Measuring the Price Movements of Used Cars and Residential Rents in the New Zealand Consumers Price Index

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

Acknowledgements
The author would like to thank Alistair Gray for discussion and input to this work.

Citation
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 Ottawa Group, Wellington, New Zealand.

Abstract

Price movements of second hand cars and residential housing rentals in the New Zealand consumers price index (CPI) are both based on large samples of data that lend themselves well to hedonic methods.

In 2001 the existing stratification method for second hand cars was replaced by a hedonic method, and this is being updated with an improved hedonic model in 2011. The rental index, which currently uses a matched-sample approach, was recently assessed using a hedonic index as a benchmark.

The longitudinal nature of the rental survey lets us control for unobserved characteristics in the hedonic model, but questions were raised about how well this formulation deals with the newly rented dwellings. We build on a result from Aizcorbe, Corrado and Doms (2003) to assert that the implicit imputations for new rental dwellings are appropriate.

Aging is an issue for both of the measures. For used cars we are able to include age of car in the hedonic model. Age of rental dwelling, however, is unobserved. We are still determining how to formulate individual aging effects into the hedonic model.

Introduction

Statistics New Zealand has used hedonics for two major components of the New Zealand consumers price index: used cars and residential rentals. Used cars contributed 2.4 percent to the New Zealand CPI at the June 2008 quarter, and rentals contributed 7 percent. The price movement of used cars has been estimated using hedonics since 2001, and a hedonic index was recently used to benchmark the existing matched-sample index for housing rentals.

The used cars index has been estimated using a hedonic model for 10 years, and the model is currently being improved by adopting a log-linear formulation and including more characteristics. With this more efficient estimation there is scope to reduce respondent burden by decreasing the large sample originally designed for a stratified estimation of average prices.

Housing rentals are currently estimated from a matched sample derived from a probabilistic area-based sample, which is regularly updated with newly lodged bonds. A hedonic index was constructed as a benchmark when concerns were raised about potential understatement from the matching procedure.

The used cars survey

Each quarter, data on the sales of approximately 3,500 cars are collected from a sample of used car dealers. Price, year of manufacture, make and model, engine size (cc rating), and odometer reading are collected for each car sold.

This sample was designed initially to support the calculation of average prices within estimation cells based on combinations of make and model, cc rating ranges, and age of car. To ensure robust estimation of averages, cells with too few observations were excluded, which resulted in use of only around 25 percent of the data collected.

Introduction of hedonic estimation in 2001

Statistics New Zealand introduced a hedonic estimation for used cars in 2001 (Krsinich, 2000). Price is modelled against quarter, region, make and model, age of car, cc rating, and odometer reading, shown in formula (1).

(1) $$P_c = \sum_k \beta_k C_{kc} + \sum_t \delta_t D_{ct} + \varepsilon_{ct}$$

Where:

P_c is the price of car c (note that there is no t term as we can assume each individual car is only sold once)

D_{ct} = 1 in the quarter t that car c is sold and 0 otherwise

The following characteristics K are included in the model for each car c sold:

- town of purchase (15 categories)
- make and model (47 categories)
- age (in years)
- size of engine (cc rating, eg 2300)
- odometer reading (in kilometres).

The linear formulation was used initially as this made the hedonic index a more efficient version of the existing estimation cell method.

Hedonic indexes are calculated for each of five broad regions of New Zealand (Auckland, Wellington, Rest of North Island, Christchurch, and Rest of South Island) and these are weighted together with population-based expenditure shares that are updated at the three-yearly CPI reviews, to get the national used car index.

To make the method operational, a rolling window of the latest eight quarters is used to estimate the hedonic model each quarter, and the derived index for the most recent quarter is linked to the previous quarter's index number.

The new hedonic estimation

In 2011 Statistics New Zealand will introduce an improved hedonic formulation for used cars. At the same time, the production process will be streamlined to make it simpler and more robust.

The hedonic model has been updated as follows:

- Price is now logged, which gives the model a better fit to the data.
- A more detailed classification of make and model is incorporated – from 47 categories to 96.
- Squared terms have been added for age of car and cc rating.
- An identifier for the used car dealer is now included.

