Paying with credit card or giving checking account number on Internet is secure	1: disagree completely– 10: agree completely	Mean: 2.7
Personal involvement		
Enjoy talking with bank teller	1: disagree completely– 10: agree completely	Mean: 5.4

[a]The baseline for this variable is market size over 2 million.
[b]The baseline for this variable is income under $20,000.
[c]The baseline for this variable is account at local bank.

NOTES

[1]See Radecki (1999) and Ernst and Young (1999).

[2]See Janssen (1999).

[3]See Hood (1999).

[4]See Bank Systems Technology, Inc. (2000).

[5]For instance, see Snell (1999a).

[6]See Kotler (1994).

[7]See Jolly (1997) for an introduction to this subject and Ferguson (1998) for background on the role of infrastructure in new payment instrument adoption. One important issue that is beyond the scope of this article is to consider products with a network nature to their use, such as the cell phone, where the first products sold are of little value since customers value having the ability to reach many others. See Good (1997) for a discussion of network externalities relating to payments.

[8]The dataset contains variables pertaining to electronic bill payment usage that are in some cases highly correlated. I used a correlation matrix to discern the relationship between responses across questions. The variables included in the final model are the responses to the questions emphasized by theory.

[9]Note, I analyzed various cutoff points for low and high users and the choice of a given cutoff point did not significantly affect the outcome from the regression.

[10]Recall that the binomial logistic model takes the functional form where the probability that a consumer uses electronic bill payment = $e^{\beta x}/1+e^{\beta x}$, where X is a vector of variables proposed to be related to electronic bill payment usage. See Greene (1993).

[11]Electronic bill payments are defined as pre-authorized ACH debits or debits by personal computer banking. Nonelectronic payments include cash, checks, and money orders.

[12]Subsequent analysis confirmed this.

[13]While the survey also asked consumers about other bills, such as membership bills and tuition payments, I exclude them from my analysis because of the small sample size of individuals with these specific bills.

[14]See Mantel (2000) for an overview of the types of initiatives which might be undertaken to address consumers' concerns with the use of electronic bill payment. In general, they tend to promote services that are similar to the Giro systems common in some European countries (for example, consumer initiated, ability to easily build in partial payments, or access to customer service).

[15]For one perspective advocating the need to look actively at this, see Mann (1999). For a second perspective advocating monitoring these types of developments but being careful to act only when there is a clear and compelling reason to do so, see Perritt (1999).

[16]See Mann (1999).

[17]For instance, individual consumers may not have an incentive to negotiate for small dollar adjustments to their accounts, even if they are completely justified, if they expect this process to require significant cost and/or time.

[18]This is a particularly important question given the work underway to promote financial relationships for unbanked consumers as part of the EFT99 legislation.

[19]The approach of allowing multiple standards to flourish is clearly what many public entities do by abstaining from getting involved in these discussions. But, when public entities do get involved, it may still be worth considering advocating multiple standards rather than just one.

REFERENCES

American Bankers Association and Dove Associates, 1999, *Study of Consumer Payment Preferences Focusing on Online and Offline Debit*, Washington: American Bankers Association.

Arndt, Johan, 1976, "Reflections on research in consumer behavior," in *Advances in Consumer Research: Volume 3*, Beverlee B. Anderson (ed.), Ann Arbor, MI: Association for Consumer Research.

Bank Administration Institute and PSI Global, 1997, *Profiting from Change in the U.S. Payments System,* Chicago: Bank Administration Institute.

Bank for International Settlements, Committee on Payment and Settlement Systems, 1999, *Retail Payments in Selected Countries: A Comparative Study*, Basel, Switzerland: Bank for International Settlements.

Bank Systems Technology, Inc., 2000, "Digitally signed payment via EFT," *Bank Systems Technology*, August.

Barron, Patrick, 1999, "Remarks," *Proceedings from the Workshop on Promoting the Use of Electronic Payments*, Chicago: Federal Reserve Bank of Chicago.

Becker, Gary S., 1965, "A theory of the allocation of time," *Economic Journal*, Vol. 75, No. 299, pp. 493–517.

_____, 1962, "Irrational behavior and economic theory," *Journal of Political Economy*, Vol. 70, No. 1, pp. 1–12.

Belk, Russel W., 1975, "Situational variables and consumer behavior," *Journal of Consumer Research*, Vol. 2, pp. 157–167.

Bezdek, Roger, 1997, "Progress report on achieving mandatory federal EFT by 1999," paper presented at the annual meeting of the American Statistical Association, Los Angeles, CA.

Bond, Phillip, and Robert Townsend, 1996, "Formal and informal financing in a Chicago ethnic neighborhood," *Economic Perspectives*, Federal Reserve Bank of Chicago, July–August, pp. 3–27.

Carow, Kenneth A., and Michael E. Staten, 1999, "Debit, credit, or cash: Survey evidence on gasoline purchases," *Journal of Economics and Business*, Vol. 51, pp. 409–421.

Celsi, Richard L., and Jerry C. Olson, 1988, "The role of involvement in attention and comprehension processes," *Journal of Consumer Research*, Vol. 15, pp. 210–224.

Chakravorti, Sujit, 2000, "Why has stored value not caught on?," Federal Reserve Bank of Chicago, Emerging Issues Series, working paper, No. S&R-2000-6.

_____, 1997, "How do we pay?," *Financial Industry Issues*, Federal Reserve Bank of Dallas, First Quarter.

Ernst and Young, National Cash Management Consulting Practice, 1999, *16th Annual Cash Management Services Survey Results Executive Summary*, New York.

Federal Reserve Bank of St. Louis, 1998, *A Summary of Consumer and Business Attitudes on Direct Deposit and Direct Payment: A National Market Research Study*, Federal Reserve Bank of St. Louis.

Ferguson, Roger, 1998, "Electronic banking: Where are the customers? What do they think? What does it mean for the Federal Reserve?," remarks at the Bank Administration Institute's Symposium on Payments System Strategy, Washington, DC, September 29.

Freeman, Scott, 1996, "The payment system, liquidity, and rediscounting," *American Economic Review*, Vol. 86, No. 5, pp. 1126–1138.

Friedman, Milton, 1957, *A Theory of the Consumption Function*, Princeton, NJ: Princeton University Press.

Good, Barbara, 1997, "Electronic money," Federal Reserve Bank of Cleveland, Financial Services, working paper, No. 9716.

Greene, William H., 1993, *Econometric Analysis*, second edition, New York: Macmillan Publishing Company.

Greenspan, Alan, 2000, "Retail payment systems," remarks before the National Automated Clearinghouse Association Annual Meeting, Los Angeles, CA, April 10.

Hancock, Diana, and David Humphrey, 1997, "Payment transactions, instruments, and systems: A survey," *Journal of Banking and Finance*, Vol. 21, No. 11-12, pp. 1573–1624.

Hershey, John C., Howard C. Kunreuther, and Paul J. H. Schoemaker, 1982, "Sources of bias in assessment procedures for utility functions," *Management Science*, Vol. 28, No. 2, pp. 936–954.

Hogarth, Jeanne M., and Kevin H. O'Donnell, 1999, "Banking relationships of lower-income families and the government trend toward electronic payment," *Federal Reserve Bulletin*, pp. 459–473.

Hood, George, 1999, "Remarks," *Proceedings from the Workshop on Promoting the Use of Electronic Payments*, Federal Reserve Bank of Chicago, Chicago, pp. 35–37.

Humphrey, David, Moshe Kim, and Bent Vale, 1998, "Realizing the gains from electronic payments: Costs, pricing, and payment choice," *Proceedings from the Federal Reserve Bank of Chicago's 34th Annual Conference on Bank Structure and Regulation*, Chicago: Federal Reserve Bank of Chicago, pp. 158–164.

Janssen, Debra, 1999, "Electronic-checking question: Who gets the voided check," *American Banker*, Vol. 164, No. 84, p. 13.

Jolly, Vijay K., 1997, *Commercializing New Technologies: Getting from Mind to Market*, Boston: Harvard Business Review Press.

Kahn, Charles M., and William Roberds, 1998, "Real-time gross settlement and costs of immediacy," Federal Reserve Bank of Atlanta, working paper, No. 98-21.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, 1991, "The endowment effect, loss aversion, and status quo bias," *Journal of Economic Perspectives*, Vol. 5, No. 1, pp. 193–206.

Kennickell, Arthur, and Myron Kwast, 1997, "Who uses electronic banking? Results from the 1995 *Survey of Consumer Finances*," *Proceedings from the Federal Reserve Bank of Chicago's Annual Conference on Bank Structure and Competition*, Chicago: Federal Reserve Bank of Chicago, pp. 56–75.

Kolsky, Richard, 1999, "Comments," *Proceedings from the Workshop on Promoting the Use of Electronic Payments*, Chicago: Federal Reserve Bank of Chicago, pp. 29–35.

Kotler, Philip, 1994, *Marketing Management: Analysis, Planning, Implementation, and Control*, eighth edition, Englewood Cliffs, NJ: Prentice Hall.

Lafferty International, 2000, "MasterCard offers online wallet," *Electronic Payments International*, July, p. 1.

MacKie-Mason, Jeff, and Christine White, 1996, "Selecting digital payment mechanisms," University of Michigan, Program on Research in the Information Economy, Ann Arbor, MI, working paper series.

Mann, Ron, 1999, "A Payments Policy for the Information Age," remarks at the University of Michigan Electronic Payments Workshop, Ann Arbor, MI.

Mantel, Brian, 2000, "Why don't consumers use electronic banking: Towards a theory of obstacles, incentives, and opportunities," Federal Reserve Bank of Chicago, occasional paper series.

McFadden, Daniel, 1980, "Econometric models for probabilistic choice among products," *Journal of Business*, Vol. 53, No. 3, pp. 13–29.

Murphy, Neil, 1991, "Determinants of household check writing: The impacts of the use of electronic banking services and alternative pricing of services," *Journal of Individual Financial Management*, Vol. 1, No. 1, pp. 35–44.

New York Clearing House, 1997, *Direct Deposit Usage Study*, New York: Wirthlin Worldwide.

Perritt, Henry H., 1999, "Remarks," *Proceedings from the Workshop on Promoting the Use of Electronic Payments*, Chicago: Federal Reserve Bank of Chicago, pp. 86–88.

