Financial Market Frictions

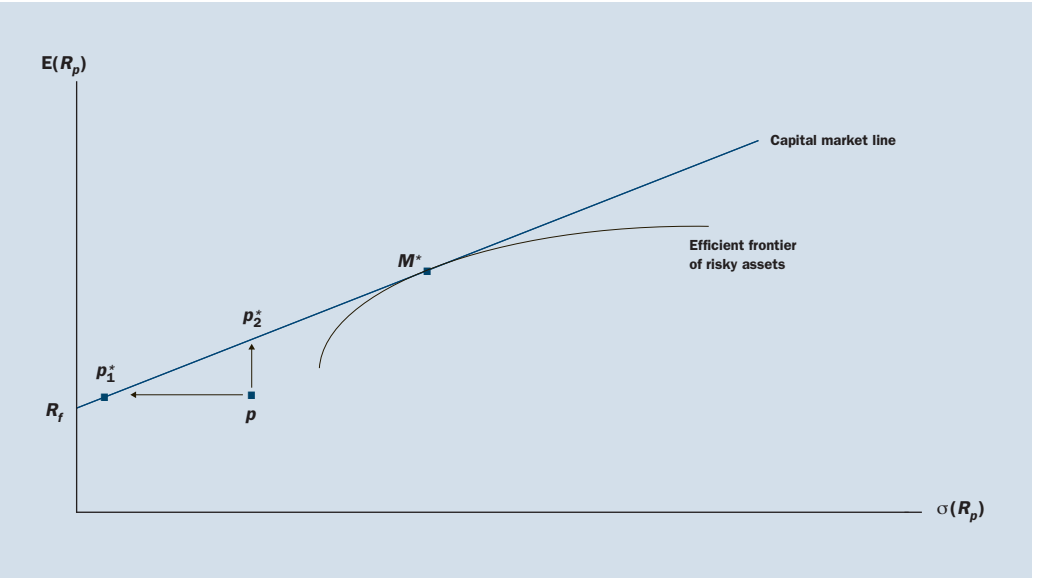
RAMON P. DEGENNARO AND CESARE ROBOTTI
DeGennaro is the SunTrust Professor of Finance at the University of Tennessee, Knoxville, and a visiting scholar in the Atlanta Fed's research department. Robotti is a financial economist and assistant policy adviser in the Atlanta Fed's research department. *They thank Tom Cunningham, Gerald P. Dwyer Jr., Paula A. Tkac, and Dan Waggoner for useful comments and discussions that clarified several points. This is a greatly revised and extended version of an article by DeGennaro (2005).*

What comes to mind when we hear the phrase "financial market frictions"? Most of us think first of taxes and transactions costs. These are obvious examples, but market frictions are diverse and widespread, affecting virtually every transaction in some way. Capital gains taxes, for example, influence decisions to trade stocks and bonds. The financial market friction need not be a monetary cost: We sometimes must stand in line to pay a lower price. New businesses must charge lower prices than companies with established reputations. Companies include stock options in their compensation packages to mitigate well-known incentives for agents to shirk and to avoid rules that trigger tax penalties for "nonperformance-based compensation" that exceeds $1 million.[1]

What is a market friction? In the context of the capital asset pricing model (CAPM), this article defines a financial market friction as anything that interferes with trade. This interference includes two dimensions. First, financial market frictions cause a market participant to deviate from holding the market portfolio. By implication, these frictions can cause a market participant to be exposed to more or less risk than she might prefer. This definition at first seems very limited but is, in fact, only as limited as the definition of the market portfolio. In this article, the term *market portfolio* means not only financial assets but also real estate, human capital, investors' time, and so on. Put differently and somewhat less obscurely, financial market frictions generate costs that interfere with trades that rational individuals make (or would make in the absence of market frictions).

This concept can be clarified within the context of the capital market line (CML) (see Figure 1). The CML shows, in two dimensions, the optimal holdings available to investors, defined by the standard deviation of the portfolio, $\sigma(R_p)$, and the expected return on the portfolio, $E(R_p)$, given that a risk-free asset exists and that investors can freely borrow and lend at that rate. Risk-averse investors prefer portfolios lying above and to the left of those lying below and to the right—they want the highest

Figure 1
The Capital Market Line and the Efficient Frontier of Risky Assets



expected return and the lowest risk. In a financial market with no frictions, investors achieve this risk-return trade-off by holding the market portfolio, M^*, and a (possibly short) position in the riskless asset, R_f. Intuitively, they hold the maximally diversified portfolio and achieve their preferred risk level by adjusting their holding of the riskless asset. This allocation dominates a portfolio of only risky assets in all cases except the point of tangency between the efficient frontier of risky assets and the CML.

In a financial market with frictions, though, investors cannot costlessly adjust their holdings. An investor holding the suboptimal portfolio p (perhaps because of an illness, inheritance, or change in employment or marital status) could lower her risk without sacrificing expected return by rebalancing to hold portfolio p_1^*. Or she could improve her expected return without accepting any more risk by rebalancing her portfolio to hold portfolio p_2^*. But rebalancing is costly or impossible in a financial market with frictions. It may pay to accept portfolio p's inferior combination of risk and expected return rather than to incur the costs of trading. For example, consider a stock investor who prefers a fifty-fifty mix of stock and bonds. If stock prices rise while bond prices do not, then the portfolio becomes overweighted with stocks and is too risky for this investor. But selling some of the equities to reestablish the fifty-fifty mix would trigger capital gains taxes. Because of this, the investor may choose to retain the unwanted risk exposure rather than incur a tax liability.

To appreciate this concept algebraically, define α_{ij} as the proportion of investor i's portfolio held in asset j, and define A_{ij} as the value of asset j held by investor i (with A_j defined as the value of asset j.) Also define M^* as the value of the market portfolio, which includes all risky assets. Then $\Sigma_j A_j = M^*$ because all assets must be held. Under the CAPM, $\alpha_{ij}^* = (A_j/\Sigma_j A_j)$ for each asset j. The CAPM tells an investor to invest α_{ij} of his portfolio in asset j, where α_{ij} equals the value of asset j relative to M^*. The result is that he holds a fraction of the market portfolio. In this paper, a financial market friction is anything that drives a wedge between α_{ij}^* and α_{ij} with expected-

utility maximizing investors, or anything that drives a wedge between the amount of risk that the investor bears and the amount that he prefers to bear given the trade-off between risk and expected return.

We make an important distinction between market financial market frictions and market inefficiencies. We assume that asset prices reflect all available public information but not necessarily all private information. Pricing errors, if they exist, are not financial market frictions. Even if an asset's price is wrong, market participants base their choices and weight their portfolios using this incorrect price. By our definition (as with most), markets can be efficient yet have frictions that interfere with trade.

Why Do We Care about Financial Market Frictions?

Financial market frictions matter for three main reasons:

- Financial market frictions can generate real costs for investors. Recognizing these costs helps us understand the total costs of transactions and decide where to place them and even whether to make them at all. The capital gains tax is an obvious example. Constantinides (1984) shows that the option to take or defer capital losses or gains has substantial value. The option's exact value—and the corresponding optimal trading strategy—depends on factors such as transactions costs, the capital gains tax rate, and the asset's volatility.
- Financial market frictions also generate business opportunities. After all, many costs are paid to someone or to some entity. Institutions that can lower costs stemming from market frictions have a competitive advantage. Until competing firms adapt, they can earn economic rents. One example from the financial markets is mutual funds, which relax wealth constraints and asset indivisibilities.[2] (See DeGennaro and Kim 1986.) Other examples are two exchange-traded funds, the American Stock Exchange's Standard and Poor's Depositary Receipts, better known as Spiders, and Nasdaq-100 Index Tracking Stock, better known as QQQs. Spiders and QQQs provide another solution to the asset indivisibility problem. Sodano (2004) reports that Spiders and QQQs are the two most actively traded securities in the world.
- Financial market frictions can and do change over time. The degree of existing market frictions varies, new ones appear, and existing frictions disappear. Bank analysts now face the daunting task of analyzing far larger and more complex institutions than existed twenty years ago, but this challenge is offset in part by a vast increase in the information and computing power now available to them. Kane (2000) shows that regulators face a similar problem: The complexity and difficulty of resolving an undercapitalized institution increases with the size of the institution, and megamergers have the capacity to shift the political calculus of a resolution, and all the financial market frictions that entails, enormously. Another change is the shift from qualitative information to quantitative information. For example, a note stating that a credit applicant has a good reputation may now be quantified as having a FICO (credit) score of 790. This tool lowers the cost of lending at a distance.

---

1. According to Section 162(m) of the Internal Revenue Code, publicly held corporations cannot deduct compensation in excess of $1 million paid to a "covered employee" from taxable income. The code makes an exception for stock option plans, though, provided that they meet certain requirements.
2. In theory, an investor can hold an infinitesimally small fraction of any asset. In practice, doing so is impossible. Economists refer to this dilemma as the asset indivisibility problem.

Market Structure

Financial market frictions, especially transactions costs, depend in part on market structure. Market structure, in turn, depends on both the risk of the traded asset and trading volume. In thin markets for risky assets, participants search for counterparties directly because the fixed costs of capital investments (including communication) are too large to be offset by the lower marginal costs of each transaction if transactions are few. As trading volume increases, markets evolve from direct search through brokered, dealer, and continuous auction markets. This evolution is a simultaneous process: As volume increases, the structure evolves, and as the structure evolves, trading volume increases. The potential size of the market determines the equilibrium structure.

As trading volume increases, it begins to make sense to invest in capital and to acquire specialized knowledge about potential buyers and sellers to facilitate trading. Stockbrokers are one example. If volume increases still further, or if risk decreases, brokers find it efficient to buy and sell on their own accounts. Although holding inventory is risky, if the asset value is sufficiently stable or if its liquidity is sufficiently high, then this risk is worth taking because holding inventory permits the dealer to make more trades in less time. For some assets, trading volume is so high that a continuous auction is possible. A good example is the secondary market for U.S. Treasury securities.

Of course, the market for some assets switches from one structure to another. The market for equities might be dominated by brokers most of the time, but at other times, dealer markets or continuous auctions might emerge. The specialist, for example, often simply crosses buy and sell orders but sometimes fills orders from his own inventory.

Some participants with expertise or investment in one type of market structure, such as real estate agents, might tend to resist changes that dilute their competitive advantage. In general, though, society tends to move from higher-cost market structures to lower-cost ones. For example, Cox and Koelzer (2000) say that the Internet has transformed the way that agents and consumers form their relationships. Housing is not a standardized commodity, so a market similar to the New York Stock Exchange is impractical. However, buyers today find it much easier to bypass a real estate broker entirely. If they do use a broker, the Internet is often the tool they use to select one. The Internet is particularly important for buyers from distant locations.

In short, as trading volume increases, markets tend to evolve from a structure with low fixed costs and high marginal costs for transactions to markets with high fixed costs and low marginal costs. Transactions costs are lower in these high-volume markets.

Can We Classify Financial Market Frictions?

The answer is yes, at least in part. The universe of financial market frictions can be partitioned in many ways. Because there are many financial market frictions, though, no structure can hope to be complete. Neither can it hope to be very precise; for any feasible partitioning, some financial market frictions can fall into more than one category. Still, providing such a structure is useful. How can this be done?

We build our structure on the economic forces underlying financial market frictions. This structure also takes a step toward identifying those entities best able to reduce the costs of market frictions. We use five primary categories: transactions costs, taxes and regulations, asset indivisibility, nontraded assets, and agency and information problems.

Transactions costs. We partition transactions costs into two categories: the costs of trade and the opportunity costs of time.

The costs of trade. The costs of trade in financial markets include postage, telephone charges, computer power, and similar real expenditures of resources. These have been declining with technological improvements. Over some periods these costs may have risen in real terms, but the costs of communication and data analysis have fallen over time. For example, the cost of an e-mail message is effectively zero. And the costs of virtually all other mechanical costs of trade have fallen. There is no reason to expect this trend to stop. For example, on March 7, 2006, the New York Stock Exchange merged with Archipelago, an electronic trading firm, and the two firms became wholly owned subsidiaries of NYSE Group Inc. This merger is likely to lower bid-ask spreads and therefore the marginal cost of trading securities for some investors.

Financial market frictions can generate real costs for investors. Recognizing these costs helps us understand the total costs of transactions and decide where to place them and whether to make them.

The opportunity costs of time. Trading requires time, which includes both search costs, or the time to gather information (including finding a trading partner), and the time to make the trade itself. Minimizing these costs represents a profit opportunity. One partial solution is to automate the process by means such as automatic electronic payments. Many investors fund their 401(k) plans this way, often via payroll deduction. Another example is dividend reinvestment plans, which let investors hold securities directly and automatically reinvest dividends (DeGennaro 2003). In all these cases, investors need to act only once to make several investments over an unspecified and possibly very long period. Other reductions in the time required to trade are sure to follow, both because technology continues to advance and because the opportunity cost of time tends to rise over time.

The future of transactions costs. Transactions costs are probably among the most familiar financial market frictions. Today, though, they might also be among the least important. Advances in communications and data-handling technology have reduced not only the costs of trade to a fraction of what they were just a few years ago but also the time needed to make trades. Together, these forces probably more than offset an increase in the opportunity cost of time itself. Vayanos (1998), for example, finds that realistically small transaction costs have negligible effects on asset returns and mainly affect the portfolio rebalancing frequency.

Taxes and regulations. The second major category in our taxonomy of financial market frictions is taxes and regulation. We use the term *regulation* loosely in this paper to encompass laws passed by legislative bodies as well as rules imposed by government agencies and industries themselves. Privately imposed rules, therefore, such as exchange-imposed trading rules, count as regulations. Taxes and regulatory costs may be either explicit or implicit. The corporate income tax is explicit: The statute imposing the tax calls it a tax, and the corporation sends funds to the government. Other taxes are implicit, such as capital requirements that insured banks must meet (Buser, Chen, and Kane 1981). In this case, the statute authorizing the capital requirements does not refer to them as taxes, and the banks do not send funds to the government to discharge the liability. But these requirements still increase the cost of doing business and operate like a tax. Regulation varies widely across jurisdictions both within the United States and internationally, as does the degree of coordination between the United States and other countries. We focus on the United States for space considerations, though the concepts are applicable to other jurisdictions.

Explicit taxes. Everyone is familiar with any number of pecuniary taxes; governments both within and outside the United States impose explicit pecuniary taxes in hundreds if not thousands of ways. Corporations pay taxes on income, which change prices.[3] Taxes can even affect the medium of exchange. For example, corporate acquisitions paid for with stock can receive more favorable tax treatment than those paid for with cash.

Individuals pay income and capital gains taxes, and these payments surely affect their investment decisions and trades. Just as surely, income taxes affect individuals' consumption decisions and their willingness to work.

The universe of financial market frictions can be partitioned in many ways, so no structure can hope to be complete or very precise.

