Banks with Something to Lose: The Disciplinary Role of Franchise Value

Rebecca S. Demsetz, Marc R. Saidenberg, and Philip E. Strahan

As protectors of the safety and soundness of the banking system, banking supervisors are responsible for keeping banks' risk taking in check. On-site examinations, off-site surveillance, and capital requirements are some of the tools that supervisors use to achieve this goal. Franchise value—the present value of the stream of profits that a firm is expected to earn as a going concern—makes the supervisor's job easier by reducing banks' incentives to take risk. In banking, sources of franchise value include efficiency, access to markets protected from competition, and valuable lending relationships. Franchise value can help reduce excessive risk taking because banks with high franchise value have much to lose if a risky business strategy leads to insolvency.

Economists studying the relationship between franchise value and risk have noted some interesting patterns over time. Most notably, Keeley (1990) documents declines in bank franchise value during the 1950s, 1960s, and 1970s, when the banking industry was experiencing deregulation and increased competition from nonbank financial institutions. He argues that this drop in franchise value led to increased risk taking in the 1980s, a decade in which the average failure rate for U.S. banks reached a fifty-year high of almost 100 per year.

In this article, we explore the relationship between franchise value and risk taking over the 1986-94 period. We extend Keeley's empirical analysis by estimating the effect of franchise value on a variety of measures of bank risk. We find an inverse relationship between franchise value and an "all-in" measure of risk based on stock-return volatility, which incorporates the risks of a bank's asset, liability, and off-balance-sheet positions as well as its leverage. We also use information from the balance sheet to determine how high-franchise-value banks reduce risk. We find that banks with more franchise value hold more capital and have less asset risk than banks with less franchise value. Though their tendency to hold risky loans is similar to that of other banks, banks with high franchise value maintain better diversified loan portfolios.

Franchise Value and Risk Taking in Banking

We define franchise value as the present value of the future profits that a firm is expected to earn as a going concern. Profits are those gains beyond what is required to cover all costs, *including* the cost of capital. Most firms in competitive environments cannot generate stable profits because competition tends to force them to lower their prices to levels just high enough to cover all costs. However, firms with access to superior technologies, such as new production processes, or scarce factors of production, such as talented managers, may have franchise value.

In banking, franchise value arises from two main sources. First, competition has been limited by regulations, giving banks greater access to profits. We term franchise value stemming from these restrictions "market-related," since differences in such franchise value vary across geographic and product markets but not across banks operating in the same geographic and product markets. Although market-related sources of franchise value were important in the 1970s, more recently that importance has diminished. Second, franchise value arises from what we term "bank-related" sources, such as efficiency differences and variations in the value of lending relationships. These bank-related factors continue to be an important source of franchise value today.

Franchise value can help reduce excessive risk taking because banks with high franchise value have much to lose if a risky business strategy leads to insolvency.

Market-Related Sources of Franchise Value

Before the 1970s, banks faced limits on geographic expansion both within states and across state borders. The Douglas Amendment to the Bank Holding Company Act of 1956 prevented bank holding companies (BHCs) from acquiring an out-of-state bank unless that bank's state explicitly permitted such acquisitions by statute. Since no state allowed such acquisitions, holding companies were in effect prohibited from operating across state lines. In addition, before 1970, about two-thirds of the states had laws restricting intrastate branching.

Both restrictions effectively limited competition within the banking industry, thereby providing banks with a greater opportunity to build franchise value. This opportunity varied across banking markets.[1] For instance, banks located in states permitting no branching faced less competition than those located in states allowing limited branching.

Competition among banks changed dramatically from the mid-1970s to the mid-1990s, however, as most of the restrictions on intrastate and interstate banking were lifted. Between 1975 and 1992, two-thirds of the states relaxed restrictions on intrastate branching. During the 1980s and early 1990s, every state but Hawaii entered into a regional or national interstate banking arrangement whereby bank holding companies could operate across state lines by owning banks in more than one state. In September 1994, the Reigle-Neal Interstate Banking and Branching Efficiency Act became law, permitting nationwide interstate banking and, with state approval, interstate branching. These changes have significantly *increased* competition within the banking industry and consequently *lowered* franchise value at many banks.

Franchise value has also declined as a result of innovation. Automated teller machines, introduced in the 1970s, increased competition by permitting banks to penetrate local markets without building full-scale branches. In the late 1970s, nonbank financial institutions such as money market mutual funds began offering close substitutes for bank products, further elevating competition and eroding bank franchise value.

Moreover, by the mid-1980s, Regulation Q interest rate ceilings were fully phased out. While this development helped banks compete with other financial intermediaries for savings, it increased competition within the banking industry.[2]

Bank-Related Sources of Franchise Value

Although regulatory and technological changes have eroded market-related sources of franchise value, bank-related sources remain important. For example, a bank's branch network can give it a competitive advantage in dealing with customers who prefer the convenience of full-service banking at a local branch. In recent years, we have also seen banks use the locational and marketing advantages of branch networks to sell financial products such as mutual funds and life insurance. Moreover, as in all businesses, some banks are simply more efficient than others. The better-managed ones derive franchise value from their ability to provide banking services less expensively than their competitors. While the removal of barriers to geographic expansion increased competition and reduced franchise value at many banks, the better-managed banks may have benefited from the opportunity to grow at the expense of their poorly managed rivals (Jayaratne and Strahan 1996).

Banks' unique relationships with many of their borrowers may also generate franchise value. Banks typically establish long-term relationships that allow them to gain private information on the characteristics and credit risks of their borrowers—information not readily available to other bank or nonbank lenders (Berger and Udell 1995; Petersen and Rajan 1995). These relationships reduce the cost of loan origination, making lending more profitable. Lending relationships continue to be an important source of franchise value.

How Does Franchise Value Affect Bank Behavior?

Firms that succeed in building franchise value will seek to preserve it. Consequently, firms with large amounts of franchise value may be predisposed to operate more safely than those with little or none. For instance, high-franchise-value banks may be more likely to hold capital in excess of that required by regulations, to limit their exposure to high-risk borrowers, and to hold well-diversified loan portfolios. In using derivatives, they may also be more likely to hedge against losses stemming from changes in interest rates and foreign exchange rates than to speculate. These strategies minimize the likelihood that such banks will lose their franchise value through insolvency.

Franchise value plays a particularly important role in banking because it helps mitigate the "moral hazard problem" associated with the federal safety net. The safety net, composed of the Federal Reserve's discount window, federal deposit insurance, and extensive supervision and regulation of banks, helps ensure the soundness of the banking system. However, this protection does not come without cost. The safety net creates a moral hazard problem by insulating bank creditors from losses, thereby limiting their incentive to restrain risk taking. Insured depositors have little motivation to keep risk in check by demanding interest rates commensurate with bank risk or by withdrawing deposits when banks become riskier. Franchise value can help lessen the moral hazard problem by increasing banks' incentives to operate safely, thereby aligning their interests with those of the deposit insurer and bank supervisor.[3]

Banks' unique relationships with many of their borrowers may also generate franchise value. . . . These relationships reduce the cost of loan origination, making lending more profitable.

An example may clarify the point. Consider the incentives facing the imaginary FirstRisk Bank, which has little capital and little or no franchise value. Its owners may decide to make high-risk loans to a high-tech start-up, knowing that if the loans are repaid, the bank will earn hefty profits.[4] If the loans default and the bank finds itself insolvent, the owners will have lost very little. (Insured depositors would have little reason to discipline FirstRisk Bank because the Federal Deposit Insurance Corporation guarantees their deposits.)

Suppose that FirstRisk Bank gets lucky. The high-tech start-up with which it has developed a strong lending relationship develops into the industry leader. The firm becomes very profitable, and FirstRisk Bank becomes profitable as a result of the lending relationship it has forged. Now FirstRisk Bank has *high* franchise value, since as a going concern it can expect strong future profits. With franchise value to lose, FirstRisk's owners are likely to rethink their aggressive lending strategy. They will probably avoid further risky lending. Moreover, they will have both the incentive and the ability to raise the bank's capital-to-assets ratio, further lowering the likelihood of losing the valuable lending relationship through insolvency.

Note that the cost of failure—as well as FirstRisk's incentive to avoid it—is particularly high if FirstRisk's profitable lending relationship cannot be transferred easily to another lending institution. In general, nontransferable franchise value increases the cost of bank failures not just to owners but also to borrowers, who may have difficulty establishing new lending relationships with other banks. To preserve the franchise value of a failed bank, the Federal Deposit Insurance Corporation typically searches for a buyer willing to assume the bank's assets and liabilities in their entirety through a purchase-and-assumption transaction.[5] When lending relationships are longstanding, however, even such a transaction is unlikely to preserve the full franchise value of the failed bank.

Franchise value plays a particularly important role in banking because it helps mitigate the "moral hazard problem" associated with the federal safety net.

FRANCHISE VALUE AND BANK CAPITAL

Franchise value is not the only force mitigating moral hazard in banking. Uninsured creditors have an incentive to monitor risk taking and to demand returns commensurate with bank risk. Bank supervisors also monitor risk taking through on-site examinations and off-site surveillance and discipline risk taking by enforcing certain rules of operation.

Perhaps the most important of these rules of operation are those dictating the maintenance of minimum capital ratios. Should a risky strategy result in insolvency, bank owners would lose their capital along with any franchise value. By requiring banks to meet capital standards, these regulations give bank owners an additional incentive to avoid excessively risky behavior.

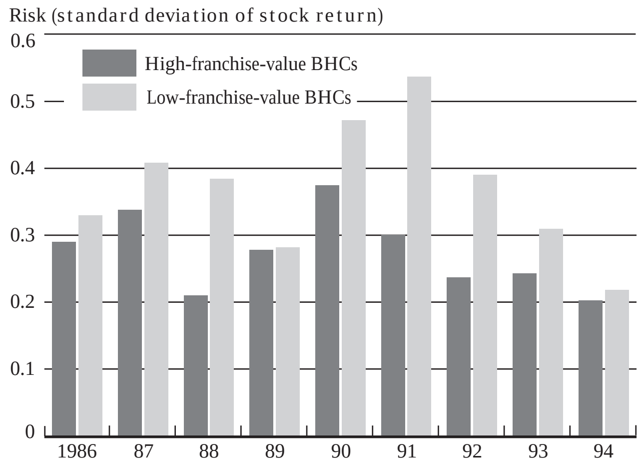
While a bank's capital position and its franchise value can each discourage risk taking, franchise value may more consistently align the incentives of the bank owner with those of the supervisor. A bank's capital position tends to vary over time in response to changes in loan demand, interest rates, and general economic conditions. In contrast, characteristics that generate franchise value, particularly those related to efficiency, are more stable. For instance, a bank with high franchise value stemming from its ability to operate as a low-cost provider will have access to profits even under poor economic conditions. This bank will have a strong incentive to avoid excessive risk taking throughout the business cycle.

When capital and franchise value are both adversely affected at a large number of institutions, the ramifications can be severe. The thrift crisis of the 1980s provides a good example. Thrift franchise value fell for many of the same reasons that franchise value fell in banking and because the development of secondary markets in mortgage securities reduced thrifts' ability to earn profits from mortgage lending. Moreover, unlike banks, thrifts faced a very large reduction in capital in the late 1970s and early 1980s because the value of their mortgage portfolios, which dominate thrift balance sheets, fell sharply in response to increased interest rates. Since the thrifts had lost much of their franchise value, owners had little incentive to rebuild their capital positions. Instead, many used fully insured deposits to increase their holdings of high-risk assets such as junk bonds and commercial real estate. This risky behavior led to a large number of thrift failures and ultimately to the taxpayer bailout of the thrift insurance fund.

In contrast, increases in franchise value should lead banks to strengthen their capital positions voluntarily. In our earlier example, we expected FirstRisk Bank to increase its capital-to-assets ratio to reduce insolvency risk

Chart 1

Risk of Bank Holding Companies with High and Low Franchise Value



Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

after franchise value rose. The bank could increase the ratio through stock sales, changes in dividend policies, or changes in the size of the balance sheet.

As we have seen, banks with high franchise value have an incentive to reduce risky behavior and strengthen their capital positions. Consequently, we expect risk at banks to be negatively related to franchise value. We look for evidence of this pattern in Chart 1, which reports the average riskiness of bank holding companies with high and low franchise value for 1986-94. We see that low-franchise-value BHCs are consistently riskier than their high-franchise-value counterparts, except in 1989, when the risk of the two groups is similar. This pattern is consistent with our expectations, but the analysis does not control for other factors that may affect risk.

Quantifying the Franchise Value/Risk Relationship

We now use a series of regressions to confirm that low-franchise-value BHCs operate with greater risk.[6] This approach also allows us to quantify the strength of the relationship between franchise value and risk.

Each of our regressions has one of seven measures of risk as its dependent variable. The independent variables include franchise value and two controls: asset size and the growth of personal income in the states where the BHC operates.[7] Asset size affects risk in two potentially offsetting ways. On the one hand, larger BHCs tend to be better diversified and hence less vulnerable to economic shocks. On the other hand, larger BHCs typically engage in riskier activities. For instance, larger BHCs generally have a larger share of their loan portfolios in relatively risky commercial and industrial loans and a smaller share in relatively safe mortgage loans. Growth in personal income is included to control for regional business cycles that can affect risk at all banks in a given area. The results of our regressions tell us whether differences in franchise value can explain differences in risk taking among BHCs of similar size in similar economic environments.

The dependent variables in our seven regressions are measured using natural logarithms. The log specification allows the estimated effect of franchise value on risk to diminish as franchise value grows and risk falls. We believe this approach is appropriate because the threat of insolvency motivates banks to reduce risk and because the likelihood of insolvency is low for banks with sufficiently low levels of risk. These banks have little incentive to reduce risk further as franchise value rises.

We estimate each of our regressions using fixed-effects and random-effects models. Because our data set follows a sample of BHCs over time, these models can be used to control for time-invariant, BHC-specific factors that may be related to risk taking but are not explicitly included in our regressions. In a random-effects regression, this is done by specifying a certain mathematical structure to the regression residuals. In a fixed-effects regression, all variables are calculated as deviations from their BHC means, so that regression results are driven by BHC-specific changes in the regression variables over time.

The advantage of the random-effects model is that cross-sectional differences in risk such as those illustrated in Chart 1 are reflected in the regression coefficients. The advantage of the fixed-effects model is that omitted BHC-specific factors related to risk taking are less likely to bias the regression coefficients. Our regressions also include time fixed effects, that is, each regression controls for changes in the average level of risk over the years in the sample period.

MEASURING FRANCHISE VALUE AND RISK

Recall that franchise value is defined as the present value of a firm's future profits—revenues in excess of all costs, including the cost of capital. One way to quantify franchise value is to look at the difference between a firm's market value and its replacement cost, where replacement cost is the expense of rebuilding the firm today:

Franchise value (FV) = market value – replacement cost.

The difference between market value and replacement cost will be large when franchise value is high, that is, when there are profits associated with the firm as a going concern.

Unfortunately, neither market value nor replacement cost can be measured directly. We approximate the market value of a BHC's assets by adding the market value

One way to quantify franchise value is to look at the difference between a firm's market value and its replacement cost.

of its equity (shares of stock outstanding times price per share) and the book value of its liabilities.[8] When a BHC purchases an asset for more than its book value, the difference between its book value and the purchase price is accounted for on the purchaser's books as goodwill. Because this difference is a component of the purchaser's franchise value, we approximate the replacement cost of a BHC's assets using the book value of its assets *minus* goodwill. Finally, we divide franchise value by assets (net of goodwill) to derive a scale-free measure:

$$\frac{FV}{(A-goodwill)} = \frac{E+L-(A-goodwill)}{(A-goodwill)} \quad ,$$

where E is the market value of equity, L is the book value of liabilities, and A is the book value of assets. Adding 1 and simplifying gives a proxy (Q) for the well-known "Tobin's q":

$$Q = \frac{E+L}{(A-goodwill)} \quad .$$

Following Keeley, we use this ratio to measure franchise value in the empirical analysis that follows.[9]

The Q ratio has the advantage of permitting comparability across BHCs of different sizes. For instance, if the market value of a BHC's assets (measured by E + L) is $520 million and the replacement cost of those assets (measured by A – goodwill) is $500 million, franchise value equals $20 million (4 percent of replacement cost) and Q equals 1.04. For a BHC with franchise value of $20 million and a replacement cost of $1 billion, Q equals 1.02, since franchise value equals only 2 percent of replacement cost. Note that measurement of Q requires information on the market value of the firm. Franchise value may be difficult to measure for firms without publicly traded stock.

In the first part of our analysis, we use risk measures that are also derived from stock market data. We start by calculating an all-in measure of risk, designed to encompass all of the BHC's risk-taking activities, including the riskiness of its assets and liabilities, its choice of off-balance-sheet activities, *and* its chosen capital-to-assets ratio. Our all-in risk measure is based on the variability of BHC stock returns over time. In particular, we calculate the standard deviation of weekly stock returns for a given BHC in a given year. Since stock returns reflect changes in the market's perceptions of future profitability, a high standard deviation in the returns indicates that the expected profits of a BHC are fluctuating rapidly—a sign that the BHC is pursuing risky activities. We discuss additional measures of risk below, including measures that separate the BHC's portfolio risk from its capital-to-assets ratio. Together, portfolio risk and capital determine all-in risk.

DATA DESCRIPTION

Our analysis is based on a sample of more than 100 BHCs with publicly traded stock. In 1993, they ranged in size from $170 million to $231 billion in assets and together held a little less than half of all U.S. banking assets. Our data set spans the 1986-94 period. Because most of the institutions in our sample operated in each of the years included in the sample period, we have 938 BHC-year observations.[10] We obtain the information needed to calculate franchise value and all-in risk from the Center for

Research in Security Prices data tapes and from regulatory reports (Y-9C reports) that contain consolidated financial statements for BHCs. Data used to calculate franchise value are from the beginning of each calendar year in the sample period. This timing helps ensure that any causal relationship runs *from* franchise value *to* BHC risk and not the other way around.

Summary statistics describing franchise value, all-in risk, and the other variables used in our analysis appear in Table 1. As the table indicates, considerable variability in risk taking exists among the BHCs in our sample. The annualized standard deviation of weekly stock returns averages 33 percent in our sample, but ranges from around 10 percent to slightly more than 180 percent. We describe several additional measures of risk below.

Our measure of franchise value averages just over 1.00, at 1.02. On average, the market value of assets for the BHCs in our sample exceeds the book value of assets by 2 percent.[11] The standard deviation of 0.03 reveals some dispersion in franchise values, but most of the BHCs in our sample have franchise values near the average. However, the minimum and maximum statistics tell us that the franchise value distribution is somewhat skewed. BHCs with franchise values below 1.00 tend to bunch up near the average; however, one BHC's market value of assets is more than 20 percent larger than its book value of assets.

Does Franchise Value Negatively Affect Risk?

The results of our first regression, in which all-in risk is regressed on franchise value, BHC size, and personal income growth, confirm that BHCs with the highest franchise value exhibit the lowest all-in risk (Table 2). The coefficient associated with franchise value is negative in both the random-effects and fixed-effects models. Moreover, the coefficient associated with franchise value is statistically significant, that is, we can be confident that franchise value is negatively related to all-in risk.[12]

The results of our first regression . . . confirm that BHCs with the highest franchise value exhibit the lowest all-in risk.

Our estimates indicate that the effects of franchise value on risk are not just statistically reliable but also eco-

Table 1
BANK HOLDING COMPANY SUMMARY STATISTICS

	Mean	Standard Deviation	Minimum	Maximum
All-in risk (annualized standard deviation of weekly stock returns)	0.33	0.19	0.10	1.81
Systematic risk	0.21	0.10	0.02	0.92
Firm-specific risk	0.25	0.17	0.08	1.56
Capital-to-assets ratio	0.06	0.01	-0.03	0.11
Loans-to-assets ratio	0.61	0.11	0.12	0.87
Commercial and industrial loans-to-assets ratio	0.18	0.07	0.005	0.40
Loan portfolio concentration[a]	0.33	0.06	0.25	0.68
Franchise value (market-to-book asset ratio)	1.02	0.03	0.96	1.22
Total assets (billions of dollars)	18.92	30.64	0.17	230.64
Growth in personal income (percent)[b]	2.05	2.03	-7.08	7.97

Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

Notes: Pooled data are from 1986 to 1994. There are a total of 938 observations.

[a] Loan portfolio concentration equals the sum of the squared shares of each loan category (real estate, consumer, commercial and industrial, and other).

