Federal Reserve Bank of New York
Research Paper No. 9614

June 1996

What Do Chain Store Sales Tell Us About Consumer Spending?

by

Ethan S. Harris and Clara Vega

Abstract

In recent years the sales reports of major retail chains have received increasing attention as timely indicators of consumer spending. Despite this attention, a close review of the literature on chain store indexes as macroeconomic indicators reveals that there is no literature! This paper fills this gap, showing how chain store data fit into the broader issues of how we measure and forecast consumer spending. We describe the linkages between chain store data and official measures of consumer spending, highlighting the key seasonality and pricing issues. We then present a battery of in-sample and out-of-sample tests to determine what are the best models for monthly forecasts of retail sales and personal consumption expenditures. While our results question the way chain store data are sometimes used in forecasting, we find strong evidence that the chain store data can significantly improve forecast accuracy and we show that the best models combine several consumer-related variables with both of the leading chain store indexes. We conclude with some practical advice on the do's and don'ts of consumption forecasting.

What Do Chain Store Sales Tell Us About Consumer Spending?

by

Ethan S. Harris and Clara Vega

As forecasters search for increasingly timely data, sales reports from major retail chains have garnered increasing attention. Available on both a weekly and monthly basis, with a lag of just a few days, indexes of chain store sales are the first information on the largest sector of the economy, consumer spending. In the last several years there has been growing attention to chain store data among forecasters and financial market participants, such that they are now regularly featured in the business press.[1] During the recent holiday season, with growing concern about the consumer sector, chain store reports moved from the business page to the front page: weak sales in December was the feature story in early January, with headlines such as "Retailers Call Sales in December Worst Since '90-'91 Recession" and "Christmas Wasn't Merry for Many Stores, and the New Year Outlook is Little Happier."[2] This attention raises an important question: while chain store reports clearly are important for industry analysts and investors, how useful are they as a macroeconomic indicator?

---

[1] Attention has grown to the point that on the same day markets can tumble in response to a strong report for one chain store index and then rally in response to the release of a weak report for another index. 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 sales 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] These headlines are from the January 5, 1996 editions of the New York Times (page A1) and the Wall Street Journal (page B1).

Despite the strong and growing attention to chain store sales, a careful review of the literature on chain store sales as an economic indicator reveals that there is no literature! In particular, while 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.[3]

This edition of the Economic Policy Review attempts to fill this gap and in the process takes a comprehensive look at short-run forecasting of consumer spending. We start with some housekeeping, explaining what "chain stores" are, how they fit into the Commerce Department's taxonomy of consumer spending and how the two major chain store indexes--Johnson Redbook and Mitsubishi--are constructed. We then review some important structural changes in the retail sector. Company consolidation, excess store space and value conscious consumers have transformed retailing, causing a shift in the composition of stores toward large discounters, putting strong downward pressure on prices, inducing a significant change in the seasonal pattern of holiday sales, and in general contributing to unduly pessimistic views of underlying consumer demand. Understanding these changes is important for forecasting and tracking consumer spending in general and for interpreting chain store data in particular.

The paper then turns to formal statistical tests of whether chain store data are useful for forecasting the growth in the official measures of retail sales and personal consumption expenditure. Using both in-sample and out-of-sample tests we compare the prediction

---

[3] In their handbooks, Rogers (1994; p. 68) Tainer (1993; p. 59, 62-63 and 68-71) and Kuwayama and O'Sullivan (1996) provide background information on the chain store data. The Mitsubishi Bank (1994) 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.

performance of models using chain store indexes to alternative models using a wide range of consumer-related economic indicators. We find that given the appropriate weights in forecast models chain store indexes can be quite useful for forecasting. They add significantly to the accuracy of in-sample and out-of-sample predictions for several measures of consumer spending and overall the best models combine economic variables with both of the major chain store indexes. Two appendices look at the sampling and statistical properties of the various measures of consumer spending and explore the predictive power of additional chain store indexes. We conclude with a recommended strategy for short-run consumption forecasting and with some practical "do's and don'ts" for forecasters.

What are "Chain Stores?"

In press reports "chain stores," "department stores," "retail chains," "broadlines" and "major retailers" are often used more or less interchangeably, so it makes sense to pin down exactly what a chain store is at the outset. Based on the official Department of Commerce taxonomy, chain stores fit into two categories: a narrow category, called "Department Stores," or a more inclusive category, "General Merchandise, Apparel and Furniture" (GAF).[4]

A "Department Store" is simply a large store (with more than 50 employees) which sells a diverse range of merchandise--household linens, dry goods, home furnishings, appliances, radios and televisions, furniture, and a general line of apparel. As Table 1 shows, the typical Department Store is quite large, with sales and employment more than ten times the average retail establishment. Thus while Department Stores comprise less than one percent of all retail

[4] US Bureau of the Census (1995), Appendix F.

establishments, they account for about 10 percent of retail sales.

Within Department Stores there are three subcategories:

- Conventional Department Stores such as Federated/Macy and May Department Stores have a single or limited number of locations;
- National Chain Department Stores such as Sears and J.C. Penny are affiliated with a company with establishments across the nation; and
- Discount Department Stores such as Wal-Mart and K-mart "convey the image of a high-volume, fast turnover outlet," with low prices and limited customer service.

Some retail companies with a national chain of outlets do not sell the variety of merchandise that a Department Store carries. Thus, in some cases "chain store" reports include stores that the Commerce Department would not classify as a "Department Store." These stores do, however, belong to a broader component of retail sales--General Merchandise, Apparel and Furniture (GAF). This broader category includes stores that compete with Department Stores in the sense that they sell one or more of the same lines of merchandise as Department Stores. Although national chains play an important role in all of GAF, most of the sector consists of smaller, local stores. As shown in Table 1, the typical Furniture or Apparel establishment is much smaller than a Department store, employing less than 10 workers and with annual sales less than $1 million.

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. This can be illustrated in a simple two step fashion (Chart 1). First, as the left panel shows, Department Stores comprise only about 14 percent of nonauto retail sales, and even if we (generously) include all of GAF in our measure, sales of chain stores are representative of a sector that comprises about 34 percent of nonauto retail sales. Second, as the right panel shows, because of the large service and motor vehicle sectors, nonauto goods comprise less than half of total personal consumption . Thus, allowing for some minor accounting adjustments, chain stores represent, directly and indirectly, somewhere between 4 and 18 percent of personal consumption.

Two Chain Store Indexes

While a number of economists have created chain store indexes in recent years, the two longest running and most watched indexes are the "Chain Store Sales Index" from Mitsubishi Bank and Wertheim Schroder, and the "Retail Sales Index" from the Johnson Redbook Service. Both of these indexes are released just a few days after the period they measure and are available on both a weekly and a monthly basis. While these indexes cover many of the same companies, they differ in four key respects:

- **Sectoral Coverage:** Johnson Redbook focuses only on companies that fit the definition of Department Stores, while the Mitsubishi index also includes major chain stores that fit the broader GAF category.

- **Type of Store:** Johnson Redbook measures total company sales, while the Mitsubishi index includes only "same-store" sales--that is, sales from locations which have been open for at least a year.

- **Sample period:** Johnson Redbook reports both monthly and weekly data back to 1983; the more recently developed Mitsubishi index reports weekly data starting in 1989, but monthly data have been reconstructed back to 1969.

- **Seasonal adjustment:** Johnson Redbook calculates a seasonally adjusted dollar value for its index by taking the official Department Store data for twelve months earlier and then applying the year-over-year growth rate estimated from its sample. Mitsubishi estimates its own seasonal factors and reports the seasonally adjusted data as an index, equal to 100 in 1977.

While seasonally adjusted data are available for both series, in press reports both chain store indexes are usually reported as percent change from a year earlier.

An Industry in Transition

Before formally testing the statistical link between chain store sales and overall consumer spending, it is important to briefly review the recent evolution of the retail sector. Retail analysts argue that three interrelated structural forces are transforming the retail sector-- the chronic excess supply of retail space, the emergence of "value conscious consumers," and the growing concentration of sales in larger companies. It turns out that there is strong circumstantial evidence to support, and quantify, these structural changes. Each of these changes has important implications for how we translate the microeconomic information on individual companies into macroeconomic implications for overall consumer demand.

Excess Capacity

While retailers have always complained that the nation is "over stored" there is evidence to support this concern. Spurred by easy lending terms and generous tax laws, commercial construction boomed in the 1980s (Chart 2).[5] This favorable investment climate changed in the 1990s, slowing the growth of commercial construction, but unlike the office building component of commercial construction, retail and wholesale construction has since rebounded and now stands near its prior peak. A number of indicators suggest that the current flow of new space is not slow enough to curb the growth in retail capacity:

- The ratio of the real stock of retail space to real GDP continues to climb (Chart 3).

---

[5] Harris et al (1994) reviews the forces behind the boom and bust in commercial construction.

- The official retail space data are confirmed by statistics reported in the industry literature: 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 store space per capita grew from 7.0 to 18.7.[6]

Although some of this increase in capacity reflects a natural process of capital deepening as the economy grows, there are also telltale signs of excess capacity:

- The pressure on space has affected the stock market performance of major retailers. 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 S&P500 fell 23 percentage points relative to the overall index.

- Financial pressures have led to an increase in bankruptcies and store closings. In the nine month period from June 1995 to February 1996 nine major retailers filled for Chapter 11 protection, up from two in 1993 and in 1994 and higher than during the last recession.[7] Short of bankruptcy, there is also evidence of an accelerated pace of store closings by on-going companies.

Apparently bankruptcies and individual store closings have not solved the overhang of space, as shuttered stores have generally reopened under new names.

Value Conscious Consumers

Not surprisingly, retailers also complain that their customers don't pay enough for their goods. Again, this complaint is not entirely specious: there is concrete evidence that the industry is under substantial price pressure. Industry analysts describe a process similar to what happened to auto dealers starting in the late 1980s when price incentives started as a temporary device for reducing inventories and ended up being almost permanently in place. In the chain store sector, as consumers have become more value conscious, retailers have increased the frequency of

---

[6] Telsey (1996) p. 28.

[7] Kernkraut (1996) p. 11.

"sales." This in turn has prompted consumers to withhold spending until items go on sale, retailers find that they can not sell at full price and "sales" are repeated with increasing frequency.

