

Statistics New Zealand - Te Tari Tatau

Retail Trade Survey - Survey design Information Paper

Retail Trade Survey

Implementation of New Survey Design

Introduction

The monthly Retail Trade Survey (RTS) was redesigned in 2003 (an information paper describing the redesign was released on 20 January 2004). The previous design operated from the May 1995 to the October 2003 month. To assist users in moving to the redesigned survey, Statistics New Zealand produced an *analytical back series* from the October 2003 month to the May 1995 month. This allowed the ongoing production of a seasonally adjusted and trend series. To provide the information necessary to produce this analytical series, a dual run of the August, September and October 2003 months was undertaken, whereby the survey was conducted on both the old and new designs.

Sample surveys require periodic redesign to ensure that the sample adequately represents the contemporary composition of the population. The redesign of the RTS also incorporated a number of methodological changes aimed at improving the reliability and quality of the results of the survey, while reducing the overall respondent load. These were consistent with changes made to the methodology of the redesigned Economic Survey of Manufacturing (QMS) and Wholesale Trade Survey (WTS) in the last two years.

Statistics New Zealand has noted that some commentators have recently expressed concern about the volatility of our retail trade statistics, especially as they relate to the movements in the seasonally adjusted series. The recently introduced redesign of the survey, which introduced a number of methodological changes, was suggested as the likely source of added volatility.

This paper summarises the methodology of the RTS and its trend and seasonal

adjustment analysis. It also checks recent volatility against that in existence prior to the redesign.

Summary of methodological changes

The scope of the RTS did not change with the redesign. It remains a monthly survey to collect data from businesses classified as undertaking retail trading or specified service activities (Australia New Zealand Standard Industrial Classification divisions of Retail Trade, Accommodation and Personal services). The basic method of selecting a sample of businesses based on industry and size, and then weighting it to represent the population of retailers, has not changed.

The methodology for calculating trends and seasonally adjusted series for component and total series, via the US Bureau of the Census X-12-ARIMA program, has not changed.

Changes to sample selection and data collection:

- A revised sample, based on current type of activity (eg whether a supermarket or clothing store) and size, was selected to represent the current population distribution of retailers.
- Where possible, small and medium sized businesses in the previous survey were rotated out of survey selection and replaced with different businesses. This is part of Statistics New Zealand's intent to minimise compliance costs for New Zealand businesses.
- The sample was split, to sample from 24 retail industries (previously 17 and published as 15). This increases the quality of statistics for industries not previously separately identified in sample selection.
- A reselection, or restratification, methodology was introduced. Businesses that change their activity or size are reclassified on the basis of their current month's industry and size. Each month, the sample is selected using the same criteria to ensure a high overlap between samples.
- Sales reported for GST are now used for small businesses, rather than collecting data directly via questionnaires.

Other changes affecting the Retail Trade Survey

At the same time, there were changes in the operating environment of the RTS as part of Statistics New Zealand's ongoing improvement of its processes.

Changes to the Business Frame maintenance

New maintenance methods were introduced to the Business Frame, Statistics New Zealand's register for selecting business survey participants, from August

2003. This was independent of the redesign of the RTS.

Processing changes

The main processing change for the new survey design involved changes to the methods of imputing data for non-respondents to the survey. Such changes are ongoing improvements, to reduce potential errors in data.

Impact of these changes

Revised sample selection and rotation

A revision of sample selection, to reclassify businesses based on type of activity (eg whether a supermarket or clothing store) and size, and then select a sample to represent the current population distribution of retailers, was the goal of the redesign. This means that for some businesses, their size classification and weighting changed from that under the previous sample design and that their chance of sample selection is now appropriate to their current size. This mathematical exercise is required regularly for any sample survey, to ensure the sample of businesses represents the current population distribution. The size identification is now based on annual GST sales data from Inland Revenue, and a count of employees, rather than on a full-time equivalent employee count, as was done previously.