Taxes can also be nonpecuniary, paid not in dollars but in effort, time, and resources. Miller and Scholes (1978) give a good example of a nonpecuniary tax.[4] They show how investors can generate deductions to offset dividends earned in order to eliminate the tax on the dividends. In practice, though, this offsetting is costly. The cost to taxpayers of an explicit tax extends far beyond the dollars remitted to the taxing authority. Taxpayers can and do take steps to minimize the amount they pay, and the costs of these steps count toward the total tax burden. Other examples are the costs of becoming informed about tax avoidance and the cost of suboptimal portfolio choices.

Implicit taxes. Privately imposed regulations (or restrictions) are easy to find. For example, the May 1, 2006, prospectus (page 11) for the RetireReady℠ Choice annuity issued by Genworth Life and Annuity Insurance Company gives surrender charges (as a percentage of purchase payments partially withdrawn or surrendered) of 6 percent for years one through four, 5 percent for year five, 4 percent for year six, and zero after that. Because underwriting contracts is costly, annuities are designed to be long-term investments, and issuers impose these fees to discourage customers from canceling the contracts after short periods. Otherwise, if investors hold such contracts for only short periods, transactions costs would harm the contract's performance. In turn, this lower performance would make the contract less attractive to those who seek a low-cost, long-term investment. Still, these restrictions do limit trading because an investor wishing to abandon this annuity and invest the proceeds elsewhere might find it too costly to do so. Although the result is that he holds a subobtimal portfolio, doing so can be preferable to paying the surrender charge.

Another example of a privately imposed regulation is short-sale restrictions. Rule 3350 of the National Association of Securities Dealers Inc. (NASD) forbids its members from short selling securities on the Nasdaq National Market System in situations that it fears might magnify price declines. Members cannot short sell at or below the best bid (the highest bid by all market makers quoting that stock) if the best bid is below the previous best bid for that stock. Such a restriction limits members' trading, thus fulfilling the definition of a financial market friction, but restrictions on short sales can also keep prices from adjusting to equilibrium levels as fast as they would otherwise. Informed traders would prefer to sell an overpriced security short, expecting to profit when the price returns to its equilibrium level. These short sales tend to eliminate the overpricing sooner. With short-sale restrictions, though, any deviation from equilibrium can persist longer. Thus, a financial market friction tracing to regulation can lead to pricing errors.

Govenment-imposed regulations can also create financial market frictions. Some of these closely parallel self-imposed regulations. For example, Rule 10a-1 under the Securities Exchange Act of 1934 governs securities registered on an exchange. Rule 10a-1's key provision is the tick test: Subject to certain exceptions, an exchange-listed security may be sold short only at a price at least as high as the last different reported price. This rule is very similar to NASD Rule 3350.

Reporting requirements are another example of nonpecuniary implicit taxes. The U.S. Securities and Exchange Commission (SEC), for example, requires numerous filings. Economists might debate the value of these reports, but no one can dispute the claim that they impose costs on businesses. The SEC's EDGAR Web site (www.sec.gov/edgar.shtml) gives some idea of how extensive this burden is. Another well-known example of a govenment-imposed reporting requirement is the Sarbanes-Oxley Act of 2002. Among this sweeping legislation's provisions are an increase in management accountability and the requirement that companies institute certain internal controls. Compliance has been expensive. Financial Executives International (2005) surveyed 217 public companies with revenues averaging $5 billion and found that the costs of compliance averaged $4.36 million per firm.

How do these compliance costs affect portfolio allocations and trades? If a corporation decides that the burden is large enough, then one option is to take the firm private because privately held companies are not required to file most of these forms. Thus, these requirements provide incentives to forgo access to the public capital markets, making it more costly for investors to hold them in their portfolios. In such a situation, the investors are imperfectly diversified, and the portfolios lie below the capital market line.

Clearly, the breadth and influence of taxes and regulations are enormous. Managing and coping with them requires a correspondingly large investment; hundreds of thousands of lawyers, accountants, and practitioners labor daily to comply with taxes and regulations in the least costly way for firms and households.

Asset indivisibility. If assets were infinitely divisible, then investors could hold an arbitrarily small portion of each asset. This practice would permit all investors, even those with little to invest, to hold the market portfolio of all investable assets. In fact, though, assets are lumpy—the minimum traded unit is finite. This means that most investors must decide whether to hold the smallest traded unit of an asset or to omit it from their portfolios. Either way, their resulting portfolios will not be invested in the same proportions as the market portfolio and thus will lie below the capital market line in Figure 1. For wealthy investors, asset indivisibility is a smaller problem than it is for less wealthy ones. In addition, a wealthy investor can hold a larger number of assets. Combined with trading costs, which usually have a fixed component, asset indivisibility makes it harder for investors of limited means to begin investing because their portfolios tend to lie farther below the capital market line. Asset indivisibilities are an important reason mutual funds and derivative securities such as Spiders and QQQs exist. By pooling funds from many investors, they permit investors to hold portfolios that more nearly approximate the market portfolio. This process is costly, though, and some indivisibilities remain because it is too expensive to eliminate them all.

---

3. Financial economists realize, of course, that corporations do not really pay taxes. Rather, they collect taxes and remit them to the government.
4. Some people might classify tax avoidance as an implicit tax rather than a nonpecuniary explicit tax. That approach would make sense, and this duality illustrates the inherent difficulty with constructing a taxonomy of financial market frictions.

Nontraded assets. Becker (2005) reports that human capital now makes up at least 70 percent of all wealth in economically advanced nations. This enormous capital stock tends to drive workers away from holding the market portfolio. For example, consider an employee of a publicly traded corporation. In a perfect market, he should hold less of his employer's stock than he otherwise would for diversification purposes because he is more likely to lose his job if his employer's stock has done poorly. The positive correlation between job loss and invesment losses magnifies risk. This strategy is unavailable to employees of privately held companies, though. In general, employees of privately held companies are forced to hold a disproportionate stake in their own human capital.

The separation of ownership and control is a financial market friction because this separation can lead to incentive problems, and financial contracts cannot handle them at zero cost.

Or are they? In a market free of frictions, an investor has an alternative to reducing his stake in his employer's shares to compensate for his increased exposure through his salary. Instead of holding fewer shares, he can sell claims on his human capital. Consider a musician. Typically, he performs and earns income over time. But suppose that instead he sells claims on his future earnings and invests the proceeds in the market portfolio, M^*. In this case, the investors who buy the claims collect pro rata shares of the funds the musician earns over time.

Selling claims against one's human capital is not as impossible as it sounds. In fact, examples are becoming increasingly common. Palacios (2002) gives one explicit example for human capital contracts for financing higher education in the United States.[5] Palacios's solution to the problem of human capital sale is impractical for at least two reasons. First, transactions costs exist. Second, and more importantly, incentive problems can remain (see the following section). We can expect financial markets to develop ways to reduce these costs, and, even now, these contracts fill an important gap in financial markets.

Financial innovation has spawned other intriguing examples. For example, in January 1997 David Bowie raised $55 million by issuing ten-year asset-backed bonds.[6] What is innovative about this issue is that future royalties from twenty-five albums that Bowie recorded before 1990 are the collateral backing these bonds. That such a performer could issue such securities serves as a good example of financial ingenuity. Similar deals were soon arranged with other artists, including James Brown (June 1999), the Isley Brothers (September 1999), and the estate of Marvin Gaye (September 2000).

Financial innovation continually removes items from the list of nontraded assets by introducing new instuments that render assets effectively tradable. In addition to the human capital examples above, recent years have seen credit-card securitizations, credit-spread derivatives, collateralized mortgage obligations, and many others. In some of these cases, bundling the assets reduces idiosyncratic risk. In others, the innovation permits unbundling the assets' risk and selling parts of it to investors who are better able to bear it (for example, credit-default swaps). This is not to say that if an asset begins to be traded, then the market friction has been eliminated. More accurately, the friction has been mitigated or exchanged for another (presumably) less onerous friction. Taking the example of human capital sales, one obvious problem is that it might not be legal to sell certain claims on future income. If not, then that legal restriction (in this article, a regulatory financial market friction) complicates the problem of an asset being nontraded. After all, traded assets are also subject to

financial market frictions. Conflicts of interest, or what economists call agency problems, are another problem with human capital sales.

Agency and information problems. Jensen and Meckling (1976) wrote the seminal paper in this area, but the concept has been known since at least Adam Smith (1776). Smith notes that the directors of large companies, who manage large amounts of other people's money, cannot be expected to exercise the same vigilance that they would exercise for their own money. He adds that negligence and inappropriate expenditures result.

Smith's insight is consistent with the familiar adage, "If you want the job done right, then do it yourself." The problem is that for all but the smallest businesses, doing it yourself is simply impossible. With size comes the separation of ownership and control because so few individuals have the wealth to own an entire company, and no one can operate a firm of any size without hiring agents to assist him.

Why is the separation of ownership and control a financial market friction? The answer is that this separation can lead to incentive problems, and financial contracts cannot handle them at zero cost. Suppose that a blues musician wishes to sell shares on the income from his future performances. The chances are good that he will find few buyers, and those who are willing to buy are almost sure to demand a large discount from what the musician views as fair market value. The reasons include adverse selection and incentives to shirk. First, the musician knows more about his ability and willingness to work than buyers, but buyers know that he knows more. This is Akerlof's (1970) familiar "lemons problem." Second, like Smith's directors, the blues musician's ability and willingness to work can be affected by the asset sale itself. Having a large sum of money might prevent the blues singer from performing with the same amount of feeling as he did without the funds—he may no longer have the blues. It is hard to imagine a contract that could costlessly eliminate this problem. This difficulty can reduce or even eliminate trading assets based on human capital because no one will pay the fair value of the musician's income stream.

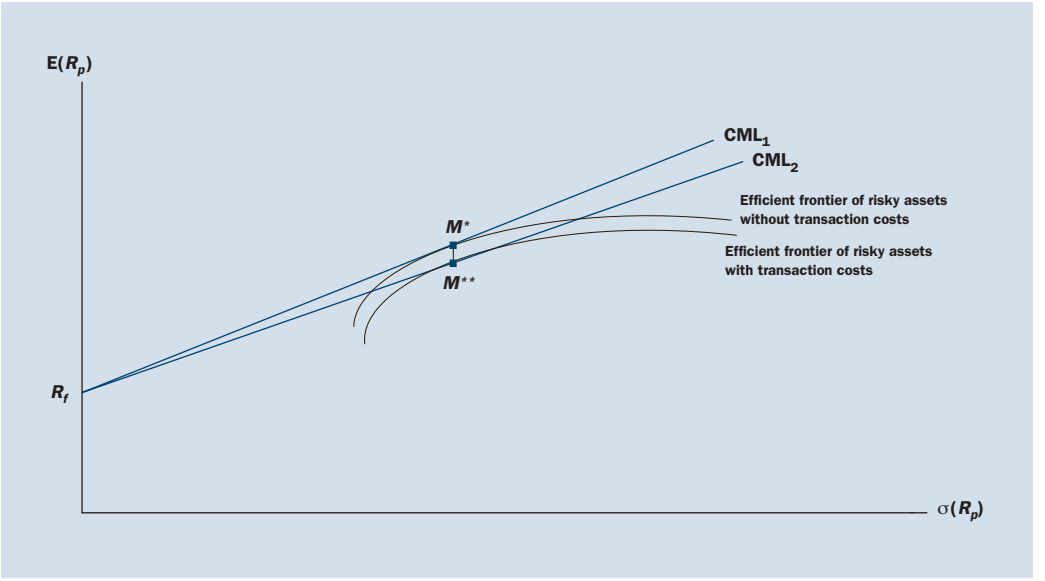
But if agency problems would hinder the musician's sales of claims against future earnings, then why were the sales of Bowie bonds successful? The answer is that the Bowie bonds were sales against future royalties from existing albums. Bowie has no ability to shirk or to reduce the quantity or quality of the albums already produced.

Other agency problems include perverse incentives to manage income. If the human capital contract is infinite or for a very long term, then sellers have an incentive to hide earnings or consume perquisites. If the contract is for a finite term, then sellers also have an incentive to delay earnings. This problem is familiar at the corporate level, where earnings management and fraud have led to the dismissal of corporate executives and even criminal charges. Bebchuk and Fried (2005) describe perhaps the most notorious example: The Federal National Mortgage Association rewarded executives for reporting high earnings but did not require them to reimburse shareholders when the earning were restated (downward) later.

Even abstracting from ownership and control, asymmetric information can also affect prices and prevent markets from clearing. The classic example is Akerlof (1970). Although he uses automobiles to illustrate his point, his insight is equally valid for financial assets. Consider initial public offerings (IPOs). Investors usually have great

5. See www.myrichuncle.com. Also see www.lumnifinance.com/, which offers human capital contracts in Chile, Columbia, and Peru, and www.career-concept.de/, which offers them in Germany.
6. The following discussion draws heavily from "Who's Who in Bowie Bonds" at www.ex.ac.uk/~RDavies/arian/bowiebonds.html.

Figure 2
The Capital Market Line in the Presence of Transaction Costs



difficulty valuing the new securities. Obviously, no recent market price is available, financial statements might be limited, and analysts rarely provide much coverage. In addition, the current owners know more about the company than potential buyers do. In the context of Akerlov's lemons problem, the owners know whether the company is a good company or a bad company. In addition, they have incentives to overstate the company's value. Investors are aware of the information problem, of course, so they assume that the company is bad and bid accordingly. In fact, unless the current owner of a good company is able to credibly certify that the company is good, he will not take a good company public. Without access to certification, the IPO market for good companies fails.

The lemons problem thus represents a profit opportunity for institutions that can evaluate IPOs and certify their value. In fact, investment banks serve that role. In addition to providing a distribution channel and advice, investment banks stake their reputations on the value of the IPO. This endorsement increases the likelihood that an IPO is good, and investors bid accordingly. The result is that good IPOs fetch higher prices than they would without certification, and trades are completed successfully.

Many researchers have applied Akerlof's insight to other markets. For example, Longhofer and Peters (2005) show that a lender's beliefs about the creditworthiness of a borrower's group (for example, his race, marital status, or educational attainment) can affect his assessment of the individual's creditworthiness. If the group's average creditworthiness exceeds the individual's, then the borrower benefits from group membership. But if the individual's creditworthiness exceeds the group's average, then the borrower suffers from group membership. The information asymmetry can work either in favor of or against various groups. Thus, imperfect information can lead to inaccurate credit decisions, in turn meaning that lenders miss some good loans and make some bad loans. The key point for our purpose is that collecting more information about individual lenders would solve this problem, but only at a cost, and

at some point the necessary information is simply not worth collecting. At least some part of the financial market friction remains.

Corporations are not immune to the lemons problem. A good example is the "pecking order" hypothesis of Myers and Majluf (1984). In that paper, management knows the correct value of the company, but investors do not. Investors know that management knows, and they know that management is issuing shares rather than borrowing or using cash to take projects. Myers and Majluf show how this information problem can cause firms to forgo profitable projects and to issue more debt and hold more cash. Because virtually any contract is subject to information asymmetry and agency problems, these financial market frictions touch virtually every area of financial economics. Jensen (1986), for example, has implications for dividend policy; with market frictions, dividends can create clienteles for high- and low-dividend stocks, depending on whether investors prefer current consumption or future consumption. In turn, these investors' preferences can drive a wedge between an investor's optimal holding and M^*. Tkac (2004) shows that investors and investment advisers have inherent conflicts of interest because they have different goals—investors want maximum returns with minimum risk, and advisers want maximum profits with minimum effort. It is difficult to imagine these types of conflicts vanishing.

