Price Indexes from Scanner Data: A comparison of different methods

Paper presented at the Ottawa Group,
Wellington, New Zealand,
4 May 2011

Frances Krsinich

Senior Research Statistician, Prices Unit, Statistics New Zealand
P O Box 2922
Wellington, New Zealand
frances.krsinich@stats.govt.nz

www.stats.govt.nz

Citation
Krsinich, F (2011, May). *Price indexes from scanner data.* Paper presented at the Ottawa Group, Wellington, New Zealand.

Abstract

The rolling window GEKS (RGEKS) method proposed by Ivancic, Diewart, and Fox (2009) can be used as a benchmark method for the production of price indexes from scanner data. It balances the invariance to chain drift of the GEKS method with preservation of characteristicity via a rolling window.

Against this RGEKS benchmark we have compared alternative methods, across some quite different scanner data sources – weekly scanner data for supermarket products and monthly scanner data for consumer electronic appliances (including computers, DVD players, and TVs).

Monthly chained Tornqvist indexes result in chain drift for many products. We also examine the performance of annually chained Tornqvist indexes and hedonic indexes controlling for product specification number – a 'time product dummy' (TPD) method. We discuss the patterns that emerge from this analysis, and what they suggest for further research.

Introduction

Statistics New Zealand had access to a research dataset of Australian supermarket data as part of a team of peer reviewers of the Australian Bureau of Statistics' scanner data research programme. We were able to use this data to test our understanding of the RGEKS method. We also compared the RGEKS indexes to monthly and annually chained Tornqvist and TPD hedonic indexes.

We then used New Zealand consumer electronics scanner data to explore how the different index methods perform and compare across some quite different products.

Data

We have access to three years of Australian supermarket scanner data, for a major supermarket chain in the greater New South Wales area. This is in the form of weekly sales and quantities by outlet, for each 'itemkey', or detailed product specification.

Table 1 shows the nine categories of products available in the data, with the corresponding number of weekly records, and individual product specifications.

Table 1
Categories in Australian supermarket scanner data

Category	Records[1]	Number of individual product specifications
Breakfast goods		
Cereals	4,400,000	350
Muesli	1,700,000	194
Carbonated soft drinks		
Bottles and cans	5,800,000	354
Mixers	1,100,000	60
Paper goods		
Food wraps, bags, and storage	2,100,000	109
Garbage bags	1,300,000	63
Paper towels	700,000	42
Tissues	1,500,000	78
Toilet rolls	1,900,000	120

1. Rounded to the nearest 100,000.

Statistics New Zealand purchases scanner data on consumer electronics from market research company GfK. This is used for expenditure weighting, sample brand shares, and sample product specification in the New Zealand consumers price index (CPI). Data is available as monthly sales and quantities by combinations of characteristics of the products. Although there are different records for different outlets, we don't have outlet identification numbers so we can't control explicitly for this aspect of the data.

The two most recent years of this data – from July 2008 to June 2010 – had highly comparable classifications and could be merged easily. We merged the most recent two years of this data for the 14 products which are shown in table 2 with the corresponding number of monthly records.

Table 2
Products in GfK consumer electronics scanner data

Category	Records(1)
Camcorders	1,600
Cathode ray TVs	800
Desktop computers	3,200
Digital cameras	6,100
DVD players and recorders	2,000
Multi-functional devices	1,800
Mobile computers	8,500
Monitors	2,300
Cordless phones	1,600
Portable media players	3,700
Printers	1,000
Plasma / flat-screen TVs	4,200
Scanners	200
Total	39,700

1. Rounded to the nearest 100.

Methods

We calculated rolling Gini-Elteto-Koves-Szulc (RGEKS), TPD hedonic, and chained Tornqvist indexes using the two sets of scanner data.

RGEKS

Ivancic, Diewart, and Fox (2010) proposed the RGEKS index, adapted from multilateral index theory, to avoid the chain drift that can occur when chained traditional indexes are calculated from scanner data, due to its high frequency and the spiking behavior of prices and quantities that results from sales.
We followed Ivancic, Diewert, and Fox (2009) in applying the RGEKS with a 13-month rolling window, based on the Tornqvist formula.

Greenlees and McClelland (2010) formulate this as follows:

For the first 13 months, the RGEKS index between time t and t-1 is given by:

$$RGEKS_{t,t-1} = GEKS_{t,t-1} = \prod_{s=1}^{13} \left[T_{ts} / T_{t-1,s} \right]^{1/13} t = 1,\ldots,13$$

For periods 14 and later, recursively compute:

$$RGEKS_{t,t-1} = \prod_{s=t-12}^{t} \left[T_{ts} / T_{t-1,s} \right]^{1/13}$$

For the Australian supermarket data we calculated average prices for each outlet-week-product specification combination and constructed the RGEKS from these.

Similarly, for the New Zealand consumer electronics data we applied the RGEKS on average prices for outlet-month-product specification combinations, after constructing pseudo product specifications by combining the categorical characteristics.

Whether the RGEKS is necessarily always the optimal index method appears to be an open question, in particular when the life cycle of a product falls within the window length (Greenlees & McClellan, 2010). Also, further work needs to be done to determine the optimal window length, which might vary with product. However, for the purposes of this paper, we have taken the RGEKS index with a 13-month window to be the benchmark against which the other index methods –TPD hedonic, monthly chained Tornqvist, and annually chained Tornqvist – are compared. We note that the choice of benchmark among these methods will not affect the comparisons between them.

Time product dummy hedonic method

We calculated hedonic indexes using a TPD approach. That is, we modelled the log of price against time and product specification. Our work to date has used the full pooled data for the regressions – that is, three years in the case of the Australian supermarket data, and two years in the case of the New Zealand consumer electronics data. In future research, and particularly with longer data series, we intend to use a rolling window to calculate the TPD index, as with the RGEKS index. We incorporate expenditure weights into the regressions to reflect the relative importance of different product specifications.

Australian supermarket data

We are unable to directly show the resulting indexes, due to confidentiality restrictions, but we can show secondary summary statistics and the relative behaviour of the various methods against the RGEKS benchmark.

The data we have is three years (May 2007 to July 2010) of weekly sales and quantities for each detailed product specification, by outlet. Table 1 shows the number of weekly records and the number of individual product specifications for each category.

We ran RGEKS (Tornqvist, 13-month window), TPD hedonic, monthly chained Tornqvist, and annually chained Tornqvist indexes on this data.

There is a high degree of turnover, or 'churn', in the product specifications being sold over the three years examined. This is one of the aspects of the scanner data that poses problems for the use of traditional index formulae as unchained indexes will become less representative of the full population of products being sold over time. Table 3 shows the percentage of the total sales of each product contributed by new product specifications at the end of three years. Toilet rolls shows the most significant contribution by new product specifications to total sales, at 55.5 percent. At the other extreme, soft drinks – bottles and cans – has only 9.6 percent of total sales being contributed by new product specifications at the end of the three years. This is likely to be due to the long-standing dominance of a few products in this market.

Table 3
Churn in itemkeys by category

Category	Contribution of new productspecs to total price after 3 years (%)
Cereals	46.7
Muesli	38.4
Soft drinks – bottles and cans	9.6
Soft drinks – mixers	16.4
Foodwraps, bags, and storage	20.3
Garbage bags	21.8
Paper towels	33.4
Tissues	22.4
Toilet rolls	55.5

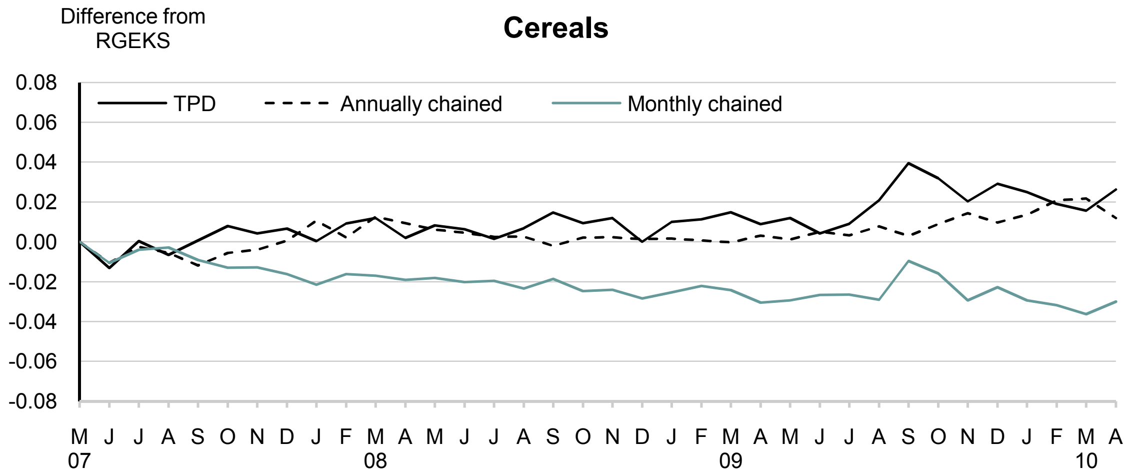
Table 4 shows the R-squared statistics from the TPD models that show how much of the variability in log of price is explained by time and product specification. These are all very high, ranging from cereal, with an R-squared of 0.82, to foodwraps, bags, and storage with an R-squared of 0.99.