[b] Growth in personal income for each BHC is computed as the asset-weighted average of the growth in real personal income for each state in which the BHC has one or more commercial bank subsidiaries.

nomically meaningful. A 1 percentage point increase in franchise value leads to a decrease in all-in risk of about 3.6 percent. This means that, on average, all-in risk at a BHC with high franchise value (equal, for example, to 5 percent of assets) would be about 18 percent lower than at a similar BHC with no franchise value.

Both asset size and personal income growth are negatively related to all-in risk, though only the coefficient on personal income growth in the random-effects specification is statistically significant. Strong economic conditions reduce variability in the stock returns of the BHCs in our sample when we control for the other variables in the regression.

FRANCHISE VALUE AND THE MIX OF BHC RISKS

We now introduce two new measures of BHC risk (Table 3). They are derived by splitting our all-in risk measure into two components: systematic risk, which reflects risks stemming from underlying economic factors that affect the banking industry as a whole (such as interest rate risk), and firm-specific risk, which reflects risks unique to particular banks (such as the industry mix of loans in a commercial and industrial loan portfolio). Systematic risk is derived by measuring the extent to which each BHC's stock return tracks those of a large sample of BHCs. Firm-specific risk is derived from the difference between all-in risk and systematic risk.[13]

Our earlier discussion suggests that BHCs would like to reduce risks across the board as franchise value rises. But it may be harder for BHCs to reduce certain kinds of risks than others. For instance, a BHC that specializes in lending to a particular industry may find it difficult to diversify into new industries. As a result, the firm-specific component of all-in risk may be less sensitive to changes in franchise value than the systematic component.

A second line of reasoning suggests that *systematic*

Table 2
RELATIONSHIP BETWEEN ALL-IN RISK AND FRANCHISE VALUE

	All-In Risk	
	Random-Effects Model	Fixed-Effects Model
Franchise value	-3.566** (0.453)	-2.898** (0.501)
Size	-0.025 (0.016)	-0.063 (0.046)
Growth in personal income[a]	-0.020** (0.007)	-0.010 (0.007)
R-squared	0.328	0.402[b]

Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

Notes: Table presents the coefficients from regressions of the log of all-in risk (annualized standard deviation of weekly stock returns) on franchise value (market-to-book asset ratio), size (log of total assets), and growth in personal income. Regressions include time fixed effects (not shown). Standard errors are in parentheses. Pooled data are from 1986 to 1994. There are a total of 938 observations.

[a] Growth in personal income for each BHC is computed as the asset-weighted average of the growth in real personal income for each state in which the BHC has one or more commercial bank subsidiaries.

[b] The reported R-squared in the fixed-effects model corresponds to a regression in which all variables are calculated as deviations from their BHC means.

* Statistically significant at the 5 percent level.

** Statistically significant at the 1 percent level.

Table 3
RELATIONSHIP BETWEEN RISK COMPONENTS AND FRANCHISE VALUE

	Systematic Risk		Firm-Specific Risk	
	Random-Effects Model	Fixed-Effects Model	Random-Effects Model	Fixed-Effects Model
Franchise value	-3.676** (0.501)	-3.061** (0.580)	-3.445** (0.506)	-2.721** (0.568)
Size	0.070** (0.016)	-0.036 (0.054)	-0.081** (0.017)	-0.074 (0.053)
Growth in personal income[a]	-0.011 (0.008)	-0.0001 (0.009)	-0.032** (0.008)	-0.021* (0.008)
R-squared	0.417	0.473[b]	0.290	0.283[b]

Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

Notes: Table presents the coefficients from regressions of the logs of systematic risk and firm-specific risk on franchise value (market-to-book asset ratio), size (log of total assets), and growth in personal income. Regressions include time fixed effects (not shown). Standard errors are in parentheses. Pooled data are from 1986 to 1994. There are a total of 938 observations.

[a] Growth in personal income for each BHC is computed as the asset-weighted average of the growth in real personal income for each state in which the BHC has one or more commercial bank subsidiaries.

[b] The reported R-squared in the fixed-effects models corresponds to a regression in which all variables are calculated as deviations from their BHC means.

* Statistically significant at the 5 percent level.

** Statistically significant at the 1 percent level.

risk may be less sensitive to changes in franchise value. Franchise value should mitigate risk taking because BHC owners fear that they will lose the value of their franchise through insolvency. However, if a BHC faces severe financial difficulties at the same time as many other BHCs, it may be more likely to receive assistance from the government, since one of the primary goals of the federal safety net is to stabilize the financial system during times of crisis. Consequently, BHCs may have little incentive to lower systematic risk, even when franchise value is high.

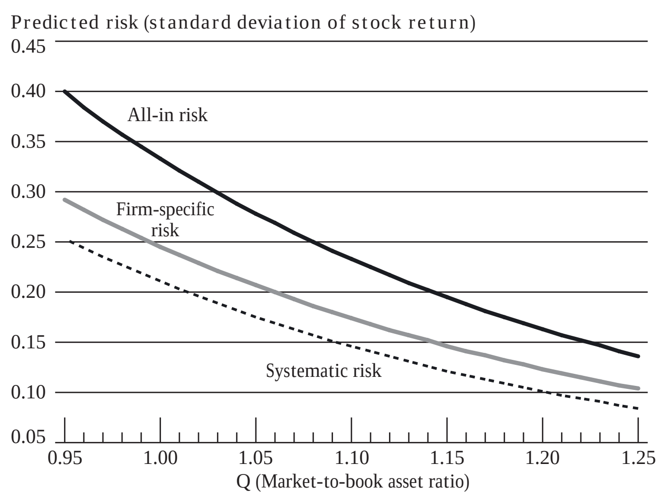
Empirically, we find that franchise value has a similar negative effect on systematic and firm-specific risk (Table 3). A 1 percentage point increase in franchise value leads to a decline of roughly 3 percent in firm-specific and systematic risk. Perhaps BHCs can adjust these two types of risk with similar ease. Alternatively, any difficulties associated with reducing firm-specific risk may simply be counterbalanced by weaker incentives to reduce systematic risk.

The relationship between personal income growth and BHC risk, particularly in the firm-specific risk regression, also proves to be negative (Table 3). Asset size, which was insignificant in Table 2, is negatively related to firm-specific risk but positively related to systematic risk in the random-effects model. This apparent inconsistency can be reconciled by noting that larger BHCs are generally better diversified than smaller ones but have lower capital-to-assets ratios and engage more intensively in certain risky activities. The negative influence of size in the firm-specific risk regression reflects the better diversification of larger BHCs. The positive influence of size in the systematic risk regression reflects differences in the mix of activities pursued by small and large BHCs. The two effects are approximately offsetting, leaving little relationship between BHC size and all-in risk.[14]

Overall, our empirical tests strongly support the hypothesized negative relationship between franchise value and risk. Analyses using measures of risk derived from BHCs' stock returns suggest that BHCs with strong profit potential, and hence with much to lose in the event of insolvency, display lower systematic risk, lower firm-specific risk, and lower overall risk. Using results reported in Tables 2 and 3, we plot the predicted level of risk against franchise value for a typical BHC in Chart 2. All three market-based measures of BHC risk fall as franchise value rises.

Chart 2

Bank Holding Company Risk and Franchise Value



Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

Note: The underlying calculations are based on coefficients from the random-effects models and sample means for the log of total assets (15.96) and growth in personal income (2.05).

How Do High-Franchise-Value Banks Reduce Risk?

Recall our example of FirstRisk Bank, which seeks to reduce its insolvency risk as its franchise value rises. FirstRisk can do so in a variety of ways. It can boost its capital-to-assets ratio by retaining earnings or issuing new equity, or it can reduce portfolio risk by steering clear of risky loans or further diversifying its loan portfolio. We now determine which type of behavior—strengthening capital, reducing portfolio risk, or both—underlies the relationship between franchise value and risk for the BHCs in our sample.

We use two approaches to explore this issue. First, we estimate regressions using the capital-to-assets ratio and three measures of portfolio risk derived from the balance sheet as dependent variables. Independent variables are the same as those used above (franchise value, BHC size, and state-level personal income growth). The results suggest that high-franchise-value BHCs have lower all-in risk because they have stronger capital positions and safer portfolios. However, since no all-encompassing measure of portfolio risk is available from the balance sheet, this approach does not allow us to quantify the effect of franchise value on overall portfolio risk.

In our second approach, we go back to using stock-return variability to measure risk, but we include the logarithm of the capital-to-assets ratio as an additional independent variable. This approach allows us to control for the effect of leverage on stock-return variability when estimating the effect of franchise value. Because capital (like portfolio risk) is chosen by the BHC, we are more comfortable estimating a regression with capital as a *dependent* variable (as in our first approach). Nevertheless, adding capital to the right-hand side of regressions with all-in risk, systematic risk, or firm-specific risk as dependent variables helps us determine whether franchise value has an effect on risk taking above and beyond its effect on capital. In other words, these regressions enable us to obtain an estimate of the effect of franchise value on overall portfolio risk.[15]

[Our] results suggest that high-franchise-value BHCs have lower all-in risk because they have stronger capital positions and safer portfolios.

The results of our first approach are reported in Table 4. We look for evidence that high-franchise-value BHCs reduce risk by: (1) increasing their capital-to-assets ratios, (2) shifting from loans in general to safer assets, (3) shifting from relatively high-risk commercial and industrial loans to less risky loans and other assets, and (4) decreasing loan portfolio concentration. We measure loan portfolio concentration by squaring and summing the shares of the loan portfolio in each of four loan groups: commercial and industrial, real estate, consumer, and other loans. The resulting concentration index ranges from zero to one, taking on higher values for portfolios concentrated in one or two of these four loan groups. Information on loans, assets, and capital was obtained from BHCs' Y-9C reports.[16]

Table 4
RELATIONSHIP BETWEEN BALANCE-SHEET RISK AND FRANCHISE VALUE

	Capital-to-Assets Ratio		Loans-to-Assets Ratio		Commercial and Industrial Loans-to-Assets Ratio		Loan Portfolio Concentration[a]	
	Random-Effects Model	Fixed-Effects Model	Random-Effects Model	Fixed-Effects Model	Random-Effects Model	Fixed-Effects Model	Random-Effects Model	Fixed-Effects Model
Franchise value	1.174**	0.648*	0.040	0.120	0.183	0.394	-0.547**	-0.524**
	(0.234)	(0.258)	(0.188)	(0.193)	(0.385)	(0.397)	(0.163)	(0.169)
Size	-0.051**	-0.079**	0.019	0.025	0.063**	0.033	-0.003	0.076**
	(0.009)	(0.025)	(0.011)	(0.018)	(0.021)	(0.037)	(0.008)	(0.016)
Growth in personal income[b]	-0.008*	-0.010**	-0.003	-0.004	-0.013*	-0.012*	-0.006**	-0.005*
	(0.004)	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)	(0.002)	(0.002)
R-squared	0.282	0.276[c]	0.032	0.134[c]	0.150	0.311[c]	0.148	0.347[c]

Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

Notes: Table presents the coefficients from regressions of the logs of the capital-to-assets ratio, the loans-to-assets ratio, the commercial and industrial loans-to-assets ratio, and loan portfolio concentration on franchise value (market-to-book asset ratio), size (log of total assets), and growth in personal income. Regressions include time fixed effects (not shown). Standard errors are in parentheses. Pooled data are from 1986 to 1994. There are a total of 938 observations. Regressions including the capital-to-assets ratio have 936 observations because capital is negative for two observations.

[a] Loan portfolio concentration equals the sum of the squared shares of each loan category (real estate, consumer, commercial and industrial, and other).

[b] Growth in personal income for each BHC is computed as the asset-weighted average of the growth in real personal income for each state in which the BHC has one or more commercial bank subsidiaries.

[c] The reported R-squared in the fixed-effects models corresponds to a regression in which all variables are calculated as deviations from their BHC means.

* Statistically significant at the 5 percent level.

** Statistically significant at the 1 percent level.

We find strong evidence that BHCs with high franchise value reduce risk by increasing their capital-to-assets ratios and by decreasing loan portfolio concentration. The franchise value coefficient in the capital regression is positive and highly significant, and the coefficient in the loan portfolio concentration regression is negative and highly significant. In contrast, we find no evidence that high-franchise-value BHCs shift from lending in general or from commercial and industrial lending in particular to safer assets, suggesting that it is costly for BHCs to adjust their lending behavior in response to changes in franchise value. It is possible, however, that the regression coefficients underestimate the effect of franchise value on lending. Since franchise value stems in part from lending relationships, BHCs that devote a greater share of their assets to lending may have higher franchise value, all else equal. This effect may counteract any negative influence of franchise value on the loans-to-assets ratio.

The results of our second approach are reported in Table 5. All-in risk, systematic risk, and firm-specific risk are each regressed on franchise value, BHC size, personal income growth, *and* the logarithm of the capital-to-assets ratio, included to control for the effect of leverage. The franchise value coefficients in Table 5 are smaller than those in Tables 2 and 3, but in all three regressions we continue to find a negative and significant coefficient on franchise value. Together, Tables 4 and 5 show that BHCs with higher franchise value have lower risk because they have stronger capital positions and safer portfolios.

Conclusion

We have argued that franchise value helps offset the incentive for firms to increase risk because firms with the ability to generate profits will act to protect their valuable franchise. The discipline introduced by franchise value is particularly important in the banking industry, where the federal safety net insulates banks from costs normally borne by risky firms.

Our empirical results support the theory that banks that are more efficient, are located in less competitive markets, or have valuable lending relationships operate more safely. We find that high-franchise-value banks hold more capital and take on less portfolio risk, leading to lower levels of overall risk. We also observe a negative rela-

Table 5
RELATIONSHIP BETWEEN RISK AND FRANCHISE VALUE CONTROLLING FOR CAPITAL

	All-In Risk		Systematic Risk		Firm-Specific Risk	
	Random-Effects Model	Fixed-Effects Model	Random-Effects Model	Fixed-Effects Model	Random-Effects Model	Fixed-Effects Model
Franchise value	-2.903**	-2.709**	-3.177**	-2.944**	-2.573**	-2.480**
	(0.417)	(0.471)	(0.493)	(0.575)	(0.462)	(0.533)
Size	-0.052**	-0.061	0.053**	-0.031	-0.116**	-0.076
	(0.014)	(0.046)	(0.015)	(0.056)	(0.014)	(0.052)
Growth in personal income[a]	-0.029**	-0.018**	-0.016*	-0.005	-0.041**	-0.030**
	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)
Capital-to-assets ratio	-0.721**	-0.656**	-0.469**	-0.425**	-0.835**	-0.766**
	(0.058)	(0.065)	(0.069)	(0.079)	(0.064)	(0.073)
R-squared	0.466	0.480[b]	0.468	0.495[b]	0.436	0.378[b]

Source: Authors' calculations, based on data from the Center for Research in Security Prices and consolidated financial statements of a sample of publicly traded BHCs.

Notes: Table presents the coefficients from regressions of the logs of all-in risk, systematic risk, and firm-specific risk on franchise value (market-to-book asset ratio), size (log of total assets), growth in personal income, and log of the capital-to-assets ratio. Regressions include time fixed effects (not shown). Standard errors are in parentheses. Pooled data are from 1986 to 1994. There are a total of 936 observations.

[a] Growth in personal income for each BHC is computed as the asset-weighted average of the growth in real personal income for each state in which the BHC has one or more commercial bank subsidiaries.

[b] The reported R-squared in the fixed-effects models corresponds to a regression in which all variables are calculated as deviations from their BHC means.

* Statistically significant at the 5 percent level.

** Statistically significant at the 1 percent level.

tionship between franchise value and systematic risk (the risk related to factors that affect the banking industry as a whole) and between franchise value and firm-specific risk (the risk unique to individual institutions).

Our results do not suggest a specific supervisory approach; however, they highlight the importance of continued monitoring of franchise value in the banking industry. When franchise value is high, banks are less inclined to take excessive risk, reducing the potential for conflicts between banks and their supervisors. This behavior holds even during periods of economic distress, when capital may be low. In contrast, the interests of banks may conflict with those of supervisors when franchise value is low, especially during periods of economic distress. As the thrift crisis of the 1980s demonstrated, institutions with low capital and low franchise value may have a strong incentive to increase risk and “go for broke.”

ENDNOTES

1. Some analysts apply the term "charter value" to these market-related sources of franchise value in banking. The term reflects the fact that investors would be willing to pay a significant amount for the right to open a bank in markets protected from competition.

2. See Edwards and Mishkin (1995) for a more complete discussion of changes in banking since the mid-1970s.

3. Marcus (1984), Keeley (1990), and Acharya (1996) show formally how franchise value can mitigate moral hazard problems in banking.

4. We do not consider agency problems that may cause the incentives of managers and owners to diverge. Demsetz, Saidenberg, and Strahan (1996) show that the effect of franchise value on risk remains important even after controlling for managers' ownership share.

5. Acharya (1996) shows that closing an insolvent bank with large amounts of franchise value can be less than optimal since much of its franchise value would be lost.

6. Our regressions are similar in structure to those estimated by Keeley, though our measures of BHC risk differ from those that Keeley analyzes, and he works with data from an earlier period. He uses two measures of risk: the interest rate paid on large certificates of deposit and the market-value capital-to-assets ratio. Keeley also estimates a somewhat more complicated system of equations than the one estimated here.

7. Since many of the BHCs in our sample operate in more than one state, we measure growth in personal income using the asset-weighted average of real personal income growth across states in which a BHC has one or more commercial bank subsidiaries.

8. Since we do not observe the market value of liabilities, our measure of franchise value will include the subsidy associated with deposit insurance, which *increases* with risk taking. Since we seek evidence of an *inverse* relationship between franchise value and risk taking, this complication makes it more difficult for us to find empirical support for the hypothesis we test.

9. An alternative measure of franchise value is the market-to-book *equity* ratio, which is highly correlated with the market-to-book *asset* ratio used here. Our empirical results are not sensitive to the use of this alternative measure of franchise value.

10. The BHCs in our data set were identified using the Bank Compustat data base. We worked with only those BHCs for which we could retrieve both stock-return data and data from regulatory reports and whose stock traded for at least thirty weeks in a given calendar year. BHCs acquired in the middle of our sample period were dropped from the sample after the date of acquisition. BHCs that acquired other firms during the sample period remained in the sample. The results presented below are qualitatively similar when we limit our analysis to BHCs that operated throughout the 1986-94 period.

11. The average value of 1.02 is statistically significantly different from 1.00.

12. Studies show that banks in protected markets operate more safely (Rhoades and Rutz 1982) and less efficiently (Berger and Hannan 1994). Initially, this behavior may induce a negative relationship between franchise value and risk since diminished competition can enhance franchise value. Over time, however, inefficient behavior will lead franchise value to decline. We tried controlling for market concentration when empirically examining the relationship between franchise value and risk and found that market concentration could not explain the negative franchise value/risk relationship that we observe.

13. We estimate a five-factor return-generating model using factor analysis, which solves for the five vectors and weights that best explain the component of returns common to the BHCs in our sample. Systematic risk is the square root of the portion of total return variance that can be explained by these five factors. Firm-specific risk is the square root of the difference between total return variance and the square of systematic risk. See Demsetz and Strahan (1995) for additional details on the construction of firm-specific and systematic risk in this sample.

14. See Demsetz and Strahan (forthcoming) for further analysis of the relationship between diversification, size, and risk at BHCs.

15. Another way to measure the relationship between franchise value and portfolio risk is to remove the effect of leverage from stock-return variability and use the resulting "deleveraged" risk measure as a dependent variable. We tried making this adjustment by multiplying all-in risk, systematic risk, and firm-specific risk by the capital-to-assets ratio and taking the log of each product. Using the resulting risk measures as dependent variables, we continued to find negative and significant coefficients on franchise value, suggesting a negative relationship between franchise value and portfolio risk. In fact, these coefficients were similar in magnitude to those reported in Table 5.

16. Loans, assets, and capital are measured at the same point in time as franchise value.

REFERENCES

Acharya, Sankarshan. 1996. "Charter Value, Minimum Capital Requirement and Deposit Insurance Pricing in Equilibrium." JOURNAL OF BANKING AND FINANCE 20: 351-75.

Berger, Allen, and Timothy Hannan. 1994. "The Efficiency Cost of Market Power in the Banking Industry: A Test of the Quiet Life and Related Hypotheses." Board of Governors of the Federal Reserve System Finance and Economics Discussion Series, no. 94-36.

Berger, Allen, and Gregory Udell. 1995. "Relationship Lending and Lines of Credit in Small Firm Finance." JOURNAL OF BUSINESS 68, no. 3: 351-82.