The Economic Significance of Financial Market Frictions

Clearly, frictions abound in modern financial markets, but how influential are these frictions in changing the behavior of market participants? Can we see evidence of the effect of frictions on stock prices or the returns on investors' portfolios? One obvious place to turn for the answer to this question of the economic significance of market frictions is the academic research literature. Unfortunately, much of the empirical research on asset pricing is conducted within a framework of frictionless asset markets. For example, researchers often assume that agents can buy and sell securities at the same price, do not face transaction costs, and are not subject to short-selling constraints when they formulate models to explain asset returns, such as the CAPM. This section presents what we do know from current research and suggests some issues that should prompt further work in this area.

As the previous section detailed, we have plenty of informal observations that frictions affect financial decisions. Consider, for example, the gap between the interest rates at which consumers can borrow and lend. This gap is a major source of profits for financial institutions and exists because intermediation (the linking of borrowers to lenders) entails costs to overcome information asymmetries. More generally, the U.S. Census Bureau reports that as of 2003, almost 6.5 million people—5.7 percent of the workforce—worked in the finance and insurance sector in the United States. Many of these workers provide some costly intermediating service between buyers and sellers of assets.

In the case of borrowing and lending, market frictions lower the rate at which consumers can lend compared to a hypothetical world in which these frictions are not present; in that case, the rate gap would be eliminated, with lending rates likely somewhat higher and borrowing rates likely somewhat lower. Frictions likely have a similar effect on investors' optimal holdings of risky assets, such as stocks. For example, in the presence of transaction costs, the efficient portion of the mean-variance frontier may shift downward, as shown in Figure 2.

Figure 2 shows the efficient frontiers of risky assets with and without transaction costs when the risk-free rate for lending and borrowing is the same (that is, no

Table
Sharpe Ratios

	Individual stocks	Aggregate stocks	International stocks
Unconstrained	0.30	0.33	0.16
Sharpe ratios	(5.63)	(7.06)	(2.54)
Constrained	0.23	0.18	0.16
Sharpe ratios	(4.43)	(3.60)	(2.54)

Note: Figures in parentheses are *T*-statistics showing the statistical significance of the Sharpe ratio estimates.
Source: Authors' calculations

frictions exist due to asymmetric information).[7] For a given standard deviation of returns, the difference between the expected returns on M^* and M^{**} reflects the transaction costs. Transaction costs, even if small, make investment more costly. For example, if an individual buys $100 worth of stock through a broker, he must pay the broker a small fee, say $5. This makes his total cash outlay equal to $105. If the stock price increases by 10 percent the next day, the stock is worth $110, but he has made only $5 on his $105 investment, which is a return of only 4.76 percent. In the presence of transactions costs, economic agents must either give up part of the expected return to maintain the same level of risk or accept higher risk in their portfolio to obtain the same expected return.

To illustrate and quantify the potential economic impact of other market frictions on the risk-return trade-off of a mean-variance investor, we present a simple empirical example in which an investor faces short-sale constraints and must hold a positive amount of each asset (short selling is the equivalent of holding a negative position). We can then compute the efficient frontiers both in the case where short-sales are allowed (no frictions) and where they are forbidden. Again using a common risk-free rate for borrowing and lending, we can compare the M and M^* portfolios available to an investor in each of these cases. Because investors prefer higher returns and, all else being equal, lower risk, we can compare these portfolios based on the level of return per unit of risk. This quantity is known as the Sharpe ratio; the higher the Sharpe ratio, the higher the return for a given level of risk. Higher Sharpe ratios mean more favorable risk-return trade-offs.

Comparing the Sharpe ratios for the unconstrained portfolio M with that of the portfolio achievable with no short selling allows us to estimate the utility cost of this market friction to investors. The lower the Sharpe ratio of M^* relative to M, the more valuable are the opportunities that are unavailable to the investor. Our empirical analysis uses three different data sets corresponding to varying degrees of aggregation of stock and bond returns: individual stocks, aggregate stocks (decile portfolios based on market value of equity), and international stocks.[8] We use three different sets of assets because we want to show that market frictions sometimes, but not always, influence the Sharpe ratio of a given portfolio.

The results of the empirical exercise are summarized in the table above, which shows unconstrained and constrained Sharpe ratios for individual stocks, aggregate stocks, and international stocks. *T*-statistics to assess the statistical significance of the difference in Sharpe ratios are in parentheses. The general picture is that the Sharpe

ratio that embeds the no-short-sale constraint is lower than the unconstrained Sharpe ratio, meaning that market frictions do indeed impose utility costs on investors by making preferable investment portfolios unattainable. These results quantify the inward shift of the efficient portion of the mean-variance frontier illustrated in Figure 2. Specifically, the deterioration in the risk-return trade-off is 30 percent for individual stock returns and 83 percent for aggregate stock returns. Notice that for international stock returns there is no deterioration in the Sharpe ratio of the tangency portfolio. The reason is that during the sample period the short-sale constraint is not binding in this latter case. In other words, the *M* portfolio that can be formed from international stocks does not include any short positions. More generally, the impact of this constraint on attainable Sharpe ratios is likely to be smaller if only a subset of assets cannot be sold short.

Market frictions may have a negligible impact on the risk-return trade-off of a given portfolio and, at the same time, a substantial effect on portfolio rebalancing frequency. In other words, transactions costs could leave the frictionless risk-return trade-off, represented by the "without transactions costs" line in Figure 2, practically unchanged. However, the constrained (with frictions) portfolio weights that guarantee the same risk-return trade-off might be very different from the unconstrained (frictionless) ones. In turn, very different portfolio weights could depend on the fact that, in the presence of transaction costs, portfolio rebalancing becomes more costly.

Consistent with this empirical example, academic studies that incorporate market frictions seem to show that investors who ignore market frictions compound the harm done by the frictions themselves. Since frictions affect the investment opportunity set of investors, investors who do not take frictions into account in their decisions can do even worse. In particular, Balduzzi and Lynch (1999) find that realistically small transactions costs tend to prompt much less rebalancing on the part of investors. They estimate that ignoring these transaction costs and rebalancing more frequently can cost investors from 0.8 percent up to 16.9 percent of wealth.[9]

So, do investors respond optimally to the existence of market frictions and, in the case of transactions costs, trade less? The answer is yes. Lo, Mamaysky, and Wang (2004) show that even small transaction costs can have a substantial effect, causing investors to refrain from trading. From an aggregate perspective, Amihud and Mendelson (1986) present some evidence that stock returns reflect the effects of market frictions. Their empirical analysis shows that the bid-ask spread affects stock returns. In particular, they find that the average returns on stocks with larger bid-ask spreads tend to be higher. This result may stem from investors' lower demand for high-transaction-cost stocks. This lower demand reduces the prices of these stocks and boosts their average return to the point where investors are willing to hold them. Investors seem to pay a price premium for the liquidity of stocks with low bid-ask spreads.

---

7. For the case of different borrowing and lending costs, see DeGennaro and Kim (1986).
8. See the appendix for a description of the data used in the empirical example.
9. Vayanos (1998) builds a general equilibrium asset pricing model with transaction costs. He shows that a stock's price may increase as transaction costs rise because an increase in transaction costs has two opposing effects on the stock's demand. While investors buy fewer shares, they hold shares for longer periods, and either effect can dominate. Vayanos finds that realistic levels of transaction costs have very small effects on asset returns but large effects on investors' trading strategies and turnover. Constantinides (1986) argues that transaction costs have only a second-order effect on equilibrium asset returns: Investors accommodate large transaction costs by drastically reducing the frequency and volume of trade.

As discussed earlier, in principle, market frictions should be explicitly considered in the context of asset pricing theories, such as the CAPM described in the first section of the paper, and in the context of optimal portfolio formation. To date, though, little empirical research regarding the impact of frictions has been done. A few papers have demonstrated that including frictions may be a fruitful avenue for further academic research. Currently, most theoretical models are rejected by the empirical data from the capital markets. He and Modest (1995) and Luttmer (1996) argue that transaction costs, short-selling constraints, and margin requirements can together reconcile popular asset pricing models such as the CAPM with the observed asset return data. Estimates presented in these papers show that the inclusion of multiple market frictions makes popular asset pricing theories such as the CAPM, the C-CAPM (consumption CAPM) and the I-CAPM (intertemporal CAPM) more consistent with the data.

To summarize, market frictions can affect the investment opportunity set available to investors, reduce investors' utility, and prompt investors to change their behavior (that is, trade less).

Summary and Conclusions

It bears repeating that because the underlying business problems remain, financial market frictions never collectively go to zero. The conflicts of interest discussed in Tkac (2004) are one example. Another example is the age-old, ongoing problem of conducting business over long distances with unknown counterparties. In the nineteenth century, negotiable banknotes were a workable solution. But negotiable banknotes are unworkable for the online payments of the twenty-first century. Yet Quinn and Roberds (2003) show that today's online payments have evolved into a form very similar to negotiable banknotes. Both provide payment finality, thus mitigating a key problem for faceless, unknown counterparties conducting business across long distances. The fundamental business problem did not change, but the specific form of the problem did. We should not be surprised that the solution did too.

We have only begun to describe the incredibly broad array of financial market frictions, leaving much ground for others to cover. As mentioned above, for example, Figure 1 assumes that investors can borrow freely at the riskless rate. In fact, though, borrowing restrictions limit the amount of leverage that an investor can take. These restrictions, of course, are market frictions. Should they be classified as a regulatory matter, tracing to limited liability? Or should they be classified as an agency or information problem? The list of financial market frictions we have ignored is of necessity very long.

This article also focuses on financial markets within the United States, leaving room for theoretical and empirical research on product markets and international trade. Nor have we addressed tariffs, for example, which are huge impediments to trade. Nor have we addressed the political arena, in which some participants attempt to circumvent certain financial market frictions while others try to maintain them. Future research could also estimate the liquidity premium due to market frictions and the composition of the optimal portfolio in the presence of a variety of trading frictions.

Finally, the success of online payment providers reminds us that financial market frictions are more than simply impediments to trade. They also represent profit opportunities. Identifying and solving these business problems remains an ongoing challenge.

Appendix
Data Description

Data used in the empirical example are monthly and are expressed in percentage per month. The one-month Treasury bill (TB) rate pertains to the shortest bill with at least one month to maturity and serves as the risk-less asset in the analysis (Ibbotson Associates, SBBI module). All rates of return are nominal.

Individual Stocks

The period considered is February 1962–October 1998. We use holding period stock returns (including dividends) of firms listed in the Dow Jones Industrial Average (DJIA). Specifically, this data set includes all of the stocks in the DJIA that have monthly return data since April 1961 (twenty-two stocks) plus eight other blue-chip stocks.[1] This set of stocks is chosen to mimic a portfolio manager's variance minimization (or tracking error minimization) problem because portfolio managers tend to trade blue-chip stocks because of their higher liquidity. All stock returns are from the Center for Research in Security Prices (CRSP), and most of them are traded on the New York Stock Exchange.

Aggregate Stocks

Data are monthly and are expressed in percentage per month. The period considered is March 1959–December 1996. We use decile portfolio returns on NYSE-, AMEX-, and Nasdaq-listed stocks. Ten stock portfolios are formed according to size deciles on the basis of the market value of equity outstanding at the end of the previous year. If a capitalization was not available for the previous year, the firm was ranked based on the capitalization on the date with the earliest available price in the current year. The returns are value-weighted averages of the firms' returns, adjusted for dividends.

The securities with the smallest capitalizations are placed in portfolio one. The portfolios on the CRSP file include all securities, excluding ADRs (American depositary receipts), that were active on NYSE-AMEX-Nasdaq for that year.

International Stocks

The period considered is April 1970–October 1998. The universe of equities includes the Morgan Stanley Capital International (MSCI) national equity indexes. The nominal returns are denominated in U.S. dollars and are calculated with dividends. All indexes have a common basis of 100 in December 1969 and are constructed using the Laspeyres method, which approximates value weighting.[2] U.S. dollar returns are calculated by using the closing European interbank currency rates from MSCI. The focus of the empirical analysis is on the four countries with the largest market capitalization: the United States, the United Kingdom, Japan, and Germany.[3]

1. The tickers of the 22 stocks that are currently in the DJIA are: T, ALD, AA, BA, CAT, C, KO, DIS, DD, EK, XON, GE, GM, HWP, IBM, IP, JNJ, MRK, MMM, MO, PG, and UTX. The other eight blue-chip stocks are: BS (Bethlehem Steel), CHV (Chevron), CL (Colgate Palmolive), F (Ford), GT (Goodyear Tire and Rubber), S (Sears, Roebuck & Co.), TX (Texaco), and UK (Union Carbide).
2. See MSCI Methodology and Index Policy for a detailed description of MSCI's indexes and properties.
3. As of 1996, the market capitalization weight for these countries is 76.2 percent of the market capitalization worldwide.

REFERENCES

Akerlof, George A. 1970. Markets for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* 84, no. 3:488–500.

Amihud, Yakov, and Haim Mendelson. 1986. Asset pricing and bid-ask spread. *Journal of Financial Economics* 17, no. 2:223–49.

Balduzzi, Pierluigi, and Anthony W. Lynch. 1999. Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics* 52, no. 1:47–78.

Bebchuk, Lucian, and Jesse Fried. 2005. Executive compensation at Fannie Mae: A case study of perverse incentives, nonperformance pay, and camouflage. John M. Olin Center for Law, Economics, and Business Discussion Paper No. 505, March.

Becker, Gary. 2005. Should the estate tax go? The Becker-Posner Blog, www.becker-posner-blog.com, May 15 posting.

Buser, Stephen A., Andrew H. Chen, and Edward J. Kane. 1981. Federal deposit insurance, regulatory policy, and optimal bank capital. *Journal of Finance* 36, no. 1:51–60.

Constantinides, George M. 1984. Optimal stock trading with personal taxes: Implications for prices and the abnormal January returns. *Journal of Financial Economics* 13, no. 1:65–89.

———. 1986. Capital market equilibrium with transaction costs. *Journal of Political Economy* 94, no. 4:842–62.

Cox, Barbara, and William Koelzer. 2000. *Internet marketing in real estate*. Lebanon, Ind.: Prentice Hall.

DeGennaro, Ramon P. 2003. Direct investments: A primer. Federal Reserve Bank of Atlanta *Economic Review* 88, no. 1:1–14.

———. 2005. Market imperfections. *Journal of Financial Transformation* 14 (August): 107–17.

DeGennaro, Ramon P., and Sangphill Kim. 1986. The CAPM and beta in an imperfect market: Comment. *Journal of Portfolio Management* 12 (Summer): 78–79.

Financial Executives International. 2005. Sarbanes-Oxley Compliance Costs Exceed Benefits. Press release, March 21.

He, Hua, and David M. Modest. 1995. Market frictions and consumption-based asset pricing. *Journal of Political Economy* 103, no. 1:94–117.

