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

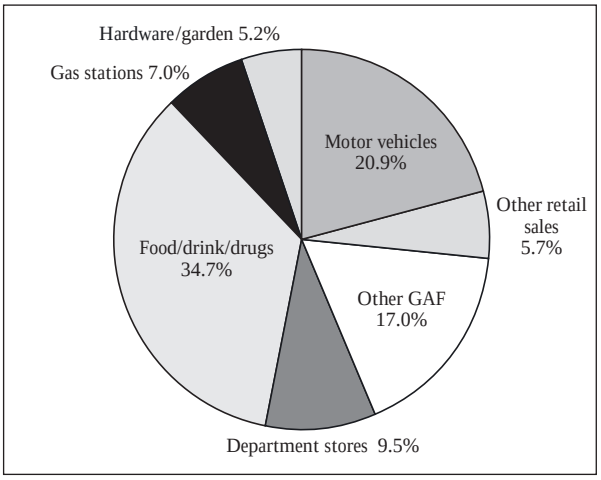
Source: Department of Commerce, Bureau of the Census (1995).

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

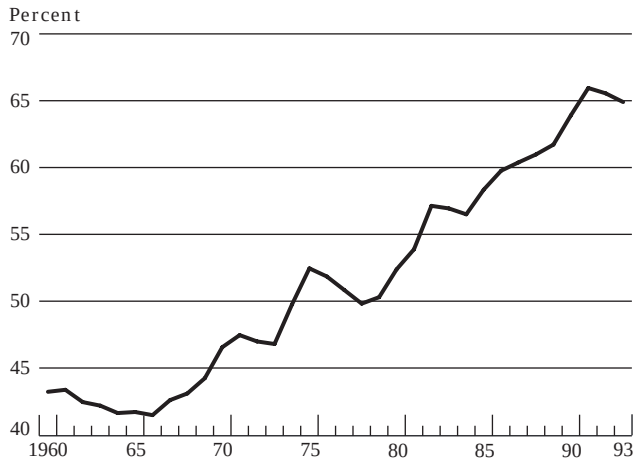
Spurred by easy lending terms and generous tax laws, commercial construction boomed in the early 1980s, with real spending roughly doubling from 1983 to 1986.[7] This favorable investment climate changed in the late 1980s, and by 1992 commercial construction had dropped below its 1983 levels. Since then, 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.

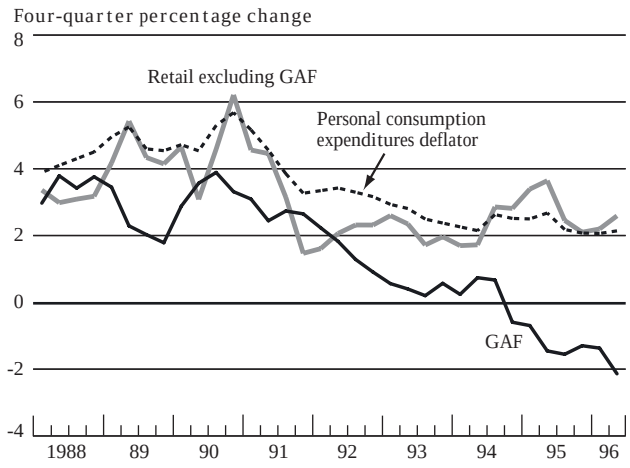
Together with the oversupply of stores, this shift in demand has also put downward pressure on prices at major retail firms. The inflation rate for goods sold at GAF stores has been consistently lower than broad measures of consumer prices such as the personal consumption deflator and has generally trailed 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