This search for "value" has had a number of important effects. Consumer demand 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 their share in total sales from 44 to 60 percent (Chart 4).

This shift in demand (and the oversupply of stores) has also put downward pressure on prices at major retail firms. As Chart 5 shows, the inflation rate for goods sold at Department stores has been consistently lower than for the overall personal consumption deflator and has generally trailed the deflator for other retail sales as well. Indeed, the weak price performance at major retailers has worsened recently: Department Store prices have actually fallen over the last year and a half, widening the inflation gap to almost 3 percentage points.[8]

A final impact of value shopping has been a shift in the seasonal pattern of Department Store sales. Chain store sales are much more seasonal than other retail sectors. Based on the latest official seasonal adjustment factors, Department store sales typically surge 78 percent above their long-run average in December, and then plunge to 27 percent below average in January. By contrast, other nonauto retail sales-- such as grocery stores, restaurants, gas stations and hardware stores-- have a mild seasonal, rising just 25 percent above normal in December and dipping about 11 percent below normal in January.

---

[8] These competitive pressures may help explain why consumer prices have been relatively subdued, despite capacity pressures in the economy.

Over the last several years, value conscious shoppers have induced substantial shift in the holiday seasonal, delaying purchases in December to take advantage of lower prices in January. In particular, comparing the last five years (1991-95) to the previous five years (1987-91), the December peak in Department Store sales has dropped from 85 percent above normal to just 78 percent above normal (Chart 6). A large portion of these sales have shifted to January: based on the same five year comparison as above, sales are now 27 percent below normal in January, rather than 31 percent below normal.

Consolidation

A third structural change in the retail sector is that larger companies are growing at the expense of smaller retailers. This is impossible to quantify precisely, but it can be illustrated by comparing sales growth for firms included in the chain store indexes to sales for the GAF sector as a whole. For example, for the five years ending December 1995 the Merrill Lynch index of total chain store sales grew at a 11.8 percent annual rate, almost double the 6.9 percent pace for GAF. A closer examination of the data suggests that most of the 11.8 percent growth in chain store sales is either due to acquisition of existing stores (4.9 percentage points) or due to sales at newly built stores (2.4 percent), while only 4.5 percentage points of sales growth is at existing stores.[9]

There are few signs that the pace of change in retailing is abating. Two recent industry

[9] This rough calculation is done as follows. Sales growth at acquired stores is calculated by subtracting GAF sales from total chain store sales; same store sales is directly measured by Merrill Lynch; and sales at newly built stores is calculated by subtracting same store sales from GAF sales. Another sign of consolidation is the fact that department stores-- which are almost all large companies-- have been capturing an increasing share of GAF sales. The share has risen to 37 percent in 1995, up from 35 percent in 1990.

trends should ensure that the process of restructuring and concentration will continue. First, there has been a rapid expansion of a new type of store with the colorful name "category killers." These "big box" stores offer a full product line in a focused category of goods. Toys "R" Us is a very early example of such a store. Second, has been the emergence of "super stores," which combine a traditional discount store with a supermarket and a variety of smaller stores under one roof.

Implications for Forecasters

It is important to distinguish two kinds of information in the chain store data: the microeconomic information on how healthy companies are and the macroeconomic information on underlying consumer demand. The structural changes in retailing have been a source of ongoing concern for the industry and have contributed to the negative tone of commentary on the consumer sector as a whole. From a macro economist's perspective, however, it is important to factor out these structural distortions and assess underlying trends in consumer demand. This suggests some practical lessons for forecasters:

- Because of declining prices, relatively modest growth in nominal Department Store sales can translate into solid growth in real sales. For example, the seemingly enimic 3.5 percent nominal growth in the year to December 1995, which barely out paced overall consumer price inflation, actually translated into about a 4.7 percent real gain.

- Similarly, because of changing seasonals, retailers have reported "disappointing" Christmas sales for three years in a row, but using appropriate seasonal adjustment factors apparently weak Department Store sales in December can translate into reasonable gains. For example, unadjusted sales in December 1995 were 75 percent above the annual average, and based on the old seasonal adjustment factor this translates into a sharp drop in sales, but based on the new factors it represented a modest gain.

- Industry consolidation and expansion has distorted the "total" and "same store" figures reported in indexes. While indexes based on total sales (such as Johnson Redbook) have exaggerated the growth in total consumer spending, indexes based on same store sales (such as Mitsubishi) have underestimated sales growth.

Is a Poor Sample Better Than No Sample?

Clearly, chain store indexes suffer a number of drawbacks as macroeconomic indicators, and it would be inappropriate to treat them as a representative sample of overall consumer spending. Nonetheless, they have two potential advantages. First, and foremost, is their timeliness. Available on a weekly basis, the indexes provide real-time updates on the progress of spending during the month, and complete monthly chain store data are available more than a week before the release of the Commerce Department's first "advance" estimate of retail sales. As the data calender for March 1996 illustrates (Table 2), the only other direct measure of consumer spending available that early in the data cycle is motor vehicle sales, and these data tend to have very different monthly patterns than the rest of retail sales.[10]

A second potential advantage is that while chain stores may be a small part of the level of retail sales, they may be particularly sensitive to the ups and downs of consumer spending. For example, a simple variance decomposition confirms that department stores account for slightly more than their "share" of the month-to-month variation in nonauto retail sales. For example, for the 1985 to 1995 period the variance of nonauto sales growth was .350, with .045 due to department stores and .317 due to other sales.[11] Thus, Department Store sales accounted for

---

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

[11] The covariance between the two subcomponents accounted for -.012 of the total variance.

roughly 13 percent (.045 divided by .350) of the monthly variation in nonauto retail sales, even though they comprise just 10 percent of the level of sales.

Thus, whether chain store sales are useful for forecasting essentially comes down to the following: is a poorly constructed, unrepresentative sample better than no sample at all? In other words, can we cull useful new information without being mislead by the noise and biases in these data?

Testing Strategy: Rattling the Chains

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 measures of consumer spending, we compare their performance to a number of alternative models and we evaluate their performance both in-sample and out-of-sample. The end result of this exercise is not only an understanding of the role of chain store indexes in consumption forecasting, but a better understanding of what other variables belong in these models.

Dependent Variables

We test the predictive power of chain store sales for five variables of interest to forecasters: Department Store sales which cover essentially the same data as the Johnson index; GAF sales which roughly match the coverage of the Mitsubishi index; advance nonauto retail sales which is what financial-sector economists are most interested in tracking;[12] the latest, fully

[12] 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 for two reasons. First, as we argued in footnote 5, motor vehicle sales follow very different monthly patterns than other retail sales. Second, unit sales of motor vehicles are available on a very timely basis and are considered more reliable than the retail sales data. In fact, the BEA uses the

revised nonauto retail sales, which presumably measures the "true" trends in the overall retail sector; and the latest, fully revised personal consumption expenditure, which is the data that feeds into the GDP accounts. In keeping with our focus on short-term forecasting these variables enter as simple monthly percent changes. In adopting this convention we reject two alternatives. First, while the business press focuses on year-over-year percent changes in the chain store indexes, we felt that this would not be very informative to forecasters (after all, the only new information in a twelve month change is the latest month). Second, we chose not to use weekly data. Not only is there no official equivalent to these data, but, as Appendices A and B show, the quality of information in the chain store indexes falls off precipitously when we move from the monthly to weekly frequency.

Information Set

For our alternative models we include data on a range of consumer-related indicators that are released prior to the advance retail sales report:

- the only other timely consumption indicator (motor vehicle sales growth),
- 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 S&P 500 (an indicator of household wealth) and an index of retail stocks in the S&P 500 (a measure of investor confidence in the industry),
- several interest rate spread variables that have proven to be useful in short-term

---

motor vehicle units data in constructing personal consumption.

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),[13]

- lags on the dependent variable (to keep this manageable we consider only the first and second lag along with the twelfth lag to capture any left over seasonality).

On a limited basis, we initially tested lags of all of these variables, but found virtually no useful additional information.

Alternative Models

All told we test eight stand-alone models and nine models which combine one or more alternative models. First, in the "ARIMA Model" we include only autoregressive and moving average terms that add significant explanatory power. This provides a pure time series alternative to the chain store data. Second, in the "Kitchen-Sink Soup Model," we shamelessly "mine" the data, throwing in every consumer-related variable, regardless of its explanatory power. Because so many variables are included and because the model is estimated in-sample, it presents a formidable (and unfair) challenge to the chain store data. Third, to get a more fair assessment of the relative value of chain store sales we construct a "Significant Model" which includes only variables that pass standard statistical tests for inclusion.[14] Fourth, because the Significant Model sometimes includes variables with perverse coefficients, we also construct a "Correct Model" which includes only economic variables that add significant explanatory power <u>and</u> have the expected sign. Fifth, sixth and seventh, we estimate a "Mitsubishi Model," a "Johnson Model" and a model which includes both chain store indexes along with a constant

---

[13] See, for example, ...

[14] Variables are selected using the Akaike information criteria.

term. Finally, we test the simplest "Back-of -the-Envelope" model: a simple average of the monthly change in the two chain store indexes. This "model" does not require regression estimation and it would be appropriate if one assumed that the indexes are representative samples of overall consumer spending.

If the chain store data are useful for tracking consumer spending sales, we would expect them to compare favorably with these alternative models, explaining a relatively large portion of the monthly growth in official measures of consumer spending, and retaining explanatory power when used in conjunction with the alternative models.

Explaining History: In-Sample Tests

Table 3 shows the results when each model is estimated using ordinary least squares regression over the period January 1985 to December 1995: The table reports the R-squares for each model-- that is, the proportion of the month-to-month variation in the dependent variable that is explained by the model-- and an F-test for the joint statistical significance of the explanatory variables. The results underscore how difficult it is to forecast retail sales: the models explain less than a third of the month-to-month variation in retail sales growth and about two-thirds of the variation in personal consumption expenditure, and they have a particularly hard time explaining the advance data.

The results for the chain store indexes are encouraging. On the one hand, the results for the Back-of-the-Envelope Model underscore the danger of taking the chain store data at face value: for this model the calculated R-squares are actually negative.[15] In other words, one would

[15] This perverse result stems from the fact that 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

be better off completely ignoring the chain store data than using this simple approach. On the other hand, if we use regression estimation to eliminate the bias and excess variance in the chain store data, they can be useful in predicting overall retail sales. The chain store variables are highly significant both individually and when used together for all five dependent variables. While the chain store models generally can not match the explanatory power of the ARIMA models or the supercharged Kitchen Sink Soup Model they generally perform on par with the others.

