

## Use of Hedonics at Statistics New Zealand

Paper presented at the Economic Measurement Group Workshop 2009  
Sydney, Australia,  
9 December 2009

Frances Krsinich

Senior Research Statistician, Prices Unit, Statistics New Zealand

P O Box 2922  
Wellington, New Zealand  
[info@stats.govt.nz](mailto:info@stats.govt.nz)  
[www.stats.govt.nz](http://www.stats.govt.nz)

**Liability statement:** Statistics New Zealand gives no warranty that the information or data supplied in this paper is error free. All care and diligence has been used, however, in processing, analysing and extracting information. Statistics New Zealand will not be liable for any loss or damage suffered by customers consequent upon the use directly, or indirectly, of the information in this paper.

**Reproduction of material:** Any table or other material published in this paper may be reproduced and published without further licence, provided that it does not purport to be published under government authority and that acknowledgement is made of this source.

## 1. Introduction

We have used hedonics at Statistics New Zealand in the context of price measurement several times over the period 1999–2009. These have been fairly non-traditional uses mainly because we used sample survey data rather than catalogue data.

For used cars, for example, we used hedonics to estimate the index in the production of the consumers price index (CPI). For rentals and house prices, we used hedonics to create benchmark indexes against which non-hedonic estimations were compared.

Our experiences have raised some points for consideration:

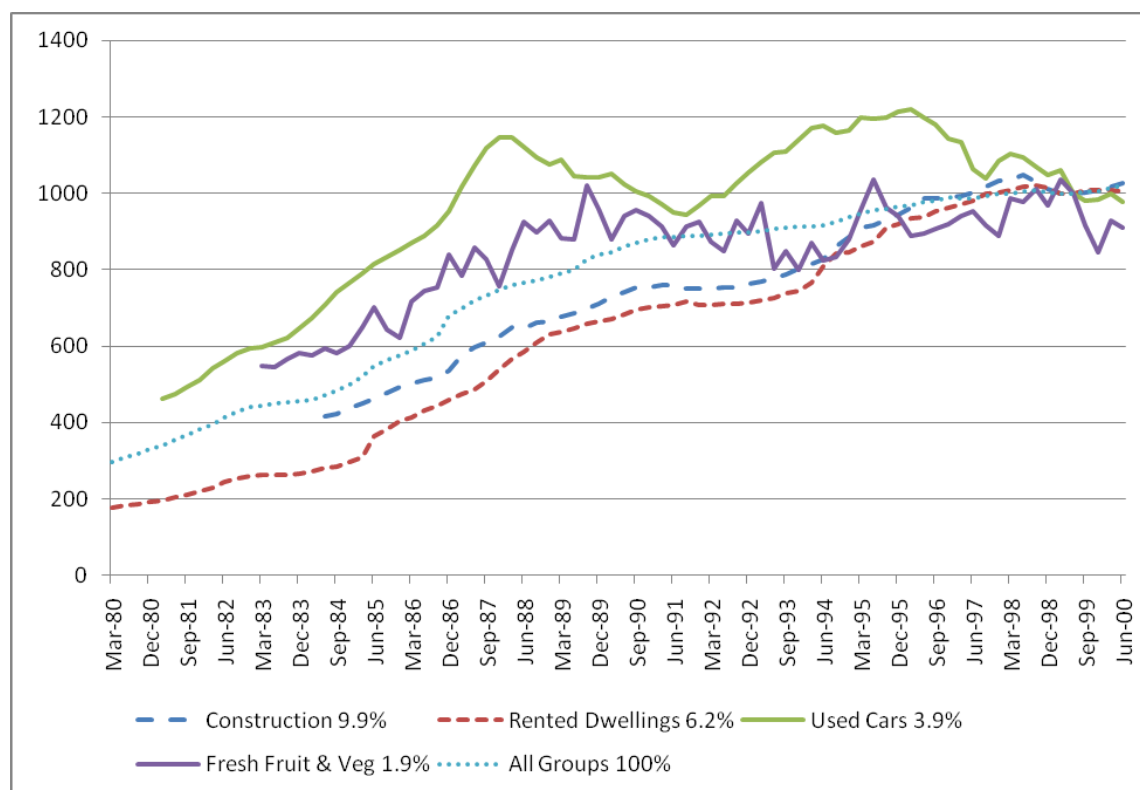
- The possibilities of using survey data rather than catalogue data. We were fortunate to have existing survey data, but consideration could be given to developing and collecting survey data for use with hedonic estimations where appropriate.
- Where individual-level characteristics are not available, but small aggregates are, the use of a 'semi-hedonic' approach using dummy variables for the small aggregates (such as suburbs, or model numbers) is likely to be a more effective and efficient approach than a stratification approach based on groupings of the same small aggregate variables.

## 2. Estimation of used cars in 2001

In 2001 we introduced a hedonic estimation for used cars (Statistics NZ, 2007) which at the time contributed 3.9 percent of the total weight of the CPI and in 2009 contributed 2.4 percent. Figure 1 shows the used cars index, along with the other two most highly weighted items of the CPI in 2000 (when the method was being developed). Fresh fruit and vegetables, a highly volatile component, is also included for comparison. The graph shows that used cars is a relatively volatile item in addition to being highly weighted.