At the same time, where possible, small and medium sized businesses selected in the previous survey were rotated from survey selection and replaced with different businesses. This is part of Statistics New Zealand's intent to minimise compliance costs for New Zealand businesses. Over half the business locations included in the previous sample were also selected in the new sample. Much of this overlap comes from large multi-location businesses that have a 100 percent chance of sample selection. The dual run operated from August 2003 to October 2003, to collect data using both the previous and new sample designs and sample selections. This allowed Statistics New Zealand to confirm the effects of changing the sample (see time series issues, below).

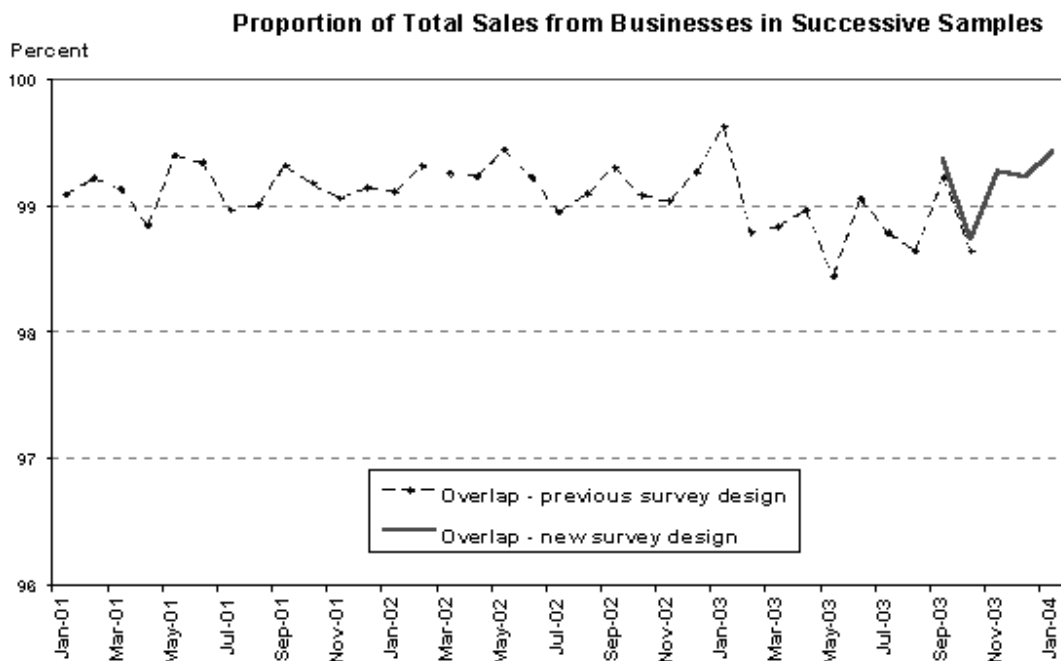
Another change was to split the sample into samples from 24 retail industries, where previously the sample was split into 17 industries and published as 15. This provides more accurate data for industries previously not separately identified in sample selection, and ensures they have appropriate representation in the total sample. These industries were defined by the industry classification used by Statistics New Zealand's National Accounts division, which is used for sample selection in other Statistics New Zealand financial surveys.

Regular reselection

The RTS has followed the methodology of other Statistics New Zealand subannual financial surveys, ie the QMS and WTS, to introduce a reselection

methodology, where the sample is restratified every month. This means that businesses are reclassified if their current industry and size has changed and the sample is reselected under the same parameters (ie the same sample size from the same part of the Business Frame). This means that rapidly growing or shrinking businesses are assessed for selection based on their new characteristics and not those at the time of the survey redesign. It also preserves a high overlap between samples, to yield reliable movements. This means the sample is a better representation of the population at any month, but that businesses have regular opportunity to enter or leave the survey or receive a different weighting for their contribution to the total. This is in addition to the ongoing process of incorporating new businesses in the sample and removing ones that have ceased operation.

While reselection does lower the overlap between successive samples, this methodology still selects mostly the same businesses from month to month. Only those changing in size are eligible for addition to, or removal from, the sample. The proportion of sales coming from businesses sampled in two successive months remains around 99 percent, as shown in the graph below.