Table 4
R-squared statistics by category

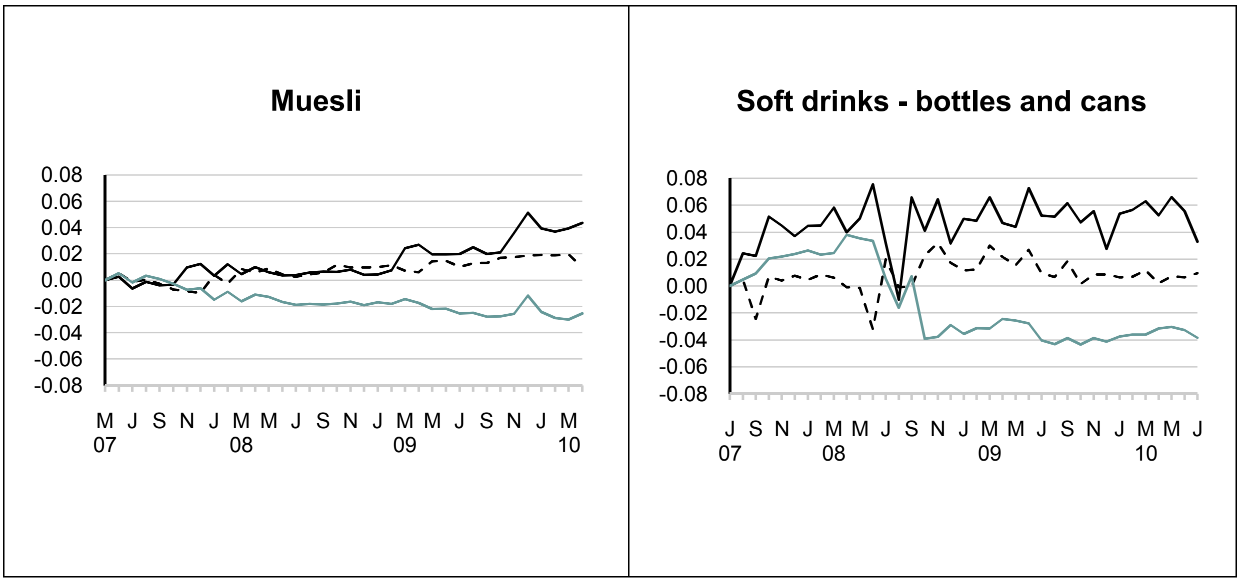
Category	R-squared
Cereals	0.82
Muesli	0.95
Soft drinks – bottles and cans	0.97
Soft drinks – mixers	0.97
Foodwraps, bags, and storage	0.99
Garbage bags	0.94
Paper towels	0.92
Tissues	0.95
Toilet rolls	0.92

As mentioned above we can't show the actual price indexes constructed from the Australian scanner data for confidentiality reasons. Instead, in Figure 1, we present the difference of each index from the RGEKS. We do this for each of the nine products. The indexes are all expressed on a base of 1 at May 2007.

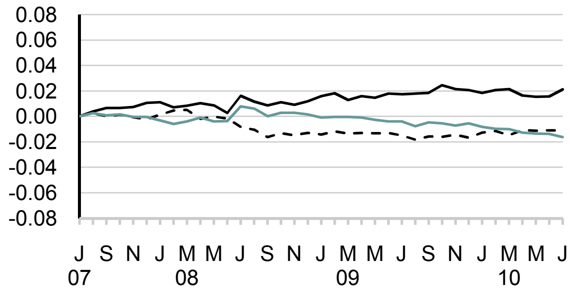
Figure 1
Australian scanner data – difference from RGEKS index of other methods



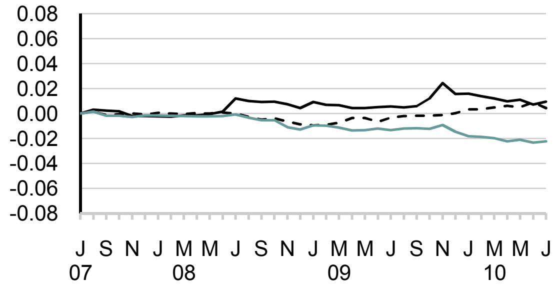
Source: Australian research scanner data



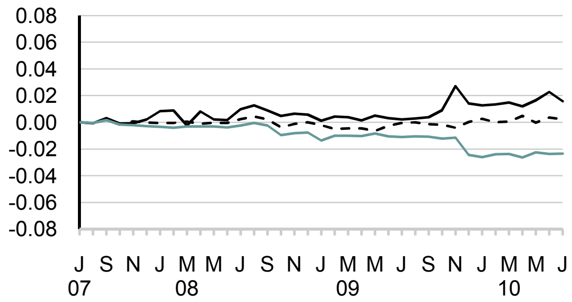
Soft drinks - mixers



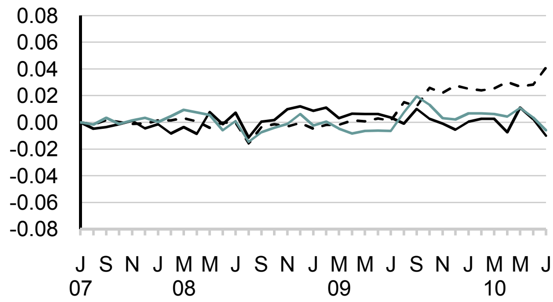
Foodwraps, bags, and storage



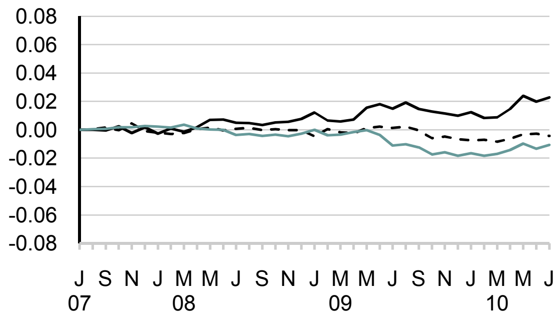
Garbage bags



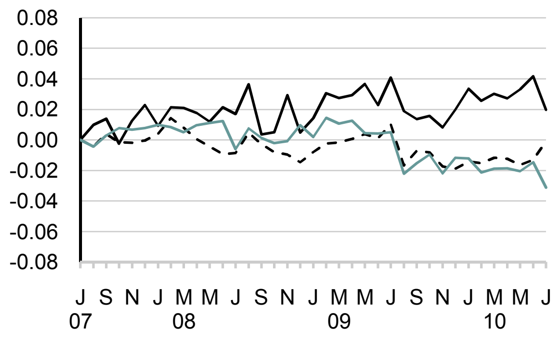
Paper towels



Tissues



Toilet rolls



In general, though not always, the monthly chained Tornqvist is biased downwards relative to the RGEKS. However, for the first year of soft drinks – bottles and cans, the monthly chained Tornqvist is above the RGEKS. For paper towels and toilet rolls the monthly chained Tornqvist is also above the RGEKS for around half of the time observed, though less so than for soft drinks – bottles and cans.

The TPD index, on the other hand, tends to be steeper than the RGEKS for most products, although an exception to this is paper towels, where it dips below the RGEKS a number of times.

Perhaps what is most notable about these results is that, although there are general patterns – that is, that the monthly chained tends to drift downwards and the TPD tends to drift upwards, with the annually chained Tornqvist usually sitting somewhere in the middle of the two – there is no pattern that is observed consistently across all the nine products examined here. This points to the need to look at these properties across a wider range of products before venturing generalisations about how the indexes are likely to behave relative to one another. This is what we intend to do using US supermarket scanner research data from IRI Marketing, which covers a wider range of product categories (around 30) over a longer time period (seven years).