Jensen, Michael. 1986. Agency costs of free cash flow, corporate finance, and takeovers, *American Economic Review* 76, no. 2:323–29.

Jensen, Michael, and William H. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics* 3, no. 4:305–60.

Kane, Edward J. 2000. Incentives for banking megamergers: What motives might regulators infer from event-study evidence? *Journal of Money, Credit, and Banking* 32, no. 3, pt. 2:671–701.

Lo, Andrew W., Harry Mamaysky, and Jiang Wang. 2004. Asset prices and trading volume under fixed transactions costs. *Journal of Political Economy* 112, no. 5:1054–90.

Longhofer, Stanley D., and Stephen R. Peters. 2005. Self-selection and discrimination in credit markets. *Real Estate Economics* 33, no. 2:237–68.

Luttmer, Erzo G.J. 1996. Asset pricing in economies with frictions. *Econometrica* 64, no. 6:1439–67.

Miller, Merton H., and Myron S. Scholes. 1978. Dividends and taxes. *Journal of Financial Economics* 6, no. 4:333–64.

Myers, Stewart C., and Nicholas S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13, no. 2:187–221.

Palacios, Miguel. 2002. Human capital contracts: "Equity-like" instruments for financing higher education. Cato Institute Policy Analysis Paper No. 462, December 16.

Quinn, Stephen F., and William Roberds. 2003. Are on-line currencies virtual banknotes? Federal Reserve Bank of Atlanta *Economic Review* 88, no. 2:1–15.

Smith, Adam. 1776. *The wealth of nations.* Edited by Edwin Cannan. New York: Modern Library, 1937.

Sodano, Salvatore F. 2004. Letter to Jonathan G. Katz, secretary, Securities and Exchange Commission, June 30.

Tkac, Paula A. 2004. Mutual funds: Temporary problem or permanent morass? Federal Reserve Bank of Atlanta *Economic Review* 89, no. 4:1–21.

Vayanos, Dimitri. 1998. Transaction costs and asset prices: A dynamic equilibrium model. *Review of Financial Studies* 11, no. 1:1–58.

When More Is Better: Assessing the Southeastern Economy with Lots of Data

PEDRO SILOS AND DIEGO VILÁN

Silos is a research economist and assistant policy adviser in the macropolicy section, and Vilán is an economist in the regional section, both in the Atlanta Fed's research department. *The authors thank Marco del Negro, John Robertson, and Ellis Tallman for very useful comments.*

Although the economy of each of the six southeastern U.S. states has unique and defining characteristics, these states' business cycles also tend to move together as if responding to some common, underlying factor. Currently, no single economic indicator exists for the economy of the Sixth Federal Reserve District as a whole, which encompasses the entire states of Georgia, Florida, and Alabama and parts of Louisiana, Mississippi, and Tennessee.[1] Rather, analyses of each southeastern state's economy are typically performed individually and then aggregated into a weighted average index. An indicator that captured the overall trend of the region's economy using information from all six states would aid in understanding the unique features of the region's business cycle. These features, when compared to those of the nation or other Federal Reserve districts, could assist in identifying crucial differences and similarities used to develop more accurate forecasts and in turn support monetary policy formulation.

This article outlines and estimates a model that provides such an indicator. We model economic activity in the Sixth District as being driven by an unobserved common factor. Economic activity is measured by a large set of time series of employment, construction, earnings, and sales tax revenues. Disaggregated information for each state is incorporated in a large model from which the common component is derived.

A thorough understanding of the dynamics behind this common factor will enable academics, policymakers, and businesspeople to make a better diagnosis of the condition of the region's economy. In addition, having one comprehensive measure of economic activity for the Sixth District will not only allow for a simplified and faster interpretation of several (sometimes contradicting) economic signals but will also make comparisons with other Federal Reserve district economies easier.

The study also seeks to compare its latent common factor model with the current practice of averaging individual states' coincident indicators. Overall we find that our indicator provides a more reasonable assessment of large idiosyncratic

shocks, such as Hurricane Katrina, than the weighted-average estimates. In other words, our model's results provide a better fit to what may be observed a priori in the data, as measured in the aggregate national income and product accounts (NIPA) from the U.S. Bureau of Economic Analysis. Moreover, the indicator provides insights about the different trajectory of the southeastern economy compared to the U.S. economy as a whole.

The Methodology

The model presented in this article is primarily based on the coincident indicator approach pioneered by Stock and Watson (1989) and is closely related to Otrok and Whiteman (1998). In the latter study, the authors develop an indicator that has since been used at the Institute for Economic Research at the University of Iowa to evaluate conditions in the Iowa economy. Because our time series are so lengthy, it is not efficient to apply that study's methodology here. Instead, we follow closely the approach used by Otrok, Silos, and Whiteman (2003), in which the estimation of the unobserved factor is done sequentially rather than in one "block" as in Otrok and Whiteman (1998). This way of sampling avoids some technical difficulties that are related to the sample size. Although our basic setup and idea are the same as Stock and Watson's, our choice of powerful simulation tools allows us to use a large cross-section of series (literally dozens) in contrast to the four or five that Stock and Watson use in their coincident index.[2]

The Setup

With standard assumptions on distributions and functional forms, we construct artificial observations of the common component using a powerful tool called Gibbs sampling (described in more detail in a later section and in the appendix).[3]

We observe n variables, denoted y_{it}, $i = 1,\ldots,n$, that reflect economic activity (employment, income, tax revenues, etc.) during period $t = 1,\ldots,T$. Each i refers to a specific data series; for example, $i = 1$ could be employment in Georgia, while $i = 2$ could be employment in Florida. There is a single common factor, F_t, that accounts for all comovement among the n variables. We assume that this factor is latent (that is, unobserved) and that it can be interpreted as an indicator of the stage of economic conditions or the business cycle in the economy being considered. Clearly, various factors could be affecting the comovement of two or more series, but in this study we are mainly interested in the common factor driving the comovement among all the variables in our data set. We assume that the relationship between any of the series and the common factor is linear:

$$(1)\quad y_{it} = \gamma_i F_t + \varepsilon_{it}.$$

The idiosyncratic dynamics (the dynamics in the individual series that are caused by something other than the common factor) are given by the errors ε_{it}. These error terms follow an autoregressive process:

$$(2)\quad \varepsilon_{it} = \varphi_{i,1}\varepsilon_{i,t-1} + \varphi_{i,t-2} + \nu_{it};\ \nu_{it} \sim N(0,\sigma_i^2).$$

Finally, the equation that governs the dynamics of the common factor has an autoregressive structure as well:

$$(3)\quad F_t = \rho_1 F_{t-1} + \rho_2 F_{t-2} + \omega_t;\ \omega_t \sim N(0,1).$$

Two identification problems arise for the model above. First, the sign for the dynamic factor and the sign of the γ_i are not independently identified. We solve this problem using two normalizations (see the appendix for details).

It should be clear from the above equations that if the common factor F_t were observed, the analysis of this system would be straightforward. In such a textbook case, equations (1) and (2) would form a series of n independent regressions in which errors have an autoregressive structure. The latent factor, however, poses some estimation difficulties. Fortunately, for such difficulties sampling methods developed in the Bayesian statistics literature can be helpful.

The final goal of the estimation is to obtain moments of interest (means, medians, standard deviations, etc.) from a density function (distribution) of the parameters and the unobserved factor given the observed time series data. Bayesian statisticians call this distribution the posterior distribution.[4] Denoting the vector of parameters by θ, the time series by Y, and the unobserved factor by F, let us write this distribution as $p(\theta,F|Y)$.

Sampling from the posterior distribution directly is generally difficult for a large number of time series, each of which is associated with several parameters, while at the same time keeping track of the unobserved factor. Fortunately, Gibbs sampling makes it possible to split this unmanageable distribution into several "sampling blocks." These sampling blocks are themselves density functions but of smaller dimension. The smaller dimension is the result of conditioning on values of parameters that belong to other blocks. For instance, in a very simple setup in which there are no unobserved factors and only two parameters, κ_1 and κ_2, our goal would be to sample from the joint posterior distribution of κ_1 and κ_2 given some data Y, $p(\kappa_1, \kappa_2|Y)$. The Gibbs sampling would allow us to sample sequentially from two conditional posteriors, $p(\kappa_1|Y,\kappa_2)$ and $p(\kappa_2|Y,\kappa_1)$. Of course, in this simple example there seems to be little computational gain from splitting the distribution. However, in problems of large dimension, Gibbs sampling could be the only feasible way of attacking a problem.

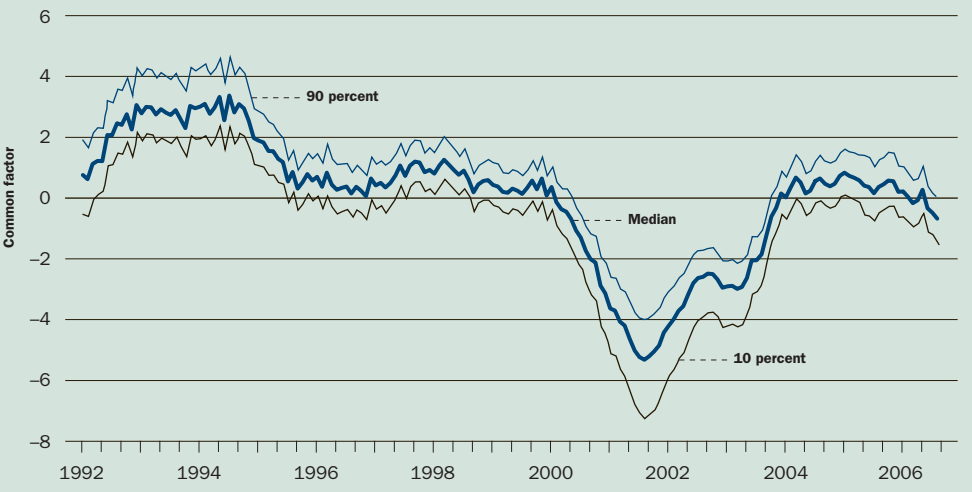
In our application the sampling blocks are as follows: The first is the distribution of the unobserved factor given θ; we have one such distribution for each point in time. The second block is the distribution of ρ_1 and ρ_2 given the unobserved factor. Finally, for each of the $i = 1,\ldots,n$ time series, we would sample γ_i, $\varphi_{i,1}$, $\varphi_{i,2}$, and σ_i^2 given the common factor. In this step we are treating the factor as observed data and therefore dealing with the simple task of obtaining n independent regressions for the "true" observed data.[5] By repeating these steps many times, starting with a guess for the parameter vector θ, the procedure generates a sample for the entire posterior distribution.

---

1. Throughout the article, the terms "Sixth Federal Reserve District" (and shortened forms), "Southeast," and "region" will be used interchangeably.
2. Crone and Clayton-Matthews (2005), using the Stock and Watson methodology, construct individual indicators for each of the fifty states using an approach similar to ours. Here, we estimate the common component jointly for all the states in the Sixth District using a larger data set (for example, we introduce sales tax data).
3. For a clear introduction to Gibbs sampling, see Casella and George (1992).
4. In general, before starting the analysis, the econometrician will combine prior information about the distribution of the unknown parameters, called the prior distribution and denoted by $p(\theta)$, with the "likelihood" of observing the data, given values for the parameters and the unobserved factor. This combination yields the posterior distribution.

   A detailed analysis of how to draw inferences from the posterior is beyond the scope of this article. See the appendix for a more technical overview than the one provided in the text.
5. This mechanism of generating data is, in essence, "data augmentation" (see Tanner and Wong 1987).

Figure 1
Distribution of the Model's Common Factor



Note: Data are from May 1992 through November 2006.

Data Description

Most of the series in this application are not seasonally adjusted. To avoid problems associated with seasonality we run the model in year-over-year growth rates. The only exception is the Georgia Purchasing Managers Index (PMI), which displays no trend, and therefore we estimate it in levels. Moreover, the data are standardized to avoid having to estimate an intercept and because we are mainly interested in comovement among variables.[6] In addition, having all series in a similar scale facilitates the estimation.

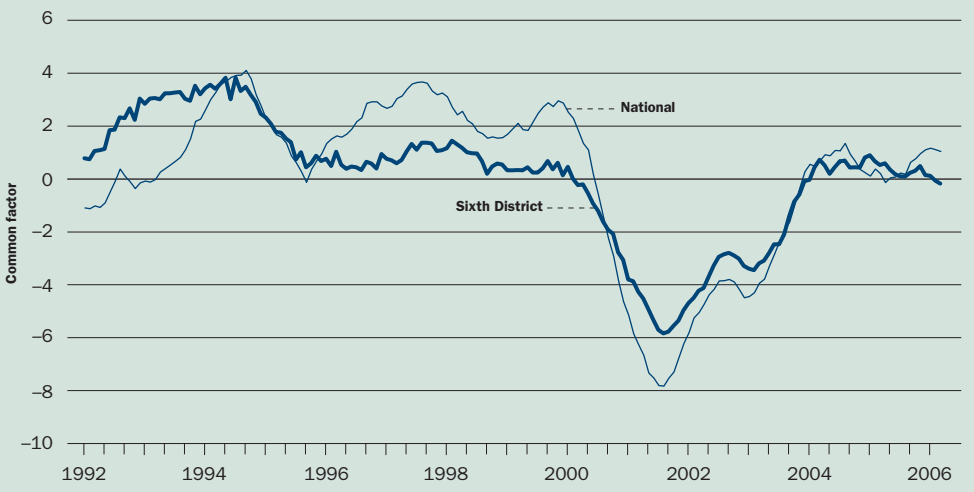
In total, we use twenty-four data series that fall into five groups—nonfarm employment, housing starts, sales tax revenues, average hourly earnings, and Georgia's PMI—described below. Data are monthly, starting in January 1991 to December 2006 (except for hourly earnings, which start in January 2001).

Employment. The employment series includes total nonfarm payroll employment for all six southeastern states. The nonfarm series include payroll data from construction, trade, transportation and utilities, information, financial activities, professional and business services, education and health services, leisure and hospitality, and government sectors. Employment figures are from the U.S. Bureau of Labor Statistics (BLS) and are monthly and seasonally adjusted.

Housing starts. Given that the housing industry accounts for about a quarter of all investment spending and around 5 percent of the overall economy, the housing starts series is considered a leading indicator. The series includes all new privately owned housing units started in each of the six states in the district. Series are seasonally adjusted annual rates from the Bank of Tokyo–Mitsubishi UFJ.

Real hourly earnings. To transform average hourly earnings (AHE) into real terms, we deflate them using the U.S. urban consumer price index (CPI).[7] In this article, earnings are for the manufacturing sector in each state. The BLS data begin in 2001 and are monthly and not seasonally adjusted.

Figure 2
The Model's Common Factor for the Sixth Federal Reserve District and National Economies



Note: Data are from May 1992 through July 2006.

Georgia PMI. The PMI report is a composite index based on five major indicators: new orders, inventory levels, production, supplier deliveries, and employment environment. The Association of Purchasing Managers surveys over 300 purchasing managers nationwide that represent twenty different industries. Georgia's PMI data, obtained from Haver Analytics, are monthly and not seasonally adjusted.