The stand alone tests suggest that there is some useful information in the chain store indexes, but is this information unique? In other words, do the chain store indexes add new information not captured in the other models? Table 4 shows how the consumer spending models fair in competition with one and another. For each combination we report both the overall explanatory power of the model as well as the coefficient (and t-statistic) of the chain store indexes. The results continue to lend support for the indexes: adding the chain store data increases the overall fit and both Johnson Redbook and the Mitsubishi index continue to have significant explanatory power in most equations. In every case, except the PCE equations, the chain store indexes consistently and significantly add to the explanatory power of the alternative models. For example, adding the Mitsubishi index to the Correct Model of Department Store sales more than doubles the model's explanatory power, from 18 percent to 40 percent. The weakest chain store results are for the PCE equations, where the economic variables do a good job of explaining sales growth and the chain store indexes always have the right sign but are statistically significant only half of the time.

---

considerably worse if the indexes are used individually.

Table 4 also reports the results of three-way races, including both of the chain store indexes simultaneously along with the alternative models. Both chain stores generally finish "in the money," although Johnson Redbook lags the Mitsubishi index in terms of explanatory power.

In general the results suggest that the best model combines the economic variables with both chain store indexes. For example, as shown in Table 5, the best model for nonauto retail sales includes the first and twelth lag on the dependent variable, and the growth in payroll employment, auto sales, gasoline prices and both of the chain store indexes. All variables in this equation are statistically significant (although the lagged dependent variable is only marginally significant) and all the coefficients have the correct sign. The chain store indexes each get a modest weight in the model so that only a substantive swing in both indexes can have a major impact on the model forecast. While the model only explains 41 percent of the variation in nonauto retail sales, that's probably about the best we can do for this volatile sector.

Real Time Tests

Thus far we have focused on in-sample comparisons of the various models. The ultimate test of these equations, however, is how they perform out-of-sample. In particular, it would be useful to know how much of a loss of predictive power occurs when we move from in-sample to out-of-sample tests, and whether there is any change in the rank order of various models.

We approximate true real time forecasting with the following three step procedure:

- variable selection: using data for the 1975-89 period we select which variables to be included in each model. We select variables using the same inclusion criteria and same menu of potential regressors as the in-sample models (see pages 13-14).[16]

---

[16] Note that we do not include the Kitchen-Sink-Soup model in this exercise on the grounds that this over-fitted model is likely to perform very poorly in these out-of-sample tests.

- recursive regressions: we successively re-estimate the model, adding one month at a time and calculating a series of one-month-ahead forecasts over the entire 1990-95 period.
- forecast evaluation: the forecasts are evaluated Mean Square Errors (MSE) and a variety of other conventional criteria.

The information used in this exercise differs from a true real time test in two respects. In one respect we use more information than a forecaster would have: we use the latest data for the independent variables, whereas in real time only preliminary data are available for some of our regressors. In another respect we use less information than a forecaster would have: by keeping the selected regressors and the starting date of the recursive regressions fixed, we do not allow for forecasters to fully adapt the structure of the model based on their forecasting experience.[17] Fortunately, neither of these departures from a strict real time forecast appears to be important in practice. For several series we do have preliminary data and substituting these did not have much impact on the results. Similarly, in additional tests we found only limited evidence of structural breaks in our models. This should not be surprising given that all of our data comes from the period after the 1973 oil shock.

Table 6 summarizes our main findings. It reports the mean square error (MSE) for 75 different forecast models: for each of our five dependent variables we test fifteen models, including seven stand-alone models, and nine models which combine two or three of the stand-alone models. Note that for each dependent variable the lowest MSE is reported with an asterisk, and for each subset of models, the lowest MSE is reported in bold.[18]

---

[17] 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.

[18] There is no single "correct" measure of forecast performance, but MSE error is the most commonly used. By squaring the out-of-sample errors it puts an extra penalty on large

While there are a lot of numbers in this table, some clear patterns emerge, confirming the findings of the in-sample tests. In particular, we find:

- for the stand-alone models, the model using both chain store indexes always has the lowest MSE;
- by contrast, the worst results are for the Back-of-the-Envelope Model, suggesting that using simple rules of thumb to forecast with these data can cause more harm than good;
- in all 45 cases adding chain store models to the alternative models reduces the MSE;
- for four out of five dependent variables, the lowest MSE is achieved by adding both chain store indexes to the alternative models; the only exception is Department store sales where the Mitsubishi index tends to dominate Johnson Redbook.
- not only does using combination models reduce the MSE, the improvement is considerable-- the MSE frequently declines by a third or more relative to the stand alone models.

Tables 7a-7e report a number of other summary measures of the out-of-sample predictive power of these models. The econometrics literature suggests a smorgasbord of alternative methods of forecast evaluation. To a large degree this diversity reflects the fact that forecasts are designed for use in a particular decision environment so that the appropriate measure of forecast accuracy depends on what kind of forecast errors are most costly to the user.[19] In addition to the MSE, the table reports four different summary statistics:

- Bias: the mean forecast error. A mean value close to zero, means that the forecast does not tend to systematically under- or over- predict the dependent variable. A "+" next to the bias estimate indicates that the bias is statistically significant.
- Average Absolute Error (AAE): the average error, regardless of sign. AAE is preferred to MSE if the forecaster does not put a disproportionate weight on large errors.

---

forecast errors.

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

- Percent Directional Errors: large econometric models are often compared based on their ability to predict turning points; for our very short-run forecasts an analogous test is to look at how frequently the model correctly predicts the direction of the dependent variable. Presumably, getting the right "handle" (positive or negative) on the predicted growth can help avoid some embarrassment for the forecaster. A good forecast should get the right direction substantially more than 50 percent of the time.

- Q-test: a test for first-order 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 period's forecast error. It is a sign that the model is missing some important information.

While tables 7a to 7e present a lot of statistics, the general conclusions are quite straight forward and strongly confirm the themes of our previous tests. The highlights are as follows:

- Using the average absolute error as the standard of evaluation, rather than mean square error, generally has very little impact on the ranking of the models. For example, it tends to favor the Johnson Redbook model in the personal consumption tests, but has virtually no impact on the rankings for the other dependent variables.

- Virtually all of the models have a modest tendency to over-predict sales growth. This bias tends to be largest in the stand-alone models and is much smaller in models that combine economic variables with both chain store indexes.

- Using the combination models also greatly reduces the most embarrassing kind of forecast error--predicting the wrong direction for sales growth. The strongest results are for the narrow definition of sales, where adding the chain store indexes to the alternative models reduces the proportion of directional errors by 3 to 10 percentage points.

- The Q-tests show some evidence of serial correlation in the forecast errors, but using the combination models generally helps in this regard as well.

Implications For Forecasters

What do our results mean in practical terms? Monthly consumer spending is very volatile, but using a combination of economic variables and the chain store data we can explain about 40 percent of the variation in various 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 points off our monthly forecast error (relative to a model that assumes no change in growth) and we correctly predict the direction of sales growth 75 to 85 percent of the time. Using these models we also avoid the pitfalls of back-of-the-envelope calculation.

Not only do chain store indexes help improve forecasts, there is evidence that this information currently is not fully exploited. In particular, private sector economists do not appear to fully account for chain store sales in their forecasts of the advance retail sales data. To show this, we used data for the 1985-1995 period on consensus forecasts of retail sales compiled each month by MMS International.[20] If forecasters fully account for the chain store indexes in making there forecasts there should be no correlation between the consensus forecast errors and the chain store indexes. In fact, while the Mitsubishi index is not correlated with the error, Johnson Redbook is, at least marginally, at the 10 percent level.[21]

We also find that chain store data have an unexpectedly long useful "shelf life" as economic indicators. In fact, 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 regressed the revision in the official retail sales growth, from the advance to the latest estimate, on the chain store indexes for the 1985-95 period. In this case it was the Mitsubishi index that turned out to

---

[20] These consensus forecasts come from a survey of several dozen market participants taken the week prior to the release of the retail sales report.

[21] The regression coefficients (and t-values) are:

$$\underset{(-1.39)}{-0.087} + \underset{(1.78)}{.059} * \text{JOHN} - \underset{(-0.07)}{.002}\ \text{MITS}.$$

be significant (at the 5 percent level).[22] It appears that the chain store data deserve longer lasting, as well as more careful, attention.

Conclusion

Our results shed light on a relatively dark corner of forecasting. We find that:

- both individual store data and the weekly indexes are of very limited value as macroeconomic indicators;

- users should also beware of the effect of changing seasonals and price discounting on chain store sales--this past December both of these factors tended to depress the nominal value of sales, exaggerating the weakness of holiday shopping;

- chain store indexes can be quite volatile and should be given the appropriate weights in forecasting--for Department Store sales weights for the Mitsubishi and Johnson indexes of .31 and .11 respectively are appropriate; for nonauto retail sales, weights of .11 and .07 are appropriate.

- a number of other indicators are useful for forecasting retail sales, including payroll employment, gasoline prices, unit sales of motor vehicles and both one and twelve lags on the lagged dependent variable.

The bottom line is simple: the chain store data are far from infallible, but focusing on the monthly indexes, giving them the right weight and combining them with economic variables results in a more accurate view of the consumer sector.

[22] The regression coefficients (and t-values) are: $\underset{(2.95)}{0.145} + \underset{(0.66)}{.018} * \text{JOHN} + \underset{(2.42)}{.066} * \text{MITS}$.

Appendix A

How Good Are These Data?

Company Reports

On the first 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 a constant reinterpretation of the data. Each report seems to focus on a different measure of sales growth: same-store sales or total sales; year-to-date or latest month; domestic sales or total company sales; calender month or "4-5-4 weeks" month; and above or below "plan." The individual company results are all over the map. For example, in January 1996, 56 companies reported by Johnson Redbook show the following extreme ranges for year-over-year sales growth: one company reported a sharp rise in total sales of 19 percent but an almost equally sharp decline in same-store sales of 9.0; the strongest company enjoyed a 112 percent sales increase and the weakest suffered a 28 percent decline; and even among the 13 largest companies, reporting more than 500 million dollars in sales, the growth rate ranged from a high of 27 percent to a low of -3 percent. Clearly these divergent company reports reflect the structural changes buffeting the retail sector.