Summarising these comparisons to RGEKS of the TPD and chained (monthly and annually) Tornqvist indexes, we took for each product an average of the differences in index points, standardised by the value of RGEKS:

$$D_m = \left(\sum_{t=1}^{36} ABS\left(I_{m,t} - I_{RGEKS,t} / I_{RGEKS} \right) \right) / 36$$

Where:

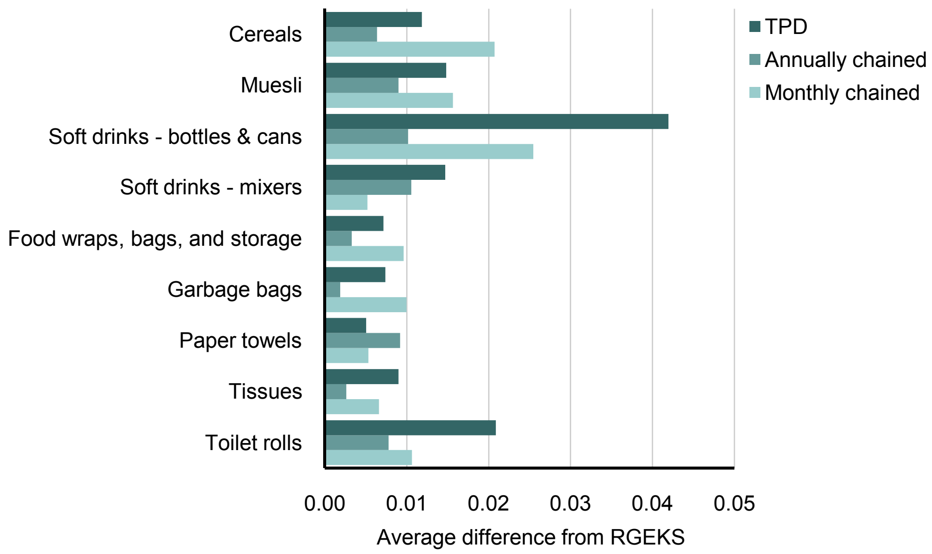
$I_{m,t}$ is the index using method m, for time t, and

m = TPD hedonic, annually chained Tornvist, or monthly chained Tonqvist

Figure 2 graphs these average differences across method and product. Looking at the results this way we again see that there is no pattern that is consistent across all the products, but monthly chained Tornqvists are on average further away from RGEKS than annually chained Tornqvists, with the exception of soft drinks – mixers and paper towels. With the exception of paper towels again, the TPD indexes are on average further away from RGEKS than the annually chained Tornqvists. This result suggests that for this type of data, annually chained Tornqvists may be quite a reasonable alternative to the RGEKS.

The TPD and monthly chained Tornqvists for soft drinks – bottles and cans, are on average the most different from the benchmark RGEKS with average differences of approximately 0.025 and 0.04 index points, respectively. This would be an interesting product to investigate further. As noted above the contribution of new product specifications in this product is very low, but this is likely to be a consequence of a very skewed distribution of contributions by individual product specifications due to some dominant brands.

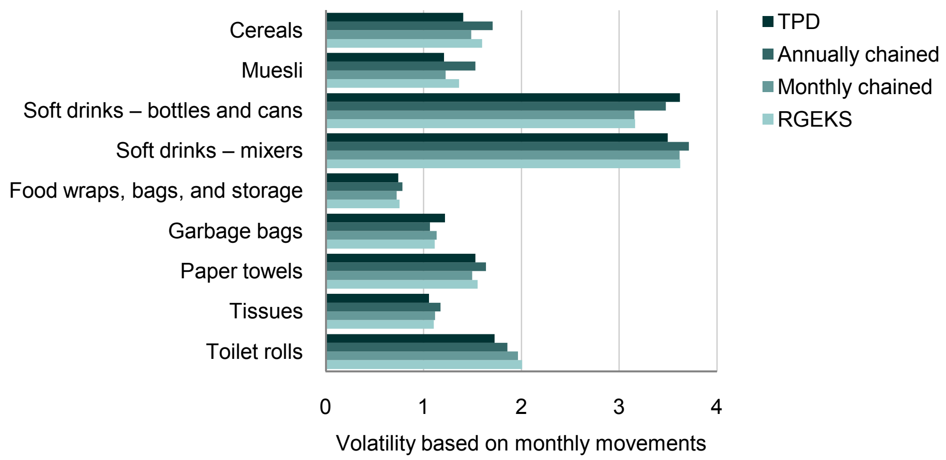
Figure 2
Average difference from RGEKS



Source: Australian research scanner data

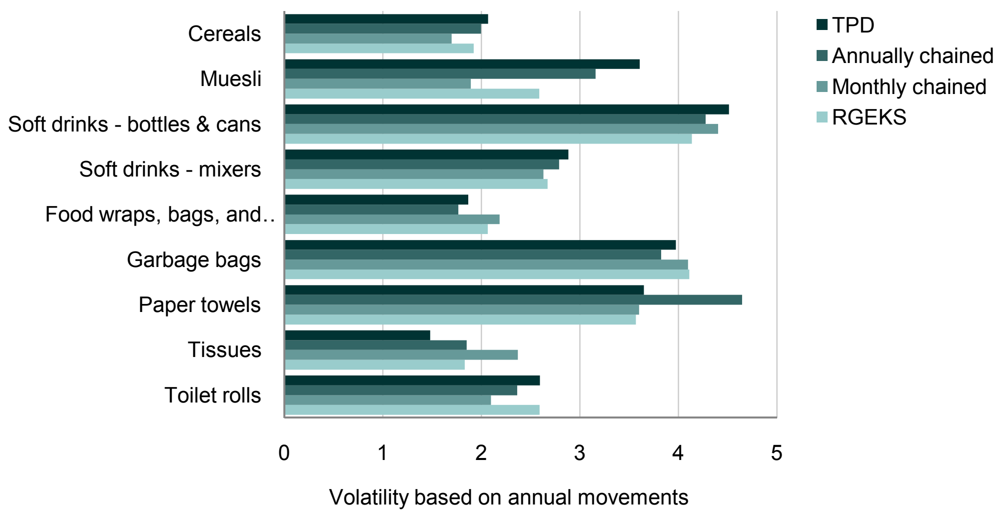
We measured the volatility of each of the four indexes on the basis of both monthly and annual average absolute percentage movements. We looked at annual percentage movements to smooth out any volatility due to seasonality in the monthly price movements. These are graphed in figures 3 and 4. None of the index methods appears to be consistently more or less volatile than the others across all these products and, in fact, it's hard to discern any pattern in these summarised measures of volatility.

Figure 3
Volatility – average absolute monthly percentage change



Source: Australian research scanner data

Figure 4
Volatility – average absolute annual percentage change



Source: Australian research scanner data

Consumer electronics data

We use scanner data on consumer electronics to inform the expenditure weighting, sample brand shares, and sample product specification in the New Zealand CPI. A number of years of data are available, but to date we have used just the two most recent years (which were very comparable and easily merged) to test the production of RGEKS, TPD, and chained Tornqvist price indexes. Primarily, we wanted to see how these methods compared with each other and whether we would observe the same kind of patterns as we had in the

Australian supermarket data (ie downwards drift in monthly chained versus upwards drift in TPD and no index method consistently more volatile than the rest) .

Note that we are not currently considering using this scanner data for price measurement for consumer electronics, as there are a number of important price-determining characteristics that are currently not on the scanner data obtained from GfK. For example, the data for DVD/Blu-ray players and recorders does not differentiate between whether the player or recorder is DVD or Blu-ray. Our main motivation here is to extend the comparison of different methods for producing indexes from scanner data on a dataset for which (unlike the Australian data) we can present the actual indexes. Also, thinking through the idea of combining categorical characteristic information to get pseudo barcodes is something that we can test with this data, which we think is a useful addition to the general research in this area.

What we found was that the scanner data indexes using all three methods are very similar to each other.

The data contains around six characteristics for each product.

For example, desktop computers have the following characteristics identified in the data: processor brand, processor number, ram (MB), storage capacity, and chipset brand.

We created product specification IDs for this data by combining together all the characteristics in the data so that each unique combination is separately identified. We treat all the characteristics as categorical variables because, even when they appear numeric, they take a relatively limited number of values.

The different products in the GFK data with the associated number of pseudo product specifications after combining all characteristics in the merged 2009/2010 data are shown in table 5.

Table 5
Number of pseudo product specifications by product

Product	Number of productspecs
Audio home systems	135
Camcorders	10
Cathode ray TVs	12
Desktop computers	308
Digital cameras	240
DVD players and recorders	40
Multi-functional devices	16
Mobile computers	591
Monitors	99
Cordless phones	78
Portable media players	81
Printers	41
Plasma / flat-screen TVs	54
Scanners	8

Note that these products don't correspond exactly to CPI categories – cathode ray TVs are no longer included in the CPI basket. None of monitors, printers, or scanners are priced separately.

