Implementation of consumer electronics scanner data in the New Zealand CPI

Paper presented at the New Zealand Association of Economists conference,
Wellington, New Zealand
3 July 2015

Frances Krsinich

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

Citation
Krsinich F (2015, July). *Implementation of consumer electronics scanner data in the New Zealand CPI.* Paper presented at the New Zealand Association of Economists conference, Wellington, New Zealand.

Abstract

From the September 2014 quarter, Statistics New Zealand began using scanner data to measure the price change of 12 consumer electronics products, including televisions, computers, and cellphone handsets. The monthly scanner data, from market research company GfK, contains comprehensive information on product characteristics. This information on characteristics allows us to use the Imputation Törnqvist rolling year GEKS (ITRYGEKS) method to derive non-revisable quality-adjusted price indexes. These indexes are free of chain drift, use all the information in the data, and reflect the implicit price movements of new and disappearing products (de Haan and Krsinich, 2014a).

We will discuss aspects of the implementation of consumer electronics scanner data in the New Zealand CPI, which is a quarterly measure, including:

- how we dealt with having only two months of the quarter available in time for production
- the derivation of a quarterly index from monthly scanner data
- the monitoring and analysis process
- communicating the new approach to users.

1. Introduction

Statistics New Zealand began using scanner data for consumer electronics products in the New Zealand Consumers Price Index from the September 2014 quarter.

Monthly-aggregated scanner data is purchased from market research company GfK for twelve consumer electronics product categories including televisions, computers, and cellphone handsets.

Price indexes are derived from the scanner data using the Imputation Törnqvist rolling year GEKS (ITRYGEKS) method of de Haan and Krsinich (2014a). The ITRYGEKS index is an extension of the RYGEKS of Ivancic *et al* (2011) which is based on time-dummy hedonic indexes to ensure that the implicit price movements of new products entering the market, and old products disappearing from the market, are reflected.

The development of production processes has been an iterative one which is still evolving. This enables unanticipated issues to be dealt with flexibly and efficiently.

The paper is structured as follows:

Section 2 gives the background to the implementation of consumer electronics scanner data in the New Zealand CPI.

Section 3 discusses the implementation, with a description at the categories of products introduced, how we dealt with partial data, and using monthly aggregated data for a quarterly index and processing time.

Section 4 discusses the decision to use the ITRYGEKS index.

Section 5 presents the iterative approach to the development of the processing system.

Section 6 outlines how the implementation of scanner data in the New Zealand CPI has been communicated to users.

Section 7 concludes.

2. Background

Statistics New Zealand has been using scanner data from market research company GfK to inform expenditure weights and sample selection since 2006. Over the last five years active research on price measurement from scanner data has included collaboration with Statistics Netherlands on the Imputation Törnqvist rolling year GEKS (ITRYGEKS) method (de Haan and Krsinich, 2014a).

The New Zealand CPI Advisory Committee of 2013 recommended that Statistics NZ pursue the use of scanner data in the NZ CPI in their Recommendation 11:

> *Recommendation 11 of the CPI Advisory Committee of 2013:*
>
> *i. That Statistics NZ endeavours to use retail transaction (scanner) data to measure price change in the CPI.*
>
> *ii. That the methods used align with international best practice.*
>
> *iii. That arrangements be in place to ensure the continuity of retail transaction data supply.*
>
> *iv. That there is a suitable contingency plan in the event that the supply of retail transaction data is unavoidably interrupted.*
>
> *v. That Statistics NZ periodically audits prices in retail outlets to maintain confidence that the retail transaction data meets quality requirements.*

Statistics New Zealand decided to incorporate consumer electronics scanner data from GfK into the New Zealand CPI from the September 2014 quarter using the ITRYGEKS method, for twelve product categories.

3. Implementation in September 2014 quarter

3.1 Product categories incorporated into the NZ CPI

From the September quarter 2014, Statistics New Zealand incorporated retail transaction data, or 'scanner data', into the CPI, to measure price change for consumer electronics products.

The 12 product categories are:

- heat pumps
- desktop computers
- laptop computers
- tablet computers
- multi-function devices
- cellphone handsets
- digital cameras
- digital camera memory cards
- television sets
- set-top boxes for television sets
- DVD, Blu-ray players and player/recorders
- home theatre and stereo systems.

In terms of expenditure weighting:

- heat pumps contribute a fifth of the ‘major household appliances’ class, which contributed 0.71 percent to the all groups CPI expenditure weight for the June 2014 quarter
- cellphone handsets contribute over 90 percent of the ‘telecommunication equipment’ class, which contributed 0.29 percent to the all groups CPI

- the remaining 10 categories contribute four-fifths of the ‘audio-visual and computing equipment’ subgroup, which contributed 1.16 percent to the all groups CPI.