**Figure 1**

**Three highest weighted items, fresh fruit and vegetables and all groups, 1980–2000**



Each quarter, about 3,500 prices are collected from a sample of 300 used car dealers. We have data on price, year of manufacture, make and model (eg Ford Laser), cc rating, odometer reading (km), and area of purchase.

The previous estimation method calculated average prices within estimation cells based on aggregations of make and model, cc rating ranges, and age of car. Cells with too few units were excluded to ensure robust estimation of cell averages. This exclusion resulted in only about 25 percent of the collected sample being used in the estimation of the index.

By introducing a hedonic estimation we were able to use a fuller set of data, and introduce continuous cc-rating, odometer reading, and area of purchase into the set of characteristics controlled for. We use a moving window of the most recent eight quarters of data to estimate the index and splice on the most recent movement to calculate the current index number.

Figure 2

Distribution of age against quarter in the used cars sample, March 1998–September 2000

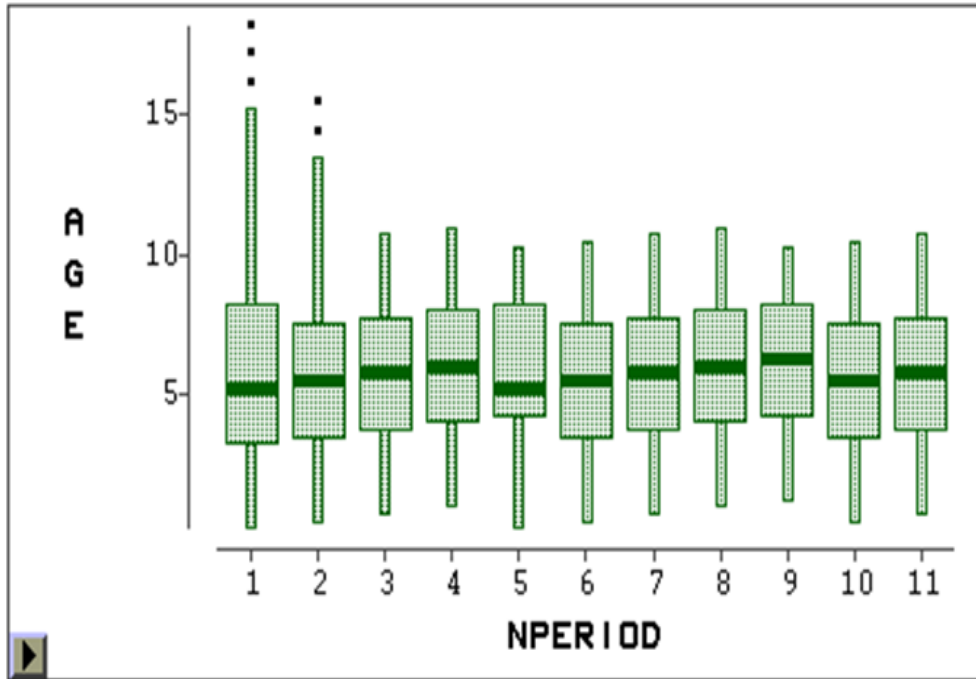
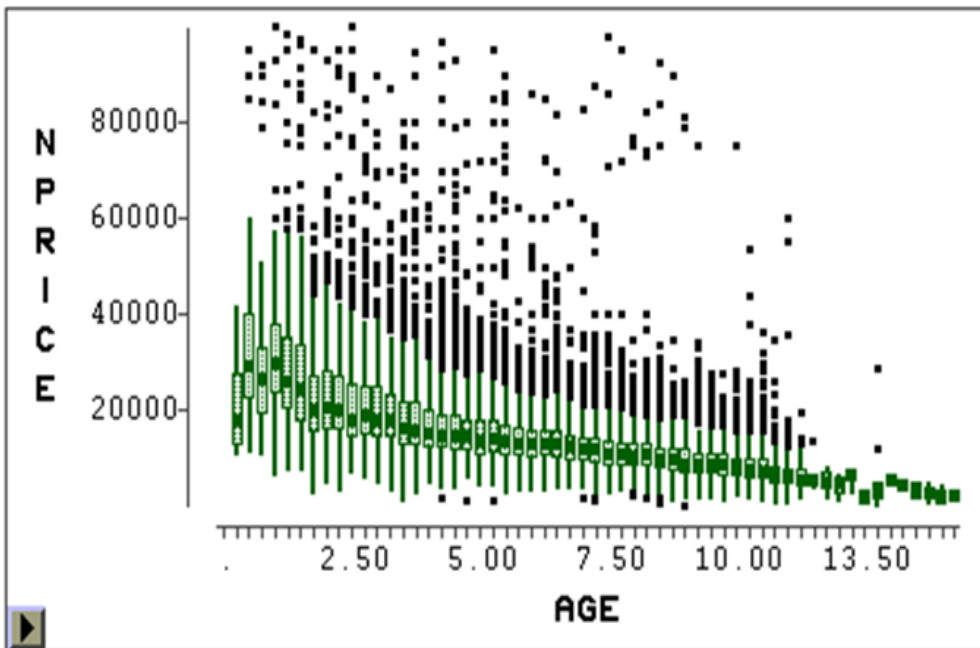