Another major problem with the raw chain store data is that it is very difficult to seasonally adjust. The data for March 1996 provided a striking example. Most company reports for March 1996 included sales for the five week ending on April 6th. Since Easter was on April 7th this meant that these figures captured all the shopping for this holiday. By contrast, 1995 the March reports only included data through April 1 and Easter fell on the 15th so that most of the

Easter shopping is excluded. This means that even on a year-ago basis the data will be distorted by changing seasonals.

These problems with the company reports underscore the danger of using anecdotal evidence to assess industry trends. They also point to the importance of getting a large representative sample and using extreme care in seasonally adjusting the data.

Chain Store Indexes

Combining these data into indexes removes some, but not all, of the idiosyncrasies of these data. Table 8 shows how erratic the growth rates for these indexes are, using two statistics: the autoregressive coefficient which shows where the indicator is subject to sharp reversals and the standard deviation of the growth rate. The weekly data are particularly erratic. 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 standard deviation of almost a percentage point, it is not unusual to see one-week percent changes of two percentage points or more.[23] The monthly data are also quite erratic: the Mitsubishi index is particularly prone to reversal, and for both indexes the standard deviation of the monthly growth rate is almost 2 percent.

A major drawback of the chain store indexes is that companies that produce them offer very little information on their construction. Thus we know very little about the sampling properties of the data, how outliers or nonresponses are handled, and how the data are seasonally

[23] Johnson Redbook only reports its weekly data on a year-ago basis so it could not be used in this table. Using the year-ago percent change data, Johnson Redbook is twice as volatile as the Mitsubishi index. Given this volatility it should not be surprising that the weekly data have a relatively week correlation with their monthly counterparts. For example, using the percent change from a year-ago, and comparing the first week of each month to the full month index, the Mitsubishi index has a correlation of only .59 and Johnson Redbook has a correlation of just .42 over the 1990-95 period.

adjusted (a major problem, particularly with the weekly data). News reports suggest that considerable judgement is used in constructing the Johnson Redbook index: "Johnson said he has learned over the years to adjust the data he receives to 'unskew it' and has developed inside sources, which he guards scrupulously. His long-running relationships with sources allow him to interpret their comments, he said."[24] In addition, because neither index is revised we can be sure that any late responses or reporting errors are never corrected. One quirk in the Johnson Redbook series illustrates the problem with not revising the data: there is a discontinuous jump in the Johnson data in January 1989 when a major revision in the official data (which is used as a benchmark in constructing the level of Johnson Redbook) was not matched by a similar level adjustment in Johnson Redbook.

Retail Sales

The official retail sales data are less erratic. In part, this is because they are constructed using sophisticated (and expensive) sampling and statistical methods which simply cannot be matched by a private firm. Nonetheless, on a month-to-month basis these data are volatile and subject to considerable revision.

Monthly data on retail sales are based on a random sample of more than 12,000 companies. The sample covers firms of all sizes, but is stratified with coverage ranging from 100 percent for major firms to 0.1 percent for the smallest firms. Because of their large size, Department Stores are heavily represented in this sample: while the sample captures less than half of overall retail sales, it captures 99 percent of the Department Store sector. These data are

---

[24] Weir (1992).

repeatedly and heavily revised. "Advance" data are released, somewhat reluctantly[25], only two weeks after the month is over, but include data from 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, annual revisions are released each spring and 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 latest nonauto retail sales data and the advance data is just 52 percent for the 1985-95 period.

The Table shows that while these data are less volatile than the chain store indexes, they are nonetheless quite variable by the standards of macroeconomic data. Despite the fact that retail sales tend to grow over time and tend to rise and fall with the business cycle, the monthly growth rates have a negative serial correlation, implying that strong growth one month tends to be reversed the next month. The standard deviations are lower than for chain store data, but still suggest considerable month-to-month variation, particularly for Department Store sales.

[25] The Census Bureau writes: "The Bureau releases "non-final" advance and preliminary data to provide government and private data users with much demanded early measures of consumer spending;" and "the advance sales estimates are based on early reporting of sales by a small subsample of the Bureau's retail survey panels." (U.S. Department of Commerce (1995) p. B8.)

Appendix B

Results For Additional Chain Store Indexes

Growing interest in the chain store data, has helped spur a cottage industry of new chain store indexes. This section briefly reviews several additional indexes that receive press coverage:

- the Goldman Sachs Monthly Comparable-Store Sales Index, an index of department, apparel specialty, discount and hard goods specialty stores that is available starting in 1988;[26]
- the Merrill Lynch Broadlines Same Store Sales Index, an index of department and general merchandise stores which is available starting in 1992; and
- the weekly versions of the Mitsubishi index and Johnson Redbook.[27]

Table 9 shows the results when the growth rate of nonauto retail sales is regressed on all six chain store indexes. Three sample periods are reported based on the availability of the various indexes. There are few surprises here. Most of the indexes are of marginal value in predicting nonauto retail sales. At one extreme, the monthly Mitsubishi index consistently outperforms the others; at the other extreme, the weekly Mitsubishi data appear to be of no use in predicting nonauto retail sales.

---

[26] The Goldman Sachs data are reported as a percent change from a year ago. Using the same methodology as Johnson Redbook's, we convert this to seasonally adjusted monthly data by taking the Goldman Sachs growth rates and applying them to the year ago level of the official GAF data.

[27] The tests on the weekly data are not strictly comparable to the other tests. For weekly data, seasonally adjusted data are not available so the independent variable is the percent change, from year-ago, of sales for the first week each month.

References (draft)

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

Granger, C.W.J. and Paul Newbold (1986). Forecasting Economic Time Series. Academic Press Inc. San Diego, California 1986.

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.

Johnson Redbook Service (1996). "Redbook Weekly Comments," February 7, 1996.

Kuwayama, Patricia Hagan and James F. O'Sullivan (ed) (1996) Global Data Watch Handbook, Second Edition, JP Morgan, New York.

Kennedy, Peter (1992). A Guide to Econometrics. The MIT Press, Cambridge Massachusetts.

Kernkraut, Steven (1996) "Beyond the Comps, Part Two," Retailing Industry Bear Sterns, February 22.

Merrill Lynch "Retailing Broadlines" January 4, 1996.

Missett, Bruce M. and Sheelagh K. McCauley (1996) Broadlines/Hardlines: Retail Monthly, Morgan Stanley, April.

Mitsubishi Bank, "Weekly Chain Store Sales Snapshot" November 1994.

Rogers, Mark R. (1994) Handbook of Key Economic Indicators Irwin Professional Publishing, New York.

Tainer, Eva (1993) Using Economic Indicators to Improve Investment Analysis. John Wiley and Sons, New York.

Telsey, Dana L. (1996) "Specialty Retail: Hard and Soft Lines," Equity Research, Bear Sterns, April.

U.S. Bureau of Census, Geographic Area Series, RC92-A-52. 1992 Census of Retail Trade. Washington, DC. 1995.

U.S. Bureau of the Census, Current Business Reports, Series BR/94-RV, Combined Annual and Revised Monthly Retail Trade, January 1985 Through December 1994, Washington DC 1995.

U.S. Bureau of the Census, Current Business Reports, Advance Monthly Retail Sales, Washington DC various issues 1985-1995.

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.

Weir, Dan "Analysts Debate Accuracy, Uses of Johnson Redbook US Retail Data," Knight-Ridder/Tribune Business News, June 1, 1992.

Table 1
Retail Establishments in 1992

	Number of Establishments* (1000's)	Average Sales per Establishment ($1000's)	Average Employees per Establishment**
GAF	463.1	2,025.9	19
General Merchandise	34.6	7,089.2	60
Department	11.0	16,946.0	156
Chain Stores	1.9	18,872.5	179
Conventional Department	2.4	20,832.4	203
Discount Stores	6.7	15,031.9	134
Other	23.6	2,495.5	15
Apparel	145.5	699.1	8
Furniture	110.1	846.8	6
Miscellaneous Shopping Goods	127.3	519.8	6
TOTAL RETAIL	1,526.2	1,241.5	12

* "A single physical location at which business is conducted."
** This is a snapshot for the week of March 12. 1992. Over the course of a year each establishment temporarily employees many more workers.

Source: 1992 Census of Retail Trade. US Department of Commerce Economics and Statistics Administration.

Table 2

Getting a Jump on the Competition:

Release Dates for February

Dates	Releases
February 13,20 and 27	**Johnson Redbook and Mitsubishi (weekly)**
February 27	Confidence Indexes (Michigan and Conference Board)
March 5	Auto and Light Truck Sales
March 5	**Retail Company Reports**
March 6	**Johnson Redbook (monthly)**
March 7	**Mitsubishi (monthly)**
March 7	Initial Claims (monthly)
March 8	Payroll Employment
March 13	Advance Retail Sales
April 3	Personal Consumption Expenditures
April 11	Preliminary Retail Sales
May 14	“Final” Retail Sales

Table 3

"In-Sample" Explanatory Power of Consumer Spending Models

					Non-Auto Retail					
	Department Store		GAF		Advance		Latest		PCE	
Models	R^2	F-test	R^2	F-test	R^2	F-test	R^2	F-test	R^2	F-test
ARIMA	0.190	10.01*	0.304	13.87*	0.159	8.09*	0.316	19.73*	0.163	25.24*
Kitchen Sink	0.241	2.28*	0.258	2.50*	0.201	1.81	0.301	3.11*	0.678	15.12*
Significant	0.183	9.58*	0.237	5.50*	0.170	4.24*	0.267	7.58*	0.664	49.91*
"Correct"	0.183	9.58*	0.237	5.50*	0.101	4.80*	0.250	8.41*	0.664	49.91*
Mitsubishi	0.285	52.04*	0.223	37.21*	0.070	9.79*	0.161	24.10*	0.011	1.49
Johnson Redbook	0.072	10.07*	0.142	21.52*	0.112	16.47*	0.121	17.89*	0.025	3.27
Both Chain Indexes	0.311	29.16*	0.303	28.08*	0.151	11.55*	0.234	19.69*	0.030	2.01
Back-of-Envelope**	-0.123		-0.334							

Note: The table reports the R-squared and F-test. In each case the sample period is January 1985 to December 1995.