The scanner data is being supplied by market research company GfK. GfK adds information about product features, or characteristics, to the data collected from retailers.

Total monthly sales values and quantities sold are available, across retailers, for each product, along with extensive information about the characteristics of each product. There are hundreds of thousands of transactions represented in the scanner data each month.

The number of characteristics in the data range from 10 for digital camera memory cards to 77 for digital cameras, as shown in table 1 for August 2014. The number of distinct products in the data at August 2014 ranges from 24 for set-top boxes for televisions, to 445 for laptop computers.

Table 1. Number of characteristics and distinct products at April 2014

Product category	Characteristics	Products
Heat pumps	27	72
Desktop computers	56	107
Laptop computers	71	445
Tablet computers	73	148
Multi-function devices	53	102
Cellphone handsets	59	392
Digital cameras	77	228
Digital camera memory cards	10	254
Television sets	62	325
Set-top boxes for television sets	49	24
DVD, Blu-ray players, and player/recorders	50	129
Home theatre and stereo systems	62	224

3.2 Incomplete data available in time for production

The New Zealand CPI is a quarterly index. One of the limitations of the GfK scanner data is that only the first two months of the quarter are available in time for production.

There are four options for how to deal with this limitation in production. We assessed each of these options against the benchmark of the index if we had all three months of the quarter incorporated.

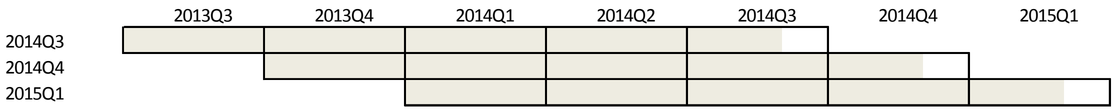
1. Estimate the ITRYGEKS index on the first 2 months of the quarter, with complete backdata feeding into the estimation (ie the 3rd month will be incorporated into the 13-month estimation window for the following quarter)
2. Estimate the ITRYGEKS index using 3 months of data, lagged by one month
3. Estimate the ITRYGEKS using only the middle month of each quarter
4. Estimate the ITRYGEKS using only the first two months of each quarter

Option 1 performed better than the other options on two summary measures:

1. The average of the absolute relative differences between each option’s movement and that of the benchmark.
2. The average of the absolute difference between each option’s movement and that of the benchmark.

Therefore this is the approach taken – we base the index on the first two months of data of the latest quarter along with complete backdata for the previous four quarter. This is shown below in figure 1.

Figure 1. Data incorporated into each quarter's index at the time of production



The figure shows that, for example, to calculate the quarterly price movement for the third quarter of 2014 (ie the top row), we use full quarterly data from each of the quarters from quarter three of 2013 through to quarter two of 2014, and the first two months of data from the third quarter of 2014.

3.3 Quarterly indexes from monthly data

The consumer electronics scanner data is aggregated to a monthly level.

The New Zealand CPI is a quarterly index, so we can either derive the quarterly index from the monthly index, or preaggregate the data to a quarterly level before deriving a quarterly index.

While it is useful to produce the monthly index as part of the monitoring and analysis processes, it is conceptually more appropriate for the quarterly index to be derived from quarterly prices and quantities, as it ensures the prices for products sold in each month of the quarter are appropriately weighted for price deflation, to produce quarterly volume indexes in the National Accounts.

In practice, the two approaches give very similar results.

4. Methodology

4.1 The need for new methods

Consumer electronics is a rapidly changing product class, so it is particularly important that we use methods that will appropriately adjust for the change in quality of what is being purchased.

Also, because consumer electronics products can have short life-cycles, it is important that new products are introduced into the index in such a way that the implicit price movement associated with their introduction is appropriate. That is, if a new product has an introductory price that is low, relative to its set of features, then this is a price decrease that the price index should reflect.

In addition to this requirement for appropriate quality adjustment, the CPI is non-revisable. This places another constraint on methods for creating price indexes from scanner data using statistical models, where the models are generally based on a past window of data – usually a year – ending with the most recent quarter.

The two key reasons why traditional index number methods do not work in the case of scanner data are:

- The high level of ‘churn’ or products appearing and disappearing in the market.

- Volatile prices and quantities due to discounting, which leads to a bias called 'chain drift' when chained superlative indexes such as a chained Törnqvist, are used to continually update the basket in the presence of high product turnover.