Demsetz, Rebecca S., Marc R. Saidenberg, and Philip E. Strahan. 1996. "Franchise Value, Ownership Structure, and Risk-taking at Banks: The Role of Nonregulatory Forces." Federal Reserve Bank of New York, unpublished paper.

Demsetz, Rebecca S., and Philip E. Strahan. 1995. "Historical Patterns and Recent Changes in the Relationship between Bank Holding Company Size and Risk." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 1, no. 2: 13-26.

———. Forthcoming. "Diversification, Size, and Risk at Bank Holding Companies." JOURNAL OF MONEY, CREDIT AND BANKING.

Edwards, Franklin, and Frederic S. Mishkin. 1995. "The Decline of Traditional Banking: Implications for Financial Stability and Regulatory Policy." Federal Reserve Bank of New York ECONOMIC POLICY REVIEW 1, no. 2: 27-45.

Jayaratne, Jith, and Philip E. Strahan. 1996. "Entry Restrictions, Industry Evolution and Dynamic Efficiency: Evidence from Commercial Banking." Federal Reserve Bank of New York Research Paper no. 9630.

Keeley, Michael. 1990. "Deposit Insurance, Risk, and Market Power in Banking." AMERICAN ECONOMIC REVIEW 80, no. 5: 1183-200.

Marcus, Alan. 1984. "Deregulation and Bank Financial Policy." JOURNAL OF BANKING AND FINANCE 8: 557-65.

Petersen, Mitchell, and Raghuran Rajan. 1995. "The Benefits of Lending Relationships: Evidence from Small Business Data." JOURNAL OF FINANCE 49, no. 1: 3-37.

Rhoades, Stephen A., and Roger D. Rutz. 1982. "Market Power and Firm Risk: A Test of the 'Quiet Life' Hypothesis." JOURNAL OF MONETARY ECONOMICS 9: 73-85.

What Do Chain Store Sales Tell Us about Consumer Spending?

by Ethan S. Harris and Clara Vega

In the last several years, reports from major retail chains have been closely watched by journalists, forecasters, and financial market participants. Interest peaked during the 1995 Christmas season, when chain store reports showing weak sales fueled growing concern about the consumer sector. Under headlines such as "Retailers Call Sales in December Worst since '90-'91 Recession," news coverage of the reports moved from the business page to the front page.[1] This attention raises an important question: While chain store reports are clearly an important measure of the health of large retail companies, are they also useful in assessing and forecasting consumer spending as a whole?

This study is the first comprehensive examination of the value of chain store data as macroeconomic indicators.[2] We begin by considering important structural changes in the retail sector and their implications for interpreting the chain store data. We then turn to formal statistical tests of the linkages between chain store data and the official measures of overall retail sales and personal consumption expenditure.

Our empirical tests provide mixed support for the use of chain store data. On the one hand, we find that weekly indexes and monthly reports from individual companies are too erratic to be useful for forecasting. On the other hand, we find that monthly chain store indexes, if given the appropriate weights in forecast models, add significantly to the accuracy of in-sample and out-of-sample predictions for several measures of consumer spending. Overall, models that combine economic variables with the two major chain store indexes provide the best forecasts.

WHAT ARE CHAIN STORES?

In press reports, the term "chain store" is used more or less interchangeably with "department store," "retail chain," "broadline," and "major retailer." To clarify how this term is generally understood, we relate it to specific categories in the U.S. Department of Commerce taxonomy of retail establishments (Table 1).[3] All chain stores could be placed in the broad Commerce Department category of *general merchandise, apparel, and furniture* (GAF). Within this category, chain stores encompass virtually all *department stores*, including *national chain department stores* such as Sears and J.C. Penney, *conventional department stores* such as Federated/Macy and May, and *discount department stores* such as Wal-Mart and Kmart. Note that the term "chain store" applies to all major department stores, even those that have a limited number of locations.

Establishments classified as department stores by the Commerce Department employ, on average, more than 150 workers and carry a diverse range of merchandise—household linens, dry goods, home furnishings, appliances, radios and televisions, furniture, and a general line of apparel. Annual sales at the typical department store run close to $17 million, more than ten times the sales of the average retail establishment. Consequently, while department stores make up less than 1 percent of all retail establishments, they account for about 10 percent of retail sales.

Not all chain stores are department stores; some fall into other subcategories of GAF—*apparel*, *furniture*, *miscellaneous shopping goods*, and *other general merchandise.* Chain stores in these categories share two features: they are large retail companies with a national chain of outlets, and they specialize in one or more of the same lines of merchandise as department stores. Examples of companies in this group are The Limited, which sells apparel, and Bed, Bath, and Beyond, which sells household linens and home furnishings.

Not included in the definition of chain stores are the smaller, local stores that make up the bulk of GAF establishments. Sales and employment at these stores are much more modest than at the chain stores: a local store selling furniture or apparel would, on average, employ less than ten workers and post annual sales of less than $1 million.

Table 1
RETAIL ESTABLISHMENTS IN 1992

Commerce Department Category	Number of Establishments (Thousands)	Average Sales per Establishment (Thousands of Dollars)	Average Employees per Establishment
Total retail	1,526.2	1,242	12
GAF	463.1	2,026	19
General merchandise	34.6	7,089	60
Department stores	11.0	16,946	156
National chain	1.9	18,873	179
Conventional	2.4	20,832	203
Discount	6.7	15,032	134
Other	23.6	2,496	15
Apparel	145.5	699	8
Furniture	110.1	847	6
Miscellaneous shopping goods	127.3	520	6
Other retail	1,063.1	900	9

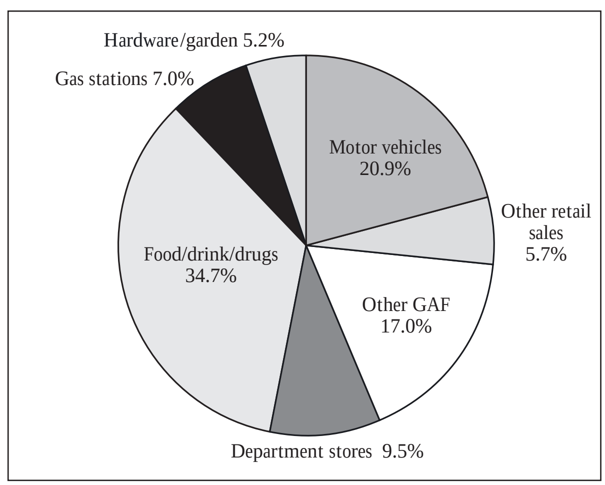
Source: Department of Commerce, Bureau of the Census (1995).

Notes: The Commerce Department defines an establishment as a "single physical location at which business is conducted." The last column of the table reports the average number of employees per establishment for the week of March 12, 1992. Over the course of the year, each establishment temporarily employs many more workers.

THE LINK BETWEEN CHAIN STORE SALES AND OVERALL CONSUMER SPENDING

Despite the attention they garner in the business press, chain store sales represent a relatively small portion of overall consumer spending (Chart 1). We noted earlier that department stores account for about 10 percent of retail sales. Even if we generously include all of GAF in our estimate of chain store sales, these stores claim only about one-

Chart 1
Retail Sales Shares in 1992



Source: Department of Commerce, Bureau of the Census (1995).

Notes: Some of the groupings in the chart combine several Commerce Department categories. "Motor vehicles" refers to sales of automotive dealers and includes sales of light trucks. "Food/drink/drugs" includes grocery stores, eating and drinking establishments, drug and proprietary stores, and liquor stores. "Hardware/garden" includes building materials, hardware, garden supplies, and sales of mobile home dealers.

fourth of retail sales. The remainder of the retail sector includes motor vehicle dealers, hardware and garden stores, gasoline service stations, grocery stores, restaurants, liquor stores, bars, and pharmacies. Furthermore, most of personal consumption expenditure is for services, with goods purchases making up just 42.9 percent of the total in 1992. Thus, allowing for some minor accounting adjustments, we calculate that chain stores represent, directly and indirectly, only 4 to 11 percent of personal consumption.[4]

Despite the attention they garner in the business press, chain store sales represent a relatively small portion of overall consumer spending.

Two Chain Store Indexes

Although a number of economists have created chain store indexes in recent years, the two longest running and most watched indexes are the Chain Store Index from the Bank of Tokyo–Mitsubishi[5] and the Retail Sales Index from the Johnson Redbook Service. Because of the proprietary nature of the indexes, only limited information is available on their construction. We provide some basic facts about the indexes here and a fuller account of what is known about them in Appendix 1.

The indexes differ in two respects. First, while the Johnson Redbook index focuses only on companies that fit the Commerce Department definition of department stores, the Mitsubishi index also includes stores that fit the broader GAF category. Second, while Johnson Redbook measures total company sales, Mitsubishi includes only "same-store" sales—that is, sales from locations that have been open for at least a year.

Both indexes are released at weekly and monthly intervals, just a few days after the period they measure. The weekly indexes provide real-time updates on the progress of spending during the month; the complete monthly data offer a summary look at monthly sales more than a week before the Commerce Department's advance estimate of retail sales. As the calendar of official release dates indicates (Table 2), the only other direct monthly measure of consumer spending available that early in the data cycle is auto and light truck sales, and these data tend to have monthly patterns very different from those of the rest of retail sales.[6]

Although the indexes provide the most timely data on the consumer sector, their early release comes at a cost: they are constructed with considerably less rigor than the official retail sales data issued by the Commerce Department. (These differences are detailed in Appendix 1.) The official data are drawn from a broad stratified sample of large and small companies; the chain store indexes, by contrast, are based on a small sample of only large companies. Irregularities in adjusting the data for seasonal variations may introduce distortions in the chain store data that are not present in the official measures. These small sample and seasonal adjustment problems are particularly evident in the weekly versions of these indexes. In addition, while the official data are frequently and heavily revised, the chain store data are essentially never revised. This reliance on a onetime sampling makes the chain store data easier to follow, but it also means that errors are never corrected.

Table 2
HOW CHAIN STORE REPORTS GET A JUMP ON THE COMPETITION: RELEASE DATES FOR AUGUST 1996 DATA

Date	Release
August 13, 20, 27; September 3-4	Johnson Redbook and Mitsubishi indexes (weekly)
August 15, 22, 29; September 5	Initial claims (weekly)
August 27	Consumer confidence indexes
September 3-5	Retail company reports Johnson Redbook index (monthly) Mitsubishi index (monthly)
September 5	Initial claims (monthly) Auto and light truck sales
September 6	Payroll employment
September 13	Advance retail sales
September 30	Personal consumption expenditures
October 11	Preliminary retail sales
November 14	Final retail sales

Notes: These releases are issued by the following agencies: initial claims, Department of Labor; payroll employment, Department of Labor, Bureau of Labor Statistics; advance, preliminary, and final retail sales, Department of Commerce, Bureau of the Census; personal consumption expenditures and auto and light truck sales, Department of Commerce, Bureau of Economic Analysis. The consumer confidence indexes are issued by the University of Michigan and the Conference Board.

Given the limits of their construction and their narrow company coverage, the chain store indexes should not be treated as representative samples of consumer spending as a whole. But does this mean that the indexes are of little use in forecasting consumer spending? To answer this question, we carry out formal tests of the statistical link between chain store sales and overall consumer spending. First, however, we consider recent structural changes in the retail sector that affect both the interpretation of the chain store data and the statistical models we devise to measure the data's predictive power.

Although the [chain store] indexes provide the most timely data on the consumer sector, their early release comes at a cost: they are constructed with considerably less rigor than the official retail sales data issued by the Commerce Department.

An Industry in Transition

Three interrelated structural forces are transforming retailing—the chronic excess supply of retail space, the emergence of value-conscious consumers, and the growing concentration of sales in larger companies.

Excess Capacity

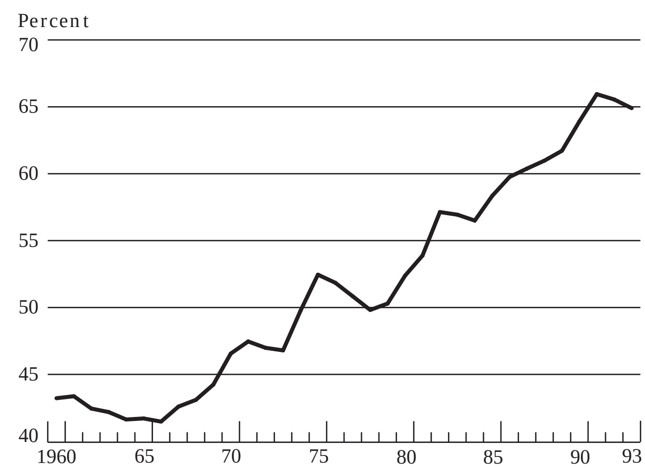
Spurred by easy lending terms and generous tax laws, commercial construction boomed in the early 1980s, with real spending roughly doubling from 1983 to 1986.[7] This favorable investment climate changed in the late 1980s, and by 1992 commercial construction had dropped below its 1983 levels. Since then, however, while the office building component of commercial construction has continued to convalesce slowly, retail and wholesale construction has recovered quickly and now stands near its earlier peak. This new surge in construction appears to be causing a rapid increase in retail capacity. Capital stock data from the Bureau of Economic Analysis show a continued rise in the stock of retail structures relative to GDP (Chart 2). Statistics reported in the industry literature provide further documentation of this trend: for example, from 1972 to 1994, the number of shopping centers in the United States tripled to 40,300, and the number of square feet of shopping center space per capita grew from 7.0 to 18.7 (Telsey 1996, p. 28).

While some of this space may lie vacant and some of the increase in capacity reflects a natural process of capital deepening as the economy grows, there are also telltale signs of excess capacity:

- The stock market performance of major retailers has suffered. Over the long run, the stocks of major retailers have generally matched the overall stock market; from March 1994 to March 1996, however, the average stock price of retail firms in the Standard and Poor's 500 index fell 23 percentage points relative to the overall index.
- Financial pressures have led to an increase in bankruptcies and store closings. Although bankruptcy rates are not very high for the retail sector as a whole, large general merchandise stores have experienced an unusually high rate of failure. According to data from Dun and Bradstreet, despite the business cycle expan-

Chart 2

Retail Structures as a Share of GDP



Source: Department of Commerce, Bureau of Economic Analysis.

Note: The chart shows the ratio of the stock of retail structures to real GDP measured in real 1987 dollars.

sion, the liabilities associated with bankruptcies in this sector climbed steadily from $0.6 billion in 1992 to almost $2 billion in 1995.

Apparently bankruptcies and individual store closings have not solved the oversupply of space; commentaries in the industry press suggest that shuttered stores have generally reopened under new names.

Value-Conscious Consumers

Not only do retailers face more competitors, they must also sell to increasingly price-conscious consumers. Pricing behavior in the GAF sector in the 1990s recalls that in the auto industry a decade earlier, when discounts introduced as a temporary device for reducing inventories became almost permanent. In the chain store sector, retailers have accommodated their more value-conscious customers by holding regular sales. Since customers have responded by deferring spending until items go on sale, retailers have been compelled to increase the frequency of the sales.

Three interrelated structural forces are transforming retailing—the chronic excess supply of retail space, the emergence of value-conscious consumers, and the growing concentration of sales in larger companies.

Consumers' search for value has had a number of important effects. Spending has steadily shifted away from conventional department stores to discount department stores. From 1988 to 1995, sales at discounters rose an average of 8 percentage points faster than sales at other department stores, driving up the discounters' share in total sales from 44 to 60 percent.

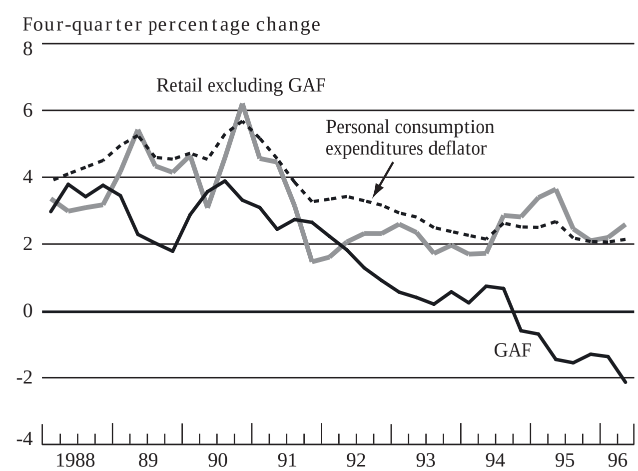
Together with the oversupply of stores, this shift in demand has also put downward pressure on prices at major retail firms. The inflation rate for goods sold at GAF stores has been consistently lower than broad measures of consumer prices such as the personal consumption deflator and has generally trailed the deflator for other retail sales as well (Chart 3). Indeed, this weak price performance has recently worsened dramatically: GAF store prices have actually fallen sharply since early 1994, widening the inflation gap to 4 percentage points.[8]

A final effect of value shopping has been a shift in the seasonal pattern of department store sales. Chain store sales are much more seasonal than sales in other retail sectors. According to the latest official seasonal adjustment factors, department store sales typically surge 78 percent above their long-run average in December, then plunge to 27 percent below average in January. By contrast, the sales of other non-auto retailers—including grocery stores, restaurants, gas stations, and hardware stores—exhibit milder seasonal patterns, rising just 25 percent above normal in December and dipping about 11 percent below normal in January.

Over the last several years, value-conscious shoppers have induced a substantial shift in the holiday seasonal pattern, delaying purchases in December to take advantage of lower prices in January. In particular, a comparison of the last five years (1991-95) with the previous five years (1987-91) shows that the December peak in department store sales has dropped from 85 percent above average to

Chart 3

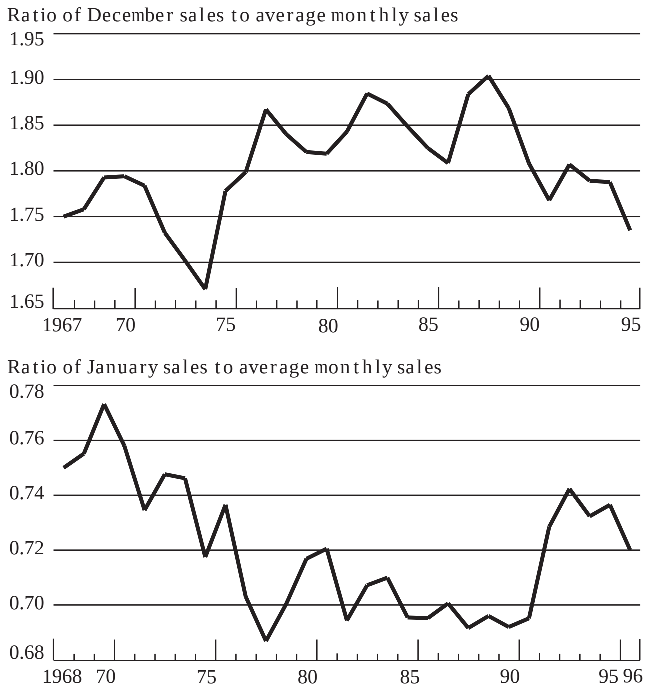
Consumer Inflation Trends



Source: Authors' calculations, based on data from the Department of Commerce, Bureau of Economic Analysis.

Chart 4

The Changing Christmas Season for Department Stores



Source: Authors' calculations, based on data from the Department of Commerce, Bureau of the Census.

Note: The average monthly sales are calculated as a twelve-month centered moving average.

just 78 percent above average (Chart 4). A large portion of these sales have shifted to January: sales for this month were 31 percent below average in 1987-91 but only 27 percent below average in 1991-96.

CONSOLIDATION

Larger retail companies are growing at the expense of their smaller counterparts. This shift is impossible to quantify precisely, but it can be illustrated by comparing sales growth for firms included in the chain store indexes—which are all large firms—to sales for the GAF sector as a whole—which includes small and large firms. For example, for the five years ending December 1995, an index of total chain store sales issued by Merrill Lynch (1996)[9] grew at an 11.8 percent annual rate, almost double the 6.9 percent pace for GAF. This relatively rapid growth stems entirely from acquisitions and new store construction: over the same period, the same-store sales in the Merrill sample actually grew more slowly than sales in the GAF sector as a whole, averaging a 4.5 percent annual rate.[10]

The pace of change in the retail sector shows little sign of abating. Two recent industry trends should ensure that the process of restructuring and concentration will continue. First, a new type of store with the colorful name "category killers" has emerged. These "big box" stores offer a full product line in a focused category of goods. Second, "super stores," which combine a traditional discount store with a supermarket and a variety of smaller stores under one roof, are gaining popularity.