Radecki, Lawrence J., 1999, "Banks' payment driven revenues," *Economic Policy Review*, Federal Reserve Bank of New York, Vol. 5, No. 2, July 1999, pp. 53–70.

Snell, Ross, 1999a, "Dropout rate is nearly 33% for web banking customers," *American Banker*, Vol. 164, No. 163, p. 1.

_____, 1999b, "Checkfree to seek deals with web portal sites," *American Banker*, Vol. 164, No. 117, p. 8.

Stavins, Joanna, 1999, "Checking accounts: What do banks offer and what do consumers value?," *New England Economic Review*, Federal Reserve Bank of Boston, March/April, pp. 3–14.

_____, 1996, "Can demand elasticities explain sticky credit card rates?," *New England Economic Review*, Federal Reserve Bank of Boston, July/August, pp. 43–54.

Stevenson, Jack, 1997, "What's in the cards? A look at the future of the U.S. payment industry," remarks at Bank Administration Institute's Payments Strategy Forum, Washington, DC.

Thaler, Richard H., 1985, "Mental accounting and consumer choice," *Marketing Science*, Vol. 4, No. 3, Summer 1985, pp. 199–214.

Wells, Kirsten, 1996, "Are checks overused?," *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol. 20, No. 7, pp. 2–12.

The effect of the run-up in the stock market on labor supply

Ing-Haw Cheng and Eric French

Introduction and summary

There are many anecdotes of people who quit their job after having their stock market wealth increase dramatically. This article assesses whether these anecdotes represent isolated incidents or whether the stock market has significantly affected U.S. labor supply. There are two main reasons why this is an important question. First, quantifying the effects of stock market fluctuations may help forecast future variation in labor force growth, employment, and unemployment. If the stock market suddenly dropped, it is possible that many people would rapidly reenter the labor market in order to rebuild enough wealth to finance their retirement. This would cause the number of potential workers in the economy to increase. If the number of new jobs grew more slowly than the number of new workers, short-term unemployment problems would result. This would exacerbate the potential unemployment problems caused by more conservative hiring practices of employers after a market downturn.

Second, we are interested in evaluating the extent to which the consumption response to variations in stock prices is consistent with economic theory. Current estimates of the marginal propensity to consume out of stock wealth, that is, the "wealth effect" often described in the popular press, range from .01 to .05. This means that each additional dollar in stock wealth increases consumption one to five cents annually. The estimate more consistent with simple economic models that posit that people eventually consume their wealth (see Poterba, 2000) is .05. If a dollar increase in stock wealth results in only a one cent increase in consumption, then 99 cents would be saved until next year. Assuming the 99 cents earns a 3 percent post-tax rate of interest, it would grow to approximately $1.02 next year. Therefore, people would not eventually consume all of their wealth, contrary to the simple economic models. If the post-tax interest rate is 3 percent, people must have a marginal propensity to consume of at least .03. Poterba (2000) suggests .04 as a reasonable lower bound.

However, these simple economic models assume that labor supply does not respond to variations in wealth. If much of the stock market wealth goes toward affording people increased leisure in addition to increased consumption of market goods, then the .01 estimate for the marginal propensity to consume market goods may be consistent with economic models that account for the effect of wealth on labor supply. People would eventually "consume" all of their wealth, but mostly in the form of increased leisure. If individuals consume three cents worth of leisure in the form of reduced earnings (that is, their earnings drop by three cents each year) in addition to a one cent increase in consumption of market goods in response to a $1 increase in wealth, then total consumption would be four cents in response to a $1 increase in wealth. This story is perfectly consistent with the theory that individuals eventually consume all their wealth.

In this article, we present estimates of the size of the increase in wealth in the U.S. economy from 1994 to 1999. Recent stock returns are high by historical standards. We also show that growth rates in stock prices are difficult to predict. Therefore, most of the recent increase in wealth caused by rising stock prices represents an unanticipated increase to national wealth. We estimate that every dollar held in stocks on December 31, 1994, resulted in $1.12 in unanticipated wealth shocks if those stocks were held until December 31, 1999. We estimate that the unanticipated component

Ing-Haw Cheng is a student at the University of Chicago and Eric French is an economist at the Federal Reserve Bank of Chicago. *The authors thank Dan Sullivan and David Marshall for helpful comments and David Marshall for data.*

of the increase in national wealth from 1994 to 1999 was $5.8 trillion in 1999 dollars.

In order to understand how many people may have been affected by the run-up in the stock market, we examine the distribution of stock market wealth in the economy. The more concentrated the distribution, the fewer people whose labor supply will be directly affected by stock market variations. Using data on the distribution of stock market wealth and on stock returns, we estimate the distribution of unanticipated increases in wealth for different groups of the population. We show that about 15 percent of all individuals aged 55 and over had an unanticipated wealth increase of $50,000 in 1999 dollars or more between December 31, 1994, and December 31, 1999.

Next, we show changes in labor force participation rates for different age groups in different years. Holding all else equal, we would expect groups with large unanticipated increases in wealth to reduce their labor force participation rates. As it turns out, this is not the case. Individuals aged 55 and above have the highest levels of stock wealth (both directly and through pensions) and, thus, have had the greatest unanticipated increases in wealth. However, labor force participation rates for individuals aged 55 and older have increased over the last five years.

In our view, one should not take these counterintuitive results as evidence against the theory that the run-up in stock market wealth has decreased labor force participation rates. Instead, we believe these results imply that the run-up in the stock market has not been the primary determinant of recent changes in labor force participation rates. There are many other reasons that labor force participation rates should be rising for older workers. For example, the strong economy has resulted in increased wages and improved employment opportunities for older workers. Importantly, too, the Social Security System has reduced the work disincentives for those eligible for Social Security benefits.

In order to understand how increases in stock market wealth affect aggregate labor supply, we use two basic approaches. First, we use estimates from two previous studies to predict the change in labor supply for a given unexpected change in wealth. Imbens et al. (1999) estimate the effect on labor supply of winning a lottery, which presumably represents an unanticipated change in wealth. Again assuming that the wealth increase is unanticipated, Holtz-Eakin et al. (1993) estimate the labor supply effect of receiving an inheritance. Both papers suggest that unanticipated increases in wealth reduce work hours and labor force participation rates. Using these estimates and the distribution of wealth, we predict the likely decline in work hours caused by the run-up in the stock market. Our estimates suggest that in the absence of the run-up in the stock market (but holding all else equal), labor force participation rates today would be .78 percentage points higher for men aged 55–64, 1.94 percentage points higher for women aged 55–64, and 1.16 percentage points higher on aggregate.

Our second approach to predicting the effect of the run-up in the stock market on labor supply is to simulate the effect using a dynamic structural model described in French (2000). French estimates the model using data on life cycle profiles for assets, hours worked, and labor force participation rates. Simulations from the model closely mimic the life cycle profiles in the data. Therefore, the model is also potentially able to closely mimic the behavioral effects of the run-up in the stock market. Our simulations imply that in the absence of a run-up in the stock market, labor force participation rates would have been 1.3 percentage points higher for men aged 65 and above and 3.2 percentage points higher for men aged 55–64. In other words, the simulation model predicts much larger behavioral responses than the estimates from other studies. We discuss why the simulation model may overestimate the behavioral responses and the estimates from other studies may underestimate the behavioral responses later in the article. Overall, our view is that the predictions from the lottery and inheritance studies form a lower bound on the effect of the stock market on labor supply and the simulation model forms an upper bound.

Lastly, we present estimates of the marginal propensity to consume leisure (also known as the marginal propensity to earn out of wealth). Recall that an estimate of the marginal propensity to consume market goods of .01 is consistent with the life cycle model only if the marginal propensity to consume leisure is at least .03. The estimates from the direct lottery and inheritance studies are in the range of .01. In other words, for every $1 increase in wealth, aggregate earnings decline one cent. The simulation model predicts a larger marginal propensity to consume leisure—about .02. In either case, the marginal propensity to consume leisure is too small to reconcile a marginal propensity to consume consumption goods of .01 with a life cycle model. Therefore, either the life cycle models are wrong or the .01 estimate of the marginal propensity to consume market goods is wrong.

Increase in national wealth in the 1990s

We provide evidence that the increase in wealth caused by the run-up in stocks was largely unanticipated and estimate the unanticipated wealth shock.

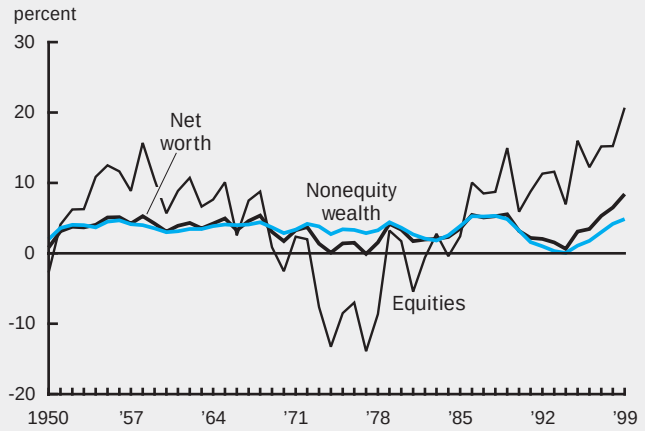
The issue of whether the increase in the level of wealth was anticipated is important. If people knew in December 1994 that five years in the future they would have higher levels of wealth, then it is possible that they would have reduced the number of hours worked in 1995–99, knowing that they would be able to finance their low levels of work because of the anticipated run-up in the stock market.[1] Therefore, we would not expect to see any correlation between stock market gains and labor supply.[2]

Figure 1 shows the growth in household net worth in the economy from 1945 to 1999. While household wealth not in equities rose at a moderate rate of 27 percent over the decade, the market value of equities increased 260 percent between December 31, 1989, and December 31, 1999. The value of equities rose by almost $9.5 trillion during the 1990s, comprising about 64 percent of the growth in wealth. Figure 1 also shows that changes in household wealth in this period are largely explained by changes in the value of equities. Figure 2 shows growth rates in the value of equities, based on Federal Reserve Board data, and growth rates in the stock market, as measured by the Center for Research in Security Prices (CRSP). The appendix describes the CRSP measure in greater detail. The two measures are almost perfectly correlated (with differences relating to treatment of dividends and American stockholdings overseas.)