The updated model is:

(2) $$\ln P_c = \sum_k \beta_k C_{kc} + \sum_t \delta_t D_{ct} + \varepsilon_{ct}$$

With P_c and D_{ct} as for formula (1), characteristics K are now:

- town of purchase (15 categories)
- dealer (approximately 300)
- make and model (96 categories)
- age (in years) and age squared
- size of engine (cc rating, eg 2300) and cc rating squared

- odometer reading (in kilometres).

R-squared statistics of models with different dependent variables (that is, price and log of price) are not comparable, but a lower bound of the improvement in the used cars hedonic model is given by comparing the R-squareds for log-linear models with the old and new sets of explanatory variables. This is shown in table 1.

Table 1

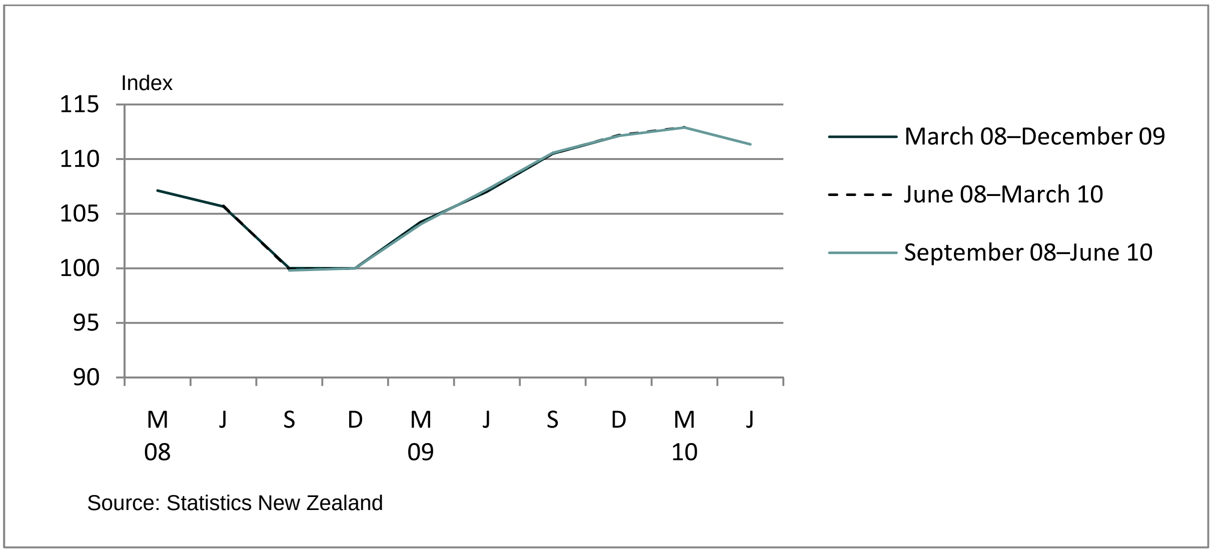
R-squared statistics for log-linear hedonic models using the old and new sets of explanatory variables

Region	Old set of explanatory variables	New set of explanatory variables
Auckland	0.76	0.82
Wellington	0.77	0.81
Other North Island	0.80	0.83
Christchurch	0.77	0.81
Other South Island	0.77	0.80

Figure 1 shows that the indexes based on successive eight-quarter windows are very stable across time, which reinforces confidence in linking on just the most recent index movement each quarter.

Figure 1

Successive eight-quarter hedonic indexes for used cars at the New Zealand level



Note: Base: Dec 2008 quarter (=100)

Future improvements to the used cars estimation

The introduction of a hedonic approach for the estimation of used cars, in particular now that the hedonic formulation has been optimised, means that there is potential for significantly reducing the sample size, and hence the respondent burden of the used cars survey. Work is now underway to determine the minimum sample size required for an accurate estimation.

Along with redesigning the sample's size and allocation, representation of the price movements of used cars sold through auction and online needs to be considered when designing the sample. These channels are not included in the current sample.

The residential rents survey

The residential rents survey uses an area-based probability sample of approximately 2,000 rental dwellings. Landlords of the selected dwellings are asked each quarter for rental information. The sample is updated every quarter with dwellings that enter the rental market within the selected areas. These are identified from bond data by reconciling dwellings with newly lodged bonds against the existing sample.