The sample reselection methodology is not having a significant impact on movements, nor is it adding extra noise to the series. This can be seen by comparison of monthly movements of total sales for all businesses over recent months, with movements of total sales for businesses in successive samples,

and with movements of total sales for businesses receiving the same weighting in successive samples.

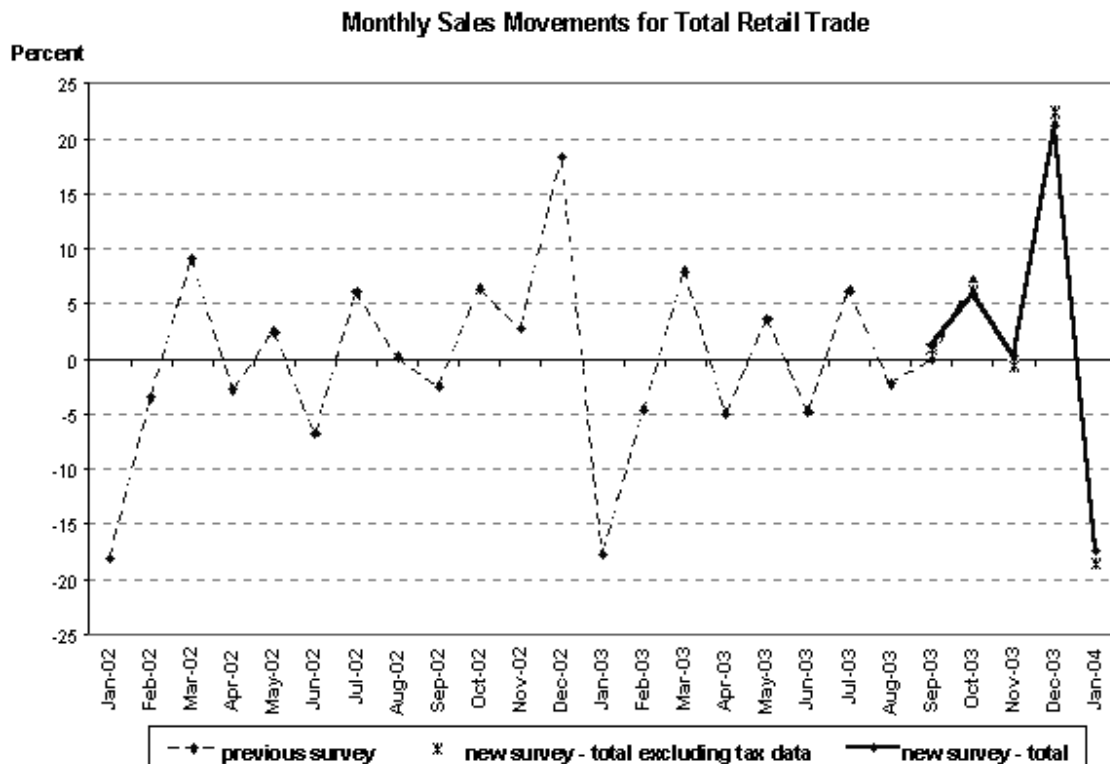
Use of administrative data for small businesses

The RTS has followed the methodology of other Statistics New Zealand subannual financial surveys (ie QMS and WTS) to use GST data for all small businesses with simple structures rather than collecting data directly via questionnaires. Again this is implemented as part of Statistics New Zealand's intent to minimise compliance costs for New Zealand businesses, especially small ones, and make full use of pre-existing alternative data.

As GST data is not available on a monthly basis for many small businesses, and not available for the latest month (due to the timing of Inland Revenue reporting and processing timeframes), Statistics New Zealand has developed a methodology that apports monthly GST data for all businesses, and forecasts data to the month required (this is a two-month ahead forecast for RTS). To provide insurance against any large shocks in current data not observed in a forecast, we also adjust the forecasted GST data by a ratio of questionnaire sales to GST sales calculated from similar businesses in the postal survey.