* F-test for significance of all explanatory variables is significant at 1 percent level.

** 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.

Table 4
"In-Sample" Explanatory Power of Combination Models

							Non-Auto Retail								
	Department Store			GAF			Advance			Latest			PCE		
Models	Chain	R^2	F-test	Chain	R^2	F-test	Chain	R^2	F-test	Chain	R^2	F-test	Chain	R^2	F-test
ARIMA and ...															
Mitsubishi	.385 (6.92)	.430	23.96*	.201 (6.56)	.433	19.26*	.096 (4.55)	.277	12.16*	.062 (3.27)	.357	17.62*	.062 (1.88)	.184	14.56*
Johnson Redbook	.184 (3.22)	.249	10.53*	.077 (3.05)	.335	12.69*	.081 (4.12)	.261	11.23*	.044 (3.28)	.351	17.14*	.066 (2.36)	.197	15.83*
Both Mitsubishi	.358 (6.77)	.453	20.93*	.188 (6.32)	.456	17.48*	.078 (3.83)	.337	12.81*	.056 (3.08)	.385	15.75*	.005 (1.36)	.208	11.23*
Johnson	.115 (2.38)			.063 (2.71)			.063 (3.36)			.038 (2.77)			.056 (1.97)		
Kitchen Sink and ...															
Mitsubishi	.339 (6.73)	.457	5.64*	.246 (5.57)	.417	4.80*	.078 (3.42)	.275	2.54*	.123 (4.57)	.410	4.66*	.047 (2.45)	.694	15.20*
Johnson Redbook	.157 (2.79)	.290	2.73*	.179 (3.94)	.347	3.57*	.076 (3.46)	.277	2.57*	.094 (3.48)	.368	3.92*	.028 (1.48)	.684	14.50*
Both Mitsubishi	.320 (6.31)	.475	5.67*	.218 (5.03)	.467	5.49*	.065 (2.86)	.326	3.04*	.109 (4.10)	.451	5.15*	.043 (2.18)	.696	14.41*
Johnson	.097 (1.95)			.137 (3.25)			.063 (2.91)			.074 (2.89)			.019 (0.99)		

Note: The table reports the R-squared and F-test on the joint significance of the explanatory variables. "Chain" is the coefficient on the chain store index, with the associated t-value in parenthesis. In each case the sample period is January 1985 to December 1995.
* F-test for significance of all explanatory variables is significant at 1 percent level.

Table 4
"In-Sample" Explanatory Power of Combination Models

		Department Store			GAF			Non-Auto Retail: Advance			Non-Auto Retail: Latest			PCE	
Models		Chain	R^2	F-test	Chain	R^2	F-test	Chain	R^2	F-test	Chain	R^2	F-test	Chain	R^2	F-test
ARIMA and ...																
Mitsubishi		.385 (6.92)	.430	23.96*	.201 (6.56)	.433	19.26*	.096 (4.55)	.277	12.16*	.062 (3.27)	.357	17.62*	.062 (1.88)	.184	14.56*
Johnson Redbook		.184 (3.22)	.249	10.53*	.077 (3.05)	.335	12.69*	.081 (4.12)	.261	11.23*	.044 (3.28)	.351	17.14*	.066 (2.36)	.197	15.83*
Both	Mitsubishi	.358 (6.77)	.453	20.93*	.188 (6.32)	.456	17.48*	.078 (3.83)	.337	12.81*	.056 (3.08)	.385	15.75*	.005 (1.36)	.208	11.23*
	Johnson	.115 (2.38)			.063 (2.71)			.063 (3.36)			.038 (2.77)			.056 (1.97)		
Kitchen Sink and ...																
Mitsubishi		.339 (6.73)	.457	5.64*	.246 (5.57)	.417	4.80*	.078 (3.42)	.275	2.54*	.123 (4.57)	.410	4.66*	.047 (2.45)	.694	15.20*
Johnson Redbook		.157 (2.79)	.290	2.73*	.179 (3.94)	.347	3.57*	.076 (3.46)	.277	2.57*	.094 (3.48)	.368	3.92*	.028 (1.48)	.684	14.50*
Both	Mitsubishi	.320 (6.31)	.475	5.67*	.218 (5.03)	.467	5.49*	.065 (2.86)	.326	3.04*	.109 (4.10)	.451	5.15*	.043 (2.18)	.696	14.41*
	Johnson	.097 (1.95)			.137 (3.25)			.063 (2.91)			.074 (2.89)			.019 (0.99)		

Note: The table reports the R-squared and F-test on the joint significance of the explanatory variables. "Chain" is the coefficient on the chain store index, with the associated t-value in parenthesis. In each case the sample period is January 1985 to December 1995.
* F-test for significance of all explanatory variables is significant at 1 percent level.

Table 5

The "Best"* Model for Non-Auto Retail Sales

LS // Dependent Variable is NRETX
Sample: 1985:01 1995:12
Included observations: 132

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.328	0.074	4.455	0
NRETX(-1)	-0.121	0.070	-1.727	0.087
NRETX(-12)	-0.224	0.065	-3.450	0.001
AUTOX	0.015	0.006	2.699	0.008
PAYX	1.063	0.299	3.551	0.001
GASPX	0.009	0.004	2.168	0.032
MITSX	0.113	0.026	4.426	0
JOHNX	0.070	0.025	2.840	0.005

R-squared	0.412	Mean dependent var.	0.424
Adjusted R-squared	0.379	S.D. dependent var.	0.615
S.E. of regression	0.484	Akaike info criterion	-1.391
Sum squared resid	29.096	Schwarz criterion	-1.216
Log likelihood	-87.495	F-statistic	12.433
Durbin-Watson stat.	2.130	Prob(F-statistic)	0
MSE	0.25		

Note: The equation is estimated with OLS for the January 1985 to December 1995 period. All variables are measured as percent changes from a month ago.
*The "Best" model is the "correct" model without the SP500 index for retailers and adding both the Mitsubishi and Johnson indexes. We dropped the SP500 index for retailers because it becomes statistically insignificant when adding both chain store indexes.

Table 6

Mean-Square Errors For One-Month Forecasts*

	Dependent Variables				
Models			Advance	Latest	Personal
Stand alone	Department	GAF	Non-Auto	Non-Auto	Consumption
ARIMA	1.442	1.074	0.391	0.494	0.315
Significant	1.292	1.022	0.288	0.514	0.222
Correct	1.292	1.022	0.260	0.504	0.240
Mitsubishi	1.068	0.855	0.265	0.348	0.256
Johnson	1.263	0.926	0.203	0.373	0.224
Both Chain Indexes	**1.002**	**0.774**	**0.198**	**0.321**	**0.218**
Back-of-Envelope	1.641	1.403	1.646	1.449	1.973
ARIMA and ...					
Mitsubishi	**0.854**	0.794	0.300	0.390	0.287
Johnson	1.156	0.912	0.262	0.375	0.272
Both chain indexes	0.880	**0.747**	**0.223**	**0.326**	**0.264**
Significant and ...					
Mitsubishi	**0.842***	0.743	0.254	0.383	0.195
Johnson	1.094	0.883	**0.172**	**0.332**	0.164
Both chain indexes	0.879	**0.701***	0.174	**0.332**	**0.156***
Correct and ...					
Mitsubishi	**0.842***	0.743	0.236	0.373	0.207
Johnson	1.094	0.883	0.172	0.375	**0.166**
Both chain indexes	0.879	**0.701***	**0.171***	**0.315***	**0.166**

Note: The "best" results for each subgroup are indicated in bold; the "best" results for each column are indicated with an asterisk.

Table 7A
Department Store Sales: Out of Sample

Models	Bias	Average Absolute Error	Mean Square Error (MSE)	Percent Correct Direction	Q-test (one lag)**
Stand alone					
ARIMA	-0.261	0.934	1.442	66.7	0.130
Significant	-0.286	0.889	1.292	69.4	**0.000***
Correct	-0.286	0.889	1.292	69.4	**0.000***
Mitsubishi	**0.018**	0.788	1.068	69.4	8.366
Johnson	-0.063	0.863	1.263	68.1	7.591
Both Chain Indexes	0.020	**0.760**	**1.002**	**73.6***	8.9419
ARIMA and ...					
Mitsubishi	0.051	0.721	**0.854**	**72.2**	0.136
Johnson	**-0.004***	0.858	1.156	70.8	0.123
Both chain indexes	0.055	**0.719**	0.880	70.8	**0.006**
Significant and ...					
Mitsubishi	-0.096	0.709	**0.842***	**72.2**	1.492
Johnson	**-0.046**	0.829	1.094	**72.2**	**0.239**
Both chain indexes	0.052	**0.698***	0.879	**72.2**	3.116
Correct and ...					
Mitsubishi	-0.096	0.709	**0.842***	**72.2**	1.492
Johnson	**-0.046**	0.829	1.094	**72.2**	**0.239**
Both chain indexes	0.052	**0.698***	0.879	**72.2**	3.116

Note: The "best" results for each subgroup are indicated in bold; the "best" results for each column are indicated with an asterisk.
** The critical value for the X^2 statistic at the 95% level is 3.4.