Average rents are estimated from the sample within strata based on bedroom number and five broad regions. For example, average rent is estimated for three-bedroom rental dwellings in the Auckland region. Movements in these averages are weighted together using regional population-based expenditure shares, which are updated at three-yearly CPI reviews to create a national index.

Table 2 shows, for 2007 and 2008, the sample each quarter, with the percentages of the dwellings that are new to the sample, and about to leave the sample – ie sample births and sample deaths. Sample births contribute an average of 6.6 percent each quarter, while sample deaths contribute on average 3.2 percent. This reflects the increasing size of the rental population in New Zealand, as a result of population growth and a shift away from home ownership.

Table 2

Rent survey sample size and contribution of sample births and deaths in 2007 and 2008

quarter	number of dwellings in sample (1)	percentage of sample births	percentage of sample deaths
07q1	1800	8.7	2.5
07q2	1950	8.5	2.9
07q3	2050	6.8	3.0
07q4	2150	7.7	3.4
08q1	2200	6.2	4.2
08q2	2200	4.0	4.4
08q3	2250	6.0	2.5
08q4	2300	5.3	2.7

Introduction of a matched sample in 2000

The rental survey was initially designed in 1998, and in 2000 it was modified by restricting the contributing sample to those dwellings existing in both the previous and current quarters – that is, a ‘matched sample’. This was adopted to ensure that differences in the composition of the sample due to newly rented dwellings and dwellings leaving the rental stock would not contaminate the estimation of price movements.

Concerns raised about the matched sample in 2008

Smith (2008), from the Reserve Bank of New Zealand, speculated that the use of a matched sample might be biasing the index downwards by removing pure price change associated with the dwellings new to the rental market (and, to a lesser extent, those leaving the rental market).

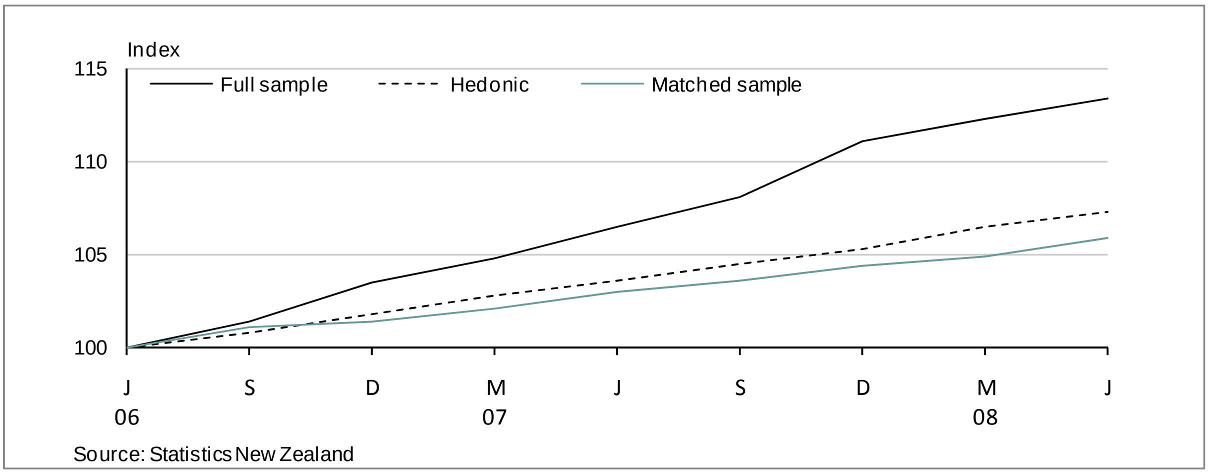
This concern was in part prompted by higher price movements observed for newly rented dwellings, a measure regularly produced by the Department of Building and Housing. There is no reason that price movements for the entire stock of rental dwellings, which is what the CPI measures, must match the price movements for newly rented dwellings. This concern was also prompted by the fact that the index produced from current quarter averages based on the full sample moves more rapidly than that from the matched sample. This is shown in figure 2. Statistics NZ decided to investigate the issue further.