This methodology, with 100 percent coverage through administrative data, has replaced the previous method of sampling a small proportion of these businesses. The latter method was subject to considerable sampling error. This new methodology has improved coverage of a volatile part of the retail sector, although the modelling procedures, which are a sophisticated method of imputation, do introduce a degree of non-sampling error to the data. Statistics New Zealand limits the contribution of this data to small businesses that contribute the bottom 10 percent of retail sales. Depending on the industry they belong to, the cut-off for determining these small businesses ranges from \$100,000 to \$2 million annual sales. Eighty percent of businesses for which tax data is used have annual sales of less than \$200,000.

Comparing monthly movements of total sales for businesses sourced from postal survey data only, with movements for all businesses over recent months, shows that this tax data methodology for small businesses is not having a significant impact on movements, nor does it add extra noise to the series. This is shown in the graph below.



Changes to Business Frame maintenance

From August 2003, Statistics New Zealand introduced new methods for maintenance of its register of businesses, which is used as a frame for selecting samples. Some key features relevant to business survey operation arising from greater use of data sourced from Inland Revenue are:

- more timely incorporation of new businesses onto the Business Frame, especially for small businesses
- a tier-based maintenance philosophy for existing businesses on the Business Frame
- replacement of survey data on full-time equivalent employees with a count of total employees engaged, sourced from Inland Revenue's employer monthly schedule.

The change of employment variable was factored into the sample redesign. The move to a more timely incorporation of new businesses onto the Business Frame means that real world patterns in businesses registering for GST with Inland Revenue (the source for the Business Frame), are likely to be reflected in a more timely manner in samples drawn from the frame.

Processing changes

Changes such as methods of imputing data for non-respondents, or methods of identifying unusual responses to exclude from imputation calculations, are ongoing improvements to reduce potential errors in data. The main change introduced was use of regression imputation, based on a ratio of reported sales to annual GST sales, for simple structure retailers in the postal survey. This is in addition to methods based on historical movements and means. Following the survey redesign, the imputation cells (ie groups of similar activity and size businesses) were redefined. This involves the level at which imputation calculations are made.

Imputation is an estimation process to substitute for actual data from retailers that adds non-sampling error to the series, regardless of methodology used. Comparison of monthly movements of total sales for responding businesses, with movements for all businesses over recent months and over several years in the previous survey, shows that imputed data can add to or reduce the extremity of movements. There has been little change to this over time.

Time series issues – creating a back series and seasonal adjustment

Linking

At least five years of data is required for seasonally adjusted and trend estimates to be produced. This is necessary to ensure that seasonal patterns can be properly identified. If a series has monthly trading day variation, then it needs to be at least seven years in length. Retail trade is known to be affected by seasonal and trading day variation. Because the redesign has meant that more retail industries are available, *analytical back time series* prior to the redesign need to be estimated if Statistics New Zealand is to provide seasonally adjusted and trend estimates for retail trade.

Creating back series

As explained above, Statistics New Zealand now publishes 24 retail industry time series (15 were published before the redesign). To create a back time series for 24 industries, unit responses from the previous survey were coded to a new retail industry, where necessary. The survey weights were applied to these reclassified unit's past responses and time series were created, going back eight years.

We expected there could be a significant difference in level between the back and the new time series, which would cause a discontinuity in the trend estimates. At the very least, the current and back series are estimated from differing samples, therefore we would expect some difference. This is what sample error measures, therefore smaller retail industries would be expected to show greater differences than larger industries. The difference at the total level is expected to be close to 1. It should be noted that many retailers, particularly

large ones, were in the old survey and continue to be in the new survey, so there is considerable overlap in the two samples.