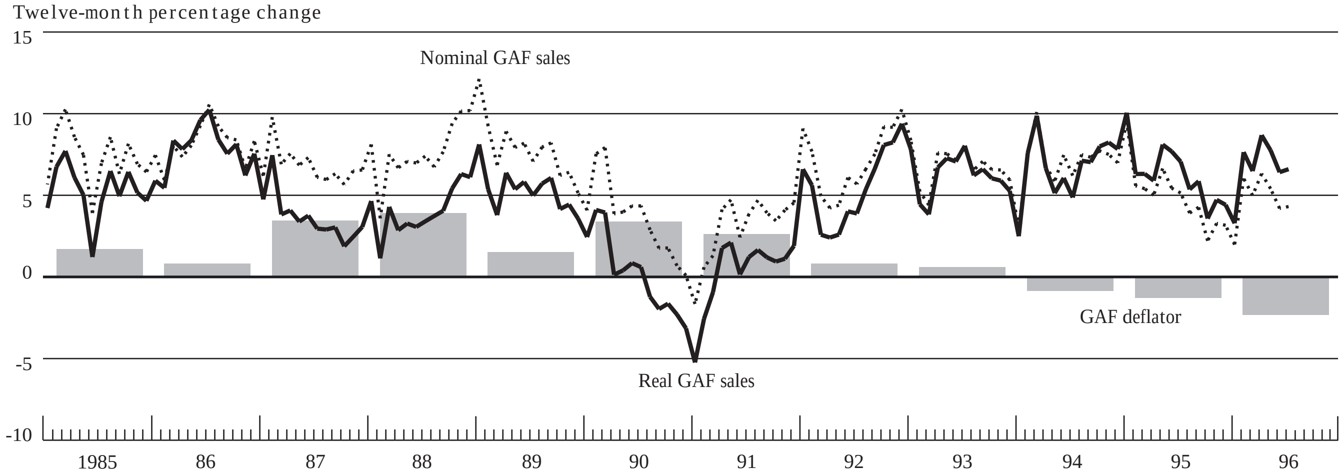
IMPLICATIONS FOR FORMAL FORECASTS AND INFORMAL COMMENTARY

Structural changes in the chain store business have made it more difficult to disentangle two kinds of information in the data: the microeconomic information on the health of individual companies and the macroeconomic information on underlying consumer demand. Retail analysts examine the recent data, see companies under competitive pressure, and infer that consumer spending is ailing. The macroeconomist's job, however, is to factor out structural distortions and assess underlying trends in consumer demand. From this perspective, the retail analysts' interpretation of chain store data has been unduly negative.

To understand how the data can be misread, consider the commentary on the recent Christmas selling seasons. In the GAF sector, the period from Thanksgiving to the end of December is vital to company profits and is often viewed as a bellwether for the year ahead. Over the past three years, despite trend growth in real, inflation-adjusted retail sales of more than 5 percent, retail analysts have repeatedly reported "disappointing" Christmas sales. The gap between Christmas commentary and macroeconomic reality reflects three structural distortions. First, analysts often focus on same-store sales as a measure of underlying demand, but rapid growth in new stores has tended to depress same-store sales, making them less representative of demand. Second, while the official retail sales data now appear to have adjusted to the sharp decline in the December seasonal increase, retail analysts continue to report "below plan" December sales. Third, because of declining prices, nominal GAF sales growth has been

Chart 5

Sales Growth in General Merchandise, Apparel, and Furniture



Sources: Department of Commerce, Bureau of the Census; Department of Commerce, Bureau of Economic Analysis.

Note: The GAF deflator is measured as a percentage change from December to December, except for the 1996 value, which is calculated from July to July.

deceptively weak. As Chart 5 shows, nominal GAF sales grew just 4 percent in the year to December 1995, barely outpacing overall consumer price inflation. With GAF prices falling at a 2 percent annual rate, however, the seemingly anemic nominal growth translates into a robust 6 percent real gain.

These structural changes can also affect the relationship between chain store sales and overall retail sales in formal statistical models. Changes in the seasonal patterns of the data, relative price shifts, and changing patterns of sales among new and old firms and firms of different sizes can have different impacts on different measures of sales. Consequently, as we will demonstrate below, a change in a chain store index in the recent period may no longer be associated with the same magnitude of change in official retail sales.

Testing the Predictive Power of Chain Store Data

We've seen that chain store indexes display several drawbacks as macroeconomic indicators. Nonetheless, they have at least one clear-cut advantage: their early release. Thus, whether chain store sales are useful for forecasting essentially comes down to the following: Is an imperfect but timely sample better than no sample at all?

To test the predictive power of the two chain store indexes, we put them through a rigorous battery of tests. We test their ability to predict a wide range of consumer spending measures, we compare their performance to structural and time series models, and we evaluate their performance both in sample and out of sample. This exercise not only clarifies the role of chain store indexes in consumption forecasting, but also highlights other variables that are useful for forecasting.

Dependent Variables

We test the power of chain store sales to predict four nominal consumption variables of interest to forecasters: GAF sales; advance non-auto retail sales;[11] the latest, fully revised non-auto retail sales; and the latest, fully revised personal consumption expenditure. The first variable roughly matches the coverage of the chain store indexes, the second is what financial sector economists are most interested in tracking, the third presumably measures the "true" trends in the overall retail sector, and the fourth is the data incorporated in the GDP accounts.

Because of our focus on short-term forecasting, most of our variables enter our models as simple monthly percentage changes. In adopting this convention, we reject two alternatives. We reject the business press practice of focusing on year-over-year percentage changes in the chain store indexes, because the year-over-year figures convey little information to forecasters (after all, the only new information in a twelve-month change is the change for the latest month). We also choose not to use weekly data. These data have no official equivalent and, as the appendixes show, the quality of information in the chain store indexes falls off precipitously when we move from the monthly to the weekly frequency.

INFORMATION SET

We compare the information in the chain store indexes with the information embodied in lags of consumer spending as well as a number of consumer-related indicators that are released before the advance retail sales report. These include:

- the only other timely consumption indicator (growth in auto and light truck sales),
- a measure of the consumer demand for home furnishings (growth in home sales, lagged one month because the data are not immediately available),
- income-type variables (payroll employment growth and initial claims for unemployment insurance),
- measures of consumer confidence (both the Michigan and Conference Board indexes),
- two measures of the stock market—the growth in the Standard and Poor's 500 index (an indicator of household wealth) and an index of retail stocks in the Standard and Poor's 500 index (a measure of investor confidence in the industry),
- two measures of price impact (the percentage change in gasoline and food prices),
- several interest rate spread variables that have proved to be useful in short-term forecasting (the difference between Treasury and commercial paper rates, the spread between corporate BAA bonds and ten-year Treasuries, and the difference between ten-year and three-month Treasuries),[12] and
- lags on the dependent variable. (To keep this exercise manageable, we consider only three lags—the first, the second, and, to capture any left over seasonality, the twelfth.)

Unless indicated otherwise, each of these variables enters our regressions contemporaneously. We initially tested lags of all variables but found that they did not add to the explanatory power of the models and did not affect the significance of the chain store indexes.

MODELS TESTED

We test six "stand-alone" models, each of which uses a different part of our information set. The *autoregressive integrated moving average* (*ARIMA*) *model* includes only autoregressive and moving average terms that add significant explanatory power. This model provides a pure time series alternative to the chain store data. The *structural model* includes every consumer-related variable that adds significant explanatory power (based on the Akaike information criteria) and whose coefficient has the economically expected sign. In addition, we estimate a *Mitsubishi model*, a *Johnson Redbook model*, and a *two-index model* that includes both chain store indexes along with a constant term. Finally, we test the simplest *"back-of-the-envelope" model*: the average of the monthly percentage changes in the two chain store indexes. This model, which assumes that the indexes are representative samples of overall consumer spending, does not require regression estimation.

In addition to these six stand-alone models, we test several combination models that integrate the chain store data with the ARIMA and structural models.[13] We also conduct a variety of tests for structural shifts in the relationships between chain store sales and overall consumer spending.

If the chain store data are useful for tracking consumer spending, we would expect them to explain a relatively large portion of the monthly growth in official measures of consumer spending and to retain explanatory power when they are used in conjunction with the ARIMA and structural models.

EXPLAINING HISTORY: IN-SAMPLE TESTS

Using ordinary least squares regressions, we estimate each stand-alone model over the period from January 1985 to December 1995.[14] Table 3 reports the R-squared for each model—that is, the proportion of the month-to-month variation in the dependent variable that is explained by the model. The results underscore how difficult it is to forecast month-to-month changes in consumer spending. At best, the models explain less than a third of the variation in retail sales growth and about two-thirds of the variation in personal consumption expenditure. The models have particular difficulty explaining the erratic advance data for non-auto retail sales.

Although none of the stand-alone models perform particularly well, the results for the index models are encouraging. To be sure, one cannot take the chain store data at face value: the calculated R-squared for the back-of-the-envelope model is actually negative, suggesting that one would be better off completely ignoring the chain store data than using this simple approach.[15] Nevertheless, if we use regression estimation to eliminate the excess volatility in the chain store data, they can be useful in predicting overall retail sales. For three of the four consumption variables—GAF sales, advance retail sales, and fully revised retail sales—the two-index models generally perform as well as the ARIMA and structural models. Additional results in Harris and Vega (1996) show that these findings are robust to a number of other specifications.

The stand-alone tests suggest that the chain store indexes contain some useful information, but is this information unique? In other words, do the chain store indexes add new information not captured in the other models? To answer this question, we test whether incorporating either the Johnson Redbook or the Mitsubishi data in the ARIMA and structural models affects the models' explanatory power (Table 4). For each of the combination models created, we report the overall explanatory power as well as the coefficient and t-statistic of the chain store indexes. The results further support the usefulness of the indexes: adding the chain store data improves the overall fit, and both the Johnson Redbook and the Mitsubishi indexes continue to have significant explanatory power in most

Adding the Mitsubishi index to the structural model of GAF sales almost doubles the model's explanatory power, from 24 percent to 40 percent.

equations. For example, adding the Mitsubishi index to the structural model of GAF sales almost doubles the model's explanatory power, from 24 percent to 40 percent (compare Tables 3 and 4). The weakest chain store results are those for the personal consumption expenditure equations. In these equations, the economic variables by themselves do a good job of explaining sales growth, and while the chain store indexes always have the right sign, they are statistically significant only half the time.

We also test three-way combination models, created by adding both chain store indexes to the ARIMA and structural models (Table 4). Both chain store indexes generally finish "in the money," with statistically significant coefficients. Note that Johnson Redbook trails Mitsubishi in both the magnitude of the coefficient and its statistical significance. The smaller coefficient is consistent with the fact that Johnson Redbook measures total store sales while Mitsubishi measures same-store sales. The lower statistical significance is consistent with the fact that Johnson Redbook is subject to greater measurement error because of its smaller sample and less sophisticated seasonal adjustment methodology (see Appendix 1).

In general, the results of our empirical tests indicate

Table 3
IN-SAMPLE EXPLANATORY POWER OF STAND-ALONE MODELS

Models	GAF Sales	Non-Auto Retail Sales: Advance Data	Non-Auto Retail Sales: Latest Data	Personal Consumption Expenditures
ARIMA	0.304**	0.159**	0.316**	0.163**
Structural	0.237**	0.101**	0.250**	0.664**
Mitsubishi	0.223**	0.070**	0.161**	0.011
Johnson Redbook	0.142**	0.112**	0.121**	0.025*
Two-index	0.303**	0.151**	0.234**	0.030
Back-of-the-envelope[a]	-0.334			

Source: Authors' calculations. Details on the explanatory variables included in each model are available from the authors.

Notes: The table reports the R-squared. In each case, the sample period is January 1985 to December 1995.

[a] For this model, the R-squared is calculated as one minus the ratio of the variance of the forecast error to the variance of the dependent variable.

* Explanatory variables are jointly significant at the 5 percent level.

** Explanatory variables are jointly significant at the 1 percent level.

that the models with the best in-sample fit combine the economic variables with both chain store indexes. For example, our recommended model for non-auto retail sales includes the first and twelfth lag on the dependent variable and the growth in auto sales, payroll employment, gasoline prices, and both chain store indexes (Table 5). All variables in this equation are statistically significant (although the lagged dependent variable is only marginally significant), and all coefficients have the correct sign. The chain store indexes each get a modest weight in the model so that only a substantive swing in an index can have a major impact on the model forecast. Although the model explains just 41 percent of the variation in non-auto retail sales, its performance is reasonably good given the volatility of monthly retail sales.

SIGNS OF STRUCTURAL CHANGE?

We suggested earlier that structural changes in the retail business have tended to bias the informal commentary on the health of consumer spending. These same changes may also have affected formal statistical models of consumer spending. In particular, in a regression of nominal retail sales growth on nominal chain store growth, we would expect to see the following changes:

- Because of the fall in the relative price of GAF goods, a given change in nominal chain store sales might be associated with a larger change in nominal retail sales, increasing the coefficient on chain store sales.
- However, because large chain stores have been capturing increasing market share, the chain store data might overstate the growth of consumer spending, lowering the coefficient on chain store sales.
- Since same-store sales grow more slowly than total store sales in a period of expansion by the major retail chains, same-store indexes such as the Mitsubishi index might have a larger coefficient than total-store indexes such as Johnson Redbook.
- Finally, because it takes several years for seasonal adjustment procedures to adapt to changes in actual seasonal patterns, the adjusted data for the early 1990s are likely

Table 4
IN-SAMPLE EXPLANATORY POWER OF COMBINATION MODELS

			Non-Auto Retail Sales					
	GAF Sales		Advance Data		Latest Data		Personal Consumption Expenditures	
Models	Coefficient	R-Squared	Coefficient	R-Squared	Coefficient	R-Squared	Coefficient	R-Squared
ARIMA and ...								
Mitsubishi	.201 (6.56)	.433	.096 (4.55)	.277	.062 (3.27)	.357	.062 (1.88)	.184
Johnson Redbook	.077 (3.05)	.335	.081 (4.12)	.261	.044 (3.28)	.351	.066 (2.36)	.197
Both indexes		.456		.337		.385		.208
Mitsubishi	.188 (6.32)		.078 (3.83)		.056 (3.08)		.005 (1.36)	
Johnson Redbook	.063 (2.71)		.063 (3.36)		.038 (2.77)		.056 (1.97)	
Structural and ...								
Mitsubishi	.249 (5.87)	.404	.069 (3.07)	.163	.124 (4.78)	.381	.045 (2.47)	.680
Johnson Redbook	.179 (4.12)	.329	.076 (3.58)	.184	.087 (3.33)	.327	.028 (1.60)	.672
Both indexes		.456		.222		.412		.683
Mitsubishi	.221 (5.32)		.055 (2.49)		.111 (4.31)		.041 (2.16)	
Johnson Redbook	.137 (3.41)		.066 (3.08)		.067 (2.69)		.019 (1.07)	

Source: Authors' calculations. Details on the explanatory variables included in each model are available from the authors.

Notes: The table reports the R-squared and the coefficients on the chain store indexes, with the associated t-value in parentheses. In each case, the sample period is January 1985 to December 1995.

to contain some residual seasonal variation, but the current data should be free of significant distortion.

These expectations are, in fact, borne out by our regression analysis. To test whether the chain store coefficient has changed over time, we split our sample at the beginning of 1990 and regress each of the consumer spending variables on the two chain store indexes (Table 6). Although Chow tests show only limited evidence of a statistically significant shift in the overall structure of these equations, in seven out of eight cases the coefficient on the chain store index increases in the second half of the sample. The results for the latest, fully revised retail sales data are most striking. The Chow test is significant, and the coefficients on both chain store indexes increase sharply in the second half of the sample. These findings provide some support for the idea that falling prices in the GAF sector caused a change in the historic relationship between chain store sales growth and overall growth in retail spending.

In general, the results of our empirical tests indicate that the [forecast] models with the best in-sample fit combine the economic variables with both chain store indexes.

Our regression results also show the effects of consolidation in the retail industry. As we saw in Table 4, the coefficient on the Johnson Redbook index is consistently smaller than the coefficient on the Mitsubishi index. In part this finding may be due to the better sampling properties of the Mitsubishi index, but it is also consistent with measurement differences in these indexes: new store con-

Table 5
RECOMMENDED MODEL FOR PREDICTING NON-AUTO RETAIL SALES

Variable	Coefficient	T-Statistic
Constant	0.328	4.455
NRET(-1)	-0.121	-1.727
NRET(-12)	-0.224	-3.450
AUTO	0.015	2.699
PAY	1.063	3.551
GASP	0.009	2.168
MITS	0.113	4.426
JOHN	0.070	2.840

R-squared	0.412
Mean dependent variable	0.424
Adjusted R-squared	0.379
Durbin-Watson statistic	2.130
Mean square error	0.250

Sources: Department of Labor, Bureau of Labor Statistics; Department of Commerce, Bureau of the Census; Department of Commerce, Bureau of Economic Analysis; *Wall Street Journal*.

Notes: This model combines the Mitsubishi and Johnson Redbook indexes with the structural model. We drop Standard and Poor's index for retailers from the structural model, however, because it becomes statistically insignificant after we add both chain store indexes. The equation is estimated with ordinary least squares for the period from January 1985 to December 1995. All variables are measured as percentage changes from a month ago. NRET(-1) and NRET(-12) are the first and twelfth lag on non-auto retail sales, AUTO is auto and light truck sales, PAY is payroll employment, GASP is gasoline prices, and MITS and JOHN are the Mitsubishi and Johnson Redbook indexes, respectively.

Table 6
EVIDENCE OF STRUCTURAL CHANGE: SPLIT SAMPLE RESULTS FOR THE TWO-INDEX MODEL

			Non-Auto Retail Sales					
	GAF Sales		Advance Data		Latest Data		Personal Consumption Expenditures	
	1985-90	1990-95	1985-90	1990-95	1985-90	1990-95	1985-90	1990-95
Constant	0.453	0.191	0.217	0.151	0.467	0.195	0.582	0.382
	(3.864)	(1.775)	(3.591)	(2.758)	(6.450)	(3.014)	(5.712)	(6.816)
Mitsubishi	0.217	0.277	0.050	0.062	0.085	0.172	0.005	0.055
	(3.766)	(4.008)	(1.672)	(1.766)	(2.389)	(4.135)	(0.098)	(1.526)
Johnson Redbook	0.108	0.243	0.058	0.101	0.047	0.157	0.051	0.048
	(1.868)	(3.813)	(1.955)	(3.106)	(1.316)	(4.102)	(1.007)	(1.441)
R-squared	0.274	0.361	0.128	0.187	0.138	0.385	0.019	0.075
Chow F-test	1.601		0.215		3.311		1.180	
Chow significance	0.192		0.886		0.022		0.320	

Source: Authors' calculations.

Notes: The table reports regression coefficients. The associated t-statistics are in parentheses.

struction by major retail chains means that measures of total store sales (Johnson Redbook index) tend to exaggerate underlying demand, while measures of existing store sales (Mitsubishi index) tend to understate demand.

Finally, we find evidence that the consumer spending data have only recently caught up with the changing Christmas seasonals. The residuals from our models suggest that December sales, particularly for the GAF sector, have indeed been significantly weaker than expected, while January sales have been significantly stronger.[16]

Overall, our findings show both a shift in and a strengthening of the relationship between chain store sales and overall consumer spending.

Real Time Tests

Thus far we have focused on in-sample comparisons of the various models. The ultimate test of these models, however, is how they perform out of sample. This section investigates how much of a loss of predictive power occurs when we move from in-sample to out-of-sample tests and whether the rank order of the models changes.

We approximate true real time forecasting with a three-step procedure. First, using data for the 1975-89 period, we select the variables to be included in each model.[17] We use the same inclusion criteria and same menu of potential regressors employed in the in-sample models. Next, we use recursive regressions to reestimate the model, adding one month at a time and calculating a series of one-month-ahead forecasts over the entire 1990-95 period. Finally, we evaluate the forecasts using mean square error (MSE) and a variety of other conventional criteria.

The information used in this exercise differs somewhat from a true real time test. In one respect, we avail ourselves of more information than a forecaster would possess. We use the latest, fully revised data for the independent variables, whereas in real time only preliminary data for some of our regressors would be available. In another respect, however, we use less information than a forecaster would possess. By keeping the selected regressors and the starting date of the recursive regressions fixed, we limit how much the model can be modified to take account of the user's forecasting experience.[18] Fortunately, for several series, we do have preliminary data and substituting these did not have much impact on the results; unfortunately, we find some evidence of structural breaks in our models. As we will see, this shortcoming creates some underprediction bias and some evidence of serial correlation in our forecast errors.