Table 7B
GAF: Out of Sample

		Average	Mean	Percent	
Models		Absolute	Square	Correct	
Stand alone	Bias	Error	Error (MSE)	Direction	Q-test (one lag)**
ARIMA	-0.208+	0.814	1.074	70.8	**0.126**
Significant	-0.250	0.796	1.022	68.1	0.234
Correct	-0.250	0.796	1.022	68.1	0.234
Mitsubishi	**-0.092**	0.718	0.855	68.1	4.708
Johnson	-0.169+	0.723	0.926	**72.2**	4.273
Both Chain Indexes	-0.116	**0.673**	**0.774**	**72.2**	
ARIMA and ...					
Mitsubishi	**-0.078***	0.696	0.794	69.4	0.109
Johnson	-0.170	0.743	0.912	66.7	0.090
Both chain indexes	-0.144	**0.668***	**0.747**	**72.2**	**0.036**
Significant and ...					
Mitsubishi	-0.139+	0.690	0.743	69.4	0.401
Johnson	-0.209	0.742	0.883	69.4	**0.004***
Both chain indexes	**-0.124**	**0.675**	**0.701***	**73.6***	0.292
Correct and ...					
Mitsubishi	-0.139+	0.690	0.743	69.4	0.401
Johnson	-0.209	0.742	0.883	69.4	**0.004***
Both chain indexes	**-0.124**	**0.675**	**0.701***	**73.6***	0.292

Note: The "best" results for each subgroup are indicated in bold; the "best" results for each column are indicated with an asterisk.
** The critical value for the X^2 statistic at the 95% level is 3.4.
+ The sign test for bias is significant at the 5 percent level

Table 7C
Advance Non-Auto Retail Sales: Out of Sample

		Average	Mean	Percent	
Models		Absolute	Square	Correct	
Stand alone	Bias	Error	Error (MSE)	Direction	Q-test (one lag)**
ARIMA	-0.275^{+}	0.476	0.391	66.7	0.122
Significant	-0.257^{+}	0.436	0.288	70.8	1.566
Correct	-0.246^{+}	0.386	0.260	70.8	**0.117**
Mitsubishi	-0.199^{+}	0.418	0.265	**73.6**	1.819
Johnson	-0.088	**0.342**	0.203	72.2	4.302
Both Chain Indexes	**-0.070**	0.345	**0.198**	**73.6**	
ARIMA and ...					
Mitsubishi	-0.218^{+}	0.428	0.300	**72.2**	**0.123**
Johnson	-0.089	0.382	0.262	68.1	1.510
Both chain indexes	**-0.042**	**0.369**	**0.223**	70.8	0.192
Significant and ...					
Mitsubishi	-0.196^{+}	0.419	0.254	72.2	1.623
Johnson	-0.047	**0.315**	**0.172**	**73.6**	1.141
Both chain indexes	**-0.029**	0.331	0.174	**73.6**	**1.038**
Correct and ...					
Mitsubishi	-0.182^{+}	0.387	0.236	72.2	**0.022**
Johnson	-0.036	**0.310***	0.172	**77.8***	0.858
Both chain indexes	**-0.007***	0.318	**0.171***	76.4	0.994

Note: The "best" results for each subgroup are indicated in bold; the "best" results for each column are indicated with an asterisk.
** The critical value for the X^2 statistic at the 95% level is 3.4.
$^{+}$ The sign test for bias is significant at the 5 percent level

Table 7D
Latest Non-Auto Retail Sales: Out of Sample

Models Stand alone	Bias	Average Absolute Error	Mean Square Error (MSE)	Percent Correct Direction	Q-test (one lag)**
ARIMA	-0.250^{+}	0.540	0.494	**72.2***	**1.063**
Significant	-0.301^{+}	0.546	0.514	**72.2***	1.980
Correct	-0.299^{+}	0.534	0.504	70.8	1.246
Mitsubishi	-0.179^{+}	0.463	0.348	**72.2***	1.711
Johnson	-0.154^{+}	0.471	0.373	70.8	2.679
Both chain indexes	**-0.129**	**0.444**	**0.321**	**72.2***	
ARIMA and ...					
Mitsubishi	-0.196	0.483	0.390	**70.8**	1.526
Johnson	-0.164^{+}	0.473	0.375	**70.8**	**0.468**
Both chain indexes	**-0.144**	**0.447**	**0.326**	**70.8**	0.482
Significant and ...					
Mitsubishi	-0.232^{+}	0.475	0.383	**72.2***	2.567
Johnson	-0.160	0.464	0.391	70.8	**0.100**
Both chain indexes	**-0.135**	**0.441**	**0.332**	**72.2***	0.102
Correct and ...					
Mitsubishi	-0.230^{+}	0.467	0.373	**72.2***	1.557
Johnson	-0.160	0.464	0.375	70.8	**0.021***
Both chain indexes	**-0.112***	**0.437***	**0.315**	**72.2***	0.024

Note: The "best" results for each subgroup are indicated in bold; the "best" results for each column are indicated with an asterisk.
** The critical value for the X^2 statistic at the 95% level is 3.4.
$^{+}$ The sign test for bias is significant at the 5 percent level

Table 7E
Personal Consumption Expenditures: Out of Sample

Models		Average Absolute	Mean Square	Percent Correct	
Stand alone	Bias	Error	Error (MSE)	Direction	Q-test (one lag)**
ARIMA	-0.340^{+}	0.446	0.315	83.3	**0.163***
Significant	-0.278^{+}	0.376	0.222	**84.7**	0.834
Correct	-0.320^{+}	0.401	0.240	83.3	2.693
Mitsubishi	-0.241^{+}	0.389	0.256	83.3	7.082
Johnson	-0.163^{+}	0.357	0.224	83.3	5.610
Both chain indexes	$\mathbf{-0.156^{+}}$	**0.350**	**0.218**	83.3	
ARIMA and ...					
Mitsubishi	-0.313^{+}	0.413	0.287	**83.3**	**0.413**
Johnson	-0.279^{+}	0.392	0.272	**83.3**	0.659
Both chain indexes	$\mathbf{-0.271^{+}}$	**0.378**	**0.264**	**83.3**	1.082
Significant and ...					
Mitsubishi	-0.245^{+}	0.347	0.195	84.7	2.874
Johnson	-0.130^{+}	0.292	0.164	86.1	**0.676**
Both chain indexes	$\mathbf{-0.119^{+}}$	**0.288**	**0.156**	**87.5***	1.975
Correct and ...					
Mitsubishi	-0.280	0.366	0.207	83.3	6.220
Johnson	-0.179^{+}	0.307	0.166	**84.7**	**2.759**
Both chain indexes	$\mathbf{-0.161^{+}}$	**0.292***	**0.154***	**84.7**	5.086

Note: The "best" results for each subgroup are indicated in bold; the "best" results for each column are indicated with an asterisk.
** The critical value for the X^2 statistic at the 95% level is 3.4.
$^{+}$ The sign test for bias is significant at the 5 percent level

Table 8

The Statistical Properties of Consumption Indicators

Variables	AR(1) Coefficient	Mean	Variance
Weekly (Percent Change From Previous Week)			
Mitsubishi	-0.380*	0.072	0.756
Monthly (Percent Change From Previous Month)			
Mitsubishi	-0.482*	0.350	3.098
Johnson	-0.148	0.550	3.233
Department Stores	-0.295*	0.481	1.471
GAF	-0.369*	0.510	0.374
Non-Auto Retail Sales ("Advance")	-0.087	0.246	0.221
Non-Auto Retail Sales ("Latest")	-0.174*	0.424	0.378
Personal Consumption Expenditures	-0.369*	0.510	0.374

Note: 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.

*Denotes significant at 5% level.

Table 9

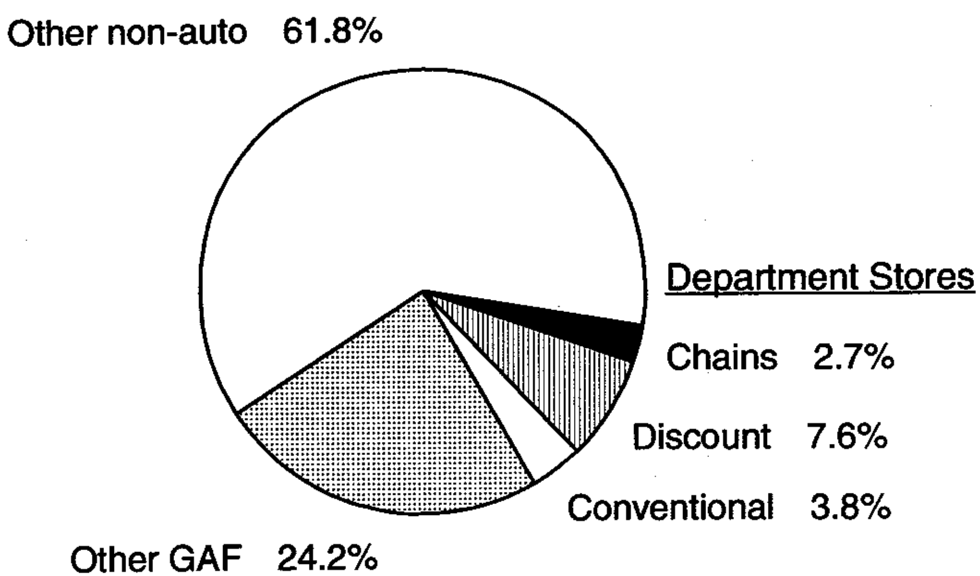
Results for Alternative Indexes

Sample: 1988:02-1995:12	Chain	R-Squared
Mitsubishi	0.154	0.250
	(5.57)	
Johnson Redbook	0.109	0.123
	(3.61)	
Goldman Sachs	0.395	0.044
	(2.07)	
Sample: 1992:07-1995:12		
Mitsubishi	0.236	0.374
	(4.82)	
Johnson Redbook	0.121	0.077
	(1.80)	
Goldman Sachs	0.374	0.076
	(1.79)	
Merrill Lynch	0.940	0.096
	(2.06)	
Sample: 1992:07-1995:12		
Weekly Mitsubishi	0.472	0.006
	(0.622)	
Weekly Johnson Redbook	0.468	0.065
	(2.043)	