Figure 3

Distribution of price against age in the used cars sample



Figures 2 and 3 demonstrate that in terms of the age of the used cars, the sample distribution changes over time; this change is associated with price. Similar graphs can be produced for all the observed characteristics in the used cars data. Clearly, some kind of adjustment is required to ensure that the estimate of price change is not affected by this changing sample composition.

Note that all figures relate to the period from the March 1998 quarter to the September 2000 quarter.

**Figure 4**

**Used cars index estimated using the estimation cell method, compared with the index calculated from unadjusted sample averages**

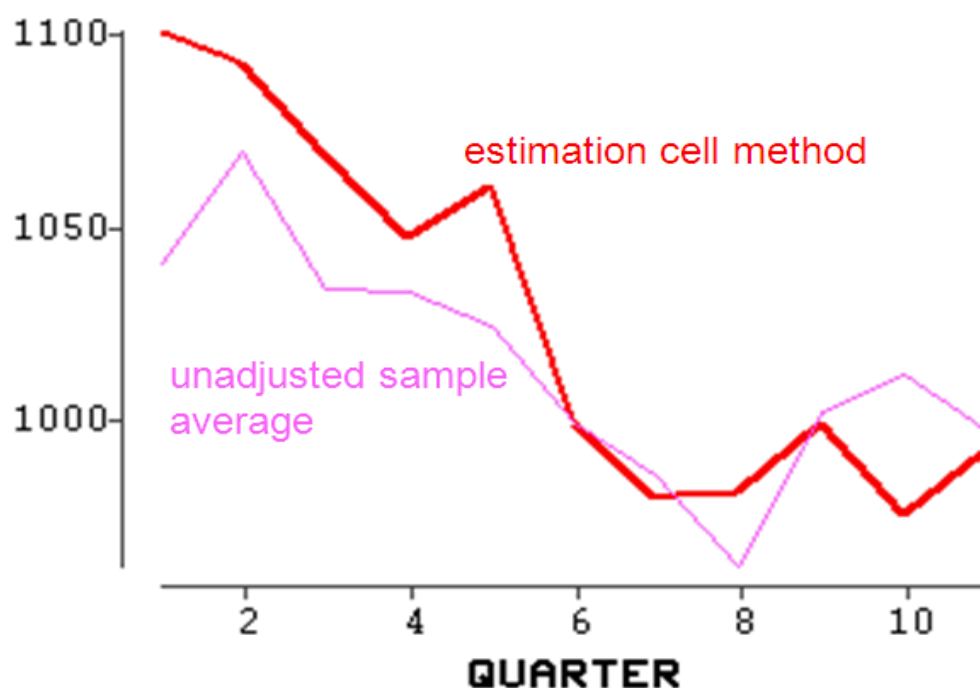


Figure 4 shows that the use of the estimation cell method results in a different index from that calculated directly from straight sample averages. This difference results from controlling for compositional change at the estimation cell level, but is also due to subsetting the sample to the more common and newer cars.

**Figure 5**

**Used cars index estimated using the estimation cell method, compared with the index calculated from the hedonic time dummy model**

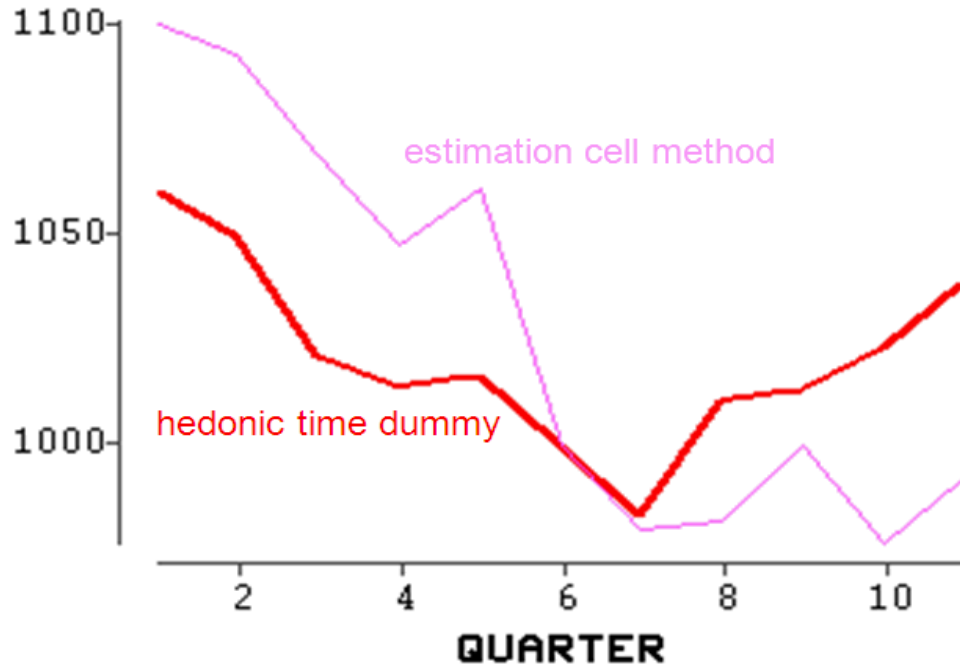


Figure 5 shows the different index obtained using hedonic estimation. This method allows the full representative sample of data to be used, and controls for more of the price determining characteristics.