Sales tax revenues. Sales tax revenues are an important indicator of each state's fiscal strength and, indirectly, of the current regional business cycle conditions. Series data, from Haver Analytics, are monthly and not seasonally adjusted.

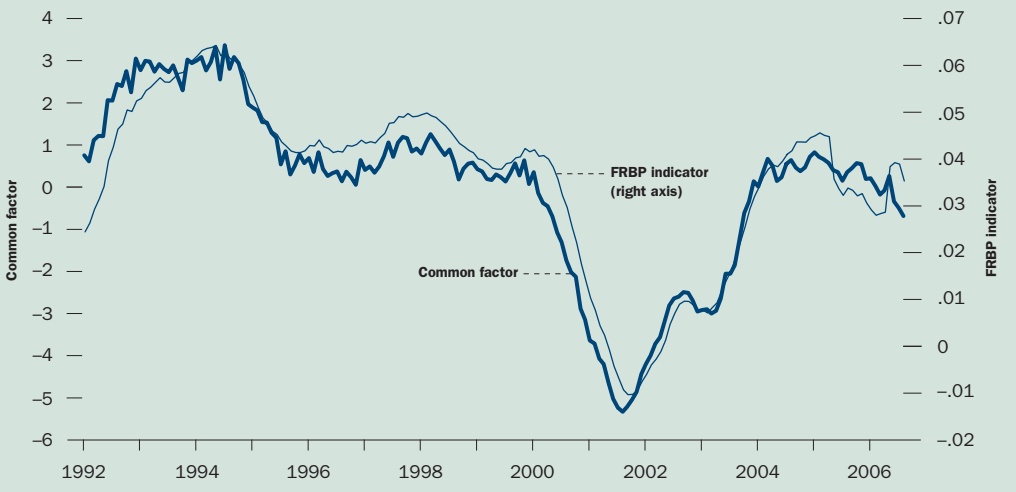
Estimation Results

To summarize the results of our estimation, we first describe the evolution of the unobserved component for the Sixth District and then compare that indicator to an analogous indicator for the U.S. economy. Finally, we compare the estimate obtained here with an indicator of economic activity constructed from series provided by the Federal Reserve Bank of Philadelphia.

Figure 1 shows the median of the common component along with the 10th and 90th percentiles. It is important to realize that the common component is a random variable, and as such it has a distribution at each point in time. The percentiles are plotted along the median to give an idea of the uncertainty of that distribution. At first glance, it is easy to distinguish the recovery from the 1991 recession during the first half of the nineties, the 1994 soft landing caused by the contraction in residential

6. "Standardization" implies that from each series we subtract its mean and then divide by the standard deviation. As a result, all series prior to estimation have a sample mean of zero and a sample standard deviation of one.
7. One can deflate earnings by the Southeast CPI provided by the BLS, but the difference in results is quantitatively insignificant.

Figure 3
The Model's Common Factor and the Federal Reserve Bank of Philadelphia Averaging Indicator for the Six Southeastern States



Note: Data are from May 1992 through December 2006.

investment, and the subsequent slowdown of economic expansion during the second half of the decade. The plunge of economic activity coincides with the recession of 2001, followed by a recovery during the 2003–05 period. One can conclude that the underlying (median) factor reflects the prior notions about the evolution of the region's economy during the past fourteen years.

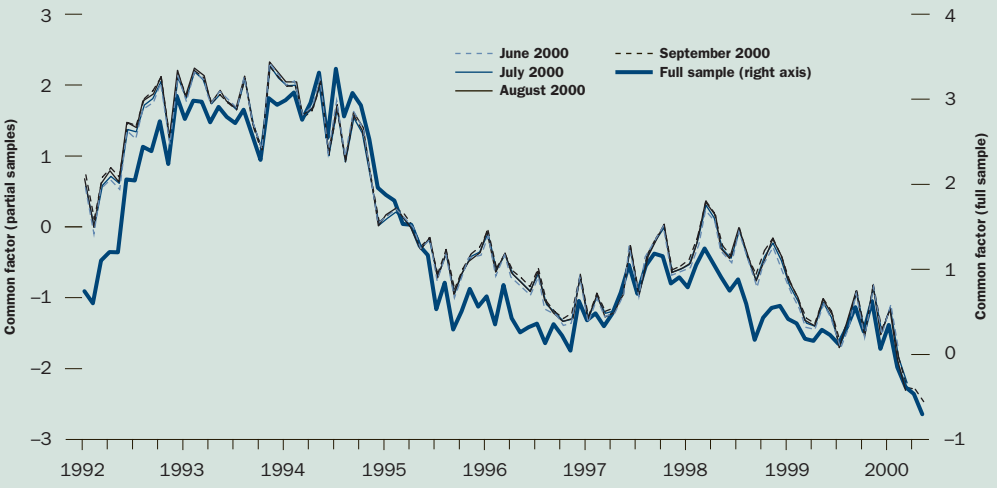
Figure 2 compares the Sixth District economy with the U.S. economy. The graph plots the common factor computed as described above along with the common factor for the national data. The national factor is computed using the same methodology and the series used by the National Bureau of Economic Research (NBER) to date the stages of the business cycle.[8] The figure clearly shows that the two series are highly correlated (the correlation is 0.86). There are, however, a few differences. First, the Sixth District seems to have benefited more than the U.S. economy during the initial recovery after the 1991 recession as illustrated by the concave (downward) shape of the Sixth District common factor compared to the concave (upward) shape of the U.S. factor.

During the dip in economic activity during 1994 and 1995, both series' declines were similar in magnitude. However, although the district economy was able to outgrow the national economy during the recovery in the years after the 1990–91 recession, apparently it did not benefit as much from the boom in the second half of the nineties. Our index reports a much stronger expansion at the national than at the district level.

Additionally, both series fell starting in 2001, although the Sixth District seems to have endured a milder slowdown in economic activity than the overall U.S. economy. Not surprisingly, the recovery following the slowdown was also less pronounced in the district's states than in the rest of the nation. In recent years both series seem to be trailing quite closely.

Figure 3 compares two indicators for the Sixth District: our dynamic factor and the averaging indicators for the six individual southeastern states obtained from the Federal Reserve Bank of Philadelphia (FRBP). The FRBP indicators are constructed with a

Figure 4
Comparison of the Model's Common Factor Estimated Using Different Samples



Note: The full sample uses data from May 1992 through December 2006, but for comparison to the partial samples the graph shows data only through September 2000.

dynamic factor model with four series for each state: employment, hours, the unemployment rate, and real wages. The weights for averaging the six states are given by the states' gross state products (GSPs), and the weighted average is shown in Figure 3.

Comparing our common component and the FRBP indicator, one can see that both series move closely together. There are, nonetheless, two differences between the estimates. The first is related to the impact of Hurricane Katrina on the district's economic activity. While the FRBP index depicts a strong drop in the aftermath of Katrina, our index describes a much smoother trend. We believe, given the relative weight that the state of Louisiana has in the district, that ours is a more accurate estimate.

The second difference in the two indicators appears at the beginning of the plunge in 2001. Our indicator starts falling a few months before the FRBP indicator. This pattern could stem from our use of housing starts data, which are traditionally a leading indicator of the business cycle. Is the early drop an artifact of estimating the model for the entire sample, or would the fall be signalled as well if the data ran only until 2000?[9] To determine whether this leading property of our indicator is a "real time" property, we estimate four additional indicators in which we vary the sample size. The first indicator ends in June 2000, the second in July 2000, the third in August 2000, and the fourth in September 2000. The results of this estimation are plotted in Figure 4, which clearly shows that the magnitude and timing of the fall for a given data point do not depend on the sample size. All four lines lie on top of each other, and the shape of the indicators is very similar to the one obtained when using the full sample.

8. These series are industrial production, real income less transfer payments, employment, and retail sales and inventories.
9. Note that our definition of real time does not take into account data revisions because we assume that the econometrician has the already revised data at the end of her sample.

Conclusion

Evaluating economic conditions generally involves keeping track of literally dozens of time series describing different aspects of an economy. Central bankers, financial institutions, and many corporations and individuals comb through data on labor, products, and factors markets to assess the current state of an economy and make judgments about its future state.

This article applies a methodology to extract a "signal" from a large array of time series representing economic activity and uses that methodology to construct an economic indicator for the Sixth Federal Reserve District. The article outlines the idiosyncrasies of the southeastern economy relative to the U.S. economy and compares the new indicator with a weighted average of indicators for individual states constructed using a similar methodology. The indicator demonstrated here should be of interest to anyone analyzing the condition of the southeastern economy.

Appendix
Constructing the Gibbs Sampler

More detailed expositions of constructing the Gibbs sampler are available in Kim and Nelson (1999) and Otrok, Silos, and Whiteman (2003), but this appendix provides a broad idea of how the algorithm is structured.

Recall that the original model was:

$$y_{it} = \gamma_i F_t + \varepsilon_{it};$$
$$\varepsilon_{it} = \varphi_{i,1}\varepsilon_{i,t-1} + \varphi_{i,t-2} + \nu_{it};\ \nu_{it} \sim N(0,\sigma_i^2);\ \text{and}$$
$$F_t = \rho_1 F_{t-1} + \rho_2 F_{t-2} + \omega_t;\ \omega_t \sim N(0,1).$$

As mentioned in the text, the sign of the factor is not identified, as one can see by observing that $-\gamma_i * (-F_t) = \gamma_i * F_t$. We handle this ambiguity by fixing the factor's coefficient of any particular time series to be positive so that for any given time period t, we are fixing the sign of the factor as well. Also, the scales of the factor loadings (the γ_is) and of the factor itself are not separately identified, as can be seen by noting that $\gamma_i/\eta * F_t * \eta = \gamma_i * F_t$ for any η. This problem is solved by normalizing the standard deviation of the innovations in the factor equation (3) to 1.

After solving these two identification issues, the first step is to set prior distributions for the parameters:

$$\gamma_i \mapsto N(\underline{\gamma},\underline{\Sigma}_\gamma);$$
$$(\rho_1,\rho_2) \mapsto N(\underline{\rho},\underline{\Sigma}_\rho)I_\rho(S);$$
$$(\varphi_{i,1},\ \varphi_{i,2}) \mapsto N(\underline{\varphi},\underline{\Sigma}_\varphi)I_\varphi(S);\ \text{and}$$
$$\sigma_i^2 \mapsto IG(\underline{\alpha},\underline{\beta}).$$

In the previous expressions we use the symbol $\mapsto$ to denote "distributed as." A normal distribution for the intercepts and an inverse gamma (IG) for the variances are typical choices in Bayesian econometric models. Also note that for the φs and the ρs we impose a stationarity restriction, represented by an indicator variable that takes the value 1 if the parameter is inside the stationarity region S and 0 otherwise.

Starting with a guess for the parameter vector $\theta = \rho_1,\rho_2,(\gamma_1,\ldots,\gamma_n),(\varphi_{1,1},\ldots,\varphi_{1,n},\varphi_{2,1},\ldots,\varphi_{2,n}),(\sigma_1,\ldots,\sigma_n)$, from the following distributions, we sequentially

(a) sample the unobserved factors, $F_t \mid \{Y\}_t$, $\theta \mapsto N(F_t^*,P_t^*)$;[1]

(b) sample the ρs, $(\rho_1,\ \rho_2) \mid \{Y\}_t,\ \{F\}_t \mapsto N(\bar{\rho},\bar{\Sigma}_\rho)$; and

(c) for $i = 1,\ldots,n$, sample $\gamma_i \mid \{Y\}_t,\ \{F\}_t,\ \sigma_i,\ \varphi_{i,2} \mapsto N(\bar{\gamma}_i,\bar{\Sigma}_{\gamma,i})$, $\varphi_{i,1},\ \varphi_{i,2} \mid \{Y\}_t,\ \{F\}_t,\ \sigma_i^2$, $\gamma_i \mapsto N(\bar{\varphi}_i,\bar{\Sigma}_{\varphi,i})$ and $\sigma_i^2 \mid \{Y\}_t,\ \{F\}_t,\ \gamma_i,\ \varphi_{i,1}$, $\varphi_{i,2} \mapsto IG(\bar{\alpha}_i,\bar{\beta}_i)$.

This sequential sampling is repeated several thousand times. At each step we condition on the previously sampled values for the parameters and the unobserved factor. We eliminated the first 1,500 draws to avoid having an influence from the initial conditions.

1. This step is the most involved. One must first apply the Kalman filter to the system in order to compute the mean and the covariance matrix for the unobserved factor at each point in time (see Kim and Nelson 1999, chap. 8, or Carter and Kohn 1994).

REFERENCES

Carter, C.K., and R. Kohn. 1994. On Gibbs sampling for state space models. *Biometrika* 81, no. 3:541–53.

Casella, George, and Edward I. George. 1992. Explaining the Gibbs sampler. *American Statistician* 46, no. 3:167–74.

Crone, Theodore M., and Alan Clayton-Matthews. 2005. Consistent economic indexes for the 50 states. *Review of Economics and Statistics* 87, no. 4: 593–603.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-space models with regime switching: Classical and Gibbs-sampling approaches with applications.* Cambridge, Mass.: MIT Press.

Otrok, Christopher, Pedro Silos, and Charles H. Whiteman. 2003. Bayesian dynamic factor models for large data sets: Measuring and forecasting macroeconomic data. University of Iowa unpublished manuscript.

Otrok, Christopher, and Charles H. Whiteman. 1998. Bayesian leading indicators: Measuring and predicting economic conditions in Iowa. *International Economic Review* 39, no. 4:997–1014.

Stock, James, and Mark Watson., 1989. The revised NBER indexes of coincident and leading economic indicators. In *NBER macroeconomics annual 1989*, edited by Olivier J. Blanchard and Stanley Fischer. Cambridge, Mass.: MIT Press.

Tanner, Martin A., and Wing Hung Wong. 1987. The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association* 82, no. 398:528–40.

Smoking: Taxing Health and Social Security

BRIAN S. ARMOUR AND M. MELINDA PITTS

Armour is a health scientist at the Centers for Disease Control and Prevention (CDC) in Atlanta. Pitts is a research economist and associate policy adviser in the regional group of the Atlanta Fed's research department. *The authors thank Ralph Caraballo, Scott Grosse, and Corinne Husten for helpful comments. This article reflects the authors' views and not those of the CDC.*

Cigarette smoking is the largest single health risk in the United States, accounting for approximately 440,000 deaths each year (U.S. Department of Health and Human Services [USDHHS] 2004b). The financial cost of smoking-attributable health care expenditures and lost productivity has been well documented (Centers for Disease Control and Prevention [CDC] 2003). In general, smokers have higher health care expenditures and more sick days than do nonsmokers (Max 2001). However, the effects of smoking-attributable mortality on income distributions are less well known.