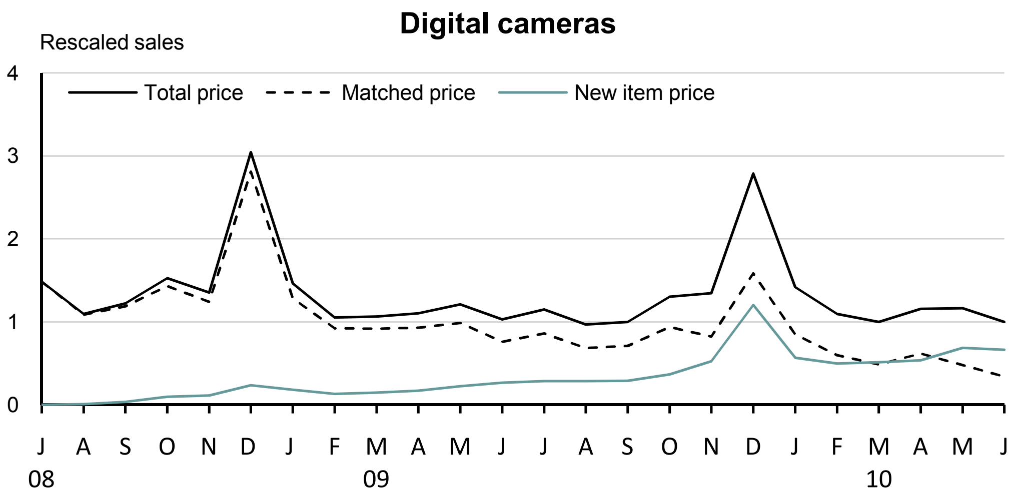
The churn in the pseudo product specifications being sold across the two years observed is quite variable. Table 6 shows the percentage of total sales contributed by new products at the end of the two-year period. For example, after two years, sales of new specifications of mobile computers contribute 99 percent of the total sales. In contrast to this, only 2 percent of the sales for DVD players and recorders are contributed by new specifications after two years. Cathode ray TVs are at the end of their life cycle and have virtually no sales by the end of the two years.

Table 6
Churn in pseudo product specifications by product

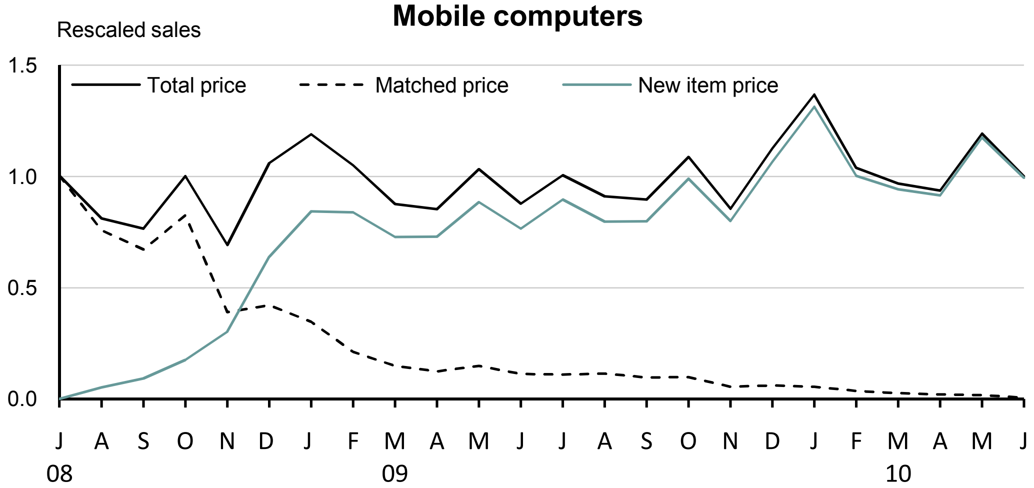
Product	Contribution of new productspecs to total price after 2 years (%)
Audio home systems	12.9
Camcorders	4.7
Cathode ray TVs	0.0
Desktop computers	72.3
Digital cameras	66.2
DVD players and recorders	2.2
Multi-functional devices	10.2
Mobile computers	99.3
Monitors	57.5
Cordless phones	2.6
Portable media players	25.3
Printers	0.5
Plasma / flat-screen TVs	79.3
Scanners	21.5

As an example of the churn over time, figure 5 shows the contribution of new and matched pseudo product specifications to the total sales for each of digital cameras and mobile computers, over the two-year period.

Figure 5
Churn of productspecs in digital cameras and mobile computers



Source: GfK Retail and Technology



Source: GfK Retail and Technology

Note: for confidentiality reasons, the y-axis has been rescaled so that total sales at the end of the two years equals 1.

Table 7 shows how well the characteristics in the data, combined into pseudo product specifications, along with time, explain the variation in log of price by the R-squared statistics from the TPD hedonic models. This ranges from just 0.13 for camcorders, and up to 0.93 and 0.92 for desktop and mobile computers, respectively.

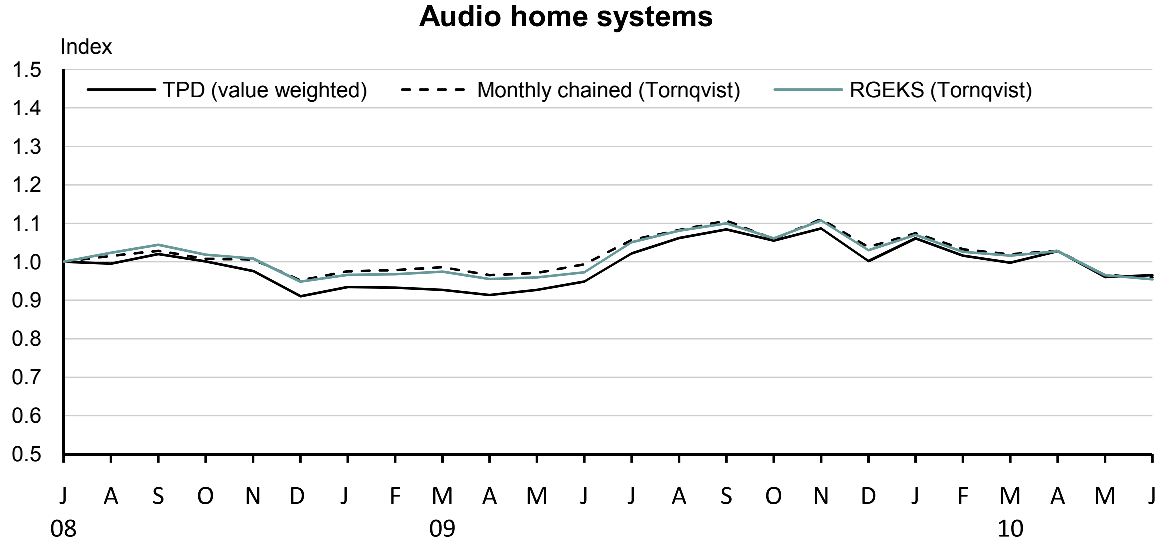
Table 7
R-squared statistics by product

Category	R-squared
Audio home systems	0.87
Camcorders	0.13
Cathode ray TVs	0.18
Desktop computers	0.93
Digital cameras	0.84
DVD players and recorders	0.71
Multi-functional devices	0.36
Mobile computers	0.92
Monitors	0.88
Cordless phones	0.74
Portable media players	0.81
Printers	0.65
Plasma / flat-screen TVs	0.48
Scanners	0.69