Figure 6

Used cars index estimated using the hedonic time dummy method on the full sample, compared with the index calculated from the 25% sample

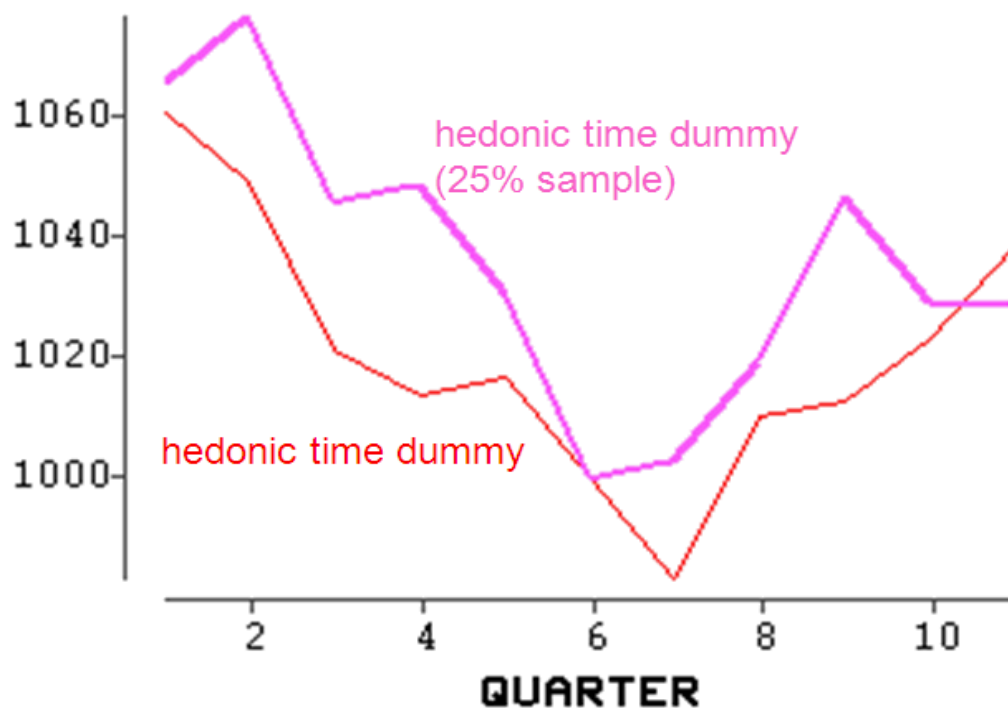


Figure 6 shows the effect of the subsetting of the data that was required for the estimation cell method. It demonstrates that part of the difference between the index calculated using the estimation cell method, and the index calculated using the hedonic time dummy method, as shown in figure 5, is due to the subsetting of the data to more common and newer cars. As the full sample is representative of the population of used cars for which we are estimating price change, it is clear that we should be using the full sample if possible, particularly if the price movement is affected significantly as shown in figure 6.

**Figure 7**

**Comparison of the hedonic time dummy method controlling for all observed characteristics versus just those incorporated in the estimation cell method (age and category)**