Figure 3 shows rates of return in the stock market over five-year horizons, with the latest being the rate of return between December 31, 1994 and December 31, 1999. Figure 3 also shows the results from a simple forecasting model (described in box 1), which uses information on five-year Treasury bond yields and stock returns from 1950 to 1999 to predict stock returns. We compute the difference between five-year total stock returns and total returns on five-year Treasury bonds, that is, the "excess" return on the stock market. Between 1950 and 1999, the excess return was 45 percent over five years. The forecast of the five-year return in the stock market is the sum of the excess return (which is assumed constant) plus the five-year Treasury bond return. The predicted five-year return has increased over time because interest rates have increased.

FIGURE 1

Five-year average growth rate of household wealth
(1999 billions of dollars)

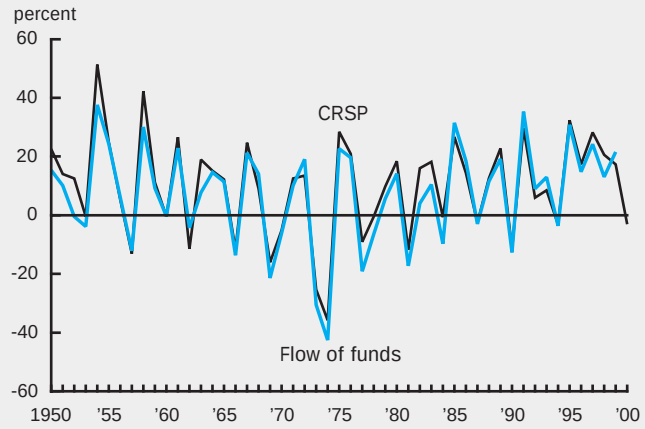


Source: Data from Board of Governors of the Federal Reserve System, *Flow of Funds Accounts*, 1946–99.

Figure 3 shows there have been very large differences between the predicted return and the five-year realized return. Over the past five years, the average annual rate of return in excess of the Treasury bond rate has been 16.4 percent. The five-year excess return was 93 percent. This is well above the historical average of a 45 percent excess return over five years. It has not been since the 1950s that there has been such a large, sustained increase in the stock market. Moreover, because stocks represented about twice as large of a share of national wealth in 1994 than in 1950, the growth in national wealth was greater in

FIGURE 2

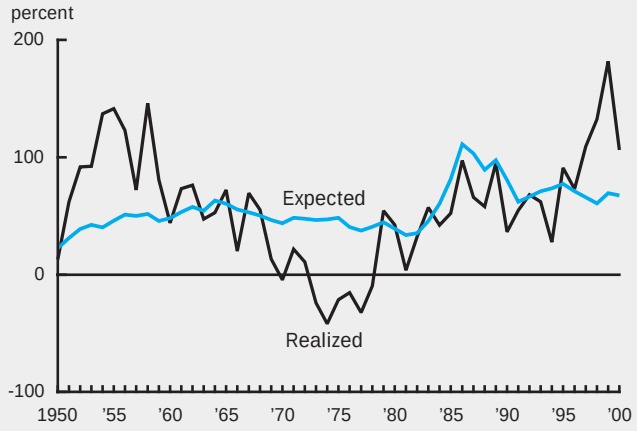
One-year return in the stock market



Sources: Board of Governors of the Federal Reserve System, *Flow of Funds Accounts*, 1950–99, and Center for Research in Security Prices, 1950–2000.

FIGURE 3

Five-year total return in the stock market



Note: In 1999 dollars.
Sources: Center for Research in Security Prices, 1946–2000, and Treasury bond data, 1946–2000.

the late 1990s than in the 1950s. Figure 3 shows that $1 invested in December 31, 1994, would have reached $2.82 in December 31, 1999, compared with a predicted level, based on the historical average, of $1.70. This means that every $1 invested in December 31, 1994, resulted in a windfall gain of $1.12 by December 31, 1999. Because stock market wealth constituted just over 15 percent of aggregate wealth in 1994, the run-up in stock market wealth resulted in national wealth being 17 percent greater in 1999 than it would have been if returns had been as expected. The run-up in the stock market represents a $5.8 trillion shock to national wealth.

It seems unlikely that people anticipated the high rates of return during the late 1990s. For one thing, why did many people not invest in stocks at all? Rates of return on risk-free assets declined in the late 1990s. If stocks were a sure bet, nobody would ever prefer bonds to stocks.

Another way of looking at the problem is to ask whether any historical relationships would predict high stock returns in the late 1990s. We investigate two relationships, described in detail in box 1. First, people might believe that if returns were high in the recent past they would continue to be high in the future. We find that since 1950 high returns over the previous four years and the previous ten years have indicated high returns in the near future. However, as figure 3 shows, the 1980s and early 1990s were not remarkably good years for the stock market. The second historical relationship we investigate is that between price/dividend ratios and stock returns. When price/dividend ratios are high, stocks are possibly overpriced and should perform poorly in the near future. We do see this pattern in stock market data from 1950 onwards. However, in 1994 price/dividend ratios were already high. Therefore, the statistical evidence indicated that stocks would perform poorly in 1995–99. If people were making forecasts according to this simple statistical model, every dollar in the stock market on December 31, 1994, would have led to a $1.46 unexpected gain in wealth by December 31, 1999. Therefore, assuming a constant excess return of stocks over bonds leads to a conservative estimate of the unexpected shock to the stock market.

Unexpected wealth changes in the population

Given a U.S. population of just under 300 million and an aggregate wealth shock of about $5.8 trillion, the run-up in the stock market from 1994 to 1999 represents an unanticipated increase in wealth of $20,000 per person. This is roughly enough to finance one year out of the labor market with no change in consumption of market goods for every individual in the U.S. However, because wealth in stocks is highly concentrated among the wealthy, we would expect the effect on labor supply to be smaller than it would be if stock market wealth were evenly distributed. Our information on the distribution of stock market wealth comes from two sources—data on non-pension stock market wealth from the *Panel Study of Income Dynamics* (PSID) and pension wealth data from other studies that used data from the Health and Retirement Survey (HRS).

The PSID is a nationally representative dataset that includes demographic information, the value of non-pension wealth held by individuals, and breakdowns of wealth into various components, including stock wealth. The stock market wealth measure includes the value of stocks in mutual funds, IRAs, and Keogh plans, in addition to directly held stocks. Juster et al. (1999) show that respondents in the PSID report over 85 percent of their stock market wealth and over 75 percent of total wealth.[3] They also find that the distribution of wealth in the PSID is extremely accurate for everyone but the wealthiest 2 percent of the population. To adjust for the slight underreporting of stock market wealth in the PSID, we multiply stock market wealth in the PSID by 1/.85 = 1.18. The PSID does not provide information on who controls the wealth within households. We assume that non-pension

BOX 1

Predicting stock returns

This section shows the method we use to compute the difference between the k-year realized and expected returns. Denote the return over the past year as r_i (for example, r_{1995} is the one-year rate of return from the end of December 31, 1994, through the end of December 31, 1995) and the gross return over a k-year horizon as $R_{t_0 \to t_k}$, where $t_k = t_0 + k$ (for example, if $k = 5$ and $t_0 = 1994$, the five-year return is $R_{1994 \to 1999}$, the return from the end of December 31, 1994, through December 31, 1999). We measure all returns and growth rates in real terms. The k period rate of return is

$$1) \quad R_{t_0 \to t_k} = \prod_{i=1}^{k} (1 + r_{t_0 + i}).$$

We can compare this to the gross k-year expected return, which we forecast as:

$$2) \quad \hat{R}_{t_0 \to t_k} = R^f_{t_0 \to t_k} + excess_{t_0 \to t_k},$$

where

$$3) \quad excess_{t_0 \to t_k} = R_{t_0 \to t_k} - R^f_{t_0 \to t_k} = \alpha + \varepsilon_t,$$

α is the constant excess rate of return, ε_t is a white-noise random variable, and $R^f_{t_0 \to t_k}$ represents the continuously compounded return on a k-year risk-free asset. The rationale for forecasting the excess stock return (that is, stock returns net the risk-free return) is that the k-year total return of a risk-free asset is known in advance at time t_0; for example, we can easily find information regarding yields on five-year Treasury bonds today, and thus compute the associated five-year holding period return (assuming the bond is held until maturity). Thus, the only variables in the forecast at time t_0 of future expected returns is the average excess return over the sample period and the five-year Treasury bond return.

The unexpected windfall for an investor at time t_1, who had M dollars in the stock market at time t_0 is thus simply

$$\Delta \hat{A}_t \equiv M(R_{t_0 \to t_k} - \hat{R}_{t_0 \to t_k}),$$

where $R^f_{t_0 \to t_k}$ is the predicted stock market return given by equations 2 and 3.

Equation 3 is an extremely simple model of forecasting excess stock returns; in fact, the predicted value $\widehat{excess}_{t_0 \to t_1}$ is simply the mean of the risk premium over the entire time span. Other models have been suggested. Cochrane (1997) recommends several possible indicators that track long-horizon market movements relatively well. In particular, he suggests that the price/dividend (P/D) ratio is a good indicator of long-horizon market movements. When P/D ratios are high, stocks are overpriced and, thus, stock prices should grow slowly.

The regression

$$4) \quad excess_{t_0 \to t_k} = R_{t_0 \to t_k} - R^f_{t_0 \to t_k} = \alpha + \beta \frac{P_{t_0}}{D_{t_0}} + \varepsilon_t$$

using excess returns over k = five-year horizons of the CRSP NYSE value-weighted portfolio from the end of 1954 to the end of 1996 (that is, using stock market information from the start of 1950 to the end of 1996) and five-year Treasury yields (see the appendix for more information) gives a point estimate of $\beta = -5.30$ with an R^2 of 0.54. Using these estimates, the predicted excess return for December 31, 1994, to December 31, 1999, is –27 percent. This is an implausible prediction given that if people expected stock returns to be lower than bond returns, nobody would invest in stocks. However, extending the regression to include 1997–99 as in-sample years cuts the point estimate to –2.79. Using these estimates, the predicted excess return for December 31, 1994, to December 31, 1999, was 12 percent. If anything, these estimates indicate that our model would actually underestimate the unexpected wealth shock during 1994–99, since our model simply predicts the market to return the mean historical excess return plus the going return on bonds, whereas more complex models predicted the market would perform poorly. In fact, our estimate for unexpected windfalls 1994–99 using equation 3 is \$1.12 on every dollar, whereas the same estimate using equation 4 and our own data is \$1.46 on every dollar (in 1999 dollars).