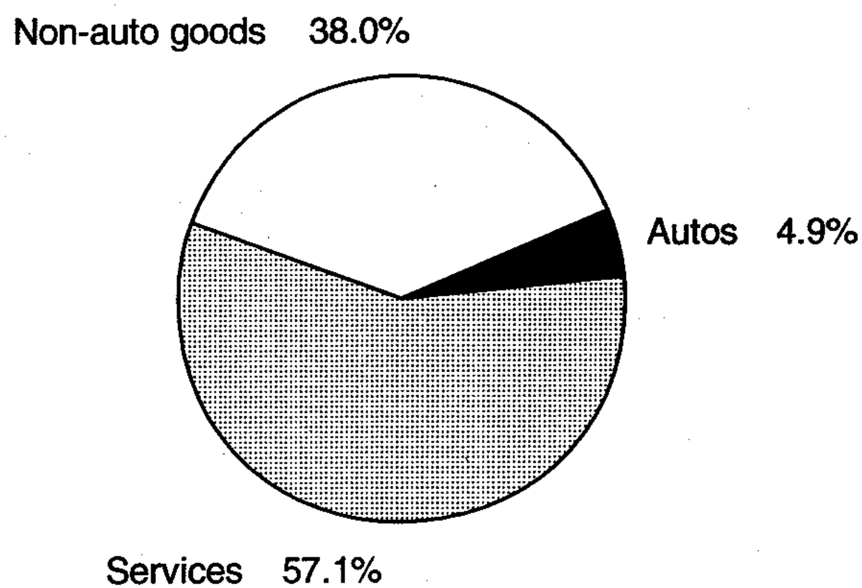
Chart 1

Sales Shares in 1992

Non-Auto Retail Sales



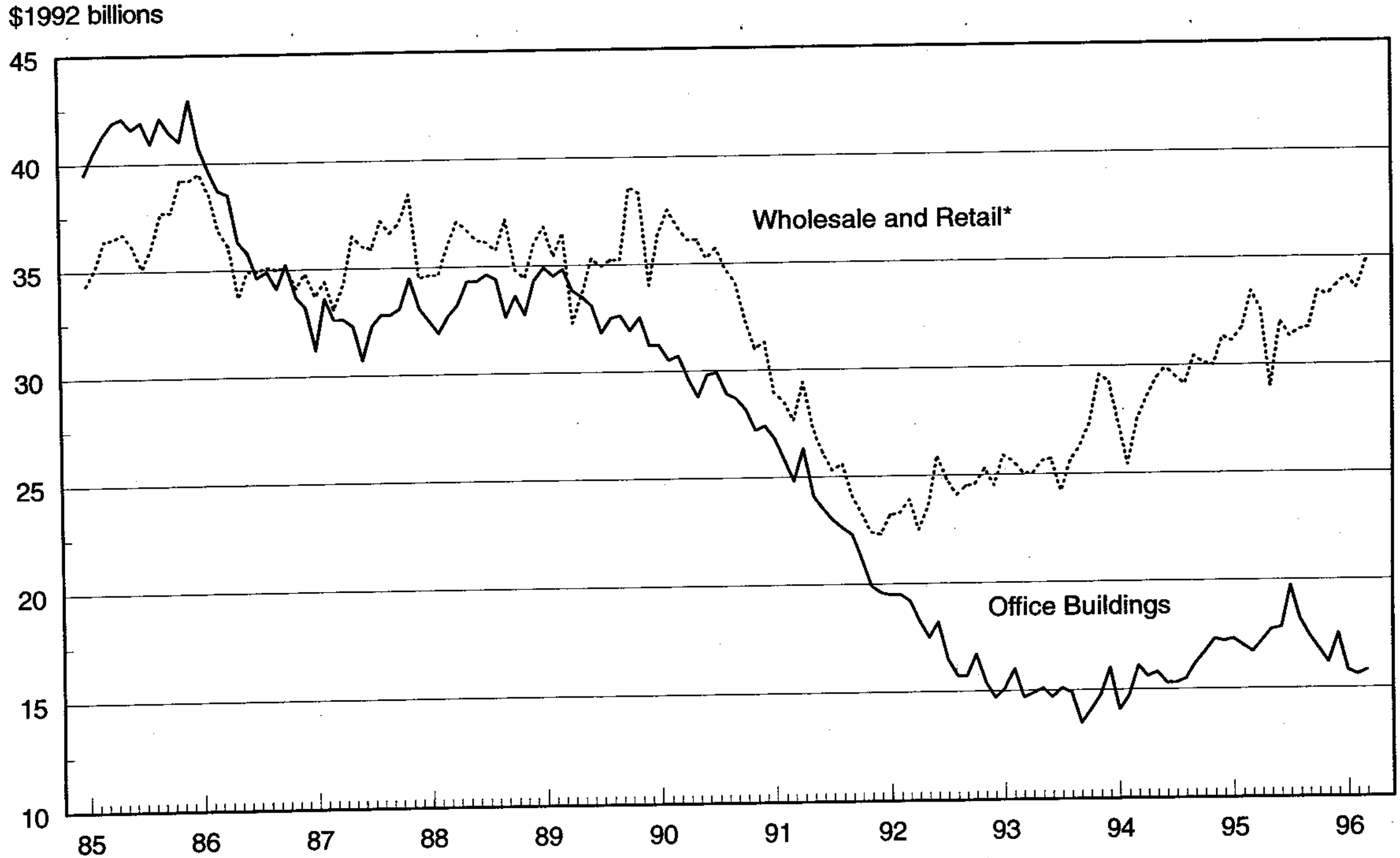
Personal Consumption Expenditures (PCE)



Note: "Autos" refers to sales of automotive dealers and includes sales of light trucks

Chart 2

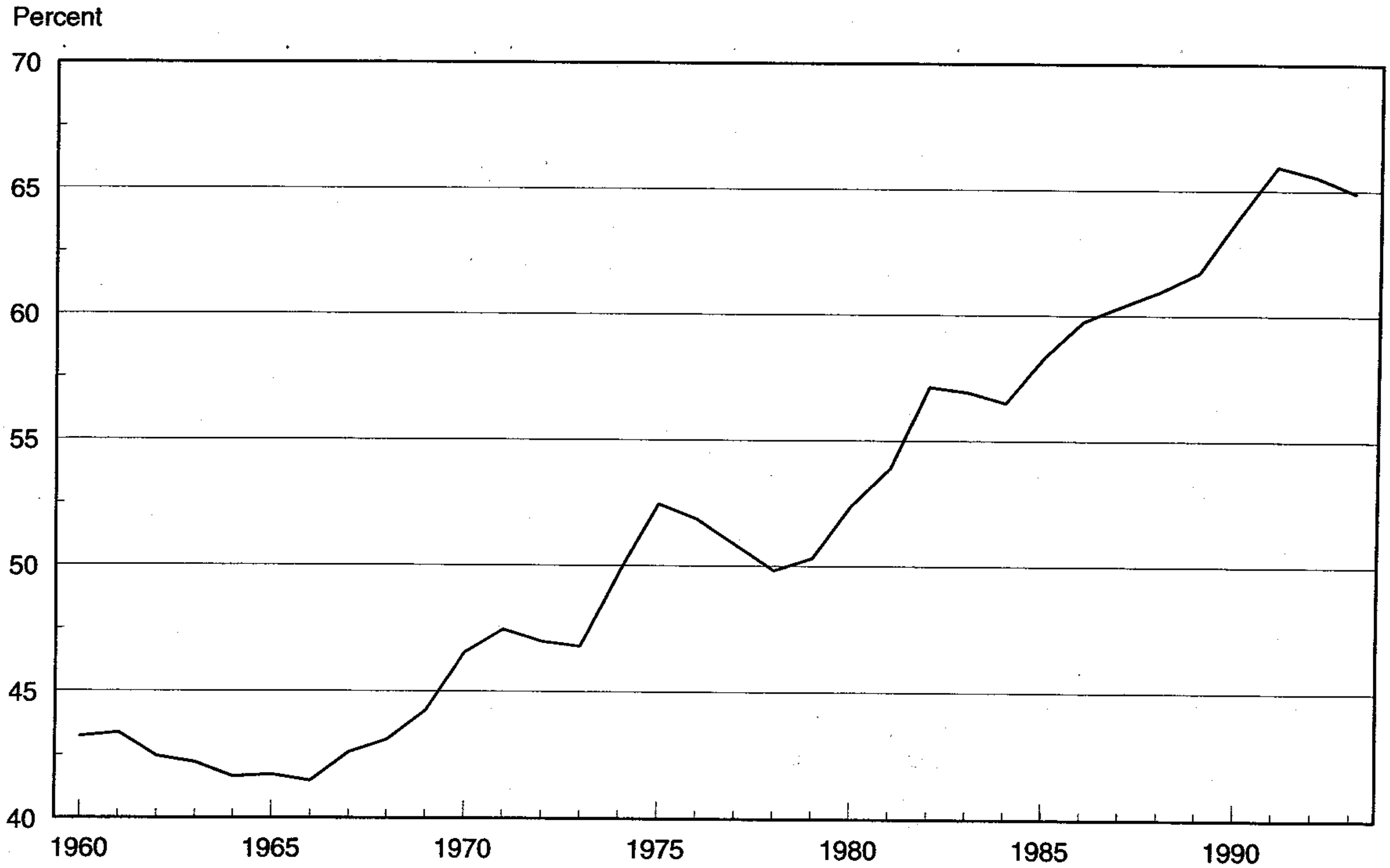
Commercial Construction



*Also includes a small portion of service trade establishments such as low rise banks and financial institutions, parking garages, auto repair garages, dry cleaners etc..