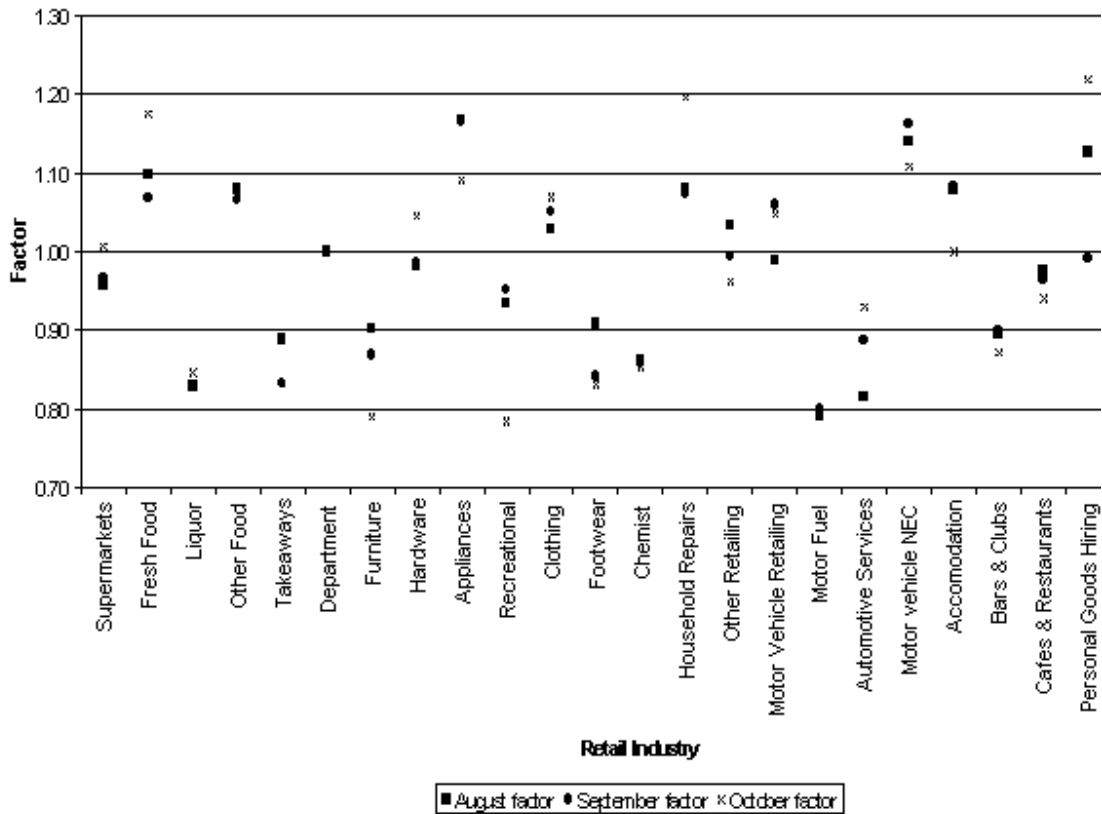
As there could be a difference between the levels of old and new retail industry series, the standard Statistics New Zealand policy of smoothing out any differences between old and new series was followed. For this, a current linking factor (ie the difference between old and new series values) is determined at the point the two series join. At some designated point in the past, usually the date of the last redesign, the linking factor is assumed to be 1. Intervening time period linking factors are a smoothed transition from 1 to the current linking factor.

To estimate the current linking factor, it is necessary to have estimates for both the old and new series. This requires a *dual run*, where both old and new survey units are surveyed and survey estimates from both groups are made. The difference between the two estimates is the current linking factor for each retail industry.

To enable Statistics New Zealand to produce ad hoc time series other than for the published retail industries, the linking factor for a particular retail industry is combined with the survey weights for the individual units in the past survey that have been classified, or reclassified, to that industry. This ensures other time series are consistent with published retail industry series. As the retail survey is designed only to provide estimates for national retail industries, back data for any other categories could be of lesser quality.

The linking factors for each retail industry and total retail trade are plotted in the following graph.

Comparison of Linking Factors By Industry



As can be seen for many of the retail industries, the linking factors do not vary much over the three months. The greatest variation occurs for the smaller retail industries, which have the largest sample errors and are therefore known to have the greatest variability. As consistent quarterly estimates are required, it was decided to use the average of the August and September linking factors as the current September quarter linking factor.

Seasonal adjustment and trend estimation

Seasonal adjustment and trend estimation in Statistics New Zealand decomposes a time series into several components. These components reflect different types of factors that would be expected to drive the movements of the series. The components estimated for retail trade series are:

Trend

This models the factors that drive the series in the longer term (ie for periods greater than a year). Typical factors would include the underlying patterns of economic growth, consumer confidence, and retail behaviour.