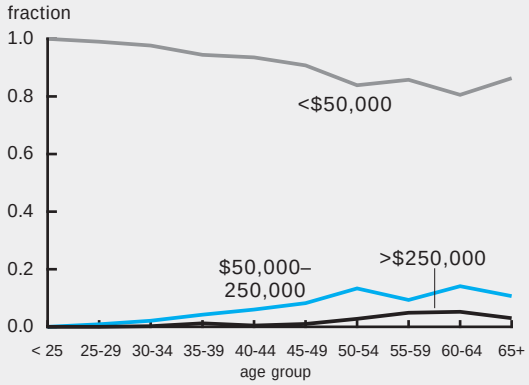
Interestingly, forecasting the following unrestricted version of equation 3 using 1954–99 data,

$$R_{t_0 \to t_k} = \alpha + \beta R^f_{t_0 \to t_k} + \varepsilon_t,$$

yields a point estimate of $\beta = 0.95$ with a standard error of 0.40 (not including adjustments for heteroskedasticity and serial correlation), so one cannot statistically reject our model where we assume $\beta = 1$.

FIGURE 4

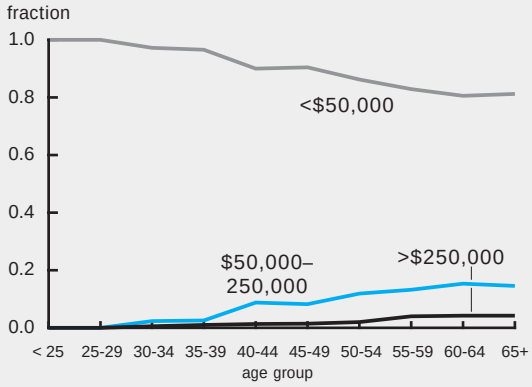
Fraction of women in stock value groups



Sources: PSID data, 1994, and authors' calculations.

FIGURE 5

Fraction of men in stock value groups



Sources: PSID data, 1994, and authors' calculations.

household wealth is split evenly among spouses, with each receiving 50 percent. Browning et al. (1994) show that consumption between husbands and wives is close to an even split, regardless of the resources brought into a household.[4] We attribute pension wealth to the individual receiving the pension.

Figures 4 and 5 show the distribution of stock market wealth by age of the individual for women and men. The data show two things. First, older individuals have more stock market wealth than younger individuals. The fraction of the population with over $50,000 in stock market wealth is less than 4 percent for households where the head is younger than 54. Second, most individuals have little stock market wealth. Even for households aged 55 and older, less than 13 percent had over $50,000 in the stock market (including stocks, mutual funds, IRAs, and Keogh plans).

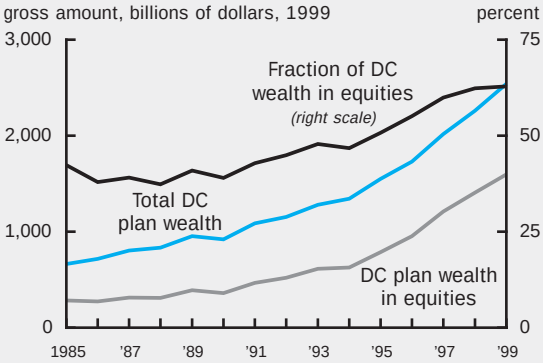
Gustman and Steinmeier (1999) show that pension wealth is very broadly held in the population and constitutes a large portion of overall wealth. Not surprisingly, the run-up in the stock market has led to an increase in pension wealth for many people. Figure 6 shows total national wealth held in defined contribution pension plans. In this type of plan, individuals contribute a portion of their income and the account's value grows by that amount plus the rate of return on the plan's portfolio of assets. Data from the Federal Reserve Board's *Flow of Funds* shows that by 1999 about 12 percent of all U.S. wealth in equities was held by defined contribution pension plans and that the amount of stock wealth in defined contribution pension plans rose 182 percent from 1994 to 1999.

The other major type of pension plan, the defined benefit pension plan, provides benefits that are specified by the employer. These benefits do not depend on the rate of return for assets in the pension fund. If there is a run-up in the stock market, the employer gets the windfall. Likewise, in the event of a market crash, the employer must make up the shortfall if pension fund reserves are low. Therefore, changes in the stock market affect the stock price of the firm holding the pension reserves but do not affect the wealth level of employees at the firm.

Gustman and Steinmeier (1999) show that in their HRS sample of older workers, 66 percent of all households are covered by a pension plan. Of the households covered, 48 percent are covered by a defined benefit plan, 21 percent are covered by a defined contribution plan, and 31 percent are covered by some combination of defined benefit and defined contribution plans. Defined contribution pension plans tend to be less generous than defined benefit pension plans and joint

FIGURE 6

Defined contribution pension wealth and equity composition



Source: Board of Governors of the Federal Reserve System, *Flow of Funds Accounts*, 1985–99.

plans. In 1992, average wealth held in defined contribution plans at age 65 was $57,000 in 1999 dollars. In contrast, the amount held in defined benefit plans was $135,000 and the amount held in combined plans was $153,000. We assume that half the wealth in combined plans is in the form of defined benefit wealth and the other half is in defined contribution wealth. This means that 35 percent of all households in the HRS held an average level of wealth in defined contribution plans equal to $69,000 at age 65 in 1994. The other 65 percent held no defined contribution wealth. Given that 40 percent of defined contribution plan pension wealth is in the form of equities, 35 percent of all elderly households would have an average of $28,000 in the stock market by age 65. Since $28,000 invested in the stock market on December 31, 1994, would have resulted in an unanticipated windfall of about $31,000 by December 31, 1999, a large number of elderly households would have received a large unexpected increase in wealth because of their pensions.

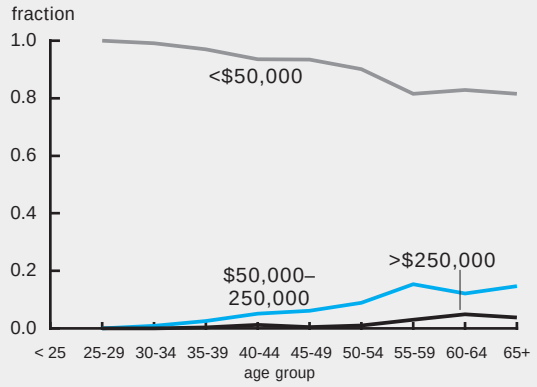
Because the PSID only has data on whether respondents were covered by a pension plan and whether they contributed to that plan in 1989, we assume that those who were contributing to a defined contribution plan in 1989 were also contributing in 1994. If the individual was not contributing in 1989, we assume that person never contributed to a defined contribution pension plan. The fraction of the population covered by a pension does not vary much by age, except for those under 35 who have lower coverage rates. We assume that individuals over 35 who are contributing to a defined contribution plan contribute a fixed amount after age 35. We assume individuals younger than 35 contribute for only one year.

Younger households would also have had windfalls from increases in stock market wealth, although the windfalls would be smaller. To calculate the amount of pension plan wealth at each age, we assume a 2.3 percent real rate of return on pension investments, the same amount of pension contributions each year, that the worker starts working at a firm that provides a defined contribution plan at age 35, and that the level of wealth in the defined contribution plan would be $69,000 at age 65, on average. For example, an individual who contributes $1,550 annually would have an imputed defined contribution wealth of $10,000 at age 40, $28,000 at age 50, and $53,000 at age 60.[5]

Given the distribution of stock market wealth in the economy and the rates of return on stocks computed in the previous section, we compute a measure of unexpected wealth increases for different segments of the population. Recall that $1 invested in December 31, 1994, would have resulted in $1.12 in unanticipated wealth gains by December 31, 1999. Figures 7 and 8 show the distribution of wealth shocks in the economy for women and men. The differences between these figures and figures 4 and 5 are twofold. First, figures 7 and 8 include information on pensions for 1994. Second, figures 7 and 8 do not describe total stock wealth but how stock wealth in 1994 became wealth shocks in 1999. These figures make two points clear. First, there is a sizable minority of individuals who received wealth shocks in excess of $50,000. Second, individuals aged 55 and older received most of the wealth shocks; 21 percent of all individuals aged 55 and older received unanticipated wealth gains in excess of $50,000. Given that most individuals had earnings below $50,000, an unanticipated wealth gain of $50,000 could replace at least one year of earnings for most individuals.

FIGURE 7

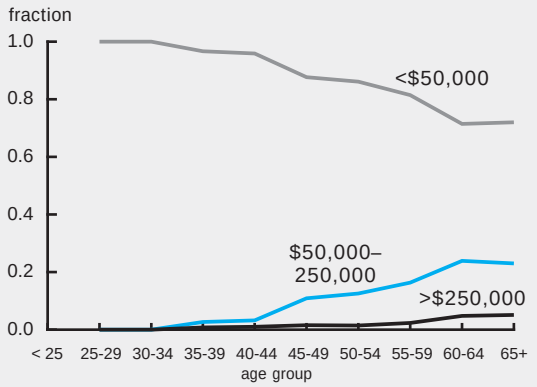
Unexpected wealth shocks for women by age



Sources: PSID data, 1989 and 1994, and authors' calculations.

FIGURE 8

Unexpected wealth shocks for men by age



Sources: PSID data, 1989 and 1994, and authors' calculations.

Changes in labor force participation rates

Estimates of the level of unexpected wealth increases show that a sizable minority of the population had large unexpected increases in wealth. Most of these increases in wealth are concentrated among individuals aged 55 and above in 1999 (or 50 and above in 1994). Therefore, we would expect this group to have the largest declines in labor force participation rates between 1994 and 1999.

Figures 9 and 10 show labor force participation rates between 1980 and 2000. For men aged 55 and above, labor force participation rates have been rising recently, following a steady decline from 1980 to 1993. For women aged 65 and above, labor force participation rates have remained steady since 1994. For women aged 55–64, labor force participation rates have been rising since 1994. These data show that the rise in the stock market has not been the dominant source of changes in labor force participation rates for individuals aged 55 and over. The trends do not support the wealth effect hypothesis.