4.2 The rolling year GEKS

Ivancic, Diewert, and Fox (2011) proposed a method for producing price indexes from scanner data that uses all the prices and quantities in the data, and is free of chain drift[1]. Called the rolling year GEKS (RYGEKS) index, it is based on the Gini, Eltetö and Köves, and Szulc (GEKS) index used for multilateral spatial price indexes such as purchasing price parities – which compare prices in different countries at a point in time.

Within some estimation window[2] the RYGEKS index between periods *t1* and *t2* is the geometric mean of all the superlative bilateral indexes (such as the Törnqvist index or, as used in Ivancic et al (2011), the Fisher index) between:

- *t1* and all the other periods in the window, and
- *t2* and all the other periods in the window.

The monthly RYGEKS based on a 13-month rolling estimation window is as follows:

For the first window, ie *T*=0 to 12, the RYGEKS index is equal to the GEKS index:

$$P_{RYGEKS}^{0T} = P_{GEKS}^{0T} = \prod_{t=0}^{12} \left(P^{0t} \times P^{tT}\right)^{1/13} \tag{1}$$

Where P^{ij} is any superlative index (eg a Törnqvist index) between periods *i* and *j*.

From *t=13* onwards, RYGEKS links on the most-recent movement from the GEKS calculated on the next window (ie from *t* =1 to 13, then from *t* = 2 to 14, and so on) as follows:

$$P_{RYGEKS}^{0,13} = P_{GEKS}^{0,12} \prod_{t=1}^{13} \left(P^{12,t} \times P^{t,13}\right)^{1/13} = \prod_{t=0}^{12} \left(P^{ot} \times P^{t,12}\right)^{1/13} \prod_{t=1}^{13} \left(P^{12,t} \times P^{t,13}\right)^{1/13}$$

$$P_{RYGEKS}^{0,14} = P_{RYGEKS}^{0,13} \prod_{t=2}^{14} \left(P^{13,t} \times P^{t,14}\right)^{1/13} \tag{2}$$

and so on.

However, a limitation of the RYGEKS method is that it does not reflect the implicit price movements of new and disappearing products entering and leaving the market.

For example, if the initial price of the latest model of a mobile phone is high relative to its set of features then this implicit price increase is not reflected in the RYGEKS index.

[1] Though it non-transitive due to the rolling window aspect of the method so has the potential to suffer from what we could call 'splice-drift' though this is likely to be relatively insignficant.
[2] The estimation window is commonly one year plus a period – ie 13 months or 5 quarters.

4.3 The imputation Törnqvist rolling year GEKS

An extension of the RYGEKS index, called the imputation Törnqvist rolling year GEKS (ITRYGEKS) was proposed in de Haan and Krsinich (2014a). Like the GEKS, the ITRYGEKS is able to utilise all the information in the data, while remaining free of chain-drift. In addition, it reflects the implicit price movements of new and disappearing products.

Unlike the RYGEKS described above, which is based on superlative indexes such as the Törnqvist or Fisher - ie the P^{ij} in equations (1) and (2) – the ITRYGEKS index is based on bilateral time-dummy hedonic indexes. The 'Törnqvist' in its name refers to the fact that it is algebraically equivalent to a Törnqvist index based on real and predicted prices, as shown below in equation (4).

A bilateral time-dummy hedonic index between any periods *0* and *t* is derived from a statistical model based on the data from periods *0* and *t*.[3]

The estimating equation for the bilateral time dummy hedonic model is:

$$\ln p_i^t = \alpha + \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik} + \varepsilon_i^t \qquad (3)$$

Where:

$\ln p_i^t$ is the log of the average monthly price for product *i* in period *t*

α is the intercept term

D_i^t has the value 1 if the observation relates to period *t* ($t \neq 0$) and the value 0 if the observation relates to period 0

z_{it} is the quantity of the kth characteristic, or feature, for product *i*.[4] By definition, this will be the same in both periods [5]

ε_i^t is an error term with an expected value of 0

Since model (3) controls for changes in the product characteristics, $\exp\delta^t$ is a measure of quality-adjusted price change between periods 0 and t.

And, when the average[6] expenditure shares are used as weights for the matched products and half of the expenditures shares for the unmatched products (in the periods they are available) the ITRYGEKS index from period *0* to *t* can be expressed as follows:

$$P_{ITRYGEKS}^{0t} = \prod_{i \in U^{0t}} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}} \prod_{i \in U_D^{0t}} \left(\frac{\hat{p}_i^t}{p_i^0} \right)^{\frac{s_i^0}{2}} \prod_{i \in U_N^{0t}} \left(\frac{p_i^t}{\hat{p}_i^0} \right)^{\frac{s_i^t}{2}} \qquad (4)$$

Where:

[3] This can be generalised to any two periods *i* and *j*

[4] For categorical characteristics, such as those on the consumer electronics scanner data, each characteristic k will be represented by a set of dummy variables corresponding to all possible values of characteristic k – that is, variables which are set to 1 (or 0) in the presence (or absence) of that value of the characteristic.