Seasonality

The seasonal component models factors that occur at very much the same time each year. Typical factors include: the total number of days in the month; Christmas; and the arrival of the new season fashions. Because consumer and retail behaviour changes over time, it is possible that the seasonal pattern can also change over time, either in shape or amplitude. Therefore we allow seasonality to evolve over time, but cannot model abrupt changes in seasonal patterns (eg changing the government financial year).

Trading day

The majority of months are not an exact number of weeks long. As there is variation in spending patterns within a week, there may be change from one month's value to the next because the number of times a particular day of the week occurs is different (eg the number of Sundays in a month).

Irregular

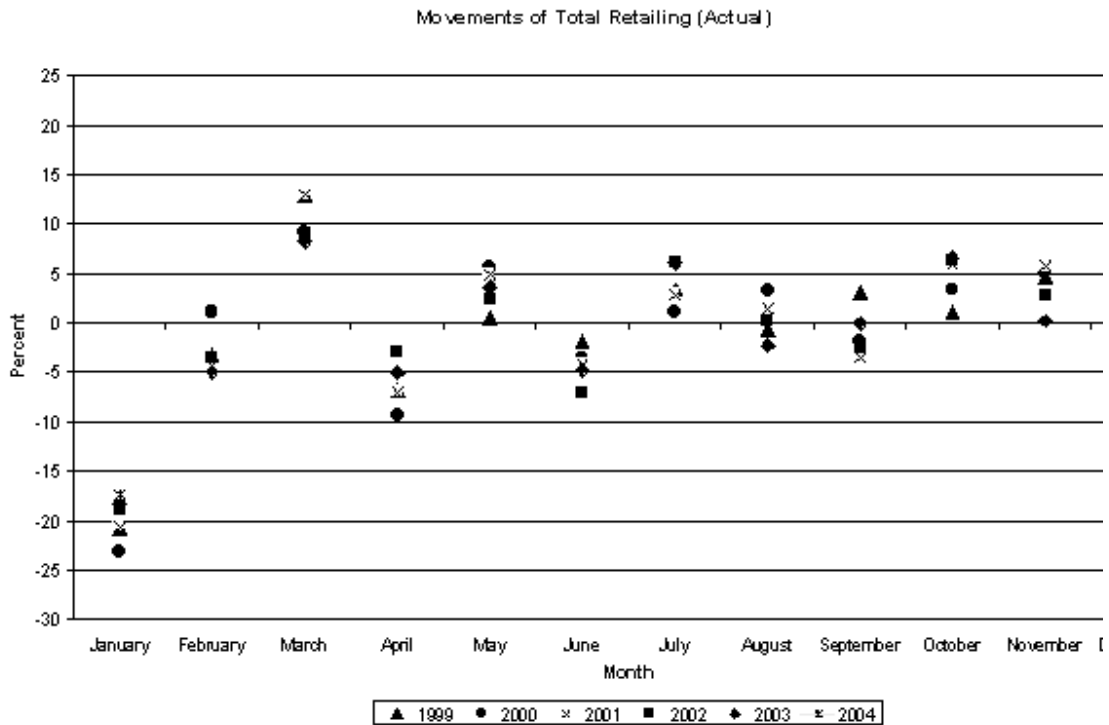
This component occurs due to the effect of unusual events and the normal variation that can arise through consumers making a range of spending decisions. Typical factors that can affect this component are climatic events or consumers holding off or bringing forward purchases to take advantage of actual or anticipated price changes.

To produce a seasonally adjusted series, Statistics New Zealand estimates the seasonal and trading day components and removes them from the original series. Therefore the seasonally adjusted series is a combination of the trend and irregular components. All components are estimated using the X-12-ARIMA seasonal adjustment package developed and used by the US Bureau of the Census.

A total can be either estimated by directly adjusting the total, or by adjusting the components and summing those adjusted series. The latter is generally better if the components have varying seasonal patterns. Retail trade totals are sums of adjusted components.