However, we argue that this should not be taken as evidence that the unanticipated increase in wealth has resulted in no change or an increase in labor force participation rates. Instead, in our view, the data provide evidence that other factors have offset the effects of the increase in the stock market. Among these factors are recent increases in wages in the economy. Moreover, the Social Security System has reduced the work disincentives for individuals 65 and older. Social Security benefit accrual is now closer to actuarially fair for individuals aged 65–70 than it was in 1994.[6] It is not clear what effect the stock market may have had on labor participation rates for individuals aged 55 and above in the absence of other factors.

Next, we look at the likely behavioral responses to the run-up in the stock market, holding all else equal.

Estimates of the effect of unanticipated wealth increases on labor supply

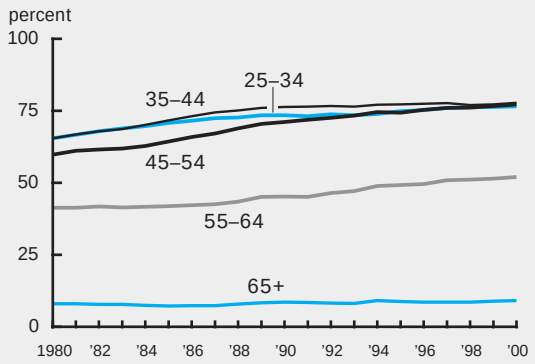
Here, we present estimates of the effect of unanticipated wealth increases on labor supply holding all else equal. Estimating the effect of an unanticipated wealth increase on labor supply is difficult because, usually, changes in wealth are anticipated or are accompanied by wage changes. The labor supply response to an anticipated wealth increase is different from the labor supply response to an unanticipated wealth increase. If the wealth change is anticipated, we expect a small labor supply response after receipt of the wealth. If people know they will receive a large sum of money tomorrow, their labor supply may not change much between today and tomorrow. They may already have reduced their labor supply in anticipation of having the wealth in the near future.

Inheritance and lottery studies

Table 1 presents estimates of the effects of inheritances on labor supply. Inheritances cause plausibly unanticipated changes in wealth. Holtz-Eakin et al. (1993) estimate the effect of receiving an inheritance on labor force participation rates. Using tax records, they observe whether reported earnings are positive (our measure of labor force participation) both before (in 1982) and after (in 1986) the receipt of an inheritance. They find fairly large effects on labor force participation and earnings. Among single households who receive a small inheritance (average of $13,000) in their sample, labor force participation rates increase from 89.9 percent to 91.1 percent (column 4), or an increase of 1.2 percent (column 5). Single households

FIGURE 9

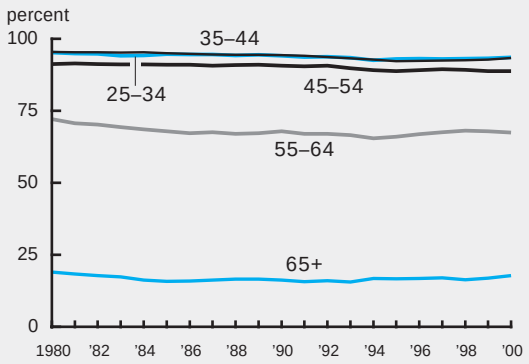
Labor force participation rates, women



Source: CPS data, 1980–2000.

FIGURE 10

Labor force participation rates, men



Source: CPS data, 1980–2000.

who receive a medium-sized inheritance (average of $120,000) show a labor force participation rate decline from 82.7 percent to 80.5 percent, or a decline of 2.2 percent. We interpret these changes in labor force participation rates to mean that in the absence of a medium-sized inheritance, labor force participation rates would have increased 1.2 percent instead of declining 2.2 percent for those who received the medium-sized inheritance. Therefore, labor force participation rates would have been 3.4 percent higher (column 6) had those individuals not received a medium-sized inheritance. Those who receive a large inheritance (average of $609,000) show a labor force participation rate decline of 9.8 percent from 75.4 percent to 65.6 percent. If these people had not received an inheritance, their labor force participation rate would have increased 1.2 percent. Therefore, receiving the inheritance potentially reduces their labor force participation by 11.0 percent. Holtz-Eakin et al. find similar results for married couples. Receiving a medium inheritance reduces average labor force participation rates within the household by 3.8 percent, and a large inheritance reduces labor participation by 4.2 percent.

Most of the individuals who received inheritances were young. Singles who received small inheritances (the youngest group) were aged 33.4, on average, and the mean age of couples who received large inheritances (the oldest group) was 44.7 years. Therefore, the sample in the Holtz-Eakin et al. study is significantly younger than the segment of the general population that has received most of the stock wealth gains. Since it is likely that large wealth gains have larger labor supply effects for those who are nearing retirement than for younger individuals, our view is that the inheritance study most likely understates the labor supply effects from the run-up in the stock market.

Joulfaian and Wilhelm (1994) find slightly smaller but similar effects using data from the PSID. Their results show that the results in Holtz-Eakin et al. are not specific to a particular dataset. Joulfaian and Wilhelm also estimate the effect of inheritances on consumption using PSID data.[7] They find that the marginal propensity to consume all goods out of inheritance wealth is about .0012. This is an order of magnitude smaller than the .01 to .05 marginal propensity to consume out of stock wealth estimated in most studies. This evidence suggests that people may anticipate inheritances and that the inheritance estimates, therefore, may underestimate the effect of unanticipated wealth changes on labor supply.

TABLE 1

Effect of inheritance on labor force participation

	Mean inheritance level	Inheritance difference	Mean pre-inheritance income	Participation rate		Participation change	Inheritance effect	Observations
	(----------Dollars----------)			(----------Percent----------)				
Single								
Small	13,359		20,863	1982	89.9	1.2		730
				1985	91.1			
Medium	119,610	106,251	23,027	1982	82.7	–2.2	–3.4	544
				1985	80.5			
Large	608,858	595,499	19,586	1982	75.4	–9.8	–11.0	358
				1985	65.6			
Married couples								
Small	13,323		60,867	1982	77.0	–0.1		1,078
				1985	76.9			
Medium	125,554	112,231	59,340	1982	73.1	–3.9	–3.8	994
				1985	69.2			
Large	597,037	583,714	66,804	1982	68.8	–4.4	–4.2	628
				1985	64.4			

Notes: Participation rate is the sum of people working divided by two multiplied by the number of households. In 1999 dollars. "Inheritance difference" is the difference between the inheritance level received and a small inheritance level. "Inheritance effect" is the difference between the participation change for a given inheritance level and the participation change for a small inheritance level. Small inheritance level is $0–43,000; medium is $43,000–255,000; and large is $255,000 and above.

Imbens et al. (1999) use data from the state lottery of Massachusetts to estimate the effect of winning the lottery on changes in hours worked and earnings. They use data on individuals who received a prize, ranging in present value from $100 to over $1,000,000.[8] A subsample of winners received a questionnaire about purchases made, labor supply, and earnings several years after they won the prize. Many of the players released their Social Security earnings records. Therefore, one can see the earnings of an individual both before and after winning the prize as measured by earnings reported to the Social Security Administration.

Unfortunately, those who won medium and large prizes included both season ticket holders and those who purchased tickets one at a time, whereas those who won a small prize included only season ticket holders. As a result, individuals who won the small prizes were much older (average age of 53.2) than individuals who won medium-sized (average age of 44.6) or large (average age of 48.5) prizes. This makes the lottery study less than perfect, although Imbens et al. attempt to overcome this problem. Moreover, sample sizes in the study are relatively small. There were a total of 496 respondents in the entire study.

Given these caveats, Imbens et al. (1999) estimate the effect of annual lottery winnings on annual labor income. Lottery winners who won a medium-sized or large prize (that is, more than several thousand dollars) received labor income for 20 years. We compute the present value of the lottery winnings and use their estimate of the effect of annual winnings on labor income to compute the effect of lottery winnings on labor income, a measure of labor force participation. Results from these computations suggest that $1 in lottery winnings reduces labor income by one cent annually. In other words, the marginal propensity to consume leisure out of wealth shocks is about .01.

Imbens et al. also find that the marginal propensity to consume leisure out of wealth shocks is greatest for individuals ages 55–65. For example, they find that for individuals younger than 55, the marginal propensity to consume leisure is .0082, whereas for individuals aged 55–65, the marginal propensity to consume leisure is .0132. This is an important point given that much of the stock market wealth is held by individuals aged 55–65. Imbens et al. also find that the marginal propensity to consume leisure is the same for both men and women. Lastly, they find that the marginal propensity to consume leisure is greater for individuals who won small amounts than for individuals who won large amounts. For example, it is .0091 for individual with almost no winnings and .0076 for individuals with close to $500,000 in winnings. It is these final two numbers that we will use to predict the labor supply response to changes in wealth.

Also somewhat interestingly from this study, the Social Security earnings records show that the labor supply response to winning a lottery is not immediate. It is several years before labor supply fully declines in response to the wealth effect. Therefore, the labor supply response to the run-up in the stock market may not be immediate either. Other studies have also found that the consumption response to changes in the stock market is not immediate (see Dynan and Maki, 2000, for example).

The lottery and inheritance studies are not the only studies of the effect of financial resources on labor supply. Blundell and MaCurdy (1999) survey a wide range of approaches to estimating the effect of income on labor supply. The majority of these studies find that increased non-labor income reduces labor supply. Assuming that income is constant over the life cycle, one can compute the annuity value of a lifetime increase in income. Given the estimates surveyed in Blundell and MaCurdy and the computed annuity value of increases in income, we estimate the change in labor supply given a change in wealth. Measured against the results from most other studies, the estimates in the lottery and inheritance papers are relatively small, although there are enormous differences in estimates from study to study. An average estimate of the effect of the annuity value of income on labor supply from Blundell and MaCurdy is about twice as large as the inheritance and lottery estimates. Therefore, our view is that the results from the inheritance and lottery surveys represent conservative estimates of the true effects of unanticipated wealth gains on labor supply.

Our interpretation of the studies

We expect unanticipated changes in wealth to lead to larger changes in labor supply for low-income workers than for high-income workers. An unanticipated $50,000 wealth change replaces two years of labor income if a worker earns only $25,000 dollars per year. In other words, this worker could retire two years earlier and still consume that same amount at each age as a result of the unanticipated $50,000 wealth change. On the other hand, if the same worker earns $50,000 per year, the $50,000 unanticipated wealth change replaces only one year of earnings.