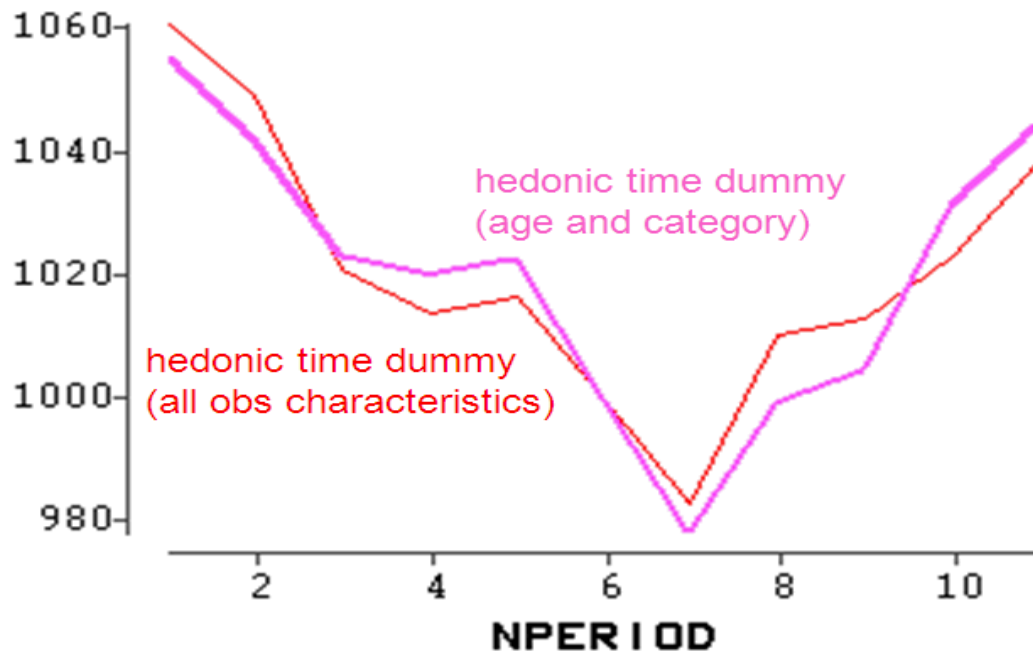
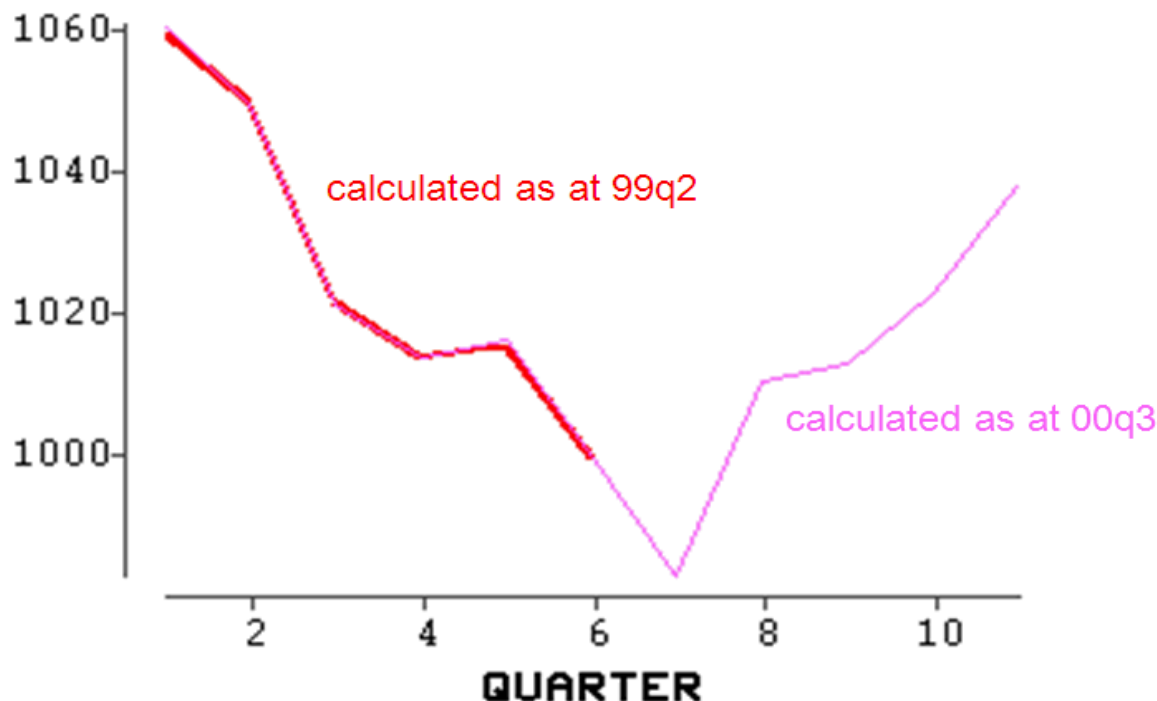


Figure 7 shows that there is some impact on the index of including all the observed characteristics in the hedonic regression model versus only those characteristics incorporated in the estimation cell method. However, the impact is less than that of the restriction of the sample shown in figure 6.

One issue with the use of hedonic estimation in production is that we cannot revise the index. Because the hedonic model uses pooled data, there is the potential for past movements to change slightly. Figure 8 shows the index calculated at two different points in time. So, in June 1999 (quarter 6 in the graph) we are using data up to that point, while for September 2000 (quarter 11) we added another five quarters of data to the pooled dataset that is being run by the hedonic regression model. As can be seen, the resulting indexes are very similar. To operationalise the method, we calculate the index each quarter using the eight most recent quarters of data, and we splice on the most recent movement to the previously published index number.