Our main findings are summarized in Table 7, which reports the MSEs for forty-eight different forecast

Table 7
MEAN SQUARE ERRORS FOR ONE-MONTH FORECASTS

	Dependent Variables			
Models	GAF Sales	Advance Non-Auto Retail Sales	Latest Non-Auto Retail Sales	Personal Consumption Expenditures
Stand-alone				
ARIMA	1.074	0.391	0.494	0.315
Structural	1.022	0.260	0.504	0.240
Mitsubishi	0.855	0.265	0.348	0.256
Johnson Redbook	0.926	0.203	0.373	0.224
Two-index	0.774	0.198	0.321	0.218
Back-of-the-envelope	1.403	1.646	1.449	1.973
ARIMA and ...				
Mitsubishi	0.794	0.300	0.390	0.287
Johnson Redbook	0.912	0.262	0.375	0.272
Both chain store indexes	0.747	0.223	0.326	0.264
Structural and ...				
Mitsubishi	0.743	0.236	0.373	0.207
Johnson Redbook	0.883	0.172	0.375	0.166
Both chain store indexes	0.701	0.171	0.315	0.166

Source: Authors' calculations. Details on the explanatory variables included in each model are available from the authors.

Note: The lowest mean square error for each column is highlighted in boldface type.

models. For each of our four dependent variables, we test twelve models—six stand-alone models and six combination models that include variables from two or three of the stand-alone models.

The results strongly support the findings of the in-sample tests. In particular, we find that of the stand-alone models, the model using both chain store indexes always has the lowest MSE. The worst results are for the back-of-the-envelope model, suggesting once again that using simple rules of thumb to forecast with these data can cause more harm than good. When we add chain store data to the stand-alone ARIMA or structural models, the MSE declines substantially—often by a third or more. Overall, the models that combine both chain store indexes with the structural models perform best.

Although MSE is the most commonly used measure of forecast performance, the econometrics literature offers a smorgasbord of alternative evaluation criteria.[19] To a large degree, this diversity reflects the fact that forecasts are designed for use in a particular decision environment: the appropriate measure of forecast accuracy will always depend on what kind of forecast errors are most costly to the user. Table 8 reports MSEs and four other measures of our models' performance in predicting non-auto retail sales:[20]

- *Bias*: the mean forecast error. A mean value close to zero indicates that the forecast does not tend to systematically under- or overpredict the dependent variable.
- *Average absolute error*: the average error, regardless of sign. The average absolute error is preferred to MSE if the forecaster does not put a disproportionate weight on large errors.
- *Percent correct direction*: the portion of the time that the forecast correctly predicts the direction of change (positive or negative) in the dependent variable. Large econometric models are often compared on the basis of their ability to predict business cycle turning points; for our very short-run forecasts, percent correct direction provides an analogous test. Presumably, getting the right "handle" (positive or negative) on the predicted growth can help avoid some embarrassment for the forecaster. A good forecast model should correctly predict the direction substantially more than 50 percent of the time.
- *Q-test*: a test for serial correlation in the forecast errors. A significant Q-test means that at any point in time, the forecast could be improved by simply looking at the previous periods' forecast errors. Such a finding indicates that the model is missing some important information.

Table 8
ADDITIONAL OUT-OF-SAMPLE PERFORMANCE MEASURES FOR MODELS OF LATEST NON-AUTO RETAIL SALES

Models	Bias	Average Absolute Error	Mean Square Error	Percent Correct Direction	Q-Test (Twelve Lags)[a]
Stand-alone					
ARIMA	-0.250[b]	0.540	0.494	72.2	36.9
Significant	-0.301[b]	0.546	0.514	72.2	24.3
Correct	-0.299[b]	0.534	0.504	70.8	26.5
Mitsubishi	-0.179[b]	0.463	0.348	72.2	16.8
Johnson Redbook	-0.154[b]	0.471	0.373	70.8	11.6
Two-index	-0.129	0.444	0.321	72.2	13.0
ARIMA and ...					
Mitsubishi	-0.196	0.483	0.390	70.8	38.0
Johnson Redbook	-0.164[b]	0.473	0.375	70.8	10.5
Both chain store indexes	-0.144	0.447	0.326	70.8	11.0
Structural and ...					
Mitsubishi	-0.230[b]	0.467	0.373	72.2	31.2
Johnson Redbook	-0.160	0.464	0.375	70.8	18.5
Both chain store indexes	-0.112	0.437	0.315	72.2	21.0

Source: Authors' calculations. Details on the explanatory variables included in each model are available from the authors.

Note: The best results for each column are highlighted in boldface type.

[a] The critical value for this chi-square statistic at the 5 percent level is 21.0.

[b] The bias is significant at the 5 percent level. This "sign test" determines whether the positive and negative forecast errors are equal in number. It is a nonparametric test of the null hypothesis that the median forecast error is zero.

The results in Table 8 confirm the themes of our previous tests. First, virtually all of the models have a modest tendency to overpredict sales growth. This bias appears to reflect the structural shift in the chain store coefficient for the second half of our sample. Fortunately, the bias is statistically insignificant in models that combine economic variables with both chain store indexes. Second, the use of average absolute error, rather than mean square error, as the standard of evaluation generally has no impact on the ranking of the non-auto retail sales models (and very little impact on the ranking of models for other dependent variables). Third, using the combination models slightly reduces the most embarrassing kind of forecast error—predicting the wrong direction for sales growth. Finally, the Q-test for the joint significance of the first to twelfth lags of the forecast errors shows some evidence of serial correlation in the forecast errors. Again, using the combination models tends to mitigate this problem.[21]

IMPLICATIONS FOR FORECASTERS

What do our results mean in practical terms? Monthly consumer spending growth is very volatile, but by using a combination of economic variables and the chain store data we can explain about 40 percent of the variation in measures of retail sales and almost 70 percent of the variation in personal consumption. Using these models, we shave about 0.2 to 0.3 percentage point off our monthly forecast error (relative to a model that assumes no change in growth), and we correctly predict the direction of sales growth 70 to 85 percent of the time. Significantly, we also avoid the pitfalls of back-of-the-envelope calculations.

By using a combination of economic variables and the chain store data . . . we shave about 0.2 to 0.3 percentage point off our monthly forecast error.

At present, forecasters do not appear to be taking full advantage of the information contained in the chain store data. In particular, private sector economists do not completely account for chain store sales in their forecasts of the advance retail sales data. To demonstrate this, we use 1985-95 data on consensus forecasts of retail sales growth compiled each month by Money Market Services International.[22] If forecasters fully account for the chain store indexes in making their forecasts, we should find no correlation between the consensus forecast errors and the chain store indexes. In fact, while the Mitsubishi index is not correlated with the errors, the Johnson Redbook index is, at least marginally, at the 8 percent significance level.[23]

The long "shelf life" of the chain store data as economic indicators may also be insufficiently appreciated. Even after the advance retail sales data are released, forecasters should continue to keep one eye on the chain store indexes. To show this, we regress the revision in the official retail sales growth—the difference between the fully revised latest estimate and the advance estimate—on the chain store indexes for the 1985-95 period. In this case, it is the Mitsubishi index that turns out to be statistically significant (at the 2 percent level).[24] It appears that the chain store data deserve longer lasting, as well as more careful, attention.

CONCLUSION

Our results underscore some of the potential pitfalls of using chain store data to forecast consumer spending. Users should be mindful of the effect of changing seasonals and price discounting on chain store sales; this past December both of these factors contributed to retail analysts' unduly negative commentary on the sector. In addition, users should recognize that both individual store data and the weekly chain store indexes are of very limited value as macroeconomic indicators. Even the monthly indexes can be quite volatile and should not be taken at face value.

Nevertheless, the problems with the chain store data may be outweighed by their usefulness as predictive tools. By focusing on the monthly indexes, giving them the right weight, and combining them with economic variables, we can achieve more accurate forecasts of consumer spending.

COMPANY REPORTS

On the first or second Thursday of each month, trading floor economists trudge into work to face perhaps the most dreaded data release—the company reports of major retailers. The results for dozens and dozens of companies scroll across computer screens over the course of the day, requiring the economists to reinterpret the data continually. Each report seems to focus on a different measure of sales growth: same-store or total, year-to-date or latest month, domestic or total company, calendar month or "four-five-four weeks" month, and above or below "plan."

The results for individual companies are all over the map. Consider, for example, the year-over-year sales growth figures reported by Johnson Redbook for a group of fifty-six companies in January 1996. One company reported a sharp rise in total sales of 19 percent but an almost equally sharp decline in same-store sales of 9 percent; the strongest company enjoyed a 112 percent sales increase, while the weakest suffered a 28 percent decline. Even among the thirteen largest companies, reporting more than $500 million in sales, the growth rate ranged from a high of 27 percent to a low of -3 percent.

The sharp divergences in company reports reflect the various structural and idiosyncratic shocks buffeting the retail sector. They also underscore both the danger of using anecdotal evidence to assess industry trends and the importance of getting a large, representative sample.

CHAIN STORE INDEXES

Combining these data into indexes removes some, but not all, of their idiosyncrasies. Table A1 compiles the available information on these indexes and compares them with the more carefully documented official retail sales data issued by the Commerce Department. Note that the table presents the weekly and monthly Johnson Redbook indexes in one column because the monthly index is simply the

Table A1
COMPARING THE INDEXES AND THE COMMERCE DEPARTMENT RETAIL SALES DATA

	Johnson Redbook Index	Mitsubishi Index		Retail Sales Data	
	Weekly and Monthly	Weekly	Monthly	Advance	Final
Sector coverage	Department stores	GAF stores	GAF stores	All retail stores	All retail stores
Company coverage	21 large companies	Not available	70-80 large companies	3,200 stratified sample	12,000 stratified sample
Reporting lag (business days)	3	3	3-9	11-14	72-75
Accounting period[a]	Fiscal	Fiscal	Fiscal	Calendar	Calendar
Type of store[b]	Total[c]	Same store	Same store	Total	Total
Seasonal adjustment	Official department store factors[d]	Piser method[e]	Modified X-11[e]	X-11 ARIMA[f]	X-11 ARIMA[f]
Revisions	None	Infrequent	Infrequent	Not applicable	Frequent
Start date	1983	1989	1969	1947	1947

[a] Fiscal months vary from firm to firm, but the most common system uses February as the start of the fiscal year, counts Saturday as the last day of each week, and allocates weeks between months on a four-five-four basis (that is, four weeks in February, five in March, etc.).

[b] "Same stores" are stores open for at least a year. The precise definition varies with the individual company's reporting system and can mean stores open for at least twelve months, fourteen months, or one full fiscal year.

[c] Johnson Redbook compiles data on a same-store basis and then grosses up the numbers to total store sales using a lagged monthly average of the ratio of total-to-same-store sales (see Johnson Redbook Service 1996).

[d] Johnson Redbook calculates a seasonally adjusted dollar value for its index by applying the year-over-year growth rate estimated from its sample to the official department store data for twelve months earlier. Thus, it implicitly uses the previous year's official seasonal factors.

[e] For an explanation of these methods, see Mitsubishi Bank (1995, p. 4) and Mitsubishi Bank (1996).

[f] For an explanation of this method, see, for example, the April 1995 issue of Department of Commerce, Bureau of the Census (1985-95b).

cumulation of the weekly data.

A major drawback of the chain store indexes is that we know very little about the sampling properties of the data and how outliers or nonresponses are handled. Because neither index is revised, we can be sure that any late responses or reporting errors are never corrected. This lack of revision can lead to some anomalies in the data. For example, the Johnson Redbook series has a discontinuity in January 1989 because a major revision in the official data (which are used as a benchmark in constructing the Johnson Redbook index) was not matched by a similar adjustment in the index.

All the chain store data are subject to major seasonal adjustment problems. The data for March 1996 provide a striking example of how the timing of fiscal calendars and holidays can severely distort the chain store data. Most company reports for March 1996 included sales for the five weeks ending on April 6. Because Easter was on April 7, these figures captured all the shopping for this holiday. By contrast, in 1995 the March reports only included data through April 1. Since Easter fell on April 15, most of the Easter shopping was excluded from the March reports. This example suggests that even monthly changes from a year ago will be distorted by changing seasonals. The Johnson Redbook index is particularly prone to adjustment errors of this kind because it uses Commerce Department seasonal factors to adjust its data, even though its sample is very different from that of the Commerce Department and its survey covers the fiscal period for each store, not the calendar month.

The weekly indexes, however, present the greatest seasonal adjustment difficulties. Because the calendar always shifts from one year to the next, the proper reference week for year-to-year comparisons of sales is often unclear. This problem is especially acute in 1996: the preceding year had fifty-three weeks, prompting companies to adopt different reference weeks for their sales comparisons.

The growth rates for these indexes have been very erratic. Table A2 uses two statistics—the autoregressive coefficient, which shows whether the indicator is subject to sharp reversals, and the variance of the growth rate—to assess the variability of the weekly and monthly data. For the Mitsubishi index, more than one-third of the growth in any week tends to be reversed in the next week. Furthermore, with a week-to-week variance of almost a percentage point, it is not unusual to see one-week annualized percentage changes of more than 50 percent.[25] The monthly Mitsubishi index is also subject to frequent reversals, and

Table A2
SUMMARY STATISTICS FOR CONSUMPTION INDICATORS

Variables	Autoregressive Coefficient	Mean	Variance
Weekly Mitsubishi (percentage change from previous week)	-0.380**	0.072	0.756
Monthly Mitsubishi (percentage change from previous month)	-0.482**	0.350	3.035
Johnson Redbook	-0.148	0.550	3.233
GAF sales	-0.277**	0.499	1.058
Advance non-auto retail sales	-0.087	0.246	0.221
Latest non-auto retail sales	-0.174*	0.424	0.378
Personal consumption expenditures	-0.369**	0.510	0.374

Notes: The table reports the first autoregressive coefficient and the sample mean and variance for each consumption indicator. For the weekly Mitsubishi index, the sample period is November 1989 to December 1995; for all monthly data, the sample period is January 1985 to December 1995.

*Significant at the 5 percent level.

**Significant at the 1 percent level.

for both indexes, the variance of the monthly growth rate is more than 3 percent.

The developers of the chain store indexes are aware of many of these problems. Chain store indexes were designed primarily for use by industry analysts, not macroeconomists—a feature that helps explain the reporting of same-store sales rather than total sales and the use of the fiscal month rather than the calendar month. Recognizing the erratic nature of the data, economists at Johnson Redbook and Mitsubishi recommend that users of their data consider long averages of the indexes. Mitsubishi reports a sixteen-week "trend" for its index and cautions users that "the best way to understand the message in our series is to view it on a week-to-week basis against its trend."

COMMERCE DEPARTMENT MEASURES OF RETAIL SALES

The official Commerce Department data are less erratic than the chain store indexes, largely because they are constructed using sophisticated (and expensive) sampling and statistical methods that simply cannot be matched by a private firm. Monthly data on retail sales are based on a random sample of more than 12,000 companies. Although the sample covers firms of all sizes, it is stratified, with coverage ranging from 100 percent for major firms to 0.1 percent for the smallest firms. Department stores are heavily represented in this sample because of their large size: while the sample captures less than half of overall retail sales, it captures 99 percent of the department store sector.

The official data are heavily and repeatedly revised. The Commerce Department releases *advance* data, somewhat reluctantly,[26] only two weeks after a month ends, but these initial estimates are based on a survey covering only about one-fourth of the full sample. The full-sample, or *preliminary*, data are reported a month later; *final* estimates are reported two months later; and annual revisions are released each spring. In addition, every five years a complete census count is made of virtually every retail establishment.

These revisions can have a substantial impact on the estimated monthly growth rates for retail sales. Thus, the reported direction of sales growth can change sign from one estimate to the next, and the correlation between the monthly growth rate for the latest non-auto retail sales data and the advance data is just 52 percent for the 1985-95 period.

While these data are less volatile than the chain store indexes, they are nonetheless quite variable by the standards of macroeconomic data (Table A2). Although retail sales tend to grow over time and to rise and fall with the business cycle, the monthly growth rates have a negative serial correlation, implying that strong growth in one month tends to be reversed the next month. The variances are lower than for chain store data, but they still suggest considerable month-to-month variation.

APPENDIX 2: RESULTS FOR ADDITIONAL CHAIN STORE INDEXES

Growing interest in the chain store data has helped spur a cottage industry of new retail sales indexes. Two new indexes that have received press coverage are:

- Goldman Sachs Monthly Comparable-Store Sales Index, an index of department, apparel, discount, and hard goods stores that was introduced in 1988; and
- Merrill Lynch Broadlines Same Store Sales Index, an index of department and general merchandise stores first released in 1992.

In this appendix, we compare the in-sample predictive power of these indexes and the weekly and monthly versions of the Johnson Redbook and Mitsubishi indexes. We regress the growth rate of the latest, fully revised non-auto retail sales on a constant term and each of six chain store indexes. The table reports our results for three sample periods and two sets of regressions—one in which variables enter as percentage changes from a month ago and one in which they enter as percentage changes from a year ago.

Tests [of the predictive power of] weekly data yield poor results. . . . Growth in sales from the first week of one month to the first week of the next month has virtually no correlation with the monthly change in non-auto retail sales.

There are few surprises here. Most of the indexes are highly significant in the relatively undemanding year-ago tests but are much less significant in explaining monthly changes in retail sales. The results support our decision to focus on the monthly versions of the Johnson Redbook and Mitsubishi indexes in the body of this article. The newer Goldman Sachs and Merrill Lynch indexes have weaker predictive power than their older counterparts. Furthermore, the tests using weekly data yield poor results. As the last two rows of the table show, growth in sales from the first week of one month to the first week of the next month has virtually no correlation with the monthly change in non-auto retail sales. Similarly poor results were obtained using the change in sales for other weeks of the month and using week-to-week sales growth within the month.

EXPLANATORY POWER OF ALTERNATIVE INDEXES

	Variables Enter As: Percentage Change from a Month Ago	Percentage Change from a Year Ago
SAMPLE: 1988:02–1995:12		
Mitsubishi	0.232**	0.230**
Johnson Redbook	0.130**	0.263**
Goldman Sachs	0.044*	0.277**
SAMPLE: 1992:07–1995:12		
Mitsubishi	0.330**	0.349**
Johnson Redbook	0.094*	0.429**
Goldman Sachs	0.079	0.006
Merrill Lynch	0.003	0.359**
SAMPLE 1990:11–1995:12		
Weekly Mitsubishi	0.001	0.010
Weekly Johnson Redbook	N.A.	0.112**

Notes: The table reports the R-squared from ordinary least square regressions of non-auto retail sales growth (latest data, fully revised) on a constant term and the percentage change in the chain store index. Goldman Sachs and Merrill Lynch report their data only as a percentage change from a year ago. Month-ago percentage changes for these indexes are constructed using the same methodology employed by Johnson Redbook: first, seasonally adjusted levels for the indexes are constructed by applying their year-over-year growth rates to the year-ago level of GAF sales; second, monthly growth rates are calculated from these monthly levels. The weekly regressions use sales for the first week of each month relative to sales for the corresponding week a month or a year earlier.

*Significant at the 5 percent level.

**Significant at the 1 percent level.

ENDNOTES

1. *New York Times*, January 5, 1996, p. A1. The chain store data have also shown an ability to move markets. For example, on March 12, 1996: "The bond market had been down by as much as a point by noon, fueled by the morning release of the Mitsubishi Bank Ltd.-Schroder Wertheim & Co. chain-store index, which showed a stronger than expected 1% rise in the week ending March 9. But Johnson Redbook weekly survey of national retail sales, released at midafternoon, showed sales down 1.5% in the first week of March compared with February. That quickly sent the 30-year price rising 5/8 point from its low, which helped reverse a 90-point plunge in the Dow Jones Industrial Average" (Vogelstein 1996).

2. Although chain store data are briefly described in books on economic indicators and in various Wall Street newsletters, there is no literature that takes a rigorous look at the usefulness of these data as macroeconomic indicators. In their handbooks, Rogers (1994, p. 68), Tainer (1993, pp. 59, 62-3, and 68-71), and Kuwayama and O'Sullivan (1996) provide background information on the chain store data. The Mitsubishi Bank (1996) briefly describes its index and presents graphs showing that smoothed year-over-year growth in its index has similar patterns to several other consumer indicators.