Hedonic benchmarking exercise

As a benchmark, a hedonic index was estimated from the rental data. This gave a very similar result to the current matched-sample index, suggesting that any bias in the current approach is not significant (Krsinich, 2009). Figure 2 shows the comparison in the context of the index that would be produced from averages based on the full sample if no matching were used – that is, if no adjustment were made for the change in sample composition due to births to, and deaths from, the sample.

Figure 2

Matched sample index compared with hedonic benchmark index and full sample index



Note: Base: June 2006 quarter (=100)

In the standard hedonic formulation, just the limited set of characteristics available in the data – that is, fine-level region and number of bedrooms – can be incorporated. The hedonic formulation used to benchmark the rental index took advantage of the longitudinal nature of the rental data by fitting dwelling-specific intercepts. In other words, a fixed-effects model was used. This meant that all characteristics of the rental dwellings fixed across time were controlled for. However, any time-varying characteristics at the dwelling level – such as renovations or deterioration of the dwelling – are not controlled for by the fixed-effects model.

It is important to note that this hedonic index would suffer a similar issue to the matched-sample index if used in production, because at least two time points are required for each dwelling in the fixed effects estimation. This means that, as with the matched sample, new dwellings will not be included in the estimation until they have been in the sample for two quarters. Because only the most recent movement is spliced into the index each quarter, any price change associated with the introduction of newly rented dwellings will be missed. However, unlike the matched-sample index, the movement due to sample deaths would not be missed.

For the benchmarking exercise, however, the hedonic index was applied retrospectively to the entire set of data. This means that only the most recent quarter's movement suffers from the issue above. For the rest of the index the hedonic index can be used as a benchmark.

Appropriateness of the fixed effects hedonic formulation

Doubts were raised about the appropriateness of the fixed effects formulation. Most seriously, given the context of the investigation, is a zero movement being implicitly imputed for new rental dwellings? If so, the benchmarking hedonic index might also suffer from any downwards bias in a similar way to the matched-sample index.

The other question was whether, and how to, incorporate the price effect of the aging of rental dwellings.

Implicit imputation for new rentals

The probability sample of the rental survey is perhaps unfamiliar in the price measurement context, where it is common for products newly introduced to a sample to be directly replacing old products that can no longer be priced. If new rental dwellings in the sample were one-to-one replacements for rentals leaving the sample, then it does seem reasonable to hypothesise that a fixed-effects model – which models the difference from the mean rent across time of each dwelling – might be linking out any quality difference between the old and new dwellings.

But in the rental probability sample, the births and deaths are representative of the underlying dynamics of the population, by construction of the area-based sample selection and quarterly maintenance. As shown in table 2, the sample size is gradually increasing, in line with the increasing population of rental dwellings.

We build on a result from Aizcorbe, Corrado & Doms (2003) to assert that the implicit imputation in the fixed effects hedonic model for a newly rented dwelling is not zero, but rather the movement from the average quality-adjusted rents of continuing dwellings in the previous quarter to the quality-adjusted rent of the new dwellings in the current quarter. This means that the movement being implicitly contributed to the index by the new rental dwelling is due to price only, rather than quality. The appendix fills in some detail between equations (1) and (5) in Aizcorbe, Corrado & Doms (2003), and then extends equation (5) to the fixed effects case.

Bias due to aging of dwellings

By construction, the matched-sample index will be biased to the extent that there is a price effect associated with aging for individual rental dwellings. For example, if there were a 1 percent decrease in price associated with each year of aging of a rental dwelling, then the matched-sample index will be biased downwards by 1 percent annually, even if the overall age distribution of the population of rental dwellings, as reflected by the continually updated sample, remains constant over time.

If the hedonic benchmark index is not affected by aging bias, it can be used to establish whether the aging bias is significant enough to cause concern.

Unfortunately, there is no information available in the rental survey data on the age of the rental dwellings, so age cannot be incorporated explicitly into the model.

At the time of writing we have not yet formulated aging effects in the fixed effects hedonic model. Our hypothesis, which we welcome discussion about, is that the index will be biased by changes in the age distribution of the sample, but not by individual aging as such. That is, if the age distribution of the population, as reflected by the sample, is reasonably stable then the hedonic estimation will be unbiased even if there is a price effect of aging at the individual rental dwelling level.