**Figure 8**

**The index calculated up to June 1999, compared with the index calculated up to September 2000**



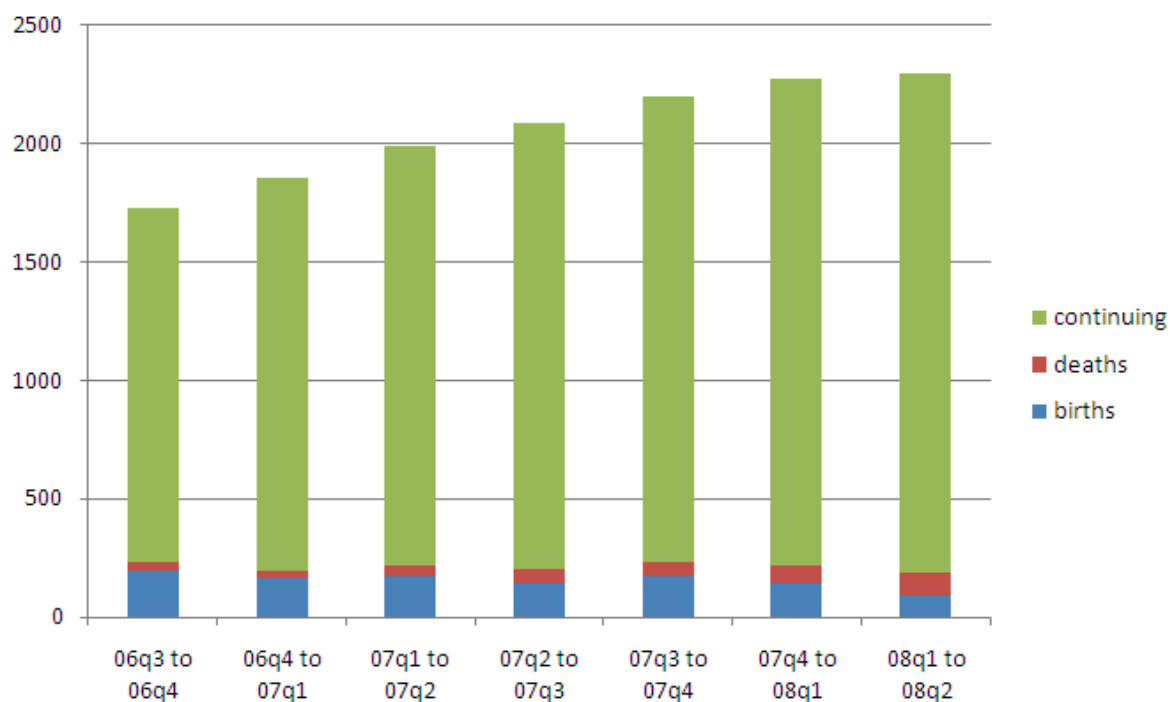
### **3. Benchmarking of the rental matched-sample method in 2009**

In 2009 we used hedonics to benchmark the performance of the matched-sample stratified estimation of the housing rentals index, which contributes about 7 percent to the CPI (Krsinich, 2009).

Every quarter we get rents for a probability-based sample of approximately 2,000 dwellings. New rental properties in the selected areas are identified from bond data and 'birthed' into the sample. These births, along with exits from the sample ('deaths'), mean that the composition of the sample is gradually changing over time. Figure 9 shows the composition of the sample in terms of those present in both quarters and entrants ('births') to, and exits ('deaths') from, the sample. Note that the sample size is increasing over time, which corresponds to the increasing rental dwelling population in New Zealand.

**Figure 9**

**Composition of the rental sample**



The current estimation method calculates average rental movements within strata based on broad region and number of bedrooms of the rental dwelling. Within these strata the movement is calculated from those that are present in the sample in both the previous and current quarters. This is a 'matched sample' approach. Figure 10 uses a simple artificial example to demonstrate how this works.

**Figure 10**

**Estimating the rental price change from matched samples**

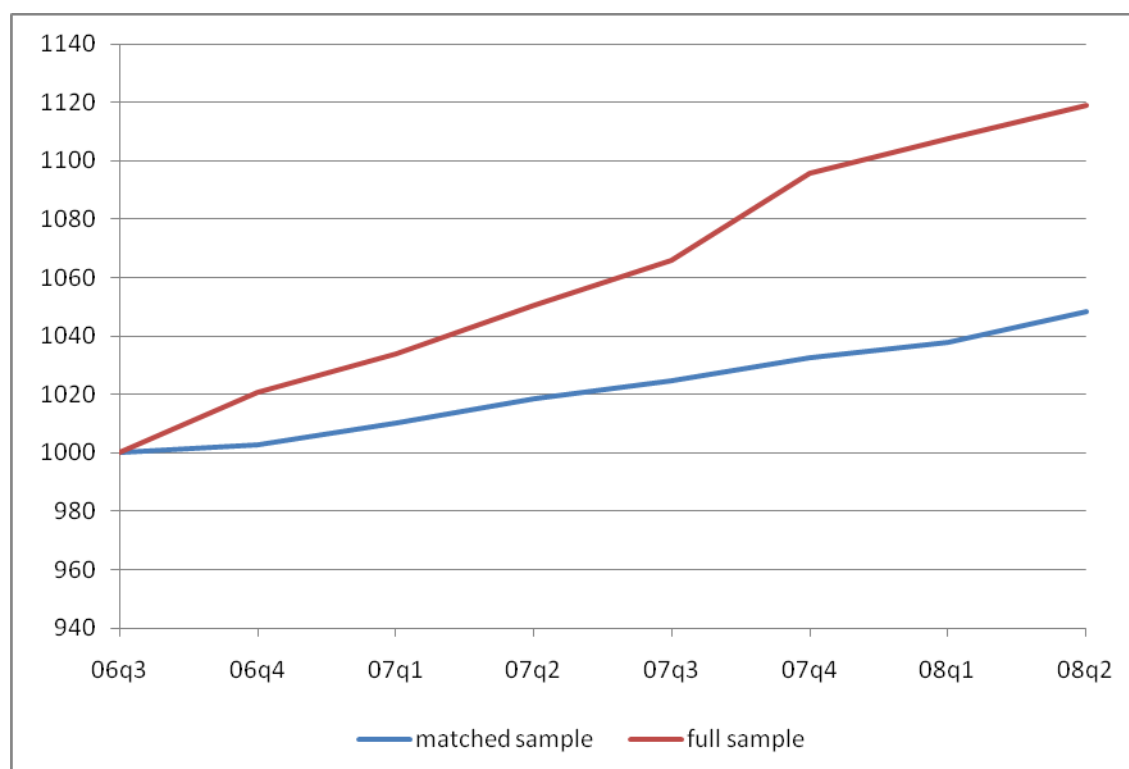
Dwelling id	Rent Q1	Rent Q2	Rent Q3
1	350	350	
2	270	270	270
3	150	160	160
4	280	280	285
5	120		
6			280
7		310	310