3. See Department of Commerce, Bureau of the Census (1995, Appendix F).

4. The Commerce Department and forecasters use the retail sales data to estimate most of the goods component of personal consumption expenditures. They must make two adjustments to the data, however. First, they net out the portion of hardware and garden store purchases made by construction companies rather than consumers. Second, they net out auto dealer sales from retail sales and substitute separately available (and better) data on unit sales of motor vehicles.

5. To be precise, the weekly series is called the Bank of Tokyo–Mitsubishi/Schroder Wertheim Weekly Chain Store Index, and the monthly series is called the Bank of Tokyo–Mitsubishi Chain Store Index.

6. Motor vehicle sales are strongly influenced by the introduction of new models and the on-again-off-again nature of price discounts. Both of these determinants have become quite erratic in recent years.

7. Harris et al. (1994) review the forces behind the boom and bust in commercial construction.

8. Competitive pressures help explain why consumer prices have been relatively subdued despite capacity pressures in the economy. On crude accounting, with GAF sales making up 11 percent of consumer spending, the 4 percent inflation shortfall would shave off almost 1/2 of 1 percentage point from overall consumer price inflation.

9. We use the Merrill Lynch index because both total store and same-store versions are readily available. See Appendix 2 for a description of this index.

10. There are other signs of consolidation. Mitsubishi Bank reports that within its chain store index, the largest companies are growing faster than the overall index. In addition, the official retail sales data show that department stores—which are almost all large companies—have been capturing an increasing share of GAF sales. Their share has risen from 35 percent in 1990 to 37 percent in 1995.

11. Here "autos" refers to auto dealers and includes sales of autos and light trucks. Forecasters usually treat motor vehicle sales separately from the rest of retail sales because motor vehicle sales follow very different monthly patterns than other retail sales and because separate data on unit sales of motor vehicles are available on a very timely basis.

12. See, for example, Bernanke (1990) and Estrella and Hardouvelis (1989).

13. Additional models are reported in our research paper, Harris and Vega (1996). In that paper, we used one additional dependent variable—department stores sales. We also tested two additional stand-alone models: (1) a "kitchen-sink soup model," in which we rigged our alternative to the chain store model, throwing in every consumer-related variable regardless of its explanatory power; and (2) a "significant model," in which we included every economic variable that met the Akaike information criteria, even if it had the wrong sign. The results for these models were very similar to the results reported here.

14. We chose this sample period so that—given the constraint of data availability—all our models could be tested over the same period. Varying the starting point of the sample did not materially affect the results.

15. This perverse result arises because the model is not estimated so that it is possible for the variance of the model error to be larger than the variance of the dependent variable. Thus, the R-squared (= 1 - var (err)/ var (dep.var.)) is negative. The results are considerably worse if the indexes are used individually.

16. Tests using seasonal dummies for the 1991-94 period showed that the under- and overpredictions were statistically significant.

17. Data limitations prevented us from using the full 1975-89 period for initializing all our models. In particular, because GAF data are available only from 1977 and the Johnson Redbook index is available only from 1983, models using these variables were based on a smaller sample. In

Note 17 continued

addition, one variable (the Treasury bill–commercial paper spread) was dropped from our list of potential regressors because it was only available starting in 1981. Finally, gasoline prices were not available before 1986; rather than drop the variable, we backfilled the data using fitted values from a regression of gas price inflation on current and lagged inflation in crude oil prices.

18. In particular, the recursive regression allows the structure of the model to evolve as new data points are added, but does not allow for abrupt structural breaks.

19. See Diebold and Lopez (1996) for a thorough review of the criteria for forecast evaluation.

20. Similar results for models of GAF sales, advance non-auto retail sales, personal consumption expenditure, and department store sales are reported in Harris and Vega (1996).

21. For most of our dependent variables, the Q-statistic tends to diminish in significance as the lag length gets smaller or larger than twelve.

22. These consensus forecasts come from a survey of several dozen market participants representing major commercial banks, brokerage firms, private consulting firms, and other institutions. The survey is taken the week before the release of the retail sales report, and the consensus is calculated as the median of the responses.

23. The regression coefficients are -0.087 + .059 * JOHN - .002 MITS. The associated t-values are -1.39, 1.78, and -0.07, respectively.

24. The regression coefficients are 0.145 + .018 * JOHN +.066 * MITS. The associated t-values are 2.95, 0.66, and 2.42, respectively.

25. Because weekly data for the Johnson Redbook index are available only on a year-ago percentage change basis, we could not use them in this table. The year-ago percentage change data do suggest, however, that the Johnson Redbook index is twice as volatile as the Mitsubishi index. Given this volatility, it should not be surprising that the weekly data have a relatively weak correlation with their monthly counterparts. For example, using the percentage change from a year ago and comparing the first week of each month to the full month index, we find that the Mitsubishi index has a correlation of only .48 and the Johnson Redbook index a correlation of just .39 over the 1990-95 period.

26. The Census Bureau explains that it "releases (non-final) advance and preliminary data to provide government and private data users with much demanded early measures of consumer spending. . . . The advance sales estimates are based on early reporting of sales by a small subsample of the Bureau's retail survey panels" (see the April 1995 issue of Department of Commerce, Bureau of the Census 1985-95a).

References

Bernanke, Ben S. 1990. "On the Predictive Power of Interest Rates and Interest Rate Spreads." Federal Reserve Bank of Boston NEW ENGLAND ECONOMIC REVIEW, November-December: 51-68.

Diebold, Francis X., and Jose A. Lopez. 1996. "Forecast Evaluation and Combination." National Bureau of Economic Research Technical Working Paper no. 192.

Estrella, Arturo, and Gikas A. Hardouvelis. 1989. "The Term Structure as a Predictor of Real Economic Activity." Federal Reserve Bank of New York Research Paper no. 8907.

Harris, Ethan S., Michael Boldin, and Mark D. Flaherty. 1994. "The Credit Crunch and the Construction Industry." In Federal Reserve Bank of New York, STUDIES IN THE CAUSES AND CONSEQUENCES OF THE 1989-92 CREDIT SLOWDOWN, pp. 301-54.

Harris, Ethan S., and Clara Vega. 1996. "What Do Chain Store Sales Tell Us about Consumer Spending?" Federal Reserve Bank of New York Research Paper no. 9614.

Johnson Redbook Service. 1996. REDBOOK WEEKLY COMMENTS, February 7.

Kuwayama, Patricia Hagan, and James F. O'Sullivan, eds. 1996. GLOBAL DATA WATCH HANDBOOK. 2d ed. New York: J.P. Morgan.

Merrill Lynch. 1996. RETAILING BROADLINES, January 4.

Mitsubishi Bank. 1995. WEEKLY CHAIN STORE SALES SNAPSHOT, November.

———. 1996. "Update to Methodology Used to Compile the Mitsubishi Bank–Schroder Wertheim Weekly Chain Store Sales Index." Mimeographed.

Rogers, Mark R. 1994. HANDBOOK OF KEY ECONOMIC INDICATORS. New York: Irwin Professional Publishing.

Tainer, Eva. 1993. USING ECONOMIC INDICATORS TO IMPROVE INVESTMENT ANALYSIS. New York: John Wiley and Sons.

Telsey, Dana L. 1996. "Specialty Retail: Hard and Soft Lines." EQUITY RESEARCH. Bear Stearns, April.

U.S. Department of Commerce. Bureau of the Census. 1985-94. "Combined Annual and Revised Monthly Retail Trade." CURRENT BUSINESS REPORTS. Washington, D.C. Various issues.

———. 1985-95a. "Advance Monthly Retail Sales." CURRENT BUSINESS REPORTS. Washington, D.C. Various issues.

———. 1985-95b. "Monthly Retail Trade Sales and Inventories." CURRENT BUSINESS REPORTS. Washington, D.C. Various issues.

———. 1995. "Geographic Area Series: United States." 1992 CENSUS OF RETAIL TRADE. Washington, D.C.

Vogelstein, Fred. 1996. "Conflicting Economic Reports on Retail Sales Create a Volatile Day for Investors in the Bond Market." WALL STREET JOURNAL, March 13, p. C21.

Determinants and Impact of Sovereign Credit Ratings

Richard Cantor and Frank Packer

In recent years, the demand for sovereign credit ratings—the risk assessments assigned by the credit rating agencies to the obligations of central governments—has increased dramatically. More governments with greater default risk and more companies domiciled in riskier host countries are borrowing in international bond markets. Although foreign government officials generally cooperate with the agencies, rating assignments that are lower than anticipated often prompt issuers to question the consistency and rationale of sovereign ratings. How clear are the criteria underlying sovereign ratings? Moreover, how much of an impact do ratings have on borrowing costs for sovereigns?

To explore these questions, we present the first systematic analysis of the determinants and impact of the sovereign credit ratings assigned by the two leading U.S. agencies, Moody's Investors Service and Standard and Poor's.[1] Such an analysis has only recently become possible as a result of the rapid growth in sovereign rating assignments. The wealth of data now available allows us to estimate which quantitative indicators are weighed most heavily in the determination of ratings, to evaluate the predictive power of ratings in explaining a cross-section of sovereign bond yields, and to measure whether rating announcements directly affect market yields on the day of the announcement.

Our investigation suggests that, to a large extent, Moody's and Standard and Poor's rating assignments can be explained by a small number of well-defined criteria, which the two agencies appear to weigh similarly. We also find that the market—as gauged by sovereign debt yields—broadly shares the relative rankings of sovereign credit risks made by the two rating agencies. In addition, credit ratings appear to have some independent influence on yields over and above their correlation with other publicly available information. In particular, we find that rating announcements have immediate effects on market pricing for non-investment-grade issues.

WHAT ARE SOVEREIGN RATINGS?

Like other credit ratings, sovereign ratings are assessments of the relative likelihood that a borrower will default on its obligations.[2] Governments generally seek credit ratings to ease their own access (and the access of other issuers domiciled within their borders) to international capital markets, where many investors, particularly U.S. investors, prefer rated securities over unrated securities of apparently similar credit risk.

In the past, governments tended to seek ratings on their foreign currency obligations exclusively, because foreign currency bonds were more likely than domestic currency offerings to be placed with international investors. In recent years, however, international investors have increased their demand for bonds issued in currencies other than traditional global currencies, leading more sovereigns to obtain domestic currency bond ratings as well. To date, however, foreign currency ratings—the focus of this article—remain the more prevalent and influential in the international bond markets.

Sovereign ratings are important not only because some of the largest issuers in the international capital markets are national governments, but also because these assessments affect the ratings assigned to borrowers of the same nationality. For example, agencies seldom, if ever, assign a credit rating to a local municipality, provincial government, or private company that is higher than that of the issuer's home country.

Moody's and Standard and Poor's each currently rate more than fifty sovereigns. Although the agencies use

Table 1
RATING SYMBOLS FOR LONG-TERM DEBT

Interpretation	Moody's	Standard and Poor's
INVESTMENT-GRADE RATINGS		
Highest quality	Aaa	AAA
High quality	Aa1	AA+
	Aa2	AA
	Aa3	AA-
Strong payment capacity	A1	A+
	A2	A
	A3	A-
Adequate payment capacity	Baa1	BBB+
	Baa2	BBB
	Baa3	BBB-
SPECULATIVE-GRADE RATINGS		
Likely to fulfill obligations, ongoing uncertainty	Ba1	BB+
	Ba2	BB
	Ba3	BB-
High-risk obligations	B1	B+
	B2	B
	B3	B-

Note: To date, the agencies have not assigned sovereign ratings below B3/B-.

Table 2
SOVEREIGN CREDIT RATINGS
As of September 29, 1995

Country	Moody's Rating	Standard and Poor's Rating
Argentina	B1	BB-
Australia	Aa2	AA
Austria	Aaa	AAA
Belgium	Aa1	AA+
Bermuda	Aa1	AA
Brazil	B1	B+
Canada	Aa2	AA+
Chile	Baa1	A-
China	A3	BBB
Colombia	Baa3	BBB-
Czech Republic	Baal	BBB+
Denmark	Aa1	AA+
Finland	Aa2	AA-
France	Aaa	AAA
Germany	Aaa	AAA
Greece	Baa3	BBB-
Hong Kong	A3	A
Hungary	Ba1	BB+
Iceland	A2	A
India	Baa3	BB+
Indonesia	Baa3	BBB
Ireland	Aa2	AA
Italy	A1	AA
Japan	Aaa	AAA
Korea	A1	AA-
Luxembourg	Aaa	AAA
Malaysia	A1	A+
Malta	A2	A
Mexico	Ba2	BB
Netherlands	Aaa	AAA
New Zealand	Aa2	AA
Norway	Aa1	AAA
Pakistan	B1	B+
Philippines	Ba2	BB
Poland	Baa3	BB
Portugal	A1	AA-
Singapore	Aa2	AAA
Slovak Republic	Baa3	BB+
South Africa	Baa3	BB
Spain	Aa2	AA
Sweden	Aa3	AA+
Switzerland	Aaa	AAA
Taiwan	Aa3	AA+
Thailand	A2	A
Turkey	Ba3	B+
United Kingdom	Aaa	AAA
United States	Aaa	AAA
Uruguay	Ba1	BB+
Venezuela	Ba2	B+

Sources: Moody's; Standard and Poor's.

different symbols in assessing credit risk, every Moody's symbol has its counterpart in Standard and Poor's rating scale (Table 1). This correspondence allows us to compare the sovereign ratings assigned by the two agencies. Of the forty-nine countries rated by both Moody's and Standard and Poor's in September 1995, twenty-eight received the same rating from the two agencies, twelve were rated higher by Standard and Poor's, and nine were rated higher by Moody's (Table 2). When the agencies disagreed, their ratings in most cases differed by one notch on the scale, although for seven countries their ratings differed by two notches. (A rating notch is a one-level difference on a rating scale, such as the difference between A1 and A2 for Moody's or between A+ and A for Standard and Poor's.)

DETERMINANTS OF SOVEREIGN RATINGS

In their statements on rating criteria, Moody's and Standard and Poor's list numerous economic, social, and political factors that underlie their sovereign credit ratings (Moody's 1991; Moody's 1995; Standard and Poor's 1994). Identifying the relationship between their criteria and actual ratings, however, is difficult, in part because some of the criteria are not quantifiable. Moreover, the agencies provide little guidance as to the relative weights they assign each factor. Even for quantifiable factors, determining the relative weights assigned by Moody's and Standard and Poor's is difficult because the agencies rely on such a large number of criteria.

Identifying the relationship between [the two agencies'] criteria and actual ratings . . . is difficult, in part because some of the criteria are not quantifiable. Moreover, the agencies provide little guidance as to the relative weights they assign each factor.

In the article's next section, we use regression analysis to measure the relative significance of eight variables that are repeatedly cited in rating agency reports as determinants of sovereign ratings.[3] As a first step, however, we describe these variables and identify the measures we use to represent them in our quantitative analysis (Table 3). We explain below the relationship between each variable and a country's ability and willingness to service its debt:

- *Per capita income*. The greater the potential tax base of the borrowing country, the greater the ability of a government to repay debt. This variable can also serve as a proxy for the level of political stability and other important factors.
- *GDP growth*. A relatively high rate of economic growth suggests that a country's existing debt burden will become easier to service over time.
- *Inflation*. A high rate of inflation points to structural problems in the government's finances. When a government appears unable or unwilling to pay for current budgetary expenses through taxes or debt issuance, it must resort to inflationary money finance. Public dissatisfaction with inflation may in turn lead to political instability.
- *Fiscal balance*. A large federal deficit absorbs private domestic savings and suggests that a government lacks the ability or will to tax its citizenry to cover current expenses or to service its debt.[4]
- *External balance*. A large current account deficit indicates that the public and private sectors together rely heavily on funds from abroad. Current account deficits that persist result in growth in foreign indebtedness, which may become unsustainable over time.
- *External debt*. A higher debt burden should correspond to a higher risk of default. The weight of the burden increases as a country's foreign currency debt rises relative to its foreign currency earnings (exports).[5]
- *Economic development*. Although level of development is already measured by our per capita income variable, the rating agencies appear to factor a threshold effect into the relationship between economic development and risk. That is, once countries reach a certain income or level of development, they may be less likely to default.[6] We proxy for this minimum income or development level with a simple indicator variable noting whether or not a country is classified as industrialized by the International Monetary Fund.

Table 3
DESCRIPTION OF VARIABLES

Variable Name	Definition	Unit of Measurement[a]	Data Sources
Determinants of Sovereign Ratings			
Per capita income	GNP per capita in 1994	Thousands of dollars	World Bank, Moody's, FRBNY estimates
GDP growth	Average annual real GDP growth on a year-over-year basis, 1991-94	Percent	World Bank, Moody's, FRBNY estimates
Inflation	Average annual consumer price inflation rate, 1992-94	Percent	World Bank, Moody's, FRBNY estimates
Fiscal balance	Average annual central government budget surplus relative to GDP, 1992-94	Percent	World Bank, Moody's, IMF, FRBNY estimates
External balance	Average annual current account surplus relative to GDP, 1992-94	Percent	World Bank, Moody's, FRBNY estimates
External debt	Foreign currency debt relative to exports, 1994	Percent	World Bank, Moody's, FRBNY estimates
Indicator for economic development	IMF classification as an industrialized country as of September 1995	Indicator variable: 1 = industrialized; 0 = not industrialized	IMF
Indicator for default history	Default on foreign currency debt since 1970	Indicator variable: 1 = default; 0 = no default	S&P
Other Variables			
Moody's, S&P, or average ratings	Ratings assigned as of September 29, 1995, by Moody's or S&P, or the average of the two agencies' ratings	B1(B+)=3; Ba3(BB-)=4; Ba2(BB)=5;...Aaa(AAA)=16	Moody's, S&P
Spreads	Sovereign bond spreads over Treasuries, adjusted to five-year maturities[b]	Basis points	Bloomberg L.P., Salomon Brothers, J.P. Morgan, FRBNY estimates

Note: S&P= Standard and Poor's; FRBNY= Federal Reserve Bank of New York; IMF= International Monetary Fund.

[a] In the regression analysis, per capita income, inflation, and spreads are transformed to natural logarithms.

[b] For example, the spread on a three-year maturity Baa/BBB sovereign bond is adjusted to a five-year maturity by subtracting the difference between the average spreads on three-year and five-year Baa/BBB corporate bonds as reported by Bloomberg L.P. on September 29, 1995.

- *Default history.* Other things being equal, a country that has defaulted on debt in the recent past is widely perceived as a high credit risk. Both theoretical considerations of the role of reputation in sovereign debt (Eaton 1996) and related empirical evidence indicate that defaulting sovereigns suffer a severe decline in their standing with creditors (Ozler 1991). We factor in credit reputation by using an indicator variable that notes whether or not a country has defaulted on its international bank debt since 1970.

QUANTIFYING THE RELATIONSHIP BETWEEN RATINGS AND THEIR DETERMINANTS

In this section, we assess the individual and collective significance of our eight variables in determining the September 29, 1995, ratings of the forty-nine countries listed in Table 2. The sample statistics, broken out by broad letter category, show that five of the eight variables are directly correlated with the ratings assigned by Moody's and Standard and Poor's (Table 4). In particular, a high per capita income appears to be closely related to high ratings: among the nine countries assigned top ratings by Moody's and the eleven given Standard and Poor's highest ratings, median per capita income is just under $24,000. Lower inflation and lower external debt are also consistently related to higher ratings. A high level of economic development, as measured by the indicator for industrialization, greatly increases the likelihood of a rating of Aa/AA. As a negative factor, any history of default limits a sovereign's ratings to Baa/BBB or below.

A high per capita income appears to be closely related to high ratings. . . . Lower inflation and lower external debt are also consistently related to higher ratings.

Three factors—GDP growth, fiscal balance, and external balance—lack a clear bivariate relation to ratings. Ratings may lack a simple relation to GDP growth because many developing economies tend to grow faster than mature economies. More surprising, however, is the lack of a clear correlation between ratings and fiscal and external balances. This finding may reflect endogeneity in both fiscal policy and international capital flows: countries trying to improve their credit standings may opt for more conservative fiscal policies, and the supply of international capital may be restricted for some low-rated countries.