Future plans for the measurement of rentals

If the fixed effects hedonic formulation is appropriate both in terms of the implicit imputation for newly rented dwellings and the pricing effects of aging, it is a valid benchmark for the current matched-sample index. This would give reassurance that, despite the known limitations of the matched-sample approach, the biases are not of practical significance.

A hedonic estimation may be considered in the future, as this would enable a more efficient estimation and therefore a reduction in sample size and respondent burden.

Appendix

The following is an extension of the Aizcorbe, Corrado and Doms (2003) formulation, showing that the implicit imputation for new rental dwellings in the fixed effects hedonic index is appropriate.

Given model m at time t with characteristics k :

(1) $$\log P_{mt} = \sum_k \beta_k C_{kmt} + \sum_{s=1}^{t} \delta_s D_{ms} + \varepsilon_{mt}$$

Note that in the second term D_{ms} is a time dummy so $D_{ms} = 1$ if m is observed at time $t = s$, and 0 otherwise, ie the 2nd term is just δ_t

So the average $\log P_{mt}$ is

(2) $$\frac{1}{M}\sum_{m=1}^{M_t} \log P_{mt} = \frac{1}{M}\sum_{m=1}^{M_t}\left(\sum_k \beta_k C_{kmt}\right) + \frac{1}{M_t}\sum_{m=1}^{M_t}\delta_t + \frac{1}{M_t}\sum_{m=1}^{M_t}\varepsilon_{mt}$$

Note that

(3) $$\frac{1}{M_t}\sum_{m=1}^{M_t}\delta_t = \delta_t$$

and taking the mean over the models of the residuals will be close to zero, as their expected value is zero, ie

(4) $$\frac{1}{M_t}\sum_{m=1}^{M_t}\varepsilon_{mt} \approx 0$$

So (2), (3), and (4) $\Rightarrow$

(5) $$\frac{1}{M_t}\left(\sum_{m=1}^{M}\left(\log P_{mt} - \sum_k \beta_k C_{kmt}\right)\right) = \delta_t$$

Hence

(6) $$\delta_t - \delta_{t-1} = \frac{1}{M_t}\sum_{m=1}^{M_t}\left(\log P_{mt} - \sum_k \beta_k C_{kmt}\right) - \frac{1}{M_{t-1}}\sum_{m=1}^{M_{t-1}}\left(\log P_{mt-1} - \sum_k \beta_k C_{kmt-1}\right)$$

If the models are the same at both time points then $M_t = M_{t-1} = M$ and $m_t = m_{t-1}$

So

(7) $$\delta_t - \delta_{t-1} = \frac{1}{M}\left(\sum_{m=1}^{M}\left(\log P_{mt} - \log P_{mt-1}\right) - \sum_{m=1}^{M}\left(\sum_k \beta_k C_{kmt} - \sum_k \beta_k C_{tmt-1}\right)\right)$$

Introduce a new model and call it N

Note that in general

(8) $$\frac{1}{n+1}\sum_{i=1}^{n+1} X_i = \frac{1}{n+1}\left(\sum_{i=1}^{n} X_i + X_{n+1}\right) = \frac{n}{n+1}\cdot\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n+1} X_{n+1}$$

and from (5), in the period that N is introduced

(9) $$\delta_t = \frac{1}{M+1}\sum_{m=1}^{M,N}\left(\log P_{mt} - \sum_k \beta_k C_{kmt}\right)$$

(9) and (8) $\Rightarrow$

(10) $$\delta_t = \frac{M}{M+1}\cdot\frac{1}{M}\sum_{m=1}^{M}\left(\log P_{mt} - \sum_k \beta_k C_{kmt}\right) + \frac{1}{M+1}\left(\log P_{Nt} - \sum_k \beta_k C_{kNt}\right)$$

and from (5) and because N is introduced at time t (and therefore not there in time $t-1$) we have

(11) $$\delta_{t-1} = \frac{1}{M}\sum_{m=1}^{M}\left(\log P_{mt-1} - \sum_k \beta_k C_{kmt-1}\right)$$