High-wage workers have been receiving most of the wealth gains from the stock market. Mean annual

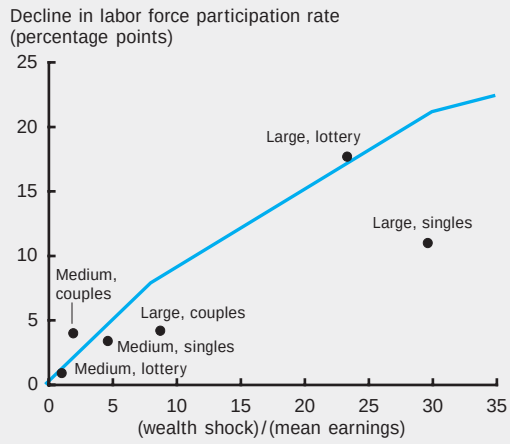
income for all households in the 1994 PSID is $36,500, but mean income is $52,900 in 1994 for households with unanticipated stock wealth gains of $50,000–$250,000 and $94,300 for households with unanticipated stock wealth gains in excess of $250,000. Therefore, $1 in unanticipated wealth gains probably has a smaller effect for individuals with large stock wealth gains than for people who receive an inheritance. We overcome this problem by measuring the labor supply response to wealth divided by income, as described below. This procedure assumes someone with twice the income of another person would need twice the unanticipated wealth gain of the other person to cause the same labor supply response.

Figure 11 uses the information in table 1 and results from the lottery study to plot changes in labor force participation rates against the amount of unanticipated wealth change divided by the mean earnings of people with that unanticipated wealth change. Labor force participation rates and unanticipated wealth shocks are relative to the reference group of small inheritance receivers in the inheritance study. Thus, the points for labor force participation rates are shown in column 6 of table 1 and the points for inheritances are shown in column 2 of table 1. The average pre-inheritance earnings for the different groups are shown in column 3. Therefore, four points in figure 11 are the four values in column 6 plotted against the values in column 2 divided by the values in column 3. We divide inheritances by two for married couples (the husband and wife each get one half), just as we divide household unanticipated wealth shocks by two for married couples in the PSID. The other two points on figure 11 are the two points previously described from the lottery study.

For example, single individuals who receive an average $120,000 inheritance ($106,000 above the reference group of those who receive a small $12,000 inheritance) have $23,000 in income before receipt of the inheritance. Therefore, the value of their unanticipated wealth gain divided by earnings is 4.6. They show a 3.4 percent drop in labor force participation. Couples who receive an average inheritance of $125,000 (or $112,000 above the reference group of couples) have an average of $59,000 in annual earnings. This results in both the husband and wife having $56,000 in inheritance wealth gain and $29,500 each in annual earnings. Therefore, the wealth shock divided by average earnings is (56,000/29,500) = 1.9. Both husbands and wives show an average decline in labor force participation of 3.8 percent.

FIGURE 11

Predicted participation rate response to wealth shocks



Note: "Medium, singles" refers to singles who received a medium inheritance; "large, singles" refers to singles who received a large inheritance; "medium, couples" refers to couples who received a medium inheritance; "large, couples" refers to couples who received a large inheritance; "medium, lottery" refers to winners of a medium-sized lottery prize; "large, lottery" refers to winners of a large lottery prize.

Source: Holtz-Eakin et al. (1993) and Imbens et al. (1999).

To use the stock market wealth gain information to predict the effect of the stock market run-up on labor force participation rates, we need a functional form for the effect of unanticipated wealth gains on labor force participation. Because we have only six data points to fit and some of the data points seem more reliable than others, we use no formal criteria to measure the functional form for how stock market gains affect labor force participation. Instead, we fit the data free-hand to an assumed functional form. We follow three guidelines. First, the functional form is "close to" the individual points in figure 11. Second, we believe that the incremental (or marginal) effect of increasing stock wealth gains on labor force participation is smaller for very high levels of stock wealth gains than for low stock wealth gains. The millionth dollar increase in stock market gains will most likely have a smaller effect on the probability that one drops out of the labor force than one's first dollar of gains. Finally, an unanticipated wealth shock that is close to zero should have a labor supply response that is close to zero. Our assumed functional form for the effect of an unanticipated wealth shock on the labor force participation rate is

$$5)\quad E\left[\Delta LFPR_{it} \,\middle|\, \left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)\right] = \beta\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right),$$

where

$$6)\quad \beta\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)=\begin{cases} .010\times\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right) \text{ if } \left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)<8 \\ .032+.006\times\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right) \text{ if } 8\le\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)<30 \\ .152+.002\times\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right) \text{ if } \left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)\ge 30, \end{cases}$$

where $E[\Delta LFPR_{it} \mid \left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)]$ is the expected change in labor force participation rates given a change in $\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)$ and $\Delta\hat{A}_{it}$ is the unexpected wealth shock of individual *i* at time *t*. Also, $\bar{E}_j$ is mean earnings for individuals in an unexpected wealth shock cell (for example, men with over $250,000 of unexpected wealth gains) in 1994. The functional form for equation 5 is plotted on figure 11.

Recall from the introduction that understanding the marginal propensity to consume leisure is of central importance to understanding the marginal propensity to consume goods. Also recall that the life cycle model predicts that the marginal propensity to consume leisure plus the marginal propensity to consume market goods should add up to at least .04. Our interest is in whether the marginal propensity to consume leisure is a large fraction of that .04 number. An attractive feature of the specification in equation 5 is that the marginal propensity to consume leisure through the labor force participation decision is parameterized. For example, equation 6 shows that if $\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)<8, E[\Delta LFPR_{it} \mid (\frac{\Delta\hat{A}_{it}}{\bar{E}_j})]=.010\times\left(\frac{\Delta\hat{A}_{it}}{\bar{E}_j}\right)$. Note that this means that the average change in labor income for individuals in wealth cell *j* is $E[\Delta LFPR_{it} \mid (\frac{\Delta\hat{A}_{it}}{\bar{E}_j})]\times E_j=.010\times(\Delta\hat{A}_{it})$, or a marginal propensity to consume leisure of .01. In other words, for every unexpected $1 gain in wealth, earnings decline by one cent. This number is below the estimates of a one- to five-cent increase in consumption that most studies have found for the effect of stock market values on consumption. Therefore, the dominant behavioral response to increases in stock market wealth is in the form of increased consumption of goods, not in reduced labor force participation that leads to reduced earnings. Because the earnings response to changes in stock market wealth is small, life cycle models that ignore the effect of wealth on labor supply are not severely biased.

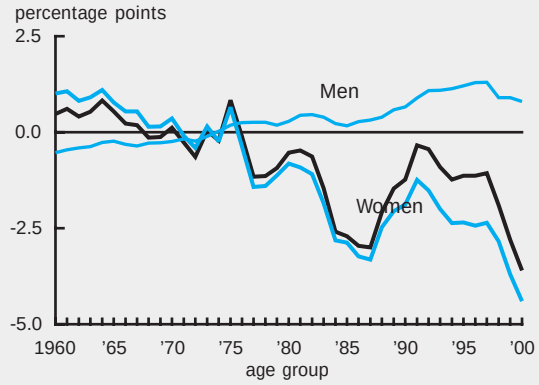
Predicted changes in labor participation due to stock market

Given the assumed labor supply function and the distribution of wealth shocks in the economy, we predict the aggregate labor supply response to the increase in stock wealth. Figure 12 shows that the predicted decline in labor force participation rates is about 1 percent for most groups, but that these estimates vary by group. For men aged 25–34, the predicted decline in labor force participation rates is .05 percent. Because older men have greater stock wealth than younger men, the predicted decline in labor force participation rates is greater for older men—.9 percent for men 55 and older.

We find larger predicted labor supply effects from the stock market for women. Women have lower earnings than men, but we assume women in married households have 50 percent of household wealth. Therefore, a $1 increase in wealth for a woman replaces a larger share of her lifetime resources than a $1 increase in wealth for a man. While this result depends critically upon the assumption of a 50/50 split of wealth for married households, most studies of income effects show larger income effects for women than for men, so we believe the results presented here are reasonable. The predicted labor supply response of women aged 25–34 to the increase in stock market wealth is a .17 percent decline in labor participation. The predicted

FIGURE 12

Change in labor force participation induced by wealth effect



Sources: Authors' calculations and PSID data.

decline becomes greater with age. For women aged 55 and over, the predicted response is a 2 percent decline in labor force participation rates.

Simulations of effect of unexpected wealth changes

Next, we describe an alternative approach to predicting the labor supply response to unanticipated changes in wealth. These results are from a dynamic model described in French (2000), which aims to accurately model the incentives individuals face over their life times. In this model, we characterize the preferences of people in the economy for consumption versus leisure, and we model how consumption and labor supply decisions by people of various ages are affected by changes in wages, wealth, taxes, and the structure of Social Security benefits. Individuals within the model choose consumption, work hours (including the labor force participation decision), and whether to apply for Social Security benefits. They are allowed to save, although assets must be non-negative. Therefore, they trade off the value of consumption in the present against the value of consumption in the future. Their annual income depends on asset income, labor income, and Social Security benefits. Individuals face federal and state income taxes as well as payroll taxes. When making these decisions, they are faced with several forms of uncertainty: survival uncertainty, health uncertainty, and wage uncertainty. The most interesting aspect of the model is the detailed modeling of the Social Security incentives to exit the labor market. Individuals who are younger than age 62 are ineligible for Social Security benefits. Once eligible for Social Security benefits, the individual faces a tradeoff of the value of receiving benefits in the present versus deferring them and receiving greater annual benefits in the future. Once the individual is drawing Social Security benefits, he or she faces the Social Security earnings test, which is a large tax on labor income above a certain threshold level.

There are seven preference parameters within the model. One parameter describes an individual's willingness to trade consumption in the present for consumption in the future. Another parameter describes an individual's willingness to trade goods for leisure. These preference parameters are estimated using data from the PSID. Given that individuals in the model face the same incentives as individuals in the data, they should behave just like individuals in the data at the true preference parameters. At the estimated parameters, the decisions of individuals in the model are very similar to those of individuals in the data. Therefore, we believe that the estimated parameters are "close to" the true preference parameters and that the model accurately describes how people behave. Consequently, we believe we can usefully apply the model to understand how the run-up in the stock market affects labor supply. We discuss the estimation of preference parameters in box 2.