Because some of the eight variables are mutually correlated, we estimate a multiple regression to quantify their combined explanatory power and to sort out their individual contributions to the determination of ratings. Like most analysts who transform bond ratings into data for regression analysis (beginning with Horrigan 1966 and continuing through Billet 1996), we assign numerical values to the Moody's and Standard and Poor's ratings as follows: B3/B- = 1, B2/B = 2, and so on through Aaa/AAA = 16. When we need a measure of a country's average rating, we take the mean of the two numerical values representing Moody's and Standard and Poor's ratings for that country. Our regressions relate the numerical equivalents of Moody's and Standard and Poor's ratings to the eight explanatory variables through ordinary least squares.[7]

The model's ability to predict large differences in ratings is impressive. The first column of Table 5 shows that a regression of the average of Moody's and Standard and Poor's ratings against our set of eight variables explains more than 90 percent of the sample variation and yields a residual standard error of about 1.2 rating notches. Note that although the model's explanatory power is impressive,

The model's ability to predict large differences in ratings is impressive. . . . A regression of the average of Moody's and Standard and Poor's ratings against our set of eight variables explains more than 90 percent of the sample variation.

Table 4
SAMPLE STATISTICS BY BROAD LETTER RATING CATEGORIES

	Agency	Aaa/AAA	Aa/AA	A/A	Baa/BBB	Ba/BB	B/B
MEDIANS							
Per capita income	Moody's	23.56	19.96	8.22	2.47	3.30	3.37
	S&P	23.56	18.40	5.77	1.62	3.01	2.61
GDP growth	Moody's	1.27	2.47	5.87	4.07	2.28	4.30
	S&P	1.52	2.33	6.49	5.07	2.31	2.84
Inflation	Moody's	2.86	2.29	4.56	13.73	32.44	13.23
	S&P	2.74	2.64	4.18	14.3	13.23	62.13
Fiscal balance	Moody's	-2.67	-2.28	-1.03	-3.50	-2.50	-1.75
	S&P	-2.29	-3.17	1.37	0.15	-3.50	-4.03
External balance	Moody's	0.90	2.10	-2.48	-2.10	-2.74	-3.35
	S&P	3.10	-0.73	-3.68	-2.10	-3.35	-1.05
External debt	Moody's	76.5	102.5	70.4	157.2	220.2	291.6
	S&P	76.5	97.2	61.7	157.2	189.7	231.6
Spread	Moody's	0.32	0.34	0.61	1.58	3.40	4.45
	S&P	0.29	0.40	0.59	1.14	2.58	3.68
FREQUENCIES							
Number rated	Moody's	9	13	9	9	6	3
	S&P	11	14	6	5	9	4
Indicator for economic development	Moody's	9	10	3	1	0	0
	S&P	10	11	1	1	0	0
Indicator for default history	Moody's	0	0	0	2	5	2
	S&P	0	0	0	0	6	3

Sources: Moody's; Standard and Poor's; World Bank; International Monetary Fund; Bloomberg L.P.; J.P. Morgan; Federal Reserve Bank of New York estimates.

the regression achieves its high R-squared through its ability to predict large rating differences. For example, the specification predicts that Germany's rating (Aaa/AAA) will be much higher than Uruguay's (Ba1/BB+). The model naturally has little to say about small rating differences—for example, why Mexico is rated Ba2/BB and South Africa is rated Baa3/BB. These differences, while modest, can cause great controversy in financial markets.

The regression does not yield any prediction errors that exceed three notches, and errors that exceed two notches occur in the case of only four countries. Another way of measuring the accuracy of this specification is to compare predicted ratings rounded off to the nearest broad letter rating with actual broad letter ratings. The average rating regression predicts these broad letter ratings with about 70 percent accuracy, a slightly higher accuracy rate than that found in the literature quantifying the determinants of corporate ratings (see, for example, Ederington [1985]).

Of the individual coefficients, per capita income, GDP growth, inflation, external debt, and the indicator variables for economic development and default history all have the anticipated signs and are statistically significant. The coefficients on both the fiscal and external balances are statistically insignificant and of the unexpected sign. As mentioned earlier, in many cases the market forces poor credit risks into apparently strong fiscal and external balance positions, diminishing the significance of fiscal and external balances as explanatory variables. Therefore, although the agencies may assign substantial weight to these variables in determining specific rating assignments, no systematic relationship between these variables and ratings is evident in our sample.

Table 5
DETERMINANTS OF SOVEREIGN CREDIT RATINGS

	Dependent Variable			
Explanatory Variable	Average Ratings	Moody's Ratings	Standard and Poor's Ratings	Moody's/Standard and Poor's Rating Differences[a]
Intercept	1.442	3.408	-0.524	3.932**
	(0.633)	(1.379)	(0.223)	(2.521)
Per capita income	1.242***	1.027***	1.458***	-0.431***
	(5.302)	(4.041)	(6.048)	(2.688)
GDP growth	0.151*	0.130	0.171**	-0.040
	(1.935)	(1.545)	(2.132)	(0.756)
Inflation	-0.611***	-0.630***	-0.591***	-0.039
	(2.839)	(2.701)	(2.671)	(0.265)
Fiscal balance	0.073	0.049	0.097*	-0.048
	(1.324)	(0.818)	(1.71)	(1.274)
External balance	0.003	0.006	0.001	0.006
	(0.314)	(0.535)	(0.046)	(0.779)
External debt	-0.013***	-0.015***	-0.011***	-0.004**
	(5.088)	(5.365)	(4.236)	(2.133)
Indicator for economic development	2.776***	2.957***	2.595***	0.362
	(4.25)	(4.175)	(3.861)	(0.81)
Indicator for default history	-2.042***	-1.463**	-2.622***	1.159***
	(3.175)	(2.097)	(3.962)	(2.632)
Adjusted R-squared	0.924	0.905	0.926	0.251
Standard error	1.222	1.325	1.257	0.836

Sources: Moody's; Standard and Poor's; World Bank; International Monetary Fund; Bloomberg L.P.; Salomon Brothers; J.P. Morgan; Federal Reserve Bank of New York estimates.

Notes: The sample size is forty-nine. Absolute t-statistics are in parentheses.
[a]The number of rating notches by which Moody's ratings exceed Standard and Poor's.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

Quantitative models cannot explain all variations in ratings across countries: as the agencies often state, qualitative social and political considerations are also important determinants. For example, the average rating regression predicts Hong Kong's rating to be almost three notches higher than its actual rating. Of course, Hong Kong's actual rating reflects the risks inherent in its 1997 incorporation into China. If the regression had failed to identify Hong Kong as an outlier, we would suspect it was misspecified and/or overfitted.

Our statistical results suggest that Moody's and Standard and Poor's broadly share the same rating criteria, although they weight some variables differently (Table 5, columns 2 and 3). The general similarity in criteria should not be surprising given that the agencies agree on individual ratings more than half the time and most of their disagreements are small in magnitude. The fourth column of Table 5 reports a regression of rating differences (Moody's less Standard and Poor's ratings) against these variables. Focusing only on the statistically significant coefficients, we find that Moody's appears to place more weight on external debt and less weight on default history as negative factors than does Standard and Poor's. Moreover, Moody's places less weight on per capita income as a positive factor.[8]

In addition to the relationship between a country's economic indicators and its sovereign ratings, the effect of ratings on yields is of interest to market practitioners. Although ratings are clearly *correlated* with yields, it is far from obvious that ratings actually *influence* yields. The observed correlation could be coincidental if investors and rating agencies share the same interpretation of a body of public information pertaining to sovereign risks. In the next section, we investigate the degree to which ratings explain yields. After examining a cross-section of yields, ratings, and other potential explanatory factors at one point in time, we examine the movement of yields when rating announcements occur.

THE CROSS-SECTIONAL RELATIONSHIP BETWEEN RATINGS AND YIELDS

In the fall of 1995, thirty-five countries rated by both Moody's and Standard and Poor's had actively traded Eurodollar bonds. For each country, we identified its most liquid Eurodollar bond and obtained its spread over U.S. Treasuries as reported by Bloomberg L.P. on September 29, 1995. A regression of the log of these countries' bond spreads against their average ratings shows that ratings have considerable power to explain sovereign yields (Table 6, column 1).[9] The single rating variable explains 92 percent of the variation in spreads, with a standard error of 20 basis points. We also tried a number of alternative regressions based on Moody's and Standard and Poor's ratings, but none significantly improved the fit.[10]

Sovereign yields tend to rise as ratings decline. This pattern is evident in Chart 1, which plots the observed sovereign bond spreads as well as the predicted values from the average rating specification. An additional plot of average corporate spreads at each rating shows that

Table 6
DO RATINGS ADD TO PUBLIC INFORMATION?

	Dependent Variable: Log (Spreads)		
	(1)	(2)	(3)
Intercept	2.105***	0.466	0.074
	(16.148)	(0.345)	(0.071)
Average ratings	-0.221***		-0.218***
	(19.715)		(4.276)
Per capita income		-0.144	0.226
		(0.927)	(1.523)
GDP growth		-0.004	0.029
		(0.142)	(1.227)
Inflation		0.108	-0.004
		(1.393)	(0.068)
Fiscal balance		-0.037	-0.02
		(1.557)	(1.045)
External balance		-0.038	-0.023
		(1.29)	(1.008)
External debt		0.003***	0.000
		(2.651)	(0.095)
Indicator for economic development		-0.723**	-0.38
		(2.059)	(1.341)
Indicator for default history		0.612***	0.085
		(2.577)	(0.385)
Adjusted R-squared	0.919	0.857	0.914
Standard error	0.294	0.392	0.304

Sources: Moody's; Standard and Poor's; World Bank; International Monetary Fund; Bloomberg L.P.; Salomon Brothers; J.P. Morgan; Federal Reserve Bank of New York estimates.

Notes: The sample size is thirty-five. Absolute t-statistics are in parentheses.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

sovereign bonds rated below A tend to be associated with higher spreads than comparably rated U.S. corporate securities. One interpretation of this finding is that although financial markets generally agree with the agencies' relative ranking of sovereign credits, they are more pessimistic than Moody's and Standard and Poor's about sovereign credit risks below the A level.

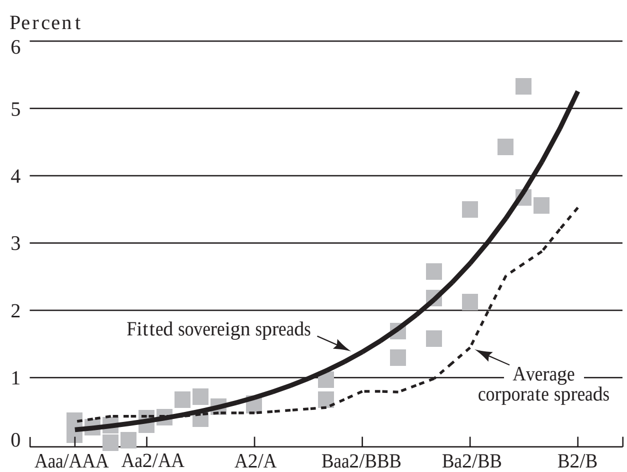
Our findings suggest that the ability of ratings to explain relative spreads cannot be wholly attributed to a mutual correlation with standard sovereign risk indicators. A regression of spreads against the eight variables used to predict credit ratings explains 86 percent of the sample variation (Table 6, column 2). Because ratings alone explain 92 percent of the variation, ratings appear to provide additional information beyond that contained in the standard macroeconomic country statistics incorporated in market yields.

In addition, ratings effectively summarize the information contained in macroeconomic indicators.[11] The third column in Table 6 presents a regression of spreads against average ratings and all the determinants of average ratings collectively. In this specification, the average rating coefficient is virtually unchanged from its coefficient in the first column of Table 6, and the other variables are collectively and individually insignificant. Moreover, the adjusted R-squared in the third specification is lower than in the first, implying that the macroeconomic indicators do not add any statistically significant explanatory power to the average rating model.

Our findings suggest that the ability of ratings to explain relative spreads cannot be wholly attributed to a mutual correlation with standard sovereign risk indicators.

Chart 1

Sovereign Bond Spreads by Credit Rating
As of September 29, 1995



Sources: Bloomberg L.P.; J.P. Morgan; Moody's; Salomon Brothers; Standard and Poor's.

Notes: The fitted curve is obtained by regressing the log (spreads) against the sovereigns' average. Average corporate spreads on five-year bonds are reported by Bloomberg L.P.

The results of our cross-sectional tests agree in part with those obtained from similar tests of the information content of corporate bond ratings (Ederington, Yawitz, and Roberts 1987) and municipal bond ratings (Moon and Stotsky 1993). Like the authors of these studies, we conclude that ratings may contain information not available in other public sources. Unlike these authors, however, we find that standard indicators of default risk provide no useful information for predicting yields over and above their correlations with ratings.

The Impact of Rating Announcements on Dollar Bond Spreads

We next investigate how dollar bond spreads respond to the agencies' announcements of changes in their sovereign risk assessments. Certainly, many market participants are aware of specific instances in which rating announcements led to a change in existing spreads. Table 7 presents four recent examples of large moves in spread that occurred around the time of widely reported rating changes.

Of course, we do not expect the market impact of rating changes to be this large on average, in part because many rating changes are anticipated by the market. To move beyond anecdotal evidence of the impact of rating announcements, we conduct an event study to measure the effects of a large sample of rating announcements on yield spreads. Similar event studies have been undertaken to measure the impact of rating announcements on U.S. corporate bond and stock returns. In the most recent and most thorough of these studies, Hand, Holthausen, and Leftwich (1992) show that rating announcements directly affect corporate securities prices, although market anticipation often mutes the average effects.[12]

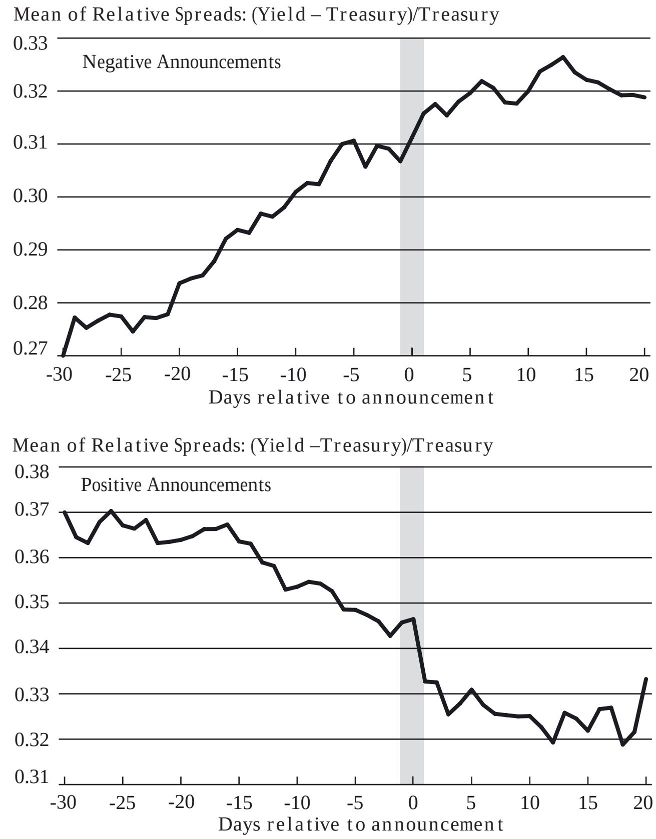
To construct our sample, we attempt to identify every announcement made by Moody's or Standard and Poor's between 1987 and 1994 that indicated a change in sovereign risk assessment for countries with dollar bonds that traded publicly during that period. Altogether, we gather a sample of seventy-nine such announcements in eighteen countries.[13] Thirty-nine of the announcements report actual rating changes—fourteen upgrades and twenty-five downgrades. The other forty announcements are "outlook" (Standard and Poor's term) or "watchlist" (Moody's term) changes:[14] twenty-three ratings were put on review for possible upgrade and seventeen for possible downgrade.

We then examine the average movement in credit spreads around the time of negative and positive announcements. Chart 2 shows the movements in relative yield spreads—yield spreads divided by the appropriate U.S. Treasury rate—thirty days before and twenty days after rating announcements. We focus on relative spreads because studies such as Lamy and Thompson (1988) suggest that they are more stable than absolute spreads and fluctuate less with the general level of interest rates.

Agency announcements of a change in sovereign risk assessments appear to be preceded by a similar change in the market's assessment of sovereign risk. During the twenty-nine days preceding negative rating announcements, relative spreads rise 3.3 percentage points on an average cumulative basis. Similarly, relative spreads fall

Chart 2

Trends in Sovereign Bond Spreads before and after Rating Announcements



Sources: Bloomberg L.P.; J.P. Morgan; Federal Reserve Bank of New York estimates.

Notes: The shaded areas in each panel highlight the period during which announcements occur. Spreads are calculated as the yield to maturity of the benchmark dollar bond for each sovereign minus the yield of the U.S. Treasury of comparable maturity. The charts are based on forty-eight negative and thirty-one positive announcements.

Table 7

LARGE MOVEMENTS IN SOVEREIGN BOND SPREADS AT THE TIME OF RATING ANNOUNCEMENTS

Country	Date	Agency	Old Rating => New Rating	Old Spread => New Spread (In Basis Points)
DOWNGRADES				
Canada	June 2, 1994	Moody's	Aaa=>Aa1	13=>22
Turkey	March 22, 1994	Standard and Poor's	BBB-=>BB	371=>408
UPGRADES				
Brazil	November 30, 1994	Moody's	B2=>B1	410=>326
Venezuela	August 7, 1991	Moody's	Ba3=>Ba1	274=>237

Sources: Moody's; Standard and Poor's; Bloomberg L.P.; J.P. Morgan.

Note: The old (new) spread is measured at the end of the trading day before (after) the announcement day.

about 2.0 percentage points during the twenty-nine days preceding positive rating announcements. The trend movement in spreads disappears approximately six days before negative announcements and flattens shortly before positive announcements. Following the announcements, a small drift in spread is still discernible for both upgrades and downgrades.

Do rating announcements themselves have an impact on the market's perception of sovereign risk? To capture the immediate effect of announcements, we look at a two-day window—the day of and the day after the announcement—because we do not know if the announcements occurred before or after the daily close of the bond market. Within this window, relative spreads rose 0.9 percentage points for negative announcements and fell 1.3 percentage points for positive announcements. Although these movements are smaller in absolute terms than the cumulative movements over the preceding twenty-nine days, they represent a considerably larger change on a daily basis.[15] These results suggest that rating announcements themselves may cause a change in the market's assessment of sovereign risk.

To move beyond anecdotal evidence of the impact of rating announcements, we conduct an event study to measure the effects of a large sample of rating announcements on yield spreads.

Statistical analysis confirms that for the full sample of seventy-nine events, the impact of rating announcements on dollar bond spreads is highly significant.[16] Table 8 reports the mean and median changes in the log of the relative spreads during the announcement window for the full sample as well as for four pairs of rating announcement categories: positive versus negative announcements, rating change versus outlook/watchlist change announcements, Moody's versus Standard and Poor's announcements, and announcements concerning investment-grade sovereigns versus announcements concerning speculative-grade sovereigns.[17] Because positive rating announcements should be associated with negative changes in spread, we multiply the changes in the log of the relative spread by -1 when rating announcements are positive. This adjustment allows us to interpret all positive changes in spread, regardless of the announcement, as being *in the direction expected given the announcement.*

Roughly 63 percent of the full sample of rating announcements are associated with changes in spread in the expected direction during the announcement period,

Table 8
DO DOLLAR BOND SPREADS RESPOND TO RATING ANNOUNCEMENTS?
Changes in Relative Spreads at the Time of Rating Announcements

	Number of Observations	Mean Change	Z-Statistic	Median Change	Percent Positive
All announcements	79	0.025	2.38***	0.020	63.3***
Positive announcements	31	0.027	2.37***	0.024	64.5**
Negative announcements	48	0.023	1.15	0.017	62.5**
Rating changes	39	0.035	2.49***	0.026	61.5**
Outlook/watchlist changes	40	0.015	0.88	0.014	65.0**
Moody's announcements	29	0.048	2.86***	0.022	69.0**
Standard and Poor's announcements	50	0.011	0.81	0.016	60.0**
Investment grade	52	0.018	0.42	0.015	53.9
Speculative grade	27	0.038	3.49***	0.026	81.5***

Notes: Relative spreads are measured in logs, that is, ln [(yield – Treasury)/Treasury)]. Changes in the logs of relative spreads are multiplied by -1 in the case of positive announcements. Significance for the percent positive statistic is based on a binomial test of the hypothesis that the underlying probability is greater than 50 percent.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

with a mean change in the log of relative spreads of about 2.5 percent. This finding is consistent with the announcement effect for U.S. corporate bonds documented by Hand, Holthausen, and Leftwich (1992). In fact, the share of responses in the expected direction is consistently above 50 percent regardless of the category of rating announcement. Moreover, the mean changes are always positive regardless of category.