Annotations: Dwelling 1 (red box), Dwelling 2 (purple box), Dwelling 3 (purple box), Dwelling 4 (purple box), Dwelling 5 (purple box), Dwelling 6 (purple box), Dwelling 7 (purple box).  
 "matched sample q1 to q2" (red text) points to Dwelling 6.  
 "matched sample q2 to q3" (purple text) points to Dwellings 2, 3, 4, 5, 6, 7.

Because the rental dwellings entering the sample tend to be higher priced than those exiting the sample, we might expect the price change estimated from the full sample to include quality change (ie a quality increase) as well as price change, and therefore to be biased upwards. Figure 11 shows that, as we would expect, the rental index calculated using the matched sample is flatter than that calculated from the full sample.

**Figure 11**

**A comparison of the rental index calculated from the full sample and the matched sample**

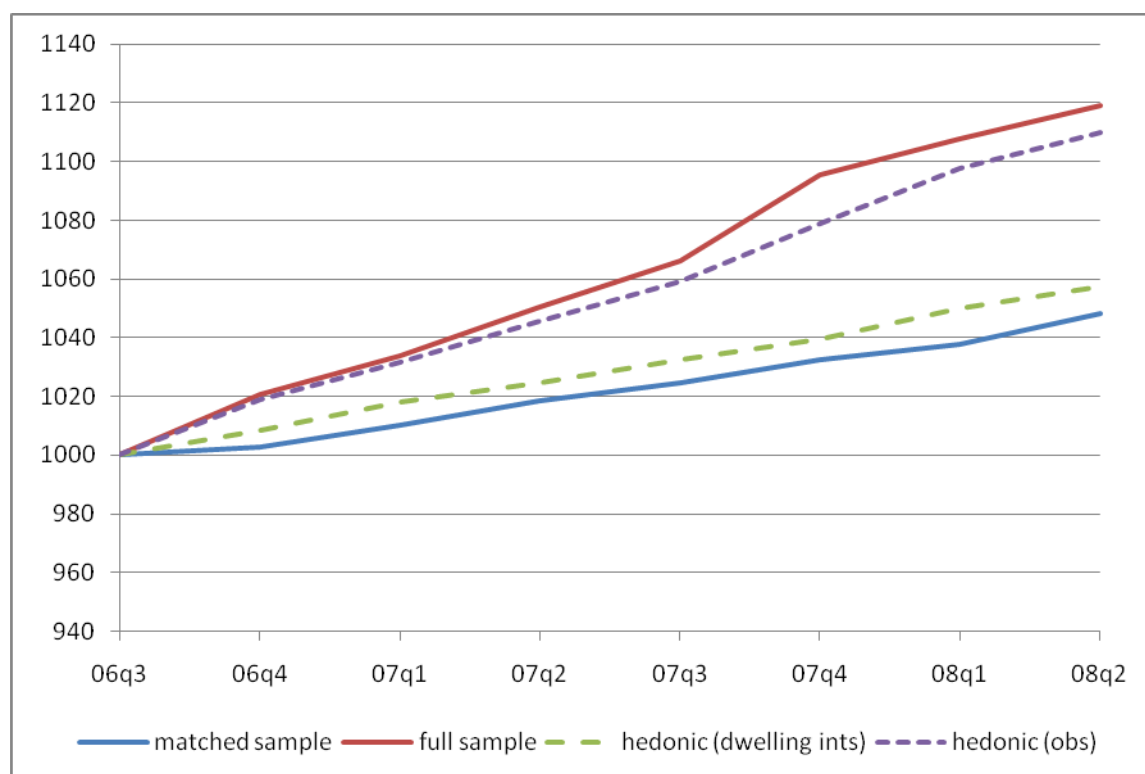


Concerns about whether this matched-sample approach was in fact introducing a downward bias to the index prompted us to do a hedonic benchmarking exercise.