The irregular component is of particular interest to our time series analysts as if any of the other components abruptly change (eg a trend break or a change in legal shopping hours) then the irregular component will increase in volatility over and above what is present in the rest of the series. Therefore increased volatility implies that there are issues in the estimation of one or more of the components that need to be investigated. It should be noted that it is entirely feasible for increased volatility to be due to more volatility in the behaviour of consumers and retailers, though this tends to be a low irregular component followed by a large one, or vice versa. Fundamental changes increase the variation in the irregular component over many months.

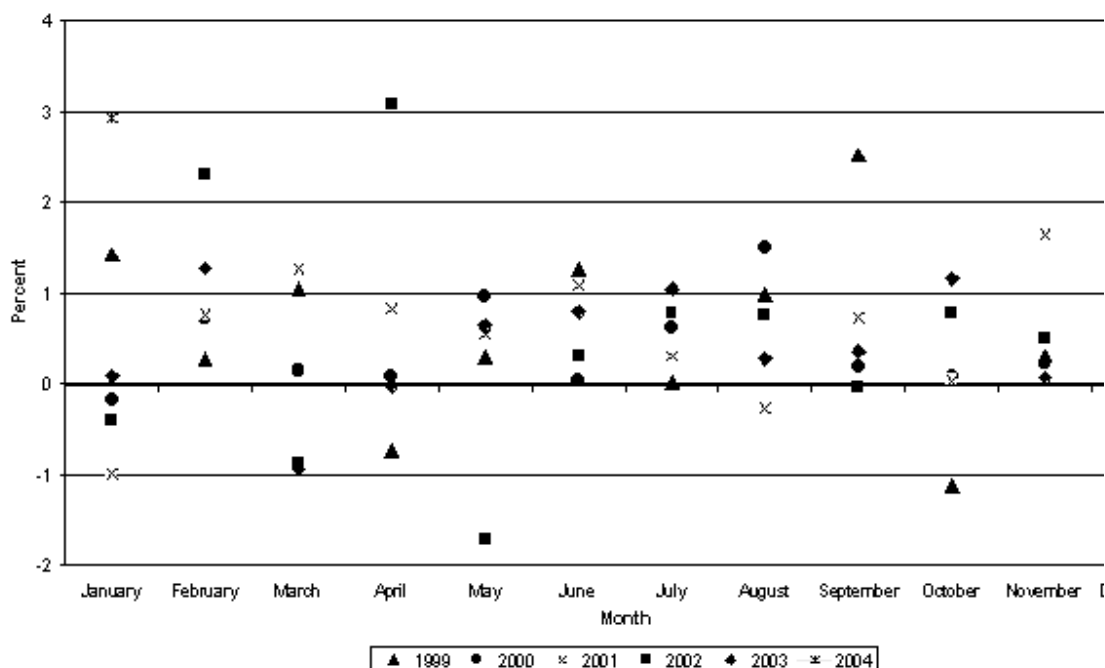
We plot the values of the month-on-month movements for the actual series of total retail trade for the last five years. As each month will have a different seasonal value, it is best to plot them by month.



It can be seen that there is a regular pattern to the movements, which arises from the seasonal variation. We cannot see trend or trading day variations from this graph. To better identify trends, we need to estimate the seasonal and trading day factors for each month.

Below is a similar graph for the seasonally adjusted series, where seasonal and trading day patterns have been removed.

Movements of Total Retailing (Seasonally Adjusted)



The first thing to note is that the seasonally adjusted series exhibits a lot less variability (-2.1% to +3.4%) than the actual (-24% to +22%). It is therefore more desirable to use the seasonally adjusted series for analysis than the actual series. The adjusted movement for January 2004 is large, but not the largest that has occurred for retail trade. A large movement in an adjusted value often precedes or follows a large movement in the opposite direction. This appears to be the case for December 2003 and January 2004 but this is not unusual behaviour in this series.

The seasonal adjustment methodology is performing as it should in identifying unusual movements in the recent retail trade series. This has occurred previously in this retail trade series and there is no evidence that the new series has a different pattern than the back-casted series, or that the seasonal pattern of retail expenditure has abruptly changed.

For technical information contact:
 Rochelle Barrow
 Christchurch 03 964 8700
Email: rochelle.barrow@stats.govt.nz