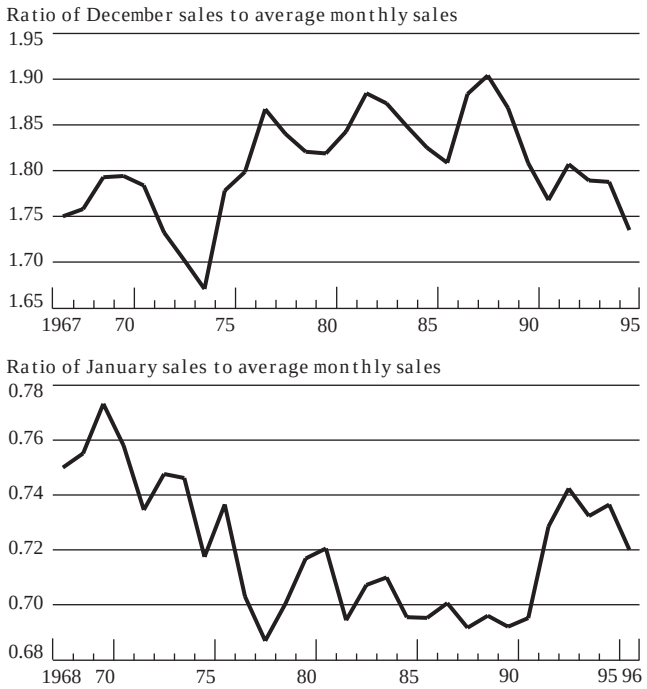
Consumer Inflation Trends



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.

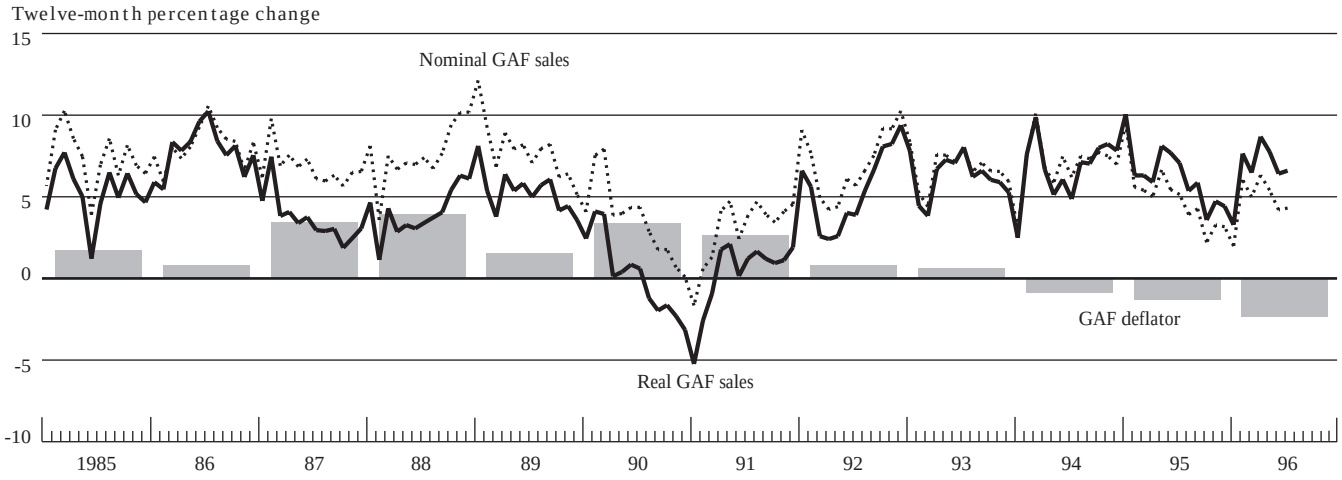
IMPLICATIONS FOR FORMAL FORECASTS AND INFORMAL COMMENTARY

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 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.

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

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

	GAF Sales		Non-Auto Retail Sales: Advance Data		Non-Auto Retail Sales: 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.

The views expressed in this article are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. The Federal Reserve Bank of New York provides no warranty, express or implied, as to the accuracy, timeliness, completeness, merchantability, or fitness for any particular purpose of any information contained in documents produced and provided by the Federal Reserve Bank of New York in any form or manner whatsoever.