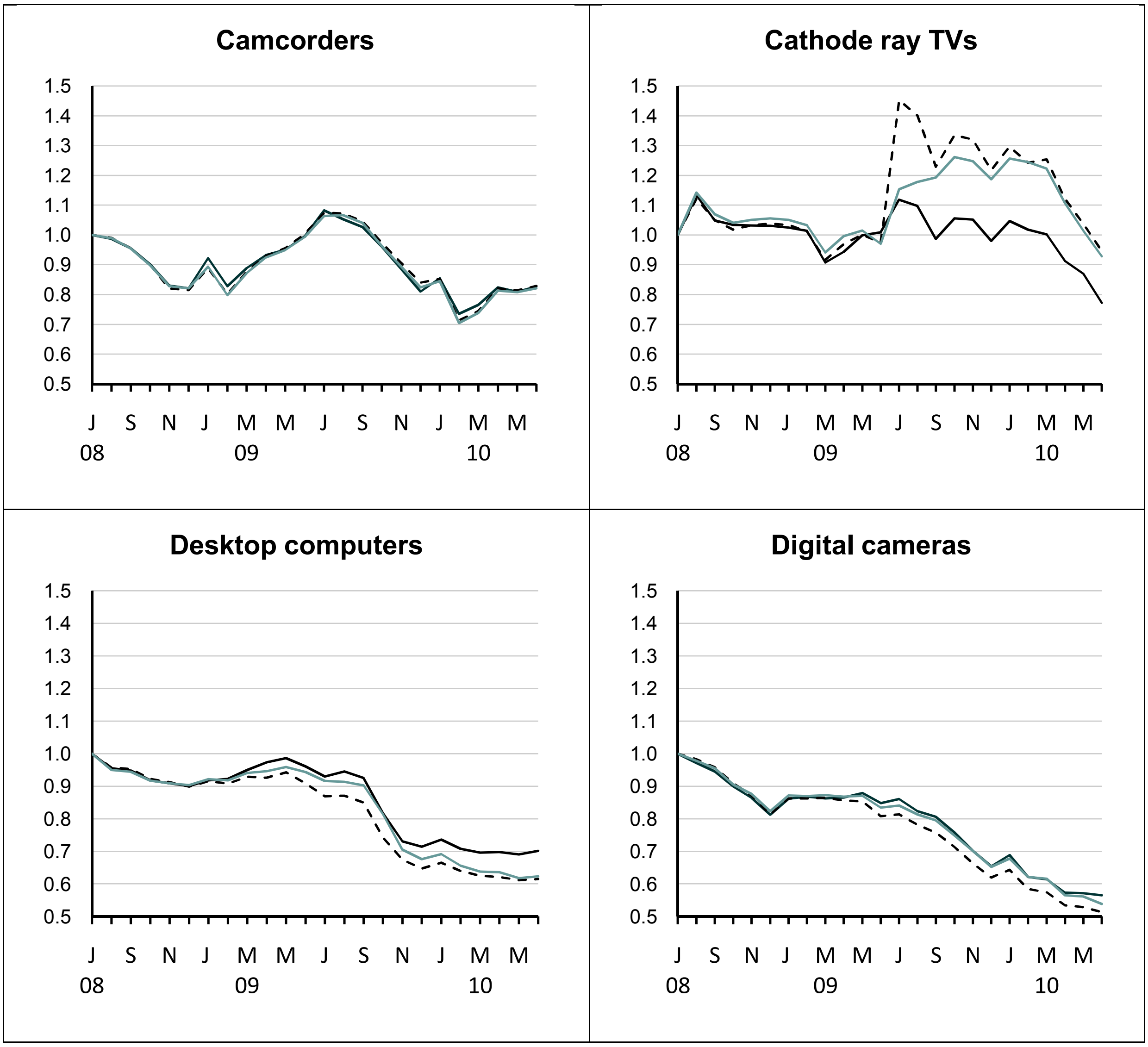
As for the Australian supermarket scanner data, we applied the RGEKS based on a Tornqvist index, with a 13-month rolling window; the TPD hedonic method and a monthly chained Tornqvist. As the series is only two years long, and given that the monthly chain drift is less obvious than for the supermarket data, we didn't run annually chained Tornqvist indexes.

Figure 6 shows the resulting indexes for the 14 consumer electronic products.