Premature death attributable to smoking may redistribute Social Security income in unanticipated ways that affect behavior and reduce the economic well-being of smokers and their dependent spouses and children (Rice et al. 1986). Knowledge of how smoking redistributes both individual and household Social Security benefits and taxes is important not only from the perspectives of informing smoking cessation efforts (Rice et al. 1986) and evaluating proposals to improve family welfare through reductions in system inequities or promotion of social adequacy but also from the standpoint of managing the Social Security System's finances. Social Security is financed by a pay-as-you-go tax levied on earnings; thus, if the harmful health effects of smoking reduce individual or household hours of work, these effects have implications for the system's funding.

Economists employ the comprehensive marginal tax rate to assess the distortionary effect of taxation on labor supply and welfare (Armour and Pitts 2004). One important component of this comprehensive marginal tax rate in the United States is the Social Security payroll tax, which is assessed on individual earnings up to the annual taxable maximum. In 2002 approximately 94 percent of all U.S. workers earned less than the annual taxable maximum of $84,900, thus incurring an Old-Age and Survivors Insurance (OASI) Social Security payroll tax at the margin.[1] For these individuals, Social Security is a benefit tax for which an extra dollar of earnings may increase their future benefits at retirement. Therefore, the net marginal Social Security tax rate (NMSSTR)—defined

as the difference between the statutory payroll tax rate and the present value of the stream of future benefits to which an additional dollar of earnings entitles the covered worker—should be used in calculating the marginal tax rate for the purpose of assessing the effect of taxation on labor supply and welfare.[2]

Studies that have used the NMSSTR to examine the distributional effects of Social Security concluded that Social Security benefit and tax rules create NMSSTRs that treat workers differently depending on age, gender, race, dependency status, earnings, insurance status, and income-related life expectancy (for example, Aaron 1977; Browning 1985; Burkhauser and Turner 1985; Feldstein and Samwick 1992; Armour and Pitts 2004). To our knowledge, no study has looked at lifestyle and the harmful health effects of an addictive habit such as smoking on NMSSTR estimation. This study contributes to the literature by examining the distributional effects of smoking-attributable mortality on NMSSTR estimation.

Methods

Social Security benefit determination. The Social Security benefits to which a covered worker is entitled at retirement depend on lifetime earnings. Average indexed monthly earnings (AIME) is the measure of lifetime earnings on which benefits are based. Earnings are indexed by multiplying a worker's taxable earnings by an indexing factor for each year after 1950 through the indexing year. The indexing year is defined as the year a worker attains age sixty. The indexing factor for each year, t, is obtained by dividing average covered worker earnings in the indexing year, $\bar{E}_{60,t}$, by average covered worker earnings at each age, a, in each year, $\bar{E}_{a,t}$. The AIME for individuals retiring in year t is

$$(1)\quad AIME = \frac{1}{n}\frac{1}{12}\sum_{t\in A}\frac{\bar{E}_{60,t}}{\bar{E}_{a,t}}E_t + \sum_{t\in B}E_t.$$

For individuals attaining age sixty-two after 1991, the AIME is based on the highest thirty-five years of earnings. However, for each year a worker is born before 1929, the number of years, n, in the computation period is reduced by one. To convert the AIME from an annual to a monthly basis, it is divided by 12. E_t denotes worker earnings in year t. The set of all years through age sixty that will be counted among the highest thirty-five or n years of earnings is denoted by A. B denotes the set of years between age sixty and the year prior to retirement in which a year of unindexed earnings replaces a year of indexed earnings in the benefit formula.

Once the AIME is determined, the primary insurance amount (PIA)—the amount of monthly benefits payable at retirement—may be calculated.[3] The benefits formula for a covered worker attaining age sixty-two in 2002 is

$$(2)\quad PIA = [0.90 \times (AIME \le \$592)] + [0.32 \times (\$592 < AIME \le \$3{,}567)] + [0.15 \times (AIME > \$3{,}567)].$$

The PIA is composed of two parts: the bend points (the dollar amounts defining the AIME bracket in the benefit formula) and the marginal replacement rate (the applicable percentage used to determine the PIA).[4]

The benefit formula illustrates one fundamental feature of the system: the progressive structure of Social Security. Low-earning workers are afforded proportionately greater benefits with a marginal replacement rate of 90 percent when compared

with average-earning and high-earning workers, whose marginal replacement rates are 32 percent and 15 percent, respectively. Because the Social Security benefit formula classifies workers into one of three earnings groups, the NMSSTR by sex and age is calculated for a representative worker in each group.

Calculation of the NMSSTR. NMSSTRs by sex, age, and earnings classification are calculated under two alternative scenarios. The first scenario uses a common mortality assumption, and the second scenario accounts for smoking-attributable mortality in calculating the NMSSTR.

The NMSSTR is $\tilde{T} = T - B_{PV}$. T denotes the OASI statutory rate, which is defined as the combined employee-employer legislated rate. The combined employee-employer tax rate was 10.6 percent in 2002.[5] This analysis assumes that the employee pays the tax.[6]

Primary beneficiary (single). The present value of the change in anticipated future benefits resulting from a $1 change in earnings is

$$(3)\quad B_{PV} = \frac{1}{n}\frac{\partial PIA}{\partial AIME}(1+g)^{\max(60-a)}\, i_{s,f,\tilde{r},t} \sum_{j=f}^{N} P_{s,t}(j|a)(1+r)^{a-j}.$$

The future benefits that an additional dollar of earnings entitles an individual to at retirement depend on the marginal replacement rate, $(\partial PIA)/(\partial AIME)$, and the age, a, at which the individual plans to retire. Workers are assumed to retire at the full benefit retirement age, f.[7] The indexing factor at each age, $(1 + g)^{\max(60-a)}$, is estimated assuming that earnings grow at a real rate of 1.1 percent.[8] The probability that an individual

1. These figures are estimated from information in USDHHS (2004a, table 4.B4).
2. While many researchers recognize the link between the payroll tax levied on an additional dollar of earnings and anticipated future benefits, their analysis typically calculates the comprehensive marginal tax rate using the Social Security statutory rate; as a consequence, their results are overstated (Browning 1985; Burkhauser and Turner 1985).
3. The benefit amount that family members may receive each month is limited. The limit varies but generally equals about 150 to 180 percent of PIA. If the sum of the benefits payable to family members exceeds this limit, their benefits will be reduced. However, any benefits paid to a surviving divorced widow or widower do not count toward this maximum amount (see USDHHS 2004a).
4. The 1977 amendments to the Social Security Act indexed the benefit formula's bend points to the growth rate in average covered earnings. The marginal replacement rates were fixed at 90, 32, and 15 percent, respectively (see USDHHS 2004a).
5. The tax rate ignores the disability insurance (DI) and health insurance (HI) contribution rates. Including both rates increases the net marginal Social Security tax rate by the statutory amount. In 2002 the combined employee-employer DI and HI rates were 1.8 and 2.9 percent, respectively (see USDHHS 2004a).
6. Brittain (1972) found that the payroll tax reduced employee earnings by the full amount of the tax.
7. The formula in equation (3) estimates the actuarial present value of anticipated future benefits relative to some benchmark retirement age. The age chosen here, f, is defined as the full benefit retirement age, which corresponds to the age at which an individual is first eligible for retirement benefits without actuarial adjustment. Following legislation implemented in the 1983 amendments to the Social Security Act, the full benefit retirement age increased two months per year, from sixty-five to sixty-six, from 2000 to 2005. Between 2005 and 2016 the full benefit retirement age will remain at sixty-six. In 2017, the full benefit retirement age is scheduled to increase two months per year and will be fixed at age sixty-seven for those attaining age sixty-two after the year 2022. The retirement age for workers with a full benefit retirement age in terms of years and months is rounded to the next full year in all calculations.
8. The economic assumptions used in the calculations are based on the 2005 Social Security Board of Trustees' best-cost estimates (USDHHS 2005).

Table 1
Net Marginal Social Security Tax Rate Estimates for Single Beneficiaries and Primary Male Beneficiaries with a Dependent Spouse by Earnings Classification and Age in 2002

	Single female			Single male			Male beneficiary and dependent spouse			
Age in 2002	Low earning	Average earning	High earning	Low earning	Average earning	High earning	Low earning	Average earning	High earning	Uninsured female
35	–4.68	5.17	8.05	–2.11	6.08	8.48	–9.75	3.37	7.21	10.6
45	–8.24	3.90	7.46	–4.94	5.07	8.01	–13.54	2.02	6.58	10.6
55	–10.34	3.15	7.11	–9.22	3.55	7.30	–19.26	–0.02	5.62	10.6
65	–15.33	1.38	6.28	–17.86	0.48	5.86	–31.37	–4.32	3.60	10.6

Note: Workers are assumed to retire at the full benefit retirement age. Low-earning workers expect a marginal replacement rate of 0.9, and average- and high-earning workers expect rates of 0.32 and 0.15, respectively. A real discount rate of 3 percent is assumed. The growth rate in real earnings is set at 1.1 percent.

of sex s and age a in year t will be eligible for benefits at age f in year $\check{t}$ ($\check{t} = t + f - a$) is denoted by $i_{s,f,\check{t},t}$.[9] The probability of an individual of sex s surviving from age a to age j is denoted by $P_{s,t}(j|a)$. N is the age at which all persons are assumed to be dead and is set at 100 in all calculations. The rate at which a worker discounts future benefits, r, is set at 3 percent in all calculations.[10]

To illustrate, consider the case of a man who is fifty-five years old in 2002 and plans to retire at age sixty-six in 2011. Because he will attain age sixty-two after 1991, the AIME is based on the highest thirty-five years of earnings. Earnings through age sixty are indexed to the growth rate in average covered earnings. Assuming that real earnings grow at a rate of 1.1 percent annually, then $(1 + g)^{\max(60-55)} = 1.056$. An additional dollar of earnings at age fifty-five increases average indexed earnings by $\$(1/35)(1.056) \approx \$0.03$.

Assuming that the fifty-five-year-old man is a lifetime average wage earner, his marginal replacement rate is 0.32, and an extra dollar of earnings at age fifty-five would increase the PIA by $\$(0.03)(0.32) \approx \$0.0097$. The present value of the change in anticipated future benefits resulting from a $1 change in earnings is $0.0097 \sum_{j=66}^{N=100} P_{f,t}(j|55)(1 + r)^{55-j}$. The discounted sum of survival probabilities for a man aged fifty-five is 7.838. Multiplying 0.076 (0.0097×7.838) by the probability that a fifty-five-year-old man will be eligible for Social Security benefits at the full benefit retirement age, 0.931, yields an estimate of $B_{PV} \approx 0.0705$. Subtracting 0.0705 from the statutory rate yields 0.0355, or 3.55 percent.

NMSSTRs for representative low-, average-, and high-earning workers by sex and select ages in 2002 are shown in Table 1. The estimates reveal that men and women at each age face an NMSSTR that is less than the statutory rate and that the NMSSTR declines with age. The age differential is the result of higher conditional survival probabilities and the fact that older workers have a shorter period over which to discount future benefits. Also, low-earning workers incur the lowest NMSSTR, as expected given the progressive nature of the benefit formula.

Across earning classes, women at most ages incur a lower NMSSTR than do men. The estimated NMSSTR for a low-earning woman aged fifty-five is 1.12 percentage points lower than the rate faced by her male counterpart (–10.34 percent compared with –9.22 percent). Gender differences in the NMSSTR are approximately 0.4 percentage points for average-earning individuals and 0.2 percentage points for high-

earning individuals aged fifty-five; this differential is attributable to the longer life expectancy of females. The NMSSTR for a woman aged sixty-five with average lifetime earnings is 0.9 percentage points higher than the rate for her male counterpart. Older women incur a higher NMSSTR because they have less of an attachment to the labor force and thus have a lower probability of being fully insured for benefits.[11]

Primary beneficiary and dependent spouse. Women who are married and do not work outside the home or fail to qualify for benefits based on their own earnings histories may qualify for dependent spouse benefits. Thus the present value of anticipated future benefits also depends on whether a primary beneficiary claims benefits for a dependent spouse.[12] A dependent spouse is entitled to an additional 50 percent of the primary beneficiary's benefit amount at retirement. In addition, if the primary beneficiary dies, the widow is entitled to 100 percent of the primary beneficiary's benefit.[13] The formula (obtained from Feldstein and Samwick 1992) for calculating the present value of the change in anticipated future benefits resulting from a $1 change in earnings for a male worker age a with a dependent spouse is shown in equation (4);

$$
\begin{aligned}
(4)\quad B_{PV} = & \sum_{j=a}^{N} P_{1,t}(j|a) - P_{1,t}(j+1|a)\,PIA(j,E_t) \sum_{j=\max(a,60)}^{N} P_{2,t}(j|a)(1+r)^{a-j} \\
& + \sum_{j=f}^{N} P_{1,t}(j|a)\,PIA(f,w)(1+r)^{a-j} \\
& + \sum_{j=f}^{N} 0.5\,P_{1,t}(j|a)P_{2,t}(j|a)\,PIA(f,\boldsymbol{E}_t)(1+r)^{a-j},
\end{aligned}
$$

where 1 = male, 2 = female, and a dependent wife is assumed to be the same age as her husband. The definitions of the other characters are identical to those for a single primary beneficiary.

The first term of equation (4) denotes the expected value of the widow's benefits conditional on the worker dying at age a. The second term denotes the expected

9. To qualify for Social Security benefits, an individual must be fully insured. The measure used to determine whether a worker is eligible for retirement benefits is quarters of coverage. Under current legislation, a worker is fully insured if he obtains one quarter of coverage for each year after 1950 (or age twenty-one, if later) and before the year he dies, becomes disabled, or attains age sixty-two (USDHHS 2001). The minimum number of quarters required to be fully insured ranges from six to forty.

   Unpublished insurance rate estimates were provided by the Social Security Office of the Actuary. The data contained projections covering the period 2002 by sex and age for the number of fully insured workers as a percentage of the total population.

10. A rate of 3 percent was chosen to approximate an individual's rate of time preference. As before, this rate was chosen on the basis of recommendations contained in USDHHS (2005).

11. The probability that a man aged sixty-five was fully insured for benefits in the year 2002 was 0.929. In comparison, the probability that a sixty-five-year-old female was fully insured was 0.741. These unpublished estimates were provided by the Social Security Office of the Actuary.

12. The Social Security Administration estimates that, of the 21.4 million women aged sixty-two and older in 2000, 8.2 million were entitled to primary benefits only, 5.9 million were dually entitled, and 7.4 million were solely entitled to benefits as a dependent spouse and failed to qualify for benefits based on their own earnings history (USDHHS 2001).

13. Widows and widowers become eligible to receive survivor benefits at age sixty. However, children and disability may lower the age of eligibility. A detailed explanation of how these criteria may affect the age that survivors may be first eligible for benefits is contained in USDHHS (2001).

value of the primary beneficiary's retirement benefit conditional on attaining the full benefit retirement age, f. The third term denotes the expected value of the dependent spouse's benefit conditional on both parties reaching the full benefit retirement age.

Because beneficiaries with a dependent spouse do not pay any additional taxes for the additional benefit, they incur a lower NMSSTR than do singles. The NMSSTR for an average-earning man aged fifty-five with a dependent spouse, assuming a discount rate of 3 percent, is –0.02 percent (see Table 1). This negative tax rate is a net marginal subsidy and is lower than the rate incurred by female dependent spouses, whose NMSSTR equals the statutory rate of 10.6 percent.