Our simulated life cycle profiles for hours, labor force participation rates, and assets match the data very well. Simulated labor force participation rates begin to decline around age 55 and decline very rapidly at the exact ages of 62 and 65, when there are the strongest Social Security incentives to exit the labor market. Given that the model fits the data very well within sample, it potentially predicts well. Further details are in French (2000).

The model is useful in that it overcomes the previous problems in the lottery and inheritance studies. Most importantly, predictions from the lottery and inheritance studies assume that two individuals of different ages should have the same labor supply response to a $50,000 wealth shock given the same income. However, our expectation is that $50,000 in wealth at age 60 would generate a larger labor supply response than $50,000 at age 30. The 30-year-old will most likely save the money toward an early retirement, whereas the 60-year-old may use the money

BOX 2

Method of simulated moments

The method of estimating preference parameters in the simulation model is called the Method of Simulated Moments. It can be described as follows. First, we estimate life cycle profiles for assets, hours worked, and labor force participation rates using *Panel Study of Income Dynamics* (PSID) data. Second, we estimate individual histories of health and wage shocks using PSID data. Third, we solve the model backwards, obtaining optimal decisions for consumption, work hours, and whether to apply for Social Security benefits for each possible level of assets, wages, health status, and potential Social Security benefits. Fourth, we simulate individual life cycle profiles for assets, hours worked, and labor force participation rates using the individual histories of health and wage shocks and the decision rules from the structural model. Finally, we aggregate the simulated profiles and the data profiles by age and compare them. Preference parameters that create simulated profiles that *look like* the profiles from the data are considered the true preference parameters. Details are in French (2000).

immediately to retire early. The lottery and inheritance studies do not address this problem.

This model also overcomes the question of whether the wealth changes in the inheritance studies are anticipated. If inheritances are anticipated, estimates from the inheritance study will be biased towards zero effect on labor supply. Therefore, the predicted decline in labor force participation rates caused by the run-up in the stock market is biased toward a zero effect. Using simulations, one can generate wealth shocks that are completely unanticipated.

Earlier, we estimated that the run-up in the stock market resulted in aggregate wealth levels being 17 percent higher than they would have been had the late 1990s been average years for the stock market. We assume that these increases in wealth are taxed at 33 percent. Therefore, we assume that every $1 in wealth results in $0.17 in pretax unanticipated wealth gains and $0.11 in post-tax unanticipated wealth gains.

Because we do not formally model rates of return as a function of wealth and age, the simulation model potentially overstates the labor supply response to the run-up in the stock market. Using PSID data, we find that older individuals and individuals with high wealth have more of their portfolios invested in stocks than younger and lower wealth individuals. This tends to overstate the effect of the run-up in the stock market on labor supply. In our model, low-income individuals receive too much unanticipated wealth; and it is low-income people who are the most prone to dropping out of the labor market. However, wealthier and older individuals usually pay higher taxes. This attenuates the problem of high wealth people having higher rates of return since more of the return is taxed away. To the extent that the model does not completely overcome this problem, we are overstating wealth shocks for low wealth (and, thus, low-income) people.

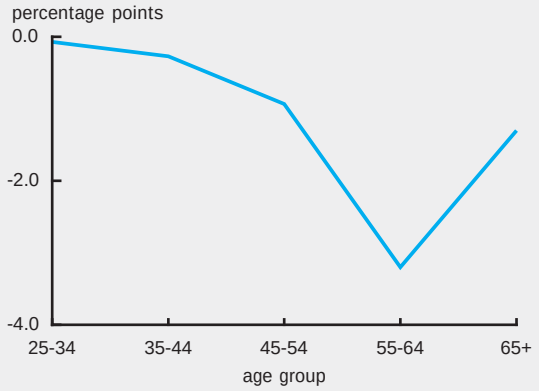
Figure 13 presents the simulated changes in labor force participation rates for men of different age groups. There are two striking differences when comparing the simulated changes against the predicted changes using the lottery and inheritance studies. The first is that the simulation study predicts much larger effects than the inheritance and lottery studies. For men aged 55–64, the simulations predict a 3.2 percentage point decline in labor force participation, whereas the inheritance and lottery studies predict only a .78 percentage point decline. As we described earlier, the inheritance and lottery studies might understate the effect of unanticipated wealth changes on labor supply. Inheritances are potentially anticipated and younger individuals usually receive inheritances. In the lottery study, the small prize group is much older than the medium and large prize groups, so the small prize winners are more likely to retire.

The second striking difference between the two sets of predictions is that our simulation only gives such large predictions for men aged 55–64. For men aged 65 and older, the simulation study predicts a 1.3 percentage point decline in labor force participation rates. This result is much closer to the results of the inheritance and lottery studies, which show a .90 percentage point increase for men aged 65 and above. The simulation study therefore provides a useful insight. Men younger than 55 are unlikely to drop out of the labor force regardless of the positive wealth shock. Most men older than 65 have already dropped out of the labor force. Men aged 55–64 are near the time when they exit the labor market. Therefore, the estimates from the other studies probably understate the effect of stock wealth on labor supply between the ages of 55–64 vis-à-vis other ages. Recall that the lottery study came to the same conclusion.

This last point is particularly important for assessing the estimates from the inheritance study. Recall that the inheritance study mostly uses information on individuals aged 35–44. Note that the predicted decline in labor force participation from the simulation study is only –.27 percent. This is only twice as large as the prediction for men aged 35–44 when using data from the inheritance study. This reaffirms our earlier point that by focusing on individuals aged 35–44, the inheritance study probably underestimates the labor supply response to changes in wealth for individuals aged 55–64. Figure 13 shows that this underestimate is likely to be significant.

FIGURE 13

Simulated change in labor force participation for men induced by wealth effect



Conclusion

In this article, we quantitatively assess the effect of the run-up in the stock market on aggregate labor supply. We arrive at our conclusions using three steps. First, we estimate the total size of the aggregate wealth shock. We find that every dollar invested in the stock market on December 31, 1994, produced on average $1.12 in stock wealth gains by December 31, 1999. Given the aggregate level of wealth in stocks in 1994, the aggregate unanticipated increase in wealth between 1994 and 1999 was $5.8 trillion, which represents an unanticipated increase in wealth of almost $20,000 per person in the U.S.

Second, we estimate the magnitude of the unanticipated wealth shock for different age groups. Using PSID data, we find that very few people younger than age 55 today benefited greatly from the run-up in the stock market. About 15 percent of all individuals aged 55 and above had unanticipated wealth increases of greater than $50,000. For most individuals, $50,000 would be more than enough to afford an additional year of retirement without any change in the consumption of market goods.

Third, we predict the effect of the run-up in the stock market on labor supply. We find that labor force participation rates for individuals aged 55 and older have increased since 1995. Increases in stock market wealth should cause reductions in labor force participation rates, all else equal. This implies that the stock market has not been the dominant factor influencing labor force participation rates from 1995 to the present. Other factors, such as rapidly rising wages, seem to be more important.

We use two approaches to predict the effect of rising stock prices on labor supply. In the first approach, we take estimates of the size of the wealth effect from other studies. Although nobody has used variation in stock prices to estimate the wealth effect on labor supply, researchers have used data on inheritances and lotteries to estimate the effect of wealth on labor supply. Using estimates from these studies and the estimated distribution of wealth shocks to different groups of people in the economy, we estimate that in the absence of a run-up in the stock market, aggregate labor force participation rates would be 1.16 percent higher today. We believe that these are conservative predictions of the stock market effect.

Our second approach is to use simulations from a model described in French (2000). We find that simulations from this model give much larger predictions of the effect of the run-up in the stock market. The predicted decline in labor force participation rates for men is over 1 percent, on average. (The model does not address the labor supply response of women.) The simulations also predict that the largest effects should be at age 55–64, when men are considering exiting the labor force. For this age group, the predicted decline in labor force participation rate is 3.2 percent. These results might overstate the effect. Therefore, we interpret the predictions based on estimates from the lottery and inheritance studies as a lower bound on the effect and the simulations as an upper bound.

Lastly, we note that the lottery, inheritance, and simulation studies imply that for every $1 in increased wealth, earnings decline by one to two cents. As we noted at the outset, total consumption of goods plus leisure must rise by at least four cents to be consistent with the life cycle model. This means that consumption of goods must rise by at least two or three cents in order to be consistent with the life cycle model. Most empirical estimates are in the range of one to five cents; as such, results at the lower end of this range are at odds with the life cycle model. Therefore, our work provides additional evidence that either the marginal propensity to consume market goods is at least 2–3 percent or that the life cycle model is not a reasonable model of consumer behavior.

APPENDIX: DESCRIPTION OF DATA

Risk-free asset data

We calculate the risk-free five-year return as the continuously compounded return on holding five-year Treasury instruments to maturity. To obtain the zero coupon rate, we use data that have been adjusted using a Fisher/Zervos technique. We obtained Fisher/Zervos estimates for 1961 to the present from the Federal Reserve Board of Governors, Division of Research and Statistics (courtesy of Mark Carey). For returns prior to 1961, we use five-year Treasury bonds. There is only a small difference in returns between the two data series (during periods of overlap).

Stock market data

Stock market annual returns (including dividends) for 1926 to the present are from the Center for Research in Security Prices (CRSP) Index Series, No. 100080, a value-weighted portfolio of all NYSE, AMEX, and NASDAQ stocks. We impute missing entries for 1999–2000 using 1970–2000 S&P 500 total return (including

dividend reinvestment) data. Values represent end of December 31 to end of December 31 returns.

Flow of funds data

Our data on the market value of equities owned by households (and related data) are from the Federal Reserve Board of Governors, *Flow of Funds Accounts of the United States*. All values are in 1999 billions of dollars and represent year-end levels. Equities in pensions include defined contributions pensions only. Equities are defined as shares of ownership in financial and nonfinancial corporate businesses, both common and preferred shares of domestic corporations, and U.S. purchases of shares of foreign corporations (including ADRs).

Price level data

We use December levels from the Consumer Price Index for all urban consumers (1999 = 100) to make our price adjustments.

PSID data

We use the 1989 and 1994 waves of PSID data in our analysis of stock and pension wealth. Our 1989 sample excludes those who do not provide an answer regarding pension status or do not respond to whether they contributed to a pension (for either husband or wife in the household). The 1994 sample includes only those families in the 1989 sample, less those who changed marital status or whose head of household had changed since 1989. We use the 1989 weights wherever applicable.