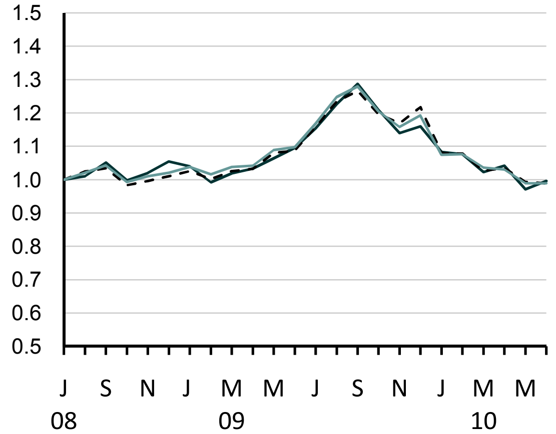
Figure 6
Indexes from GFK scanner data



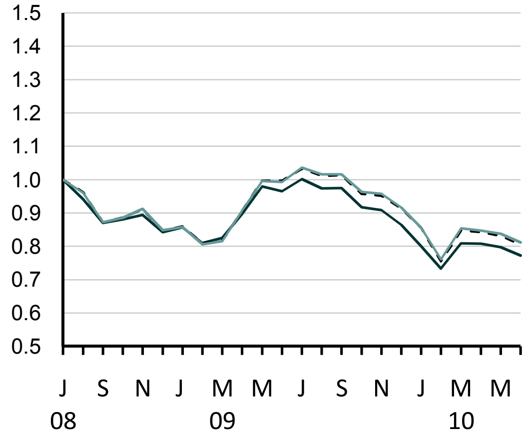
Source: GfK Retail and Technology



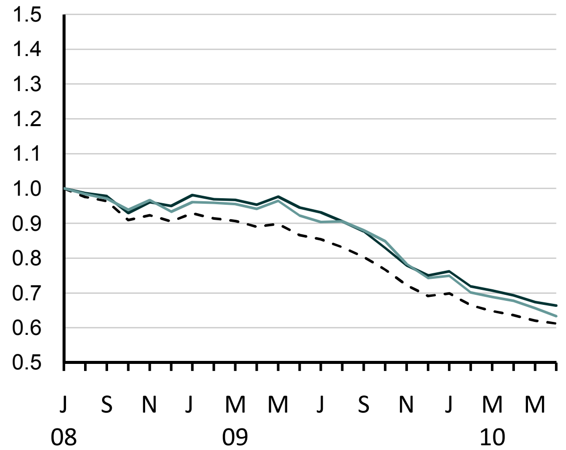
DVD players and recorders



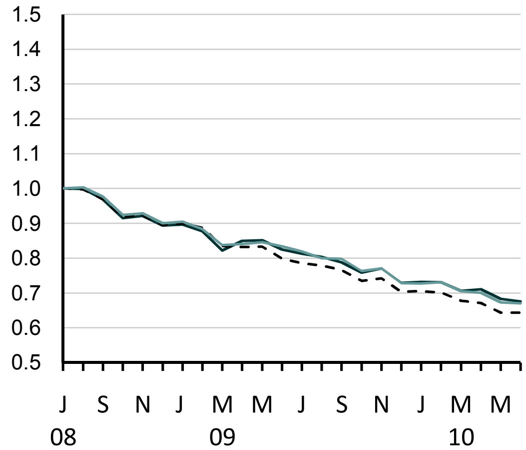
Multi-functional devices



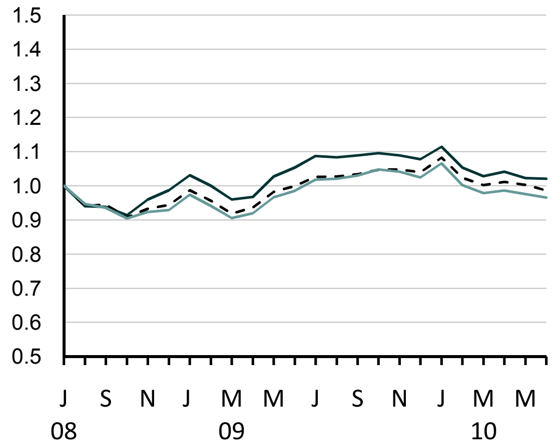
Mobile computers



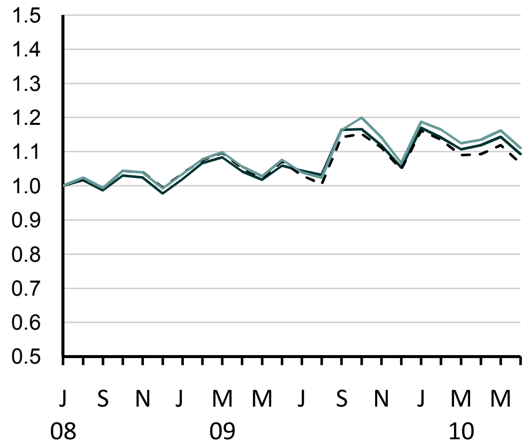
Monitors

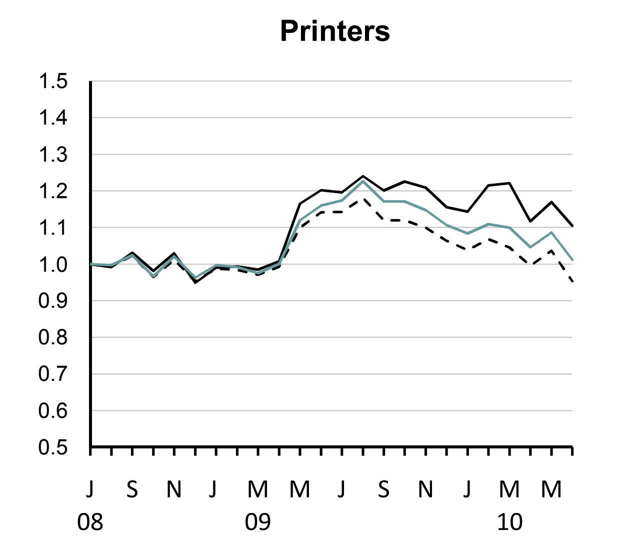


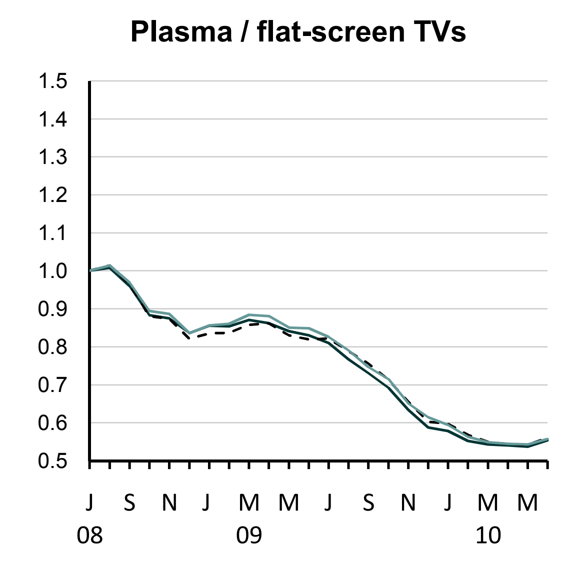
Cordless phones

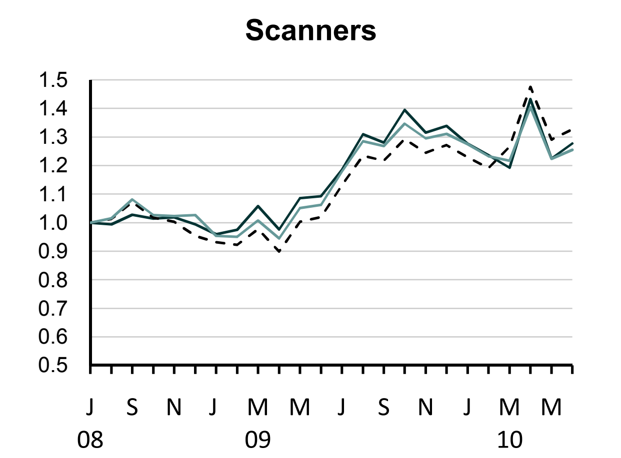


Portable media players







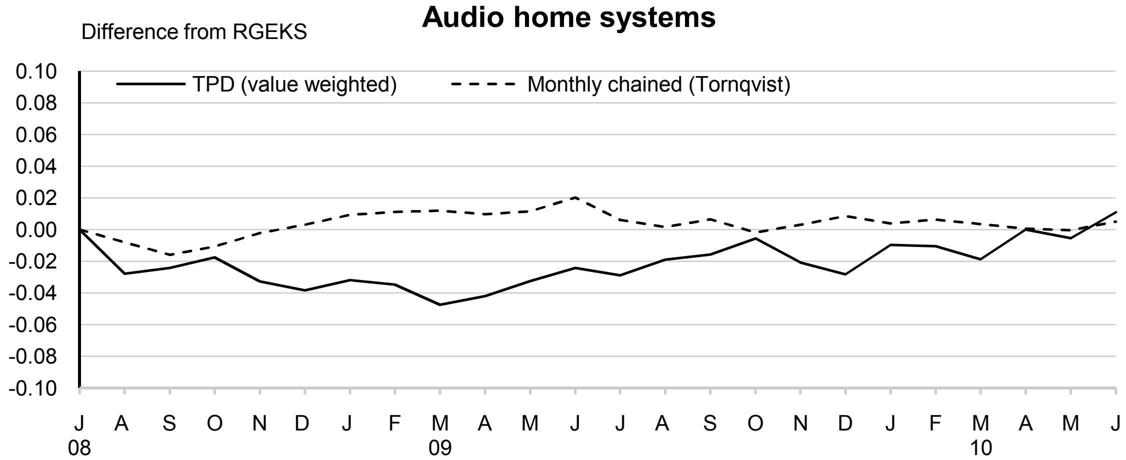


The RGEKS tends to sit between the monthly chained Tornqvist and the TPD indexes, although not always.

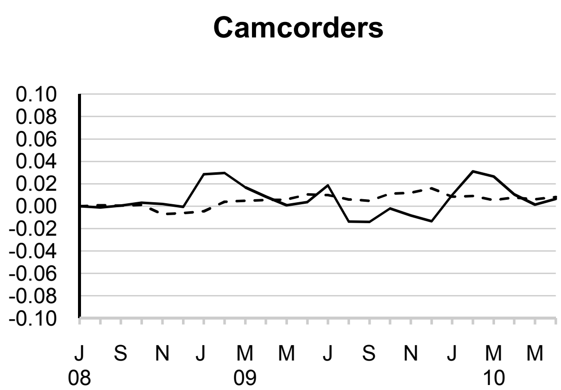
Cathode ray TVs is an interesting example. This product is right at the end of its life cycle, with very few different product specifications being sold and with prices plummeting. As might be expected, under these extreme conditions, the different index methods can give quite divergent results.

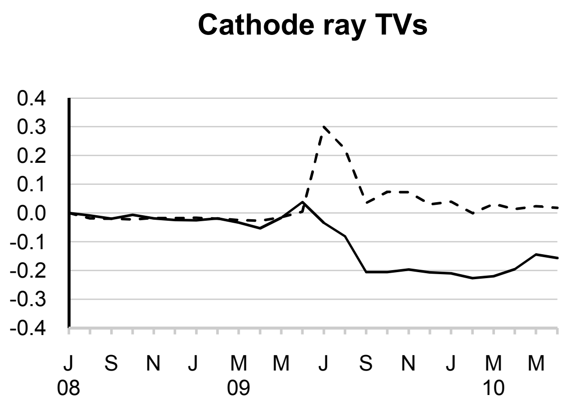
As we did for the Australian supermarket data, we graph in figure 7 the difference from the RGEKS of the other methods – in this case just the TPD and monthly chained Tornqvist. This enables us to see the differences between the index methods more clearly.