Premature death attributable to smoking may redistribute Social Security income in unanticipated ways that affect behavior and reduce the economic well-being of smokers and their dependents.

Smoking-attributable mortality. The progressivity of the Social Security benefit formula is based on a common mortality assumption. However, the literature contains evidence that smoking reduces life expectancy (USDHHS 2004b). Life tables published by the National Center for Health Statistics are used to construct and account for differences in life expectancy among current and former smokers as well as people who have never smoked in determining NMSSTRs. The approach utilizes the mortality ratios of Thun et al. (1997) and current and former smoking prevalence estimates for persons aged thirty-five through sixty-four made available by the CDC (2007). The method of estimation is described below.

Estimates of the total number of survivors, l_a, by sex, s, and exact age, a, are shown in Table 2. The probability of an individual of sex s surviving from age a to age j is $P_s(j|a) = l_j/l_a$. The mortality rate at each age is calculated by subtracting survival probabilities at each age from 1.

The mortality ratio, which is the ratio of one group's death rate to that of the population, was used to split the table into three categories: current smokers, former smokers, and those who never smoked. The mortality ratio (M) by smoking status (SS) at each age (a) is $M_{SS,a} = q_{SS,a}/q_{T,a}$. The mortality rate for the total population is $q_{T,a}$, and $q_{SS,a}$ denotes the mortality rate by smoking status. For example, the mortality rate for current smokers by sex and exact age is calculated as $q_{CS,a} = M_{CS,a} \times q_{T,a}$. For persons aged twenty-one through thirty-five, the mortality ratio for male and female current and former smokers was assumed to be 1. For men aged thirty-five and older, the mortality ratios for current smokers and former smokers were 2.30 and 1.46, respectively. For female current and former smokers aged thirty-five and older, the mortality ratios were 1.92 and 1.30, respectively.[14]

To determine the number of survivors by smoking class, we initially assumed that 23.2 percent of men were current smokers and 34.3 percent were former smokers. For women, we assumed that 18.7 percent were current smokers and 22.9 percent were former smokers.[15] We subtracted mortality rates by sex for current smokers from 1 and multiplied by the number of current smokers that survived to age $a - 1$ to estimate the number of current smokers by sex surviving to age a. The number of surviving former smokers by sex and age was calculated in a similar manner. The number of people who have never smoked of sex s surviving to age a was estimated by subtracting the number of current and former smokers from the total number of survivors. The number of survivors at each age in the three smoking classes, as shown in

14. Mortality ratios for current and former smokers were obtained from Thun et al. (1997).
15. Smoking prevalence data for current and former smokers were obtained from the CDC (2007).

Table 2
Life Tables Used in Net Marginal Social Security Tax Rate Estimation of Survivors by Smoking Status

Age in 2002	Total population	Current smoker	Former smoker	Never smoked	Age in 2002	Total population	Current smoker	Former smoker	Never smoked
Females					**Females**				
20	98,922	18,538	22,604	57,780	61	90,138	15,614	20,076	54,448
21	98,877	18,530	22,593	57,754	62	89,374	15,360	19,854	54,159
22	98,827	18,520	22,582	57,725	63	88,552	15,089	19,617	53,846
23	98,781	18,512	22,571	57,698	64	87,657	14,796	19,359	53,502
24	98,736	18,503	22,561	57,672	65	86,680	14,479	19,079	53,122
25	98,688	18,494	22,550	57,644	66	85,631	14,143	18,779	52,709
26	98,639	18,485	22,539	57,615	67	84,512	13,788	18,460	52,264
27	98,589	18,476	22,528	57,586	68	83,281	13,402	18,110	51,768
28	98,539	18,466	22,516	57,557	69	81,982	13,001	17,743	51,238
29	98,483	18,456	22,503	57,524	70	80,556	12,567	17,342	50,647
30	98,424	18,445	22,490	57,489	71	79,026	12,109	16,914	50,004
31	98,362	18,433	22,476	57,453	72	77,410	11,633	16,464	49,313
32	98,296	18,421	22,461	57,415	73	75,666	11,130	15,982	48,554
33	98,225	18,407	22,444	57,373	74	73,802	10,604	15,470	47,729
34	98,148	18,393	22,427	57,328	75	71,800	10,051	14,924	46,824
35	98,064	18,363	22,402	57,299	76	69,639	9,470	14,340	45,828
36	97,970	18,329	22,374	57,267	77	67,366	8,877	13,732	44,757
37	97,869	18,293	22,344	57,232	78	64,935	8,262	13,088	43,585
38	97,759	18,253	22,311	57,195	79	62,372	7,636	12,416	42,320
39	97,640	18,210	22,276	57,153	80	59,621	6,989	11,704	40,928
40	97,500	18,160	22,234	57,105	81	56,681	6,327	10,954	39,400
41	97,355	18,109	22,192	57,055	82	53,660	5,680	10,195	37,785
42	97,194	18,051	22,144	56,999	83	50,324	5,002	9,371	35,951
43	97,023	17,990	22,093	56,940	84	47,075	4,382	8,585	34,109
44	96,830	17,921	22,036	56,873	85	43,542	3,751	7,747	32,045
45	96,627	17,849	21,976	56,802	86	39,919	3,151	6,909	29,859
46	96,405	17,770	21,910	56,724	87	36,246	2,595	6,083	27,569
47	96,176	17,689	21,843	56,644	88	32,571	2,090	5,281	25,201
48	95,928	17,602	21,769	56,557	89	28,943	1,643	4,516	22,784
49	95,654	17,505	21,689	56,460	90	25,411	1,258	3,800	20,354
50	95,364	17,403	21,603	56,357	91	22,024	936	3,141	17,947
51	95,059	17,297	21,513	56,249	92	18,828	675	2,549	15,604
52	94,724	17,179	21,415	56,130	93	15,862	471	2,027	13,364
53	94,380	17,060	21,314	56,007	94	13,158	317	1,578	11,264
54	93,989	16,924	21,199	55,866	95	10,737	205	1,200	9,332
55	93,572	16,780	21,077	55,716	96	8,613	127	892	7,594
56	93,095	16,616	20,937	55,542	97	6,785	75	646	6,064
57	92,629	16,456	20,801	55,372	98	5,245	42	455	4,747
58	92,084	16,270	20,642	55,172	99	3,977	23	312	3,642
59	91,491	16,069	20,469	54,953	100	2,954	12	208	2,735
60	90,826	15,845	20,275	54,706					

(continued)

Table 2 (continued)

Age in 2002	Total population	Current smoker	Former smoker	Never smoked	Age in 2002	Total population	Current smoker	Former smoker	Never smoked
Males					**Males**				
20	98,436	22,778	33,724	41,934	61	83,612	16,028	26,805	40,779
21	98,299	22,746	33,677	41,875	62	82,483	15,530	26,276	40,677
22	98,157	22,714	33,629	41,815	63	81,255	14,998	25,705	40,552
23	98,021	22,682	33,582	41,757	64	79,946	14,442	25,101	40,403
24	97,882	22,650	33,534	41,698	65	78,556	13,865	24,463	40,228
25	97,746	22,618	33,488	41,640	66	77,071	13,262	23,788	40,021
26	97,614	22,588	33,443	41,584	67	75,501	12,641	23,081	39,779
27	97,479	22,557	33,396	41,526	68	73,809	11,989	22,326	39,494
28	97,352	22,527	33,353	41,472	69	72,012	11,318	21,532	39,162
29	97,225	22,498	33,309	41,418	70	70,087	10,622	20,692	38,773
30	97,091	22,467	33,263	41,361	71	68,039	9,908	19,809	38,322
31	96,954	22,435	33,216	41,302	72	65,864	9,180	18,884	37,800
32	96,813	22,403	33,168	41,242	73	63,621	8,461	17,945	37,215
33	96,678	22,371	33,122	41,185	74	61,202	7,721	16,949	36,532
34	96,526	22,336	33,070	41,120	75	58,680	6,989	15,930	35,761
35	96,367	22,251	32,990	41,125	76	56,028	6,262	14,878	34,887
36	96,196	22,161	32,905	41,131	77	53,251	5,549	13,802	33,901
37	96,016	22,065	32,815	41,136	78	50,398	4,865	12,722	32,811
38	95,823	21,963	32,719	41,141	79	47,454	4,211	11,637	31,606
39	95,610	21,851	32,612	41,147	80	44,370	3,582	10,533	30,255
40	95,381	21,731	32,498	41,152	81	41,252	3,003	9,452	28,797
41	95,128	21,598	32,373	41,157	82	38,102	2,475	8,399	27,228
42	94,859	21,458	32,239	41,163	83	34,798	1,982	7,335	25,481
43	94,577	21,311	32,099	41,167	84	31,719	1,578	6,388	23,753
44	94,266	21,150	31,945	41,171	85	28,478	1,207	5,435	21,836
45	93,929	20,976	31,778	41,175	86	25,296	897	4,548	19,851
46	93,569	20,791	31,600	41,178	87	22,212	646	3,739	17,828
47	93,171	20,587	31,404	41,179	88	19,266	449	3,015	15,803
48	92,755	20,376	31,199	41,180	89	16,494	300	2,381	13,812
49	92,296	20,144	30,974	41,178	90	13,925	193	1,840	11,893
50	91,809	19,900	30,735	41,174	91	11,585	118	1,388	10,078
51	91,286	19,639	30,480	41,167	92	9,490	69	1,022	8,399
52	90,722	19,360	30,205	41,157	93	7,648	38	732	6,877
53	90,138	19,073	29,921	41,144	94	6,059	20	510	5,529
54	89,505	18,765	29,614	41,126	95	4,715	10	345	4,360
55	88,850	18,449	29,298	41,103	96	3,601	4	226	3,371
56	88,102	18,092	28,938	41,072	97	2,698	2	143	2,553
57	87,369	17,746	28,586	41,037	98	1,982	1	88	1,894
58	86,542	17,360	28,191	40,991	99	1,426	0	52	1,374
59	85,644	16,945	27,764	40,935	100	1,005	0	29	975
60	84,637	16,487	27,287	40,863					

Note: "Survivors" refers to the number of persons by smoking status reaching age a during the year among the stationary population.

Source: Constructed from life tables published by the National Center for Health Statistics

Table 3
Net Marginal Social Security Tax Rate Estimates for Single Primary Beneficiaries by Sex, Smoking Status, Earnings Classification, and Age in 2002

Age in 2002	Current smoker			Former smoker			Never smoked		
	Low earning	Average earning	High earning	Low earning	Average earning	High earning	Low earning	Average earning	High earning
				Females					
35	–0.14	6.78	8.81	–2.87	5.81	8.35	–6.84	4.40	7.69
45	–2.83	5.83	8.36	–6.09	4.67	7.82	–10.78	3.00	7.04
55	–4.77	5.13	8.04	–8.13	3.94	7.48	–12.86	2.26	6.69
65	–9.84	3.33	7.19	–13.13	2.16	6.64	–17.62	0.57	5.90
				Males					
35	3.96	8.24	9.49	0.73	7.09	8.96	–7.68	4.10	7.55
45	2.21	7.62	9.20	–1.61	6.26	8.57	–11.16	2.86	6.97
55	–0.91	6.51	8.68	–5.37	4.92	7.94	–15.69	1.25	6.22
65	–8.84	3.69	7.36	–13.68	1.97	6.55	–23.50	–1.53	4.92

Note: Workers are assumed to retire at the full benefit retirement age. Low-earning workers expect a marginal replacement rate of 0.9, and average- and high-earning workers expect rates of 0.32 and 0.15, respectively. A real discount rate of 3 percent is assumed. The growth rate in real earnings is set at 1.1 percent.

Table 2, is then used to calculate the probability that a person age a will survive to age j. For each smoking class, the survival probabilities are in turn used to calculate B_{PV}.

NMSSTRs for single primary beneficiaries that account for smoking-attributable mortality by age, gender, and earnings class are shown in Table 3. As expected, a comparison of the results in Tables 1 and 3 reveals that a smoker's shorter life expectancy increases the NMSSTR at each age. A single male current smoker aged fifty-five with lifetime average earnings faces a net tax rate of 6.51 percent, which is approximately 3 percentage points higher than the rate estimated under the common mortality assumption (3.55 percent). The NMSSTR for a single male former smoker aged fifty-five with average lifetime earnings is 4.92 percent, which is approximately 1.4 percentage points higher than the rate estimated under the common mortality assumption. The NMSSTR for a single man aged fifty-five who never smoked with average lifetime earnings is 1.25 percent—5.3 percentage points lower than the rate for a current smoker and 3.7 percentage points lower than the rate for a fifty-five-year-old former smoker of the same age.

A single female current smoker aged fifty-five with lifetime average earnings faces an NMSSTR of 5.13 percent, which is approximately 1.4 percentage points lower than the rate estimated for a fifty-five-year-old current smoking man with lifetime average earnings. The gender differential in NMSSTRs for both current and former smokers at each age is larger than the differential estimated under the common mortality assumption. In addition, sixty-five-year-old female current and former smokers now incur a lower NMSSTR than do their male counterparts. These gender differences result from males smoking at higher rates than females and having a higher smoking-attributable mortality risk.

As shown in Table 4, a fifty-five-year-old male current smoker with lifetime average earnings and a dependent spouse who also smokes incurs an NMSSTR of 3.17 percent, which is more than 3 percentage points higher than the rate estimated under the common mortality assumption (–0.02). In addition, this rate is 1.69 percentage

Table 4
Net Marginal Social Security Tax Rate Estimates for Male Primary Beneficiaries with a Dependent Spouse by Earnings Classification, Smoking Status, and Age in 2002

	Primary beneficiary current smoker Dependent spouse			Primary beneficiary former smoker Dependent spouse			Primary beneficiary never smoked Dependent spouse		
Age in 2002	Currrent smoker	Former smoker	Never smoked	Current smoker	Former smoker	Never smoked	Current smoker	Former smoker	Never smoked
				Low earner					
35	–4.39	–6.41	–8.68	–5.56	–7.14	–8.97	–13.77	–15.42	–17.44
45	–6.53	–8.57	–10.87	–8.53	–10.18	–12.10	–18.21	–20.01	–22.22
55	–10.30	–12.38	–14.71	–13.31	–15.05	–17.05	–24.26	–26.21	–28.57
65	–20.90	–22.95	–25.18	–24.76	–26.51	–28.47	–35.42	–37.38	–39.68
				Average earner					
35	5.27	4.55	3.75	4.86	4.29	3.64	1.94	1.35	0.63
45	4.51	3.78	2.97	3.80	3.21	2.53	0.36	–0.28	–1.07
55	3.17	2.43	1.60	2.10	1.48	0.77	–1.80	–2.49	–3.33
65	–0.60	–1.33	–2.12	–1.97	–2.59	–3.29	–5.76	–6.46	–7.28
				High earner					
35	8.10	7.77	7.39	7.91	7.64	7.34	6.54	6.26	5.93
45	7.75	7.40	7.02	7.41	7.14	6.82	5.80	5.50	5.13
55	7.12	6.77	6.38	6.62	6.33	5.99	4.79	4.46	4.07
65	5.35	5.01	4.64	4.71	4.42	4.09	2.93	2.60	2.22

Note: Workers are assumed to retire at the full benefit retirement age. Low-earnings workers expect a marginal replacement rate of 0.9, and average- and high-earning workers expect rates of 0.32 and 0.15, respectively. A real discount rate of 3 percent is assumed. The growth rate in real earnings is set at 1.1 percent.

points higher than the rate incurred by a fifty-five-year-old male former smoker with lifetime average earnings and a dependent spouse who formerly smoked (1.48 percent) and approximately 5.5 percentage points higher than the rate incurred by a fifty-five-year-old male who never smoked with lifetime average earnings and a dependent spouse who never smoked (–2.49 percent).