(10) and (11) $\Rightarrow$

(12) $$\delta_t - \delta_{t-1} = \frac{M}{M+1}\cdot\frac{1}{M}\sum_{m=1}^{M}\left(\log P_{mt} - \sum_k \beta_k C_{kmt}\right) - \frac{1}{M}\left(\sum_{m=1}^{M}\log P_{mt-1} - \sum_k \beta_k C_{kmt-1}\right)$$
$$+\frac{1}{M+1}\left(\log P_{Nt} - \sum_k \beta_k C_{kNt}\right)$$

But $1 = \left(\frac{M}{M+1} + \frac{1}{M+1}\right)$ so

(13) $$\delta_t - \delta_{t-1} = \frac{M}{M+1}\left[\frac{1}{M}\sum_{m=1}^{M}\left(\log P_{mt} - \sum_k \beta_k C_{kmt}\right) - \frac{1}{M}\sum_{m=1}^{M}\left(\log P_{mt-1} - \sum_k \beta_k C_{kmt-1}\right)\right] +$$
$$\frac{1}{M+1}\left[\left(\log P_{Nt} - \sum_k \beta_k C_{kNt}\right) - \frac{1}{M}\sum_{m=1}^{M}\left(\log P_{mt-1} - \sum_k \beta_k C_{kmt}\right)\right]$$

This means that the change in price between $t-1$ and t is a weighted average of the price change of the continuing items and the new item, with the implicit imputation for the change in price of the new item being the difference between the average quality-adjusted price for the continuing items in the previous ($t-1$) period, and the quality adjusted price of the new item in the current (t) period.

If models' characteristics are constant over time in our dataset, so that $C_{kmt} = C_{kmt-1} = C_{km}$ for all t then we can rewrite (1) as

(14) $$\log P_{mt} = \sum_k \beta_k C_{km} + \delta_t + \varepsilon_{mt}$$

Then if we use fixed effects to control for quality differences across models, that is, if we assume $\sum_k \beta_k C_{km} = \beta_m$ then we can rewrite (14) as

(15) $$\log P_{mt} = \beta_m + \delta_t + \varepsilon_{mt}$$

and substituting β_m and β_N for $\sum_k \beta_k C_{kms}$ and $\sum_k \beta_k C_{kNs}$ $s \in (t-1, t)$ in the derivations above, from (2) to (13) we get

(16) $$\delta_t - \delta_{t-1} = \frac{M}{M+1}\left[\frac{1}{M}\sum_{m=1}^{M}(\log P_{mt} - \beta_m) - \frac{1}{M}\sum_{m=1}^{M}\log P_{mt-1} - \beta_m\right] + \frac{1}{M+1}\left[(\log P_{Nt} - \beta_N) - \frac{1}{M}\sum_{m=1}^{M}(\log P_{mt-1} - \beta_m)\right]$$

$$\Rightarrow \delta_t - \delta_{t-1} = \frac{M}{M+1}\left[\frac{1}{M}\sum_{m=1}^{M}(\log P_{mt} - \log P_{mt-1})\right] + \frac{1}{M+1}\left[(\log P_{Nt} - \beta_N) - \frac{1}{M}\sum_{m=1}^{M}(\log P_{mt-1} - \beta_m)\right]$$

which has the same interpretation as equation 13. That is, the implicitly imputed price change for a newly rented dwelling is the change between the average of the quality-adjusted prices of the continuing rentals in the previous period, and the quality-adjusted price of the new item in the current period. Which is what we want, in the case of the rental estimation.

References

Aizcorbe, A, Corrado, C, & Doms, M (2003, June). When do matched-model and hedonic techniques yield similar price measures? (Federal Reserve Bank of San Francisco working paper no. 2003-14). Retrieved 24 March 2011 from www.frbsf.org.

Krsinich, F (2000). *The hedonic regression parameter method for calculating the used cars index*. Unpublished document, Statistics New Zealand.

Krsinich, F (2009, July). Using hedonic regression to assess the housing rentals component of the New Zealand consumers price index. Paper presented at the New Zealand Association of Economists conference, Wellington. Retrieved 24 March 2011 from stats.govt.nz.

Smith, M (2008). *Measuring rents in the CPI*. Unpublished document, Reserve Bank of New Zealand.

Statistics NZ (2007). How used-car price changes are measured in the CPI. Retrieved 24 March 2011 from stats.govt.nz.