Tests of statistical significance do suggest some differences between categories, however. Most strikingly, by both the mean change and percent positive measures, rating announcements have a highly significant impact on speculative-grade sovereigns but a statistically insignificant effect on investment-grade sovereigns. (By contrast, Hand, Holthausen, and Leftwich find that rating announcements have a significant impact on both investment-grade and speculative-grade corporate bonds.) Table 8 also reveals that the mean change statistics are not significant for negative announcements,[18] outlook/watchlist announcements, and Standard and Poor's announcements, although the percent positive statistics are significant for those categories. Because the statistical inferences for certain categories are ambiguous, and because the various categories overlap, we employ a multiple regression to sort out which categories of rating announcements imply meaningfully different effects on spreads.

We run a regression of the change in relative spreads against four indicator variables that take on the value 1 (or 0) depending on whether (or not) the rating announcements involve actual rating changes, positive events, Moody's decisions, or speculative-grade sovereigns (Table 9, column 1). As might be expected from Table 8, the estimated coefficients are all positive. Only the coefficients on the Moody's and speculative-grade indicator variables, however, are statistically significant.[19] Thus, the multiple regression indicates that the immediate impact of

Table 9

WHAT DETERMINES REACTIONS TO RATING ANNOUNCEMENTS?

Weighted Regressions of Changes in Relative Spreads on Explanatory Factors

	(1)	(2)	(3)	(4)	(5)
Constant	-0.02	-0.01	-0.03*	-0.02	-0.02
	(0.97)	(0.39)	(1.73)	(1.11)	(1.4)
Positive announcements	0.01	0.01	0.00	0.01	0.01
	(0.72)	(0.53)	(0.11)	(1.02)	(0.34)
Rating changes	0.02	0.01	0.01	0.00	-0.01
	(1.04)	(0.81)	(0.58)	(0.13)	(0.37)
Moody's announcements	0.03*	0.03	0.03*	0.02	0.02
	(1.8)	(1.61)	(1.92)	(1.53)	(1.51)
Speculative grade	0.03**	0.03**	0.03**	0.03*	0.03**
	(1.98)	(2.25)	(2.24)	(1.67)	(2.33)
Change in relative spreads from day -60 to day -1	–	-0.05	–	–	-0.06
		(0.98)			(1.1)
Rating gap indicator	–	–	0.04**	–	0.03*
			(2.34)		(1.7)
Other rating announcements from day -60 to day -1	–	–	–	0.05**	0.05**
				(2.42)	(2.15)
Adjusted R-squared	0.05	0.03	0.10	0.11	0.12

Notes: Absolute t-statistics are in parentheses. Relative spreads are measured in logs, that is, ln [(yield – Treasury)/Treasury]. Changes in the logs of relative spreads are multiplied by -1 in the case of positive announcements. Variables are weighted in the regressions by the inverse of the standard deviation of daily change in the log of relative spreads from day -100 to day -10.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

an announcement on yield spreads is greater if the announcement is made by Moody's or if it is related to speculative-grade credit. By contrast, the impact of announcements does not appear to rely on the distinction between rating changes and outlook/watchlist changes or the distinction between positive and negative announcements.

We have established the impact of certain rating announcements on dollar bond spreads, but a second question arises: to what extent does anticipation by the market dilute the impact of these announcements? The presence of many well-anticipated events in our dataset could obscure highly significant responses to unanticipated announcements—including, perhaps, announcements by Standard and Poor's or announcements concerning investment-grade sovereigns.[20]

To pursue this issue, we construct three proxies for anticipation—changes in relative spreads, rating gaps between the agencies, and other rating announcements—all of which measure conditions before the announcement. The first proxy measures the change in relative spread (in the direction of the anticipated change) over the sixty days preceding the event. Prior movements in the relative spread may reflect the market's incorporation of information used by the agency in making the announcement. The second proxy indicates the sign of the gap between the rating of the agency making the announcement and the other agency's rating. An announcement that brings one agency's rating into line with the other's may be expected by market participants. In our regressions, the rating gap equals 1 (0) if the announcement moves the two agencies' risk assessments closer together (further apart). The third proxy is an indicator variable that equals 1 if another rating announcement of the same sign had occurred during the previous sixty days. This proxy is motivated by considerable evidence that rating announcements tend to be positively correlated—that is, positive announcements are more likely to be followed by positive announcements than by negative announcements and vice versa.[21]

Contrary to our expectations, . . . the impact of one agency's announcement is greater if the announcement confirms the other agency's rating or a previous rating announcement.

We use each of the anticipation proxies in turn as a fifth explanatory variable in a multiple regression that includes the four indicator variables for actual rating changes, positive events, Moody's decisions, or speculative-grade sovereigns. A final regression adds all three anticipation proxy variables simultaneously to the basic regression (Table 9, columns 2-5).

Our earlier results are robust to the addition of the proxy variables. Announcements by Moody's and announcements pertaining to speculative-grade sovereigns continue to have a larger impact than announcements by Standard and Poor's or announcements pertaining to investment-grade sovereigns. (Note, however, that the statistical significance of the differences between the effects of the different rating agencies declines below the 10 percent level in three of the four new specifications.)

Contrary to our expectations, however, the results reported in Table 9 suggest that market anticipation does not reduce significantly, if at all, the impact of a sovereign rating announcement. The estimated coefficient on the change in the relative spreads variable has the expected negative sign, but it is not statistically significant. Moreover, the estimated coefficients on both the rating gap and the other rating announcement indicators are unexpectedly positive and highly significant. According to these two measures, the impact of one agency's announcement is greater if the announcement confirms the other agency's rating or a previous rating announcement.

Conclusion

Sovereign credit ratings receive considerable attention in financial markets and the press. We find that the ordering of risks they imply is broadly consistent with macroeconomic fundamentals. Of the large number of criteria used by Moody's and Standard and Poor's in their assignment of sovereign ratings, six factors appear to play an important role in determining a country's rating: per capita income, GDP growth, inflation, external debt, level of economic development, and default history. We do not find any systematic relationship between ratings and either fiscal or current deficits, perhaps because of the endogeneity of fiscal policy and international capital flows.

Our analysis also shows that sovereign ratings effectively summarize and supplement the information contained in macroeconomic indicators and are therefore strongly correlated with market-determined credit spreads. Most of the correlation appears to reflect similar interpretations of publicly available information by the rating agencies and by market participants. Nevertheless, we find evidence that the rating agencies' opinions independently affect market spreads. Event study analysis broadly confirms this qualitative conclusion: it shows that the announcements of changes in the agencies' sovereign risk opinions are followed by bond yield movements in the expected direction that are statistically significant.

Although our event study results largely corroborate the findings of corporate sector studies, a few of our observations are surprising and invite further investigation. Our finding that the impact of rating announcements on spreads is much stronger for below-investment-grade than for investment-grade sovereigns is one puzzle. Another surprising result is that rating announcements that are more fully anticipated, at least by our proxy measures, have, if anything, a larger impact than those that are less anticipated.

In sum, although the agencies' ratings have a largely predictable component, they also appear to provide the market with information about non-investment-grade sovereigns that goes beyond that available in public data. The difficulty in measuring sovereign risk, especially for below-investment-grade borrowers, is well known. Despite this difficulty—and perhaps because of it—sovereign credit ratings appear to be valued by the market in pricing issues.

ENDNOTES

1. Although many studies have attempted to quantify the determinants of corporate and municipal bond ratings (see, for example, Ederington and Yawitz 1987; Moon and Stotsky 1993), our study is the first to quantify the determinants of the sovereign ratings assigned by Moody's and Standard and Poor's. Earlier researchers in the area of sovereign risk evaluated other measures of risk or presented a qualitative assessment of sovereign credit ratings. For example, Feder and Uy (1985) and Lee (1993) analyzed ordinal rankings of sovereign risk based on a poll of international bankers reported semiannually in *Institutional Investor*. Taylor (1995) discussed the importance of some of the same variables we examine, but he did not attempt to measure their individual and collective explanatory power.

2. Cantor and Packer (1995) provide a broad overview of the history and uses of sovereign ratings and the frequency of disagreement between Moody's and Standard and Poor's.

3. These variables also correspond closely to the determinants of default cited in the large academic literature on sovereign credit risk. See, for example, Saini and Bates (1984) and McFadden et al. (1985). This literature, focused largely on developing countries, estimates the importance of select variables in determining the probability that sovereign bank loans will default within one year. We do not, of course, analyze every variable considered in this literature. International reserves, a good indicator of short-term distress for developing economies, are unlikely to be helpful in explaining sovereign ratings, which measure default risk over a multiyear horizon for both developed and developing economies. We therefore do not consider this variable in our analysis.

4. Because of data limitations, we use central government debt as our measure of fiscal balance, although a more satisfactory measure would be the consolidated deficits of the federal, state, local, and quasi-public sectors.

5. Debtors undoubtedly care about a country's total debt burden, not just its foreign currency debts. Nonetheless, Moody's stresses that foreign currency obligations are generally given greater weight than total external liabilities in their sovereign ratings (Moody's 1991, p. 168).

Other measures of debt burden are also likely to be important, but they are not available for both developed and developing countries. Two such variables are net foreign assets and debt-servicing costs, both of which can be measured in domestic and foreign currencies. Although we do not measure these two factors directly, they are correlated with variables we do measure—net foreign assets represent the accumulation of past current account surpluses, and foreign currency debt service is roughly proportional to foreign currency debt. The maturity of external liabilities is another important debt-related variable of interest, but it is not generally available for most countries.

6. Countries with higher levels of development may also be less inclined to default on their foreign obligations because their economies are often substantially integrated with the world economy. As a result, developed economies are particularly vulnerable to the legal rights of creditors to disrupt trade or seize assets abroad. According to one strand of the theoretical literature on sovereign debt, the possibility of recourse to direct sanctions is a necessary condition for sovereign lending (Bulow and Rogoff 1989).

7. Although this estimation technique suffers from the limitation that ratings are treated as cardinal variables, it is the only feasible approach given that we have just forty-nine jointly rated sovereigns and sixteen potential rating categories. We found that the simple linear specification of the rating variable worked considerably better than nonlinear alternatives such as logarithmic or exponential functions. We also tried unsuccessfully to estimate the relationships with ordered probit techniques, relying only on the ordinal properties of credit ratings. Because of the large number of rating categories and the relatively few sovereign rating assignments, our attempts to implement this approach were hindered by a failure of the maximum likelihood estimates to converge. In a similar study of corporate ratings, Ederington (1985) suggests that with larger sample sizes, inferences drawn from ordered probits are likely to be similar to, and perhaps slightly more accurate than, those drawn from least squares regressions. In contrast, in their study of corporate bond ratings, Kaplan and Urwitz (1979) argue that linear least squares estimators perform better out of sample than those estimators derived from ordered probits.

8. These results were confirmed by ordered-probit regressions for rating differences. Although not reported here, the results of the probit regressions are available from the authors on request.

9. The relationship between ratings and yields is nonlinear; hence, we report our preferred specification of the natural logarithm of yields against ratings. This specification eliminates heteroskedasticity in the residuals as measured against rating levels.

10. Specifically, we included ratings from one agency at a time or selected either the higher or the lower of the two ratings for each country. We also tried adding two dummy variables to the average rating regressor: one that indicated whether or not the two agencies disagreed and, separately, one that indicated the identity of the agency with the higher rating.

11. This conclusion holds whether or not the sovereign is investment grade: separate regressions for investment-grade and speculative-grade subsamples look very similar to the full-sample regressions.

12. Because bond data are less readily available, event studies on stock

prices dominate the corporate rating literature. The event studies using bond data that precede Hand, Holthausen, and Leftwich (1992) focus solely on monthly observations and conclude that bond prices are not affected by rating changes (Weinstein 1977; Pinches and Singleton 1978; Wakeman 1984; Ederington and Yawitz 1987). A more recent study by Hite and Warga (1996) also uses monthly bond price data, but finds a significant announcement effect for downgraded firms.

13. We obtained the bond yield data by searching the daily time series data on Euro, Yankee, Global, and Brady bonds reported by Bloomberg L.P. and J.P. Morgan and made available to us by J.P. Morgan. For our event study, we used Bloomberg data for fifteen countries (Argentina, Australia, Belgium, Brazil, Canada, Colombia, Denmark, Finland, Ireland, Italy, Malaysia, New Zealand, Sweden, Thailand, and Turkey) and fifty-seven rating announcements. We use J.P. Morgan data for seven countries (Argentina, Brazil, Colombia, Hungary, the Philippines, Turkey, and Venezuela) and twenty-three rating announcements.

14. Standard and Poor's always indicates whether a sovereign has a positive, negative, or stable outlook, and many of its rating announcements report a change in this outlook alone. The agency also occasionally places a sovereign on review for probable upgrade or downgrade. Moody's does not indicate an outlook per se; however, it frequently places sovereigns on its watchlist for upgrades and downgrades.

15. Compare a 0.5 daily percentage point change during the announcement window for negative announcements with an average daily change of 0.1 for the preceding twenty-nine days. Similarly, compare a 0.7 daily percentage point change during the announcement window for positive announcements with an average daily change of 0.1 for the preceding period.

16. In the calculation of statistical significance, we control for potential heteroskedasticity with a procedure used by Mikkelson and Partch (1986) and Billet, Garfinkel, and O'Neal (1995). For each group of announcements, we calculate weighted (standardized) means in which the weights equal the inverse of the standard deviation of the relevant daily changes in the logged relative bond spread calculated during the ninety-day period ending ten days before the announcement day. The Z-statistic for significance is the standardized mean times the square root of the number of announcements.

17. To be consistent with the log-linear relationship between ratings and spreads depicted in Chart 1, we report mean and median changes to the log of the relative spread, although the results are not particularly sensitive to this aspect of the specification.

18. By contrast, most studies using stock market data find a significant price reaction to downgrades but not to upgrades (Goh and Ederington 1993).

19. Because the average absolute errors of the regression are larger when ratings are lower, we employ weighted least squares to control for this source of heteroskedasticity.

20. Hand, Holthausen, and Leftwich (1992) find that Standard and Poor's announcements that corporate ratings are under review have significant market impact only when announcements classified by the authors as "expected" are excluded from the sample.

21. Of the 109 sovereign rating announcements between 1987 and 1994 that were followed by a rating change, 86 were followed by a change in the same direction. (Similarly, Altman and Kao [1991] have shown that corporate rating changes are often followed by further changes in the same direction.) Of the 79 rating announcements in our sample, 36 were preceded by a rating gap in the implied direction of the announcement. In 20 cases, other rating announcements in the same direction had been made in the preceding sixty days.

Altman, Edward, and Duen Li Kao. 1991. "Corporate Bond Rating Drift:

An Examination of Rating Agency Credit Quality Changes Over Time." New York University-Salomon Brothers Working Paper S-91-40.

Billet, Matthew. 1996. "Targeting Capital Structure: The Relationship Between Risky Debt and the Firm's Likelihood of Being Acquired." JOURNAL OF BUSINESS. Forthcoming.

Billet, Matthew, Jon Garfinkel, and Edward O'Neal. 1995. "Insured Deposits, Market Discipline, and the Price of Risk in Banking." Federal Deposit Insurance Corporation, manuscript.

Bulow, Jeremy, and Kenneth Rogoff. 1989. "Sovereign Debt: Is to Forgive to Forget?" AMERICAN ECONOMIC REVIEW 79, no. 1: 43-50.

Cantor, Richard, and Frank Packer. 1994. "The Credit Rating Industry." Federal Reserve Bank of New York QUARTERLY REVIEW 19, no. 2 (winter): 1-26.

———. 1995. "Sovereign Credit Ratings." Federal Reserve Bank of New York CURRENT ISSUES IN ECONOMICS AND FINANCE 1, no. 3 (June).

Eaton, Jonathan. 1996. "Sovereign Debt, Repudiation, and Credit Terms." INTERNATIONAL JOURNAL OF FINANCE AND ECONOMICS 1, no. 1 (January): 25-36.

Ederington, Louis. 1985. "Classification Models and Bond Ratings." FINANCIAL REVIEW 4, no. 20 (November): 237-62.

Ederington, Louis, and Jess Yawitz. 1987. "The Bond Rating Process." In Edward Altman, ed., HANDBOOK OF FINANCIAL MARKETS. New York: John Wiley & Sons: 23-57.

Ederington, Louis, Jess Yawitz, and Brian Roberts. 1987. "The Information Content of Bond Ratings." JOURNAL OF FINANCIAL RESEARCH 10, no. 3 (fall): 211-26.

Feder, G., and L. Uy. 1985. "The Determinants of International Creditworthiness and Their Implications." JOURNAL OF POLICY MODELING 7, no. 1: 133-56.

Goh, Jeremy, and Louis Ederington. 1993. "Is a Bond Rating Downgrade Bad News, Good News, or No News for Stockholders?" JOURNAL OF FINANCE 48, no. 5: 2001-8.

Hand, John, Robert Holthausen, and Richard Leftwich. 1992. "The Effect of Bond Rating Agency Announcements on Bond and Stock Prices." JOURNAL OF FINANCE 47, no. 2: 733-52.

Hite, Gailen, and Arthur Warga. 1996. "The Effect of Bond Rating Changes on Bond Price Performance." Unpublished paper.

Horrigan, J. 1966. "The Determination of Long-Term Credit Standing with Financial Ratios." EMPIRICAL RESEARCH IN ACCOUNTING 1966, JOURNAL OF ACCOUNTING RESEARCH 4 (supplement): 44-62.

Kaplan, Robert, and Gabriel Urwitz. 1979. "Statistical Models of Bond Ratings: A Methodological Inquiry." JOURNAL OF BUSINESS 52, no. 2: 231-61.

Lamy, Robert, and G. Rodney Thompson. 1988. "Risk Premia and the Pricing of Primary Issue Bonds." JOURNAL OF BANKING AND FINANCE 12, no. 4: 585-601.

Lee, Suk Hun. 1993. "Are the Credit Ratings Assigned by Bankers Based on the Willingness of LDC Borrowers to Repay?" JOURNAL OF DEVELOPMENT ECONOMICS 40: 349-59.

McFadden, Daniel, Richard Eckaus, Gershon Feder, Vassilis Hajivassiliou, and Stephen O'Connell. 1985. "Is There Life After Debt? An Econometric Analysis of the Creditworthiness of Developing Countries." In Gordon Smith and John Cuddington, eds., INTERNATIONAL DEBT AND THE DEVELOPING COUNTRIES. Washington, D.C.: World Bank.

Mikkelson, W., and M. Partch. 1986. "Valuation Effects of Security Offerings and the Issuance Process." JOURNAL OF FINANCIAL ECONOMICS 15, no. 1/2: 31-60.

Moody's Investors Service. 1991. GLOBAL ANALYSIS. London: IFR Publishing.

———. 1995. SOVEREIGN SUPRANATIONALS CREDIT OPINIONS, September.

Moon, C.G., and J.G. Stotsky. 1993. "Testing the Differences between the Determinants of Moody's and Standard and Poor's Ratings." JOURNAL OF APPLIED ECONOMETRICS 8, no. 1: 51-69.

Ozler, Sule. 1991. "Evolution of Credit Terms: An Empirical Examination of Commercial Bank Lending to Developing Countries." JOURNAL OF DEVELOPMENT ECONOMICS 38: 79-97.

REFERENCES (*Continued*)

Pinches, G., and J. Singleton. 1978. "The Adjustment of Stock Prices to Bond Rating Changes." JOURNAL OF FINANCE 33, no. 1: 29-44.

Saini, K., and P. Bates. 1984. "A Survey of the Quantitative Approaches to Country Risk Analysis." JOURNAL OF BANKING AND FINANCE 8, no. 2: 341-56.

Standard and Poor's. 1994. "Sovereign Rating Criteria." EMERGING MARKETS, October: 124-7.

Taylor, Joseph. 1995. "Analyzing the Credit and Sovereign Risks of Non-U.S. Bonds." In Ashwinpaul C. Sondhi, ed., CREDIT ANALYSIS OF NONTRADITIONAL DEBT SECURITIES. New York: Association for Investment Research, pp. 72-82.

Wakeman, L. 1984. "The Real Function of Bond Rating Agencies." In Michael Jensen and Clifford Smith, eds., THE MODERN THEORY OF CORPORATE FINANCE. New York: McGraw-Hill.

Weinstein, M. 1977. "The Effect of a Rating Change Announcement on Bond Price." JOURNAL OF FINANCIAL ECONOMICS 5, no. 3: 329-50.