Results and Discussion

As previous studies have shown, we find that Social Security treats single people and dual-income couples less equitably than single-income couples. This study's results add to previous findings by showing that NMSSTRs also vary by smoking status.[16] The higher tax rates that smokers incur may reduce their labor supply.[17] Given that Social Security is financed by a payroll tax on earnings, any reduction in the labor supply will have implications for the system's funding. However, the aggregate effect of smoking on the OASI Trust Fund's finances would depend on how smoking redistributes benefits from smokers to people who never smoked and the resulting labor supply response to changes in marginal tax rates.

While Social Security has reduced poverty among elderly Americans, young widows are at increased risk of living in poverty because of the premature death of their spouse (Redja 1994; Engelhardt and Gruber 2004; Sevak, Weir, and Willis 2004). Many individuals who smoke die prematurely. Approximately 536,000 adults in the

United States under age sixty-five died of smoking-attributable illnesses between 1997 and 2001.[18] Widows with no children under age sixteen in their care who were married to fully insured workers who died prematurely may be ineligible for Social Security benefits until they reach age sixty. Estimates suggest that 15 percent of women aged fifty-four, too young to qualify for Social Security benefits, fall into poverty following the death of their husband (Sevak, Weir, and Willis 2004).[19] As a result, it has been suggested that Social Security is failing to live up to one of its primary goals—providing adequate survivors insurance for older low-earning Americans (Gustman and Steinmeier 2002). One proposal to improve Social Security's adequacy is to lower the eligibility age for widows from sixty years to fifty-five years (Redja 1994).[20] In addition to the establishment of private accounts, two of the three plans proposed by the President's Commission to Strengthen Social Security (2001) recommended an increase in benefits for low-earning widows and widowers.

Because low-earning workers are more likely to smoke and smokers are more likely than people who have never smoked to die prematurely, an unintended distributional effect of enacting proposals that would reduce widows' retirement age or increase retirement benefits among low-earning widows and widowers would be to redistribute benefits from people who have never smoked to smokers, thus benefiting behavior that is detrimental to health. As with life insurance, perhaps this unintended effect could be offset by smokers' paying a higher premium, in this case a smoker's insurance tax rate. The revenue generated from a tax levied on current smokers could be added to the OASI Trust Fund and used to reduce financial hardship currently faced by young widows and widowers by paying increased benefits or paying benefits at an earlier age. In addition, the higher tax penalty associated with smoking may increase cessation. The aggregate impact of such a change on the various trust fund finances would be a valuable addition to the debates surrounding the system's solvency and ways to reduce poverty among widows and widowers.

16. It has been suggested that premature deaths attributable to smoking save Social Security money (Shoven, Sundberg, and Bunker 1987). One should not infer from these results that because smokers incur a higher NMSSTR they pay more than their fair share to Social Security; the higher NMSSTR may cause smokers to reduce their labor supply and thereby reduce Social Security contributions. In addition, Social Security disability payments to persons with smoking-attributable diseases and payments to dependents and survivors of deceased smokers will offset reductions in future system liabilities that stem from smoking-attributable death.
17. In addition to reducing hours of work, an increase in taxes may decrease labor force participation. Specifically, smoking may lead to a reduction in labor supply through early retirement. Retirement studies have typically used average life expectancy by age as opposed to predictions based on health status in their analysis (Social Security Advisory Council 1997). Those smokers in poor health who retire early may be responding to financial incentives that are masked in analyses that use average life expectancies.
18. These estimates are unpublished and were estimated from Smoking-Attributable Mortality Morbidity and Economic Cost (SAMMEC) data maintained by the Office on Smoking and Health at the CDC. SAMMEC estimates are available at <http://apps.nccd.cdc.gov/sammec/>.
19. We do not know how many widows under age sixty are ineligible for benefits. However, we do know that in the year 2000, 45,680 widows received benefits because they had a child under age sixteen in their care (USDHHS 2001, table 5.F1).
20. It is unclear why age fifty-five is recommended. Widows under age fifty-five whose eligibility is based solely on age would continue to be ineligible for Social Security benefits, and the system would fail to live up to one of its main goals of providing adequate retirement security. Additional information on proposals aimed at changing Social Security survivorship benefits and poverty among widows is available from Anzick and Weaver (2001).

Table 5
Net Marginal Social Security Tax Rate Estimates for Average Earner Primary Beneficiaries and Dependents by Sex, Smoking Status, and Age in 2002

	Primary beneficiary						Smoking status of male primary beneficiary and dependent spouse		
	Single female			Single male					
Age in 2002	Current smoker	Former smoker	Never smoked	Current smoker	Former smoker	Never smoked	Both current smokers	Both former smokers	Both never smoked
				2.2 percent discount rate					
35	5.83	4.56	2.69	7.69	6.23	2.35	5.34	4.30	0.50
45	5.08	3.68	1.64	7.19	5.59	1.52	4.68	3.35	–0.95
55	4.76	3.42	1.50	6.28	4.54	0.45	3.49	1.83	–2.88
65	3.42	2.18	0.47	3.85	2.08	–1.58	0.07	–1.84	–6.32
				3.0 percent discount rate					
35	7.07	6.17	4.87	8.42	7.35	4.59	5.67	4.77	1.38
45	6.18	5.11	3.57	7.84	6.58	3.44	4.97	3.76	–0.19
55	5.54	4.44	2.89	6.81	5.35	1.95	3.73	2.17	–2.28
65	3.88	2.79	1.32	4.20	2.61	–0.62	0.24	–1.60	–5.94
				3.7 percent discount rate					
35	7.87	7.21	6.25	8.90	8.09	6.02	5.90	5.09	1.96
45	6.95	6.10	4.89	8.30	7.28	4.77	5.18	4.06	0.34
55	6.13	5.20	3.90	7.22	5.95	3.06	3.92	2.43	–1.83
65	4.25	3.28	1.97	4.49	3.04	0.13	0.38	–1.42	–5.63

Note: Workers are assumed to retire at the full benefit retirement age. Average-earning workers expect a marginal replacement rate of 0.32. The growth rate in real earnings is set at 1.1 percent. Estimates account for smoking-attributable mortality and taxation of benefits.

As in previous studies, these results are limited in that they are based on hypothetical workers; thus, the relative importance of various economic assumptions and differences is an empirical question.[21] Because analysis with money flows over time may be sensitive to the choice of discount rate, selective results shown in Tables 1, 3, and 4 for workers with average lifetime earnings were reestimated under alternative discount rate assumptions. As shown in Table 5, a lower discount rate reduces the NMSSTR at each age.[22]

Although the calculations presented are complex, they oversimplify the Social Security program in a number of ways. First, we focus on OASI and ignored the DI and HI components of Social Security. Second, we ignore benefits for dependent children of young widows or widowers. Third, we ignore the possibility of divorce and remarriage. Fourth, the employer portion of the payroll tax is tax exempt, and given the progressive nature of income taxation, this exemption disproportionately benefits higher-earning individuals. Thus, the NMSSTR for high-earning individuals may be lower than the estimates reported. Fifth, smoking prevalence is held constant across earnings classes. Because lower-earning individuals have a higher smoking prevalence than do higher-earning individuals, low-earning individuals' NMSSTRs may be higher than the rates reported whereas average- and high-earning individuals may have NMSSTRs that are lower than the rates reported.

A final potential limitation to our results is that the mortality risk measures used to account for the mortality difference among current and former smokers are adjusted

for sex and age only. Other risk factors such as educational status, diet, and alcohol consumption that are correlated with smoking were unaccounted for in the mortality risk measure that was used. As a consequence, the NMSSTR estimates may overstate the tax penalty associated with smoking (Shoven, Sundberg, and Bunker 1987; Thun et al. 1997). However, this limitation may not pose too great a problem because evidence in the literature suggests that when behavioral and demographic factors correlated with smoking were taken into account, the higher mortality risks faced by smokers did not change much (Malarcher et al. 2000; Thun et al. 1997).

Conclusion

The analyses reveal that smokers will incur higher net marginal tax rates than people who never smoked and may reduce their labor supply.[23] Any reduction in labor supply among smokers will have implications for the system's funding. Knowledge of the distributional effects of smoking on Social Security is important not only from the standpoint of the system's funding but also from the perspective of informing smoking cessation efforts (Rice et al. 1986). People can avoid higher net marginal tax rates by never smoking or reduce them by quitting smoking. Finally, smoking status should be considered in assessing Social Security legislative proposals designed to reduce system inequities or promote social adequacy—in particular, amendments designed to reduce poverty among young widows and widowers. Failure to do so may unintentionally promote behavior that is detrimental to health.

21. However, this methodology is the best one can do since the actual data are unavailable (Garrett 1995). For a discussion of the usefulness of results based on hypothetical worker data, see Leimer (1995).
22. The calculations shown in Tables 1, 3, and 4 ignored the personal income tax bracket at which Social Security retirement benefits will be taxed during retirement. Thus, the estimates shown in Table 5 assumed that Social Security benefits will be subject to a federal income tax rate of 15 percent. For a single male current smoker aged fifty-five, assuming a discount rate of 3 percent, taxation of benefits increased his NMSSTR by 0.3 percentage points (6.51 percent versus 6.81 percent).
23. The evidence is mixed on the impact of Social Security on the labor supply although the predominant research in this area has focused on the labor supply responses of older workers (Krueger and Meyer 2002).

REFERENCES

Aaron, Henry J. 1977. Demographic effects on the equity of social security benefits. In *Economics of Public Service*, edited by Martin S. Feldstein and Robert P. Inman. New York: Macmillan.

Anzick, Michael A., and David A. Weaver. 2001. Reducing poverty among elderly women. Social Security Administration Working Paper No. 8, January.

Armour, Brian S., and Melinda M. Pitts. 2004. Incorporating insurance rate estimates and differential mortality into the net marginal Social Security tax rate calculation. *Public Finance Review* 32, no. 6:588–609.

Brittain, John A. 1972. The incidence of Social Security payroll taxes. *American Economic Review* 61, no. 1:110–25.

Browning, Edgar K. 1985. The marginal Social Security tax on labor. *Public Finance Quarterly* 13, no. 3:227–51.

Burkhauser, Richard V., and John A. Turner. 1985. Is the Social Security payroll tax a tax? *Public Finance Quarterly* 13, no. 3:253–67.

Centers for Disease Control and Prevention (CDC). 2003. Cigarette smoking-attributable morbidity—United States, 2000. *Morbidity and Mortality Weekly Report* 52, no. 35:300–3.

———. 2007. Smoking-attributable mortality, morbidity, and economic costs (SAMMEC): Adult SAMMEC and maternal and child health (MCH). SAMMEC software. <http://apps.nccd.cdc.gov/sammec/login.asp> (July 20, 2007).

Engelhardt, Gary V., and Jonathan Gruber. 2004. Social Security and the evolution of elderly poverty. National Bureau of Economic Research Working Paper No. 10466, May.

Feldstein, Martin, and Andrew A. Samwick. 1992. Social Security rules and marginal tax rates. *National Tax Journal* 45, no. 1:1–22.

Garrett, Daniel M. 1995. The effects of differential mortality rates on the progressivity of Social Security. *Economic Inquiry* 33, no. 3:457–75.

Gustman, Alan L., and Thomas L. Steinmeier. 2002. The new Social Security commission personal accounts: Where is the investment principle? National Bureau of Economic Research Working Paper No. 9045, July.

Krueger, Alan B., and Bruce D. Meyer. 2002. Labor supply effects of social insurance. National Bureau of Economic Research Working Paper No. W9014, June.

Leimer, Dean R. 1995. A guide to Social Security money's worth issues. *Social Security Bulletin* 58, no. 2:3–20.

Malarcher, Ann M., Jane Schulman, Leonardo Epstein, Michael J. Thun, Paul Mowery, Ben Pierce, Luis Escobedo, and Gary A. Giovino. 2000. Methodological issues in estimating smoking-attributable mortality in the United States. *American Journal of Epidemiology* 152, no. 6:573–84.

Max, Wendy. 2001. The financial impact of smoking on health-related costs: A review of the literature. *American Journal of Health Promotion* 15, no. 5:321–31.

President's Commission to Strengthen Social Security. 2001. Strengthening Social Security and creating personal wealth for all Americans. December. <www.commtostrengthensocsec.gov/reports/Final_report.pdf> (June 14, 2007).

Redja, George E. 1994. *Social insurance and economic security*. Englewood Cliffs, N.J.: Prentice Hall.

Rice, Dorothy P., Thomas A. Hodgson, Peter Sinsheimer, Warren Browner, and Andrea N. Kopstein. 1986. The economic costs of the health effects of smoking, 1984. *Milbank Quarterly* 64, no. 4:489–547.

Sevak, Purvi, David R. Weir, and Robert J. Willis. 2004. The economic consequences of a husband's death: Evidence from HRS and AHEAD. *Social Security Bulletin* 65, no. 3:31–44.

Shoven, John B., Jeffrey O. Sundberg, and John P. Bunker. 1987. The Social Security cost of smoking. National Bureau of Economic Research Working Paper No. 2234, May.

Social Security Advisory Council. 1997. Report of the 1994–1996 Advisory Council on Social Security Volume I: Findings and Recommendations. <www.ssa.gov/history/reports/adcouncil/report/toc.htm>. (July 23, 2007).

Thun, M.J., C. Day-Lally, D.G. Myers, E.E. Calle, W.D. Flanders, B.P. Zhu, and M.M. Namboodiri. 1997. Trends in tobacco smoking and mortality from cigarette use in Cancer Prevention Studies I (1959 through 1965) and II (1982 through 1988). In *Changes in cigarette-related disease risks and their implication for prevention and control. Smoking and tobacco control monograph* 8. National Institutes of Health Publication No. 97-4213. Washington, D.C.: NIH.

U.S. Department of Health and Human Services (USDHHS). 2001. Annual statistical supplement to the *Social Security Bulletin*. Washington, D.C.: Social Security Administration.

———. 2004a. Annual statistical supplement to the *Social Security Bulletin*. Washington, D.C.: Social Security Administration.

———. 2004b. The health consequences of smoking: A report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Public Health Service, CDC, Center for Chronic Disease Prevention and Health Promotion, and the Office on Smoking and Health.

———. 2005. The 2005 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Fund. Washington, D.C.: Social Security Administration.