Figure 7
Difference from RGEKS of other indexes – GFK scanner data

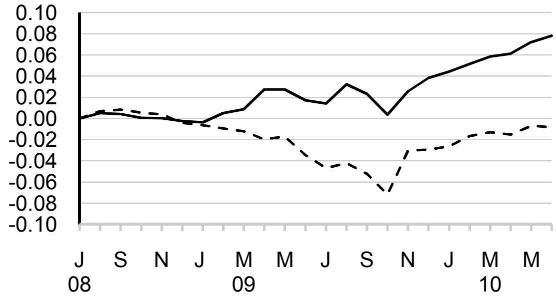


Source: GfK Retail and Technology

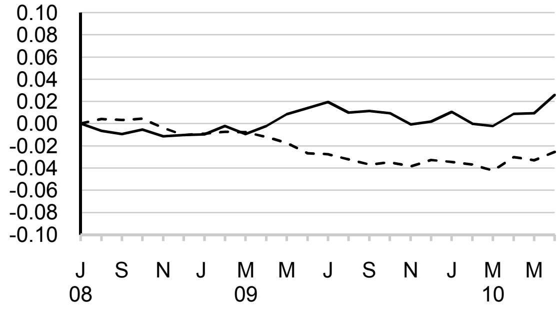




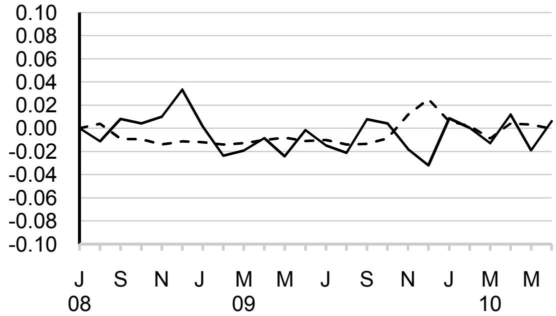
Desktop computers



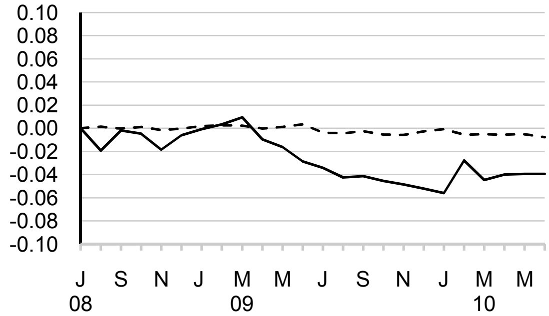
Digital cameras



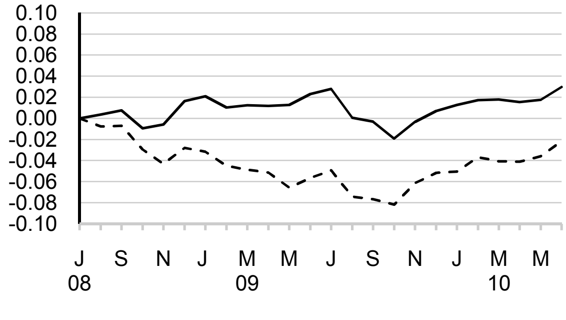
DVD players and recorders



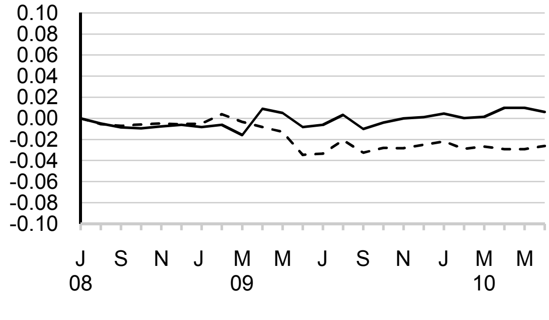
Multi-functional devices

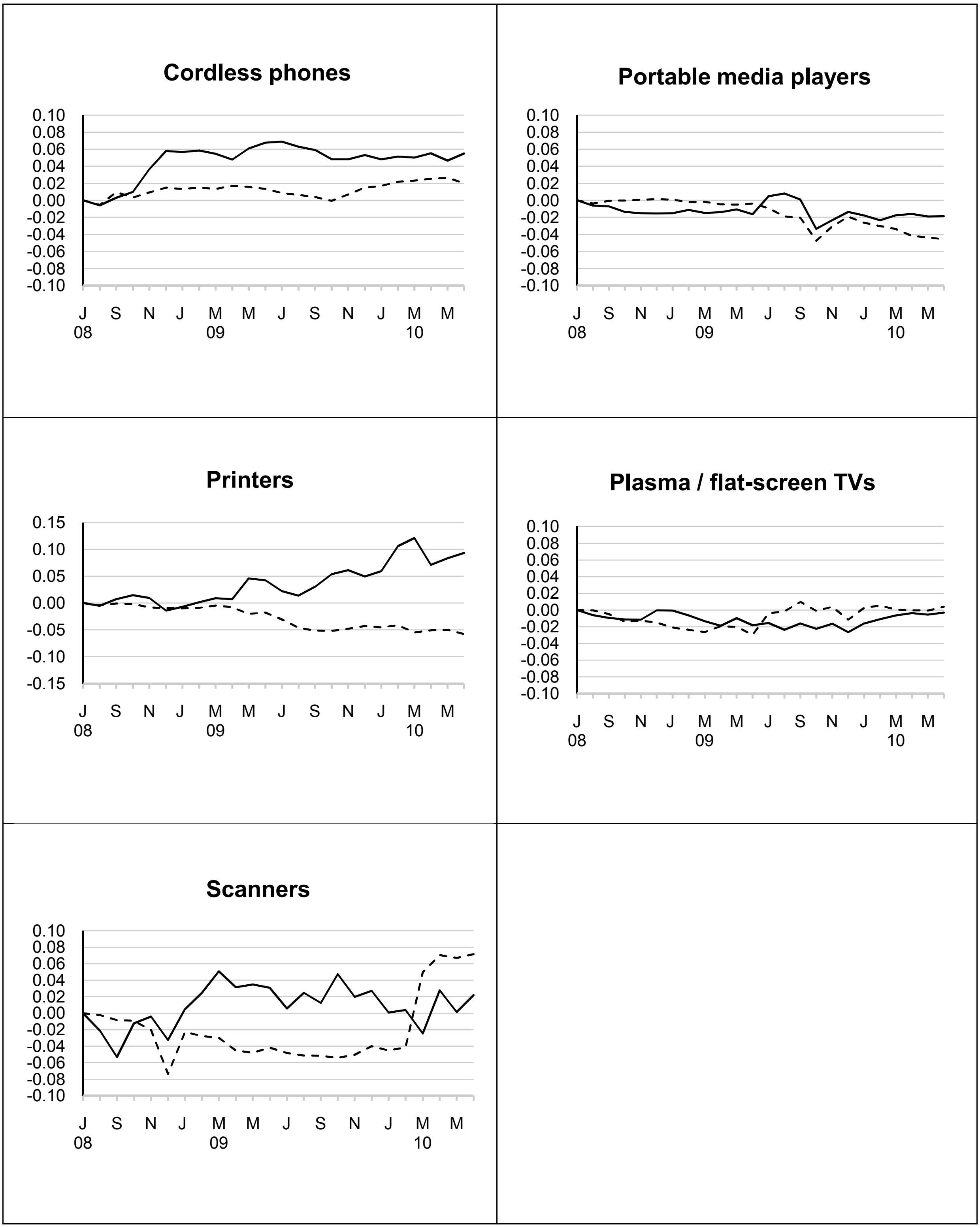


Mobile computers



Monitors





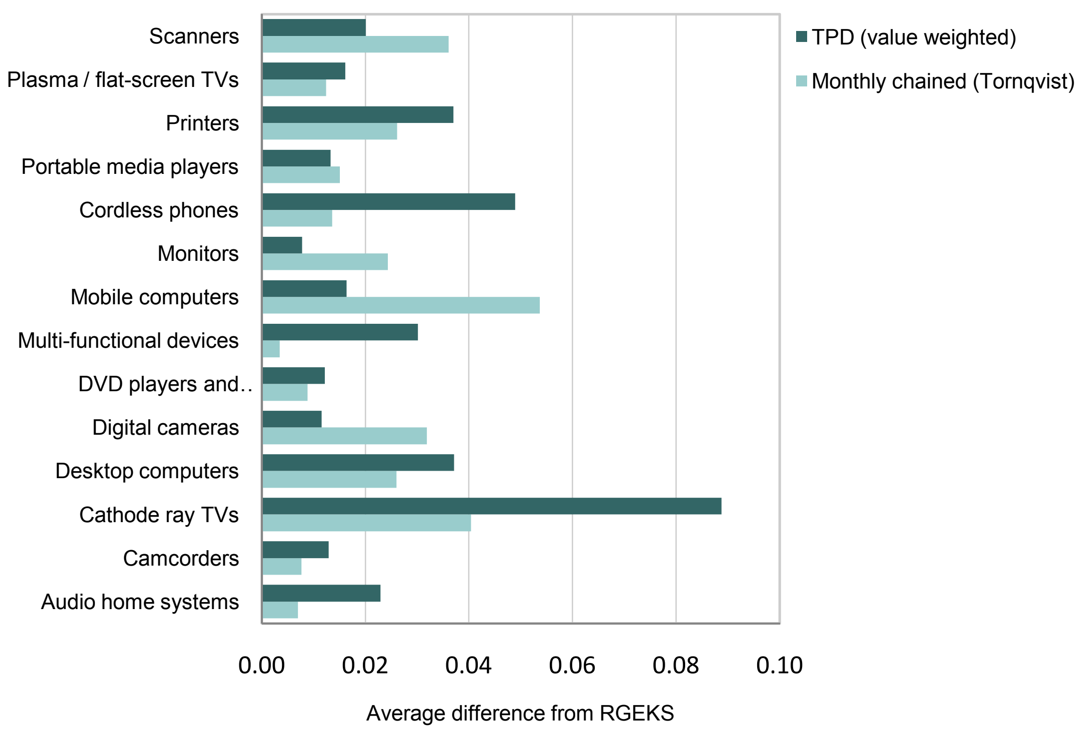
Note: Cathode ray TVs and printers have different scales on the y-axis.

Across these 14 products it is difficult to draw any general conclusions about how the TPD and chained monthly indexes move relative to the RGEKS. For five of the products (desktop computers, digital cameras, mobile computers, cordless phones, and printers) we see TPD showing steeper movements than the monthly chained Tornqvist – which was largely the case for the supermarket data, but this pattern is reversed for three of the products (audio home systems, cathode ray TVs, and multi-functional devices) and mixed for the rest.

Perhaps the strongest conclusion these comparisons support is that, for these products, the RGEKS, TPD hedonic, and monthly chained Tornqvist give comparable results – the differences are generally well within 10 percent over 2 years – except in exceptional circumstances such as cathode ray TVs.

As for the Australian supermarket data we summarise the difference from RGEKS by the average absolute difference standardised by the value of the RGEKS index. This is shown in figure 8. For eight of the 14 products – that is just over half – the TPD hedonic index is on average, more different from the RGEKS than the monthly chained Tornqvist is.