[5] A product corresponds to a distinct set of characteristics, or features. So, any change in characteristic would result in a different product.

[6] Across both time periods.

U^{0t} is the set of matched products *i* with respect to periods *0* and *t*

U_D^{0t} is the set of 'disappearing' products *i* with respect to periods *0* and *t* – that is, products that exist in period *0* but not in period *t*

U_N^{0t} is the set of 'new' products *i* with respect to periods *0* and *t*

s_i^t is the expenditure share of product *i* in period *t*

Formula (4) demonstrates that the ITRYGEKS index is equivalent to a Törnqvist index for the matched products and, for new and disappearing products, the ITRYGEKS index applies a Törnqvist formula to prices predicted from bilateral time-dummy hedonic models for the period in which there is no price available. The derivation of (4) is given in de Haan and Krsinich (2014).

The decision to use the ITRYGEKS index for the consumer electronics scanner data was made on the basis that, at the time, this was a well documented and understood method.

Since that decision was made, it has become clear that either of the multilateral time-dummy (TD) hedonic index or the fixed-effects window-splice (FEWS) index would also be feasible methods to use for this consumer electronics scanner data in production.

4.4 The potential to use either TD or FEWS indexes

Recent research by de Haan and Krsinich (2014a) has shown that the multilatral time dummy (TD) hedonic index is a geometric form of a quality-adjusted unit value (QAUV) index which, empirically, is shown to be virtually the same as the harmonic QAUV.

Krsinich (2014) has shown that the FEWS index can produce fully quality-adjusted nonrevisable price indexes without explicitly incorporating any characteristics, by leveraging off the longitudinal information in the data.

However, there is currently no incentive to use the FEWS index over methods such as the ITRYGEKS or TD which require explicit information on the full set of characteristics. This is because there is confidentialising in the data where model names are overwritten with a ‘tradebrand’ identifier when more than a certain percentage of that model are sold through just one retailer. Therefore the full set of characteristics in the GfK data is required to derive the product identifiers which would be required for the FEWS index.

5. Developing the processing system

5.1 An iterative approach to the processing system

The approach to developing the processing system for price measurement from consumer electronics scanner data has been an iterative one. Processes for analysis and monitoring have continued to evolve over the time that scanner data has been in production as some unanticipated issues have arisen. For example, changes in the coding of characteristics can cause spurious movements in the index if not identified, as they are interpreted as a complete replacement of products with new products at the time of the change. Identifying changes in coding practices is non-trivial with such a large number of characteristics across all product categories.

With such a new development as the use of scanner data for price measurement, the flexibility and adaptability of a user-written prototype system is a risk-averse and efficient approach to take. Once the system has all the necessary analysis, monitoring and estimation components bedded into a streamlined process, it will be appropriate to incorporate it as part of a more formally developed processing system for Prices ‘big data’ more generally. This system will also incorporate future use of supermarket scanner data; postal survey data such as that for used cars which use a hedonic estimation method; overseas trade data such as televisions and mobile phones; and web-scraped online data. It will be built with the flexibility to incorproate other future big-data sources.

5.2 Monitoring and analysis processes

Data is extracted from GfK's 'startrack' system into excel files, which are then imported to SAS. Before the production of quarterly price indexes using the ITRYGEKS method, there are a number of monitoring and analysis processes run, which include:

- Checks that sets of characteristics have remained the same since last quarter
- Time series of monthly quantities sold for each product category
- Time series of monthly average prices (unadjusted for quality change) for each product category
- Identification of outliers, in terms of both movements and levels for the latest quarter, and incorporating information from the full longitudinal record.
- Time series of distributions across key characteristics, for each product category, in terms of both quantities sold and expenditure shares
- Monthly ITRYGEKS quality-adjusted price indexes
- Time series of turnover statistics – ie counts of new products, disappearing products, and matched products between each pair of adjacent months

If one of the monitoring or analysis outputs suggests further investigation, other processes can then be invoked:

- The quarterly (or monthly) price indexes can be run for subsets of the data. For example, we could run the quarterly ITRYGEKS for different brands of television, to determine whether an unusual price movement is contributed to by all brands, or whether one particular brand is driving the result.
- We can run the price indexes excluding ouliers – either full longitudinal records of outlying products, or particular outlying values of products. So, if outlier detection process identifies a potential error, we can test the impact on the index of outlier removal.