The longitudinal nature of the survey data meant we were able to fit a hedonic model with dwelling-specific intercepts, controlling for all characteristics fixed over time at the

dwelling level. Figure 12 shows the index calculated using the full and matched sample under the current stratified estimation approach, compared with benchmark hedonic indexes. 'Hedonic (obs)' is the 'standard' hedonic index which controls for the observed characteristics only (region to the meshblock level, and number of bedrooms) and does not adjust for inter-dwelling correlation over time. 'Hedonic (dwelling ints)' is the hedonic estimation which includes dwelling-specific intercepts – a 'fixed effects' model – and therefore controls for all characteristics (both observed and unobserved) that are fixed over time at the dwelling level.

**Figure 12**  
**Comparison of rental indexes using different estimation methods**



The current matched sample estimation tracks the benchmark hedonic index well, implying that the current method is performing well in controlling for the compositional change in the sample due to entering and exiting rental dwellings. However, more thought is required about the effect of time-varying characteristics, such as depreciation and renovations.

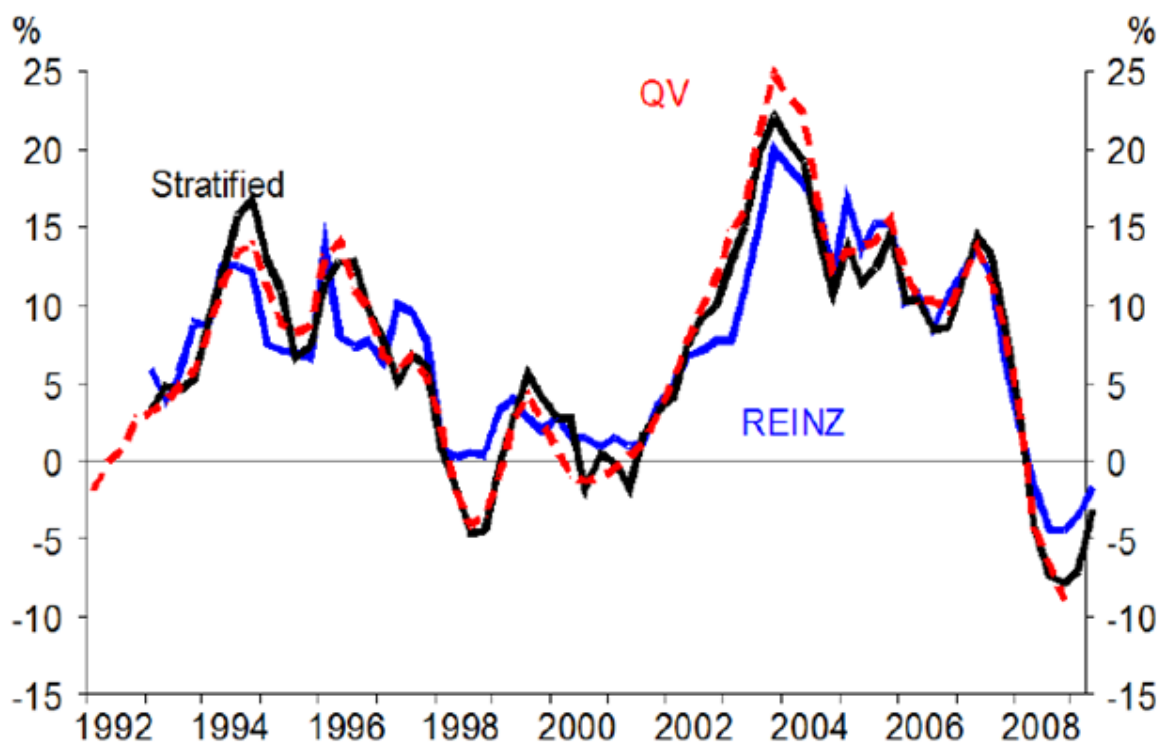
#### 4. Benchmarking of the stratified House Price Index method in 2009

The Reserve Bank of New Zealand (RBNZ) recently developed a new method for the Real Estate Institute of New Zealand (REINZ) to use in the measurement of house prices (McDonald and Smith, 2009). Previously, the REINZ measure has been a national median which, although very timely, is likely to be affected by compositional change. That is, the median house price measure will have been affected by both price and quality changes between months. The RBNZ proposed a stratification method, which calculates medians within suburbs grouped into deciles, on the basis of their historical sale prices.

The Quotable Value house price index uses the Sales Price Appraisal Ratio (SPAR) method to adjust for compositional change and was used as a benchmark in the development of the REINZ stratification method. Figure 13 shows that the stratified REINZ tracks the QV measure much more closely than the REINZ median.

**Figure 13**

**Annual percentage change in house price measures**