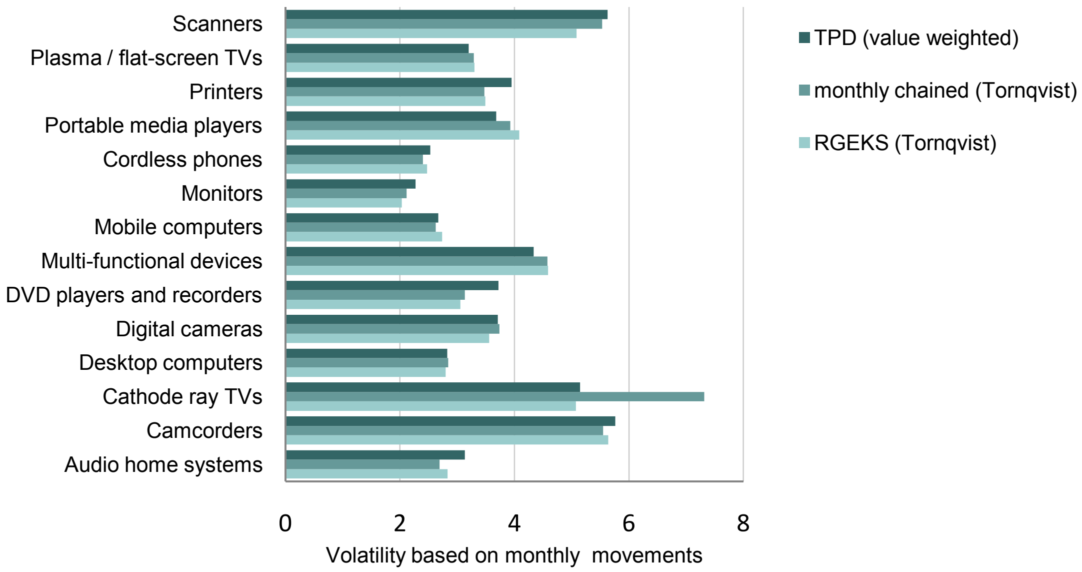
Figure 8
Average difference from RGEKS



Source: GfK Retail and Technology

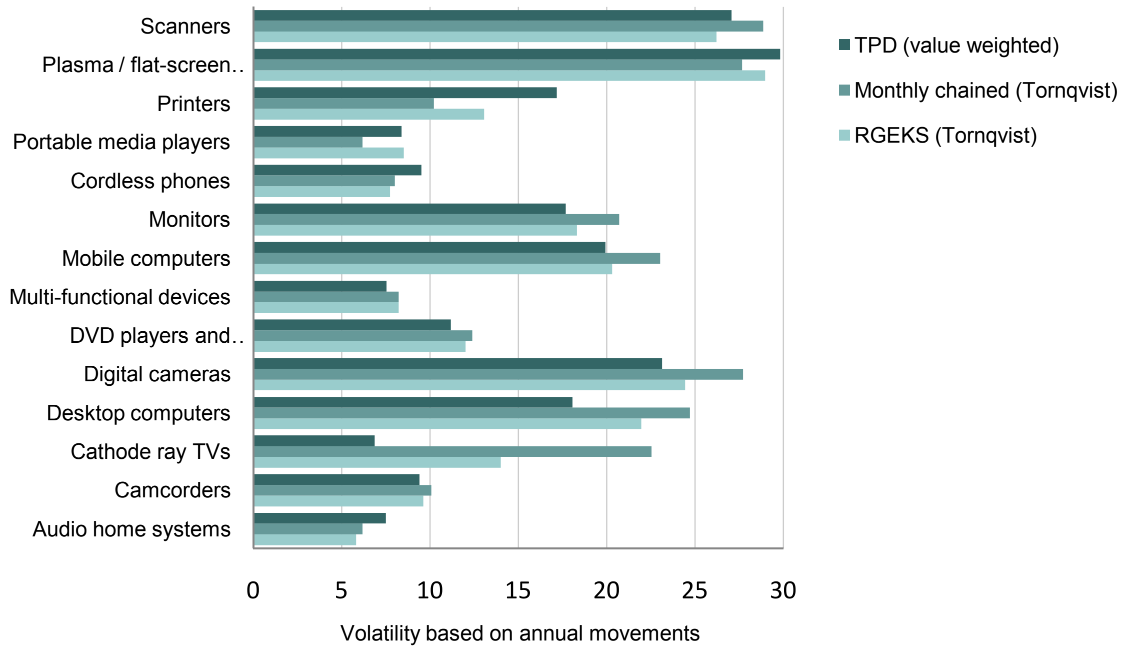
Figures 9 and 10 show the volatility of all three of the indexes, in terms of both monthly and annual percentage change. Again, none of the methods is consistently more or less volatile than the rest.

Figure 9
Volatility – average absolute monthly percentage change



Source: GfK Retail and Technology

Figure 10
Volatility – average absolute annual percentage change



Source: GfK Retail and Technology

US scanner data

We recently purchased the IRI Marketing research dataset, which gives weekly totals and quantities for seven years for 30 supermarket product categories. We plan to investigate the properties of RGEKS, TPD, and annually chained Tornqvist over this longer time series and larger set of products.

Areas for further research

The window length of the RGEKS

Of interest is the length of the RGEKS window. We have defaulted to using a 13-month window for a monthly index (ie a year plus one period) as suggested by Ivancic et al (2009), but we believe that this should be further investigated before adopting scanner data for price measurement in New Zealand. The optimal window length may be different for different products, and how to define or measure the optimal window length is a question for further research. It seems to be sensible to use at least a year, to allow for changes in seasonal availability. We note that Greenlees and McClelland (2010) show that when the product life cycle is less than the window length of the RGEKS, then the desirable properties of the RGEKS might break down. A key determining feature of the window length appears to be the notion of 'characteristicity'. We want the window to be long enough that the number of price comparisons is optimised, while not so long that the underlying price structure is non-homogeneous. This can be seen as analogous to considerations behind smoothing window lengths in time-series analysis.

Whether chain drift is influenced more by sales behaviour or behaviour at the start and/or end of the product's life cycle

Some early work we did indicated that the chain drift using a monthly chained Tornqvist on the Australian supermarket data is significantly less when the data is subset to those items that exist at both the start and the end of the series – that is, removing those that are introduced or discontinued within the period of observation. This seems to suggest that the chain drift may be driven more by the uncharacteristic price and quantity behaviour at the start and/or end of the product life cycle than by the price and quantity spiking due to sales throughout the product life-cycle. This could be tested more explicitly by identifying products that enter or leave within the period and removing just the first or last few months of data – that is, adjacent to the start and/or end of their life cycle. We haven't investigated this further as it isn't crucial to our aims at the moment. We are unlikely to want to use a monthly chained Tornqvist given the existence of viable alternative methods.

Quality adjustment bias

Quality adjustment bias will be an issue for any of these indexes. That is, if a change in specification (ie leading to a new barcode) coincides with a price change, then this price change will not be shown. This is something that manual quality adjustment is better able to cope with than these automated methods. More research is required around how to measure the significance of this bias.

Using expenditure shares as weights rather than expenditures

To date, we have used total expenditures in the TPD hedonic models to represent the relative importance of the different products. Other research has used expenditure shares. We plan to run the TPD indexes using shares at the weekly and monthly level and compare these to our existing results to see whether there are significant differences, but this has not yet happened at time of writing. Given that we are already controlling for month in the models, the use of expenditures rather than weekly expenditure shares will be giving more weight to weeks within the month that have a higher expenditure within the month, but it seems unlikely that this will make a significant difference to the resulting index.

Calculating the TPD hedonic index using a rolling window

So far, we have estimated the TPD hedonic indexes using the data pooled across all months rather than using a rolling window as for the RGEKS. This is likely not to make much difference for the examples looked at to date, which have windows of observation of three years (Australian supermarket data) and two years (New Zealand consumer electronic data). However, we would like to rerun the results using 13-month windows for greater comparability with the RGEKS. For the US scanner data, which has a length of seven years, this will be important to allow the effect of price determining characteristics to change over time.

Conclusions

For the data we've looked at to date – three years of Australian supermarket data, and two years of consumer electronics data – TPD and chained Tornqvist methods may provide viable alternatives to the RGEKS as they generally give results that are very close to the RGEKS. It might be desirable to use something relatively simple and transparent, such as an annually chained Tornqvist, as we can more readily explain it to users, and it may also be more straightforward to put into production than the RGEKS.

The R-squared statistics from theTPD indexes show us that, for the Australian supermarket data, a very high proportion of the variability in the log of price (over 90 percent for most of the products looked at) is being explained by the product specifications and time. The characteristics available in the New Zealand consumer electronics data explain far less of the price variability for some products, but we have potential for more extensive characteristics if we were to consider using this data for price movement estimation in the future.

For the consumer electronics data, we show that one approach to constructing indexes with categorical characteristics data is to combine all the characteristics to create pseudo product codes and then use these in the same way as the barcodes in the supermarket data, for RGEKS, TPD hedonic indexes, and more traditional index methods. Obviously this approach will only be as good as the completeness of the price determining characteristics captured in the data.

By examining the relative behaviour of the different methods over longer time periods and with a broader range of products, as we plan to do with the US scanner data, we may start to see some more definite patterns emerge, which will help to guide our conclusions about which method we want to adopt if we are to incorporate scanner data into the production of the New Zealand CPI.

References

De Haan, J & van der Grient, H (2010). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics, 161*(1), March 2011.

Ivancic, L, Diewert, WE, & Fox, KJ (2009). *Scanner data, time aggregation and the construction of price indexes*. (Discussion Paper 09-09). Department of Economics, The University of British Columbia, Vancouver Canada, V6T 1Z1.

Greenlees, JG & McClelland, R (2010, August). *Superlative and regression-based consumer price indexes for apparel using US scanner data*. Paper presented at the conference of the International Association for Research in Income and Wealth, St. Gallen, Switzerland.