6. Communicating the new approach to users

A number of different forums were used to communicate the move to scanner data for measuring price change for consumer electronics in the CPI.

Statistics New Zealand releases a quarterly electronic newsletter called 'Price Index News'. In the April 2014 newsletter (Statistics New Zealand, 2014a) we signalled our intention to introduce scanner data into the CPI

A 'sources and methods' document titled 'Measuring price change for consumer electronics using scanner data' was released on the Statistics NZ in November 2014 (Statistics New Zealand, 2014b) and this was later linked to by the January 2015 edition of the Price Index News (Statistics New Zealand, 2015).

A number of short articles were written to communicate the development to different groups: in the newsletter of the New Zealand Association of Economists (NZAE, 2014); the newsletter of the International Association of Survey Statisticians (IASS, 2015); and the 'Expert Data Users' newsletter of Statistics NZ (2014c).

By explaining the development in both 'plain english' and with links to technical detail, the information needs of a range of different audiences have been catered for.

7. Conclusion

Scanner data was incorporated into the New Zealand CPI for consumer electronics products, from the September 2014 using the ITRYGEKS method.

In addition to obtaining suitable data, and determining an appropriate index methodology, some practical issues arising from the timeliness of the data and appropriate levels of aggregation needed to be addressed.

An iterative approach to the development of a processing system has enabled the analysis and monitoring processes to evolve in response to unanticipated issues, such as the potential for coding of characteristics to change over time.

A number of different forums were used to communicate the new development to different user groups, at different levels of technical detail.

The production processes for consumer electronics scanner data in the New Zealand CPI are a prototype system which will inform the more formal development of a Prices ‘big data’ processing system that will encompass a range of data sources including supermarket scanner data, trade data, telecommunications bills data and online data.

References

de Haan, J, & Krsinich, F (2014a). *Scanner data and the treatment of quality change in non-revisable price indexes*. Journal of Business & Economic Statistics, 32:3, 341-358.

de Haan, J, & Krsinich, F (2014b, August). *Time-dummy hedonic and quality-adjusted unit value indexes: Do they really differ?* Paper presented at the conference of the Society for Economic Measurement, Chicago.

Ivancic, L, Diewert, WE, & Fox, KJ (2011*). Scanner data, time aggregation and the construction of price indexes.* Journal of Econometrics, 161(1), 24–35.

IASS (2015) Survey Statistician Newsletter Jan 2015. Available at http://isi-iass.org/home/services/the-survey-statistician/

Krsinich, F (2011, May). *Measuring the price movements of used cars and residential rents in the New Zealand consumers price index*. Paper presented at the 12th meeting of the Ottawa Group, Wellington, New Zealand. Available at http://www.stats.govt.nz/ottawa-group-2011/agenda.aspx

Krsinich, F (2013, May). *Using the rolling year time-product dummy method for quality adjustment in the case of unobserved characteristics*. Paper presented at the 13th meeting of the Ottawa Group, Copenhagen. Available at http://www.dst.dk/da/Sites/ottawa-group/agenda.aspx

Krsinich, F (2014, May). *The FEWS index: Fixed effects with a window splice.* Paper presented at the meeting of the group of experts on consumer price indices, Geneva, Switzerland. Available at http://www.unece.org/index.php?id=33341#/

NZAE (2014) Asymmetric Information (newsletter of the New Zealand Association of Economists) issue 41 Dec 2014 Available at http://www.nzae.org.nz/wp-content/uploads/2015/01/Asymmetric-Information-Issue-51-Nov-2014.pdf

Statistics New Zeland (2011). *Updating the CPI used cars index hedonic model*. Available at http://www.stats.govt.nz/browse_for_stats/economic_indicators/prices_indexes/used-cars-hedonic-model.aspx

Statistics New Zealand (2013). Consumers price index advisory committee 2013 discussion paper. Available at http://www.stats.govt.nz/browse_for_stats/economic_indicators/CPI_inflation/2013-cpi-review-advisory-committee.aspx.

Statistics New Zealand (2014a). *Price Index News: April 2014*. Available from http://www.stats.govt.nz/tools_and_services/newsletters/price-index-news/Apr%20-14.aspx

Statistics New Zealand (2014b). *Measuring price change for consumer electronics using scanner data*. Available from www.stats.govt.nz.

Statistics New Zealand (2014c) Expert Data Users newsletter November issue: Available at http://a.smartmailpro.com/webv/c4jvzbh2dh#electronic

Statistics New Zealand (2015) *Price Index News: January 2015*. Available at http://www.stats.govt.nz/tools_and_services/newsletters/price-index-news/jan-15.aspx