Juster et al. (1999) show that the 1989 PSID accounts for approximately 85 percent of household stock wealth in the *Survey of Consumer Finances* (SCF). Limiting the sample to only those who match between 1989 and 1994 (less those who experienced a change in marital status or change in household head during those years) results in a 30 percent higher mean stock wealth than the full 1994 sample (using 1994 weights, when both are scaled by 1/0.85 = 1.18), so we adjust by scaling down 1994 stock wealth by a total factor of 1.18/1.30 = 0.91. In order to analyze men and women separately, we assume that allocation of non-pension stock wealth in married households is 50 percent to each spouse.

To analyze 1989 pension wealth, we use a simplified model assuming constant lifetime accrual amounts (in 1999 dollars) with a real return of 2.3 percent per year. Using figures from Gustman and Steinmeier (1999) and assuming that half the wealth in combined pension plans is in the form of defined benefit wealth, the average level of wealth in defined contribution plans per household is $69,000 (1999 dollars). Assuming that it is at age 65 when the amount is $69,000, and that it is at age 35 when the worker starts contributing, we compute a schedule of pension wealth at each age. We give workers less than 35 years old one year's worth of pension accrual and workers older than 65 the maximum amount ($69,000). Assuming a 5 percent contribution rate, we use our imputed annual accrual level to impute the associated level of annual earnings. We then assign each worker (who has a defined contribution plan) a level of pension wealth equal to the previously calculated pension wealth level (associated with their age), scaled by the ratio of their 1988 earnings (from the PSID) to the imputed mean level of earnings from the HRS. To analyze the amount of pension wealth in stocks, we assume that stocks comprise 50 percent of pension wealth. To find pension wealth in 1994, we use the previously calculated schedule of pension wealth to assign a new level of pension wealth based on the individual's new age, again using 1988 earnings to scale the amount.

We make a number of other imputations to account for shortcomings in the PSID data. First, since the 1989 PSID data contains only 1988 earnings (not the pre-retirement earnings level, which is the earnings level in question), we impute the level of earnings of pre-retirement work as follows. For individuals who are covered by a pension and work more than 1,000 hours, we do not modify their level of earnings; if they work less than 1,000 hours, we take their pre-retirement earnings as the earnings in the data plus the mean earnings of those who work more than 1,000 hours, less the mean earnings of those who work less than 1,000 hours. To obtain a person's 1994 earnings (required for the computation of labor supply elasticities), we look to the person's earnings in the 1995 PSID and proceed in a similar fashion.

The 1989 PSID pension question does not allow retired persons to indicate whether they were covered by a pension while they were working; therefore, we take positive pension income as an indicator for a pre-retirement pension. To find whether that pension is a defined contribution pension, we perform the following procedure. If a person indicates pension coverage in response to the direct PSID question regarding pension coverage, we take the response to whether they contribute toward that pension as given. If a person indicates no pension coverage, but is receiving positive pension income, we assign the person a random number (according to a uniform [0,1] distribution); if that number is less than the probability of

having a defined contribution plan (given age and pension coverage), we assume the pre-retirement pension is a defined contribution plan (otherwise not). We calculate the probability of having a defined contribution plan as follows:

Pr(DC | age, pension = yes) = Pr(DC | age) / Pr(Pension | age),

where

Pr(DC | age) = Pr(PSID DC = yes | age) + [Pr(PSID DC = yes | PSID Pen = yes) × Pr(receiving pension income | age)].

Lastly, to find the people who are covered by defined contribution plans in 1994, we simply carry over those who were covered in 1989, since the 1994 PSID data release at this time does not include any questions regarding defined contribution pensions.

NOTES

[1]This effect would be ambiguous, however, as people may have wished to work more hours in 1995–99 in order to generate more wealth. Increased wealth could in turn be invested in the stock market.

[2]One caveat to this article is that it is not clear why the stock market rose in the first place. Our analysis assumes that the stock market rose for reasons unrelated to future productivity growth—perhaps financial markets have become more efficient. If the stock market rose because of beliefs about increasing productivity in the future, then there are three effects that we do not consider here. First, people should believe that wages will rise rapidly in the future because of increased productivity. Not only would stockholders feel wealthier, but so would all individuals who believe that they will be working in the future. If this is true, our analysis underestimates the true wealth shock to the economy and, thus, underestimates the true effect of the run-up in the stock market on labor supply. Second, any change in future beliefs about productivity is likely to be accompanied by wage changes in the present and near future. This potentially increases hours worked as incentives for work are greater. This offsets the wealth effect. Third, interest rates should rise if people believe productivity will rise, because higher productivity leads to higher demand for capital. If interest rates are relatively high (as they are today), people should work more hours today so that they can develop greater wealth that will earn a high rate of return. Again, this offsets the wealth effect. Any rapid change in stock prices will likely be accompanied by these three additional effects, if the change in prices reflects changing beliefs about future productivity. Therefore, it is not clear how labor supply would respond to a large stock market change in the future. However, we believe that we have increased understanding of the effect of the stock market on labor supply by focusing on the direct effect.

[3]This is done by comparing the PSID to another dataset, the *Survey of Consumer Finances*, and assuming that respondents in the *Survey of Consumer Finances* report 100 percent of their assets. The *Survey of Consumer Finances* is considered to have extremely high quality data on wealth, although the respondents probably report slightly less than 100 percent of their assets.

[4]They do note that when the husband's income accounts for 75 percent of total household income instead of 50 percent, the husband's share of consumption rises by about 2 percentage points. This shows that assuming an even split is not perfect but is roughly correct. They also note that consumption of women's clothing is slightly higher than men's, but again assuming an even split of resources is roughly correct.

[5]Using the previously described procedure to estimate defined contribution wealth, we aggregate defined contribution wealth in our PSID sample up to the national level. In other words, we take aggregate wealth in the PSID and multiply it by the ratio of U.S. households to PSID households. We compare this estimate to defined contribution wealth in the *Flow of Funds*. We find that our PSID defined contribution measure is 15 percent greater than the *Flow of Funds* measure.

[6]See *Social Security Bulletin Annual Statistical Supplement*, 1997, p. 60.

[7]Unfortunately, the PSID measures only food consumption not total consumption. The consumption results show that the marginal propensity to consume food out of inheritances is about .1 percent, far lower than the 1 percent to 5 percent marginal propensity to consume out of changes in stock wealth that most studies find (Parker, 1999; Ludvigson and Steindel, 1999; and Dynan and Maki, 2000). Because food is a necessity, the marginal propensity to consume food is lower than the marginal propensity to consume all consumption goods. For example, Attanasio and Weber (1995) show that for every 1 percent increase in food consumption, total consumption rises about 1.2 percent.

[8]All calculations assume that the after-tax real interest rate is 2.3 percent and the inflation rate is 3.3 percent.

REFERENCES

Attanasio, Orazio, and Guglielmo Weber, 1995, "Is consumption growth consistent with intertemporal optimization? Evidence from the Consumer Expenditure Survey," *Journal of Political Economy*, Vol. 103, No. 6, pp. 1121–1157.

Blundell, R., and T. MaCurdy, 1999, "Labor supply," *Handbook of Labor Economics*, Orley Ashenfelter and David Card (eds.), Amsterdam: Elsevier, North Holland.

Board of Governors of the Federal Reserve System, 2000, *Flow of Funds Accounts of the United States: Flows and Outstandings, Fourth Quarter 1999*, Washington: Board of Governors of the Federal Reserve System.

Browning, Martin, François Bourguignon, Pierre-Andre Chiappori, and Valerie Lechene, 1994, "Income and outcomes: A structural model of intra-household allocation," *Journal of Political Economy*, Vol. 102, No. 6, pp. 1067–1096.

Cochrane, John, 1997, "Where is the market going? Uncertain facts and novel theories," *Economic Perspectives*, Federal Reserve Bank of Chicago, Vol. 21, November–December, pp. 3–37.

Dynan, Karen, and Dean Maki, 2000, "Does stock market wealth matter for consumption?," Board of Governors of the Federal Reserve System, mimeo.

French, Eric, 2000, "The effects of health, wealth, and wages on labor supply and retirement behavior," Federal Reserve Bank of Chicago, working paper, No. 00-02.

Gustman, Alan, Olivia Mitchell, Andrew Samwick, and Thomas Steinmeier, 1998, "Evaluating pension entitlements," Dartmouth College, mimeo.

Gustman, Alan, and Thomas Steinmeier, 1999, "Effects of pensions on savings: Analysis with Data from the Health and Retirement Study," *Carnegie-Rochester Conference Series on Public Policy*, pp. 271–324.

Holtz-Eakin, Douglas, David Joulfaian, and Harvey Rosen, 1993, "The Caragie conjecture: Some empirical evidence," *Quarterly Journal of Economics*, Vol. 108, May, pp. 413–435.

Imbens, Guido, Donald Rubin, and Bruce Sacerdote, 1999, "Estimating the effect of unearned income on labor supply, earnings, savings, and consumption: Evidence from a survey of lottery players," National Bureau of Economic Research, working paper, No. 7001.

Joulfaian, David, and Mark Wilhelm, 1994, "Inheritance and labor supply," *Journal of Human Resources*, Vol. 29, No. 4, pp. 1205–1234.

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

Kennickell, Arthur, Martha Starr-McCluer, and Brian Surette, 2000, "Recent changes in family finances: Results from the 1998 *Survey of Consumer Finances*," *Federal Reserve Bulletin*, January, pp. 1–29.

Ludvigson, Sydney, and Charles Steindel, 1999, "How important is the stock market effect on consumption?," *Economic Policy Review*, Federal Reserve Bank of New York, July, pp. 29–51.

Parker, Jonathan, 1999, "Spendthrift in America? On two decades of decline in the U.S. savings rate," *NBER Macroeconomics Annual*, pp. 317–370.

Poterba, James, 2000, "Stock market wealth and consumption," *Journal of Economic Perspectives*, Vol. 14, No. 2, pp. 99–118.

U.S. Social Security Administration, 1997, *Social Security Bulletin Annual Statistical Supplement*, Washington.

Index for 2000

To order copies of any of these issues, or to receive a list of other publications, please telephone (312)322-5111 or write to:

Federal Reserve Bank of Chicago
Public Information Center, P.O. Box 834
Chicago, IL 60690-0834

Economic Perspectives and other Bank publications are available on the World Wide Web at www.frbchi.org.