Chart 3

Retail Structures as a share of GDP



Note: Both series are measured in real 1987 dollars

Chart 4

The Growing Share of Discounters

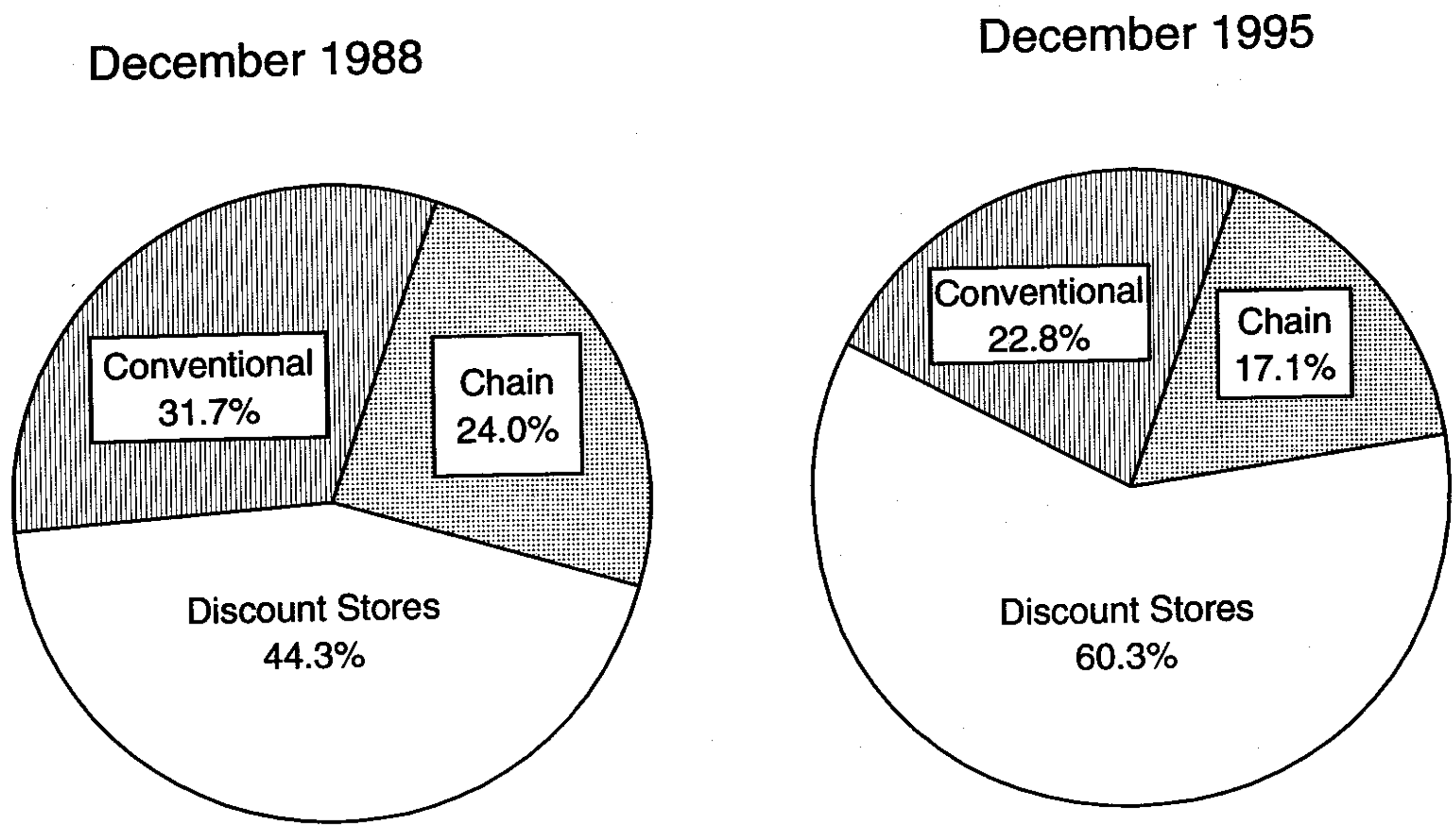


Chart 5

Consumer Inflation Trends

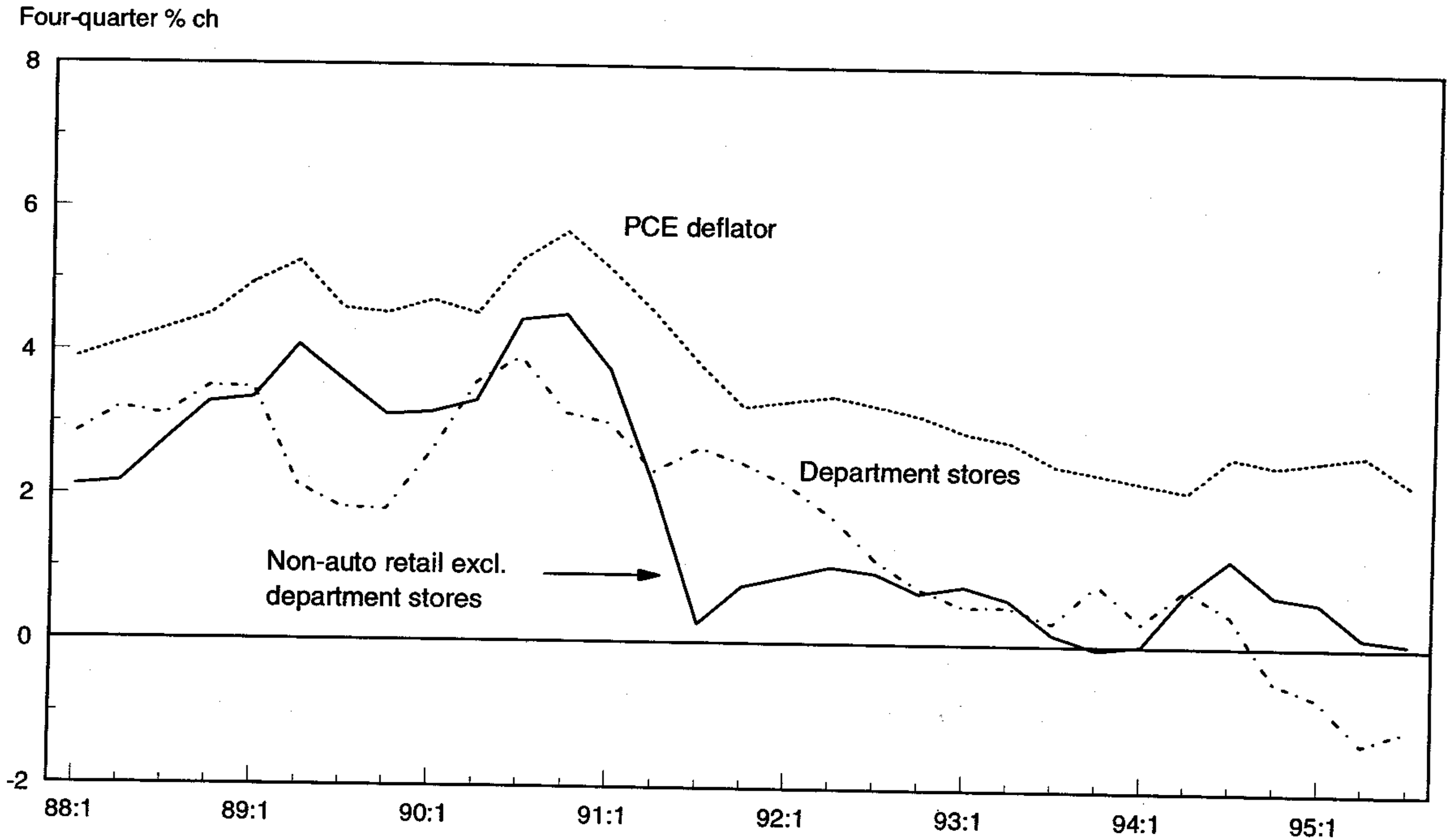
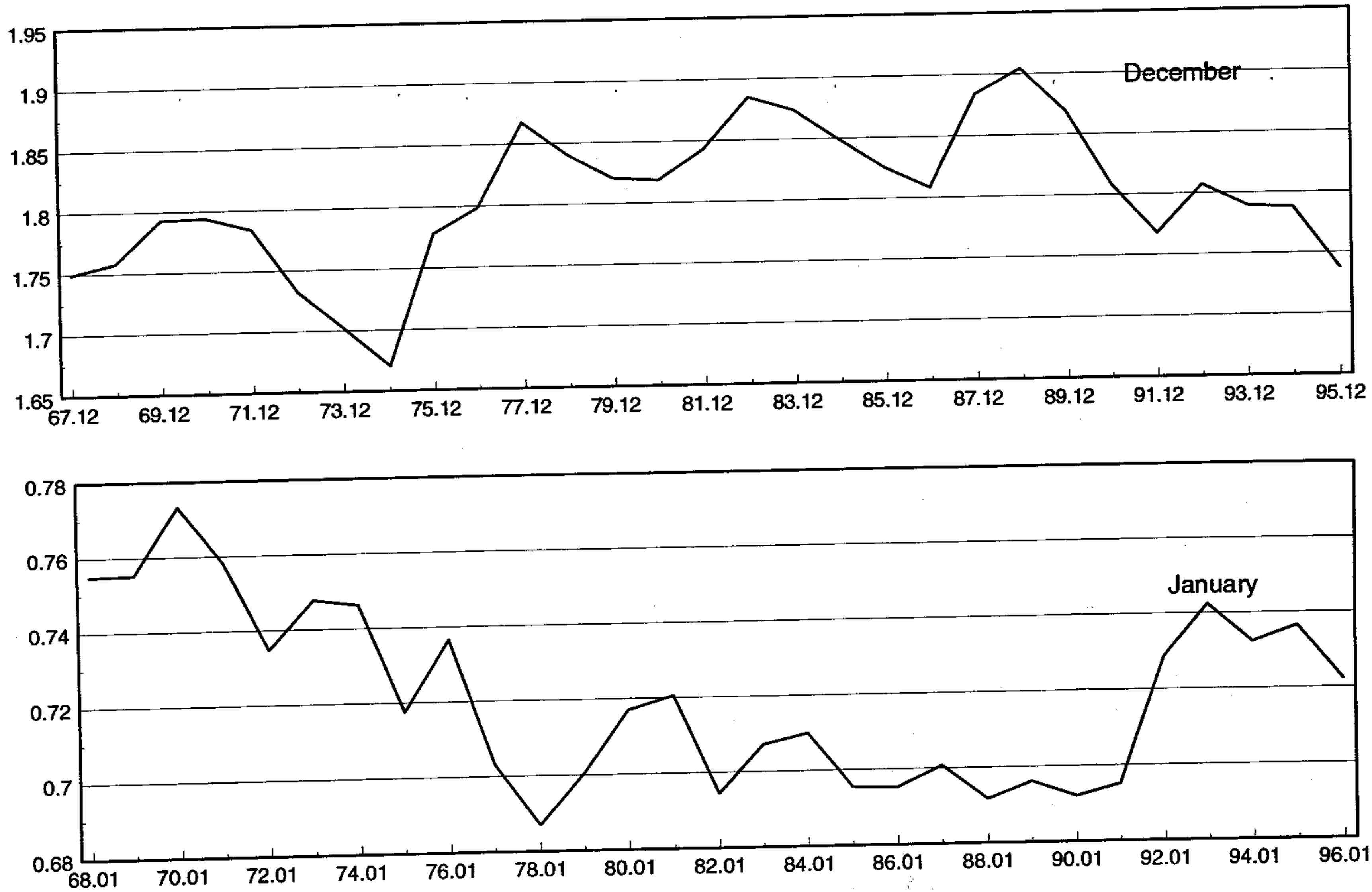


Chart 6

The Changing Christmas Season for Department Stores*



* The chart shows the ratio of December and January to their respective 12-month centered moving average sales. The latest value is estimated assuming continued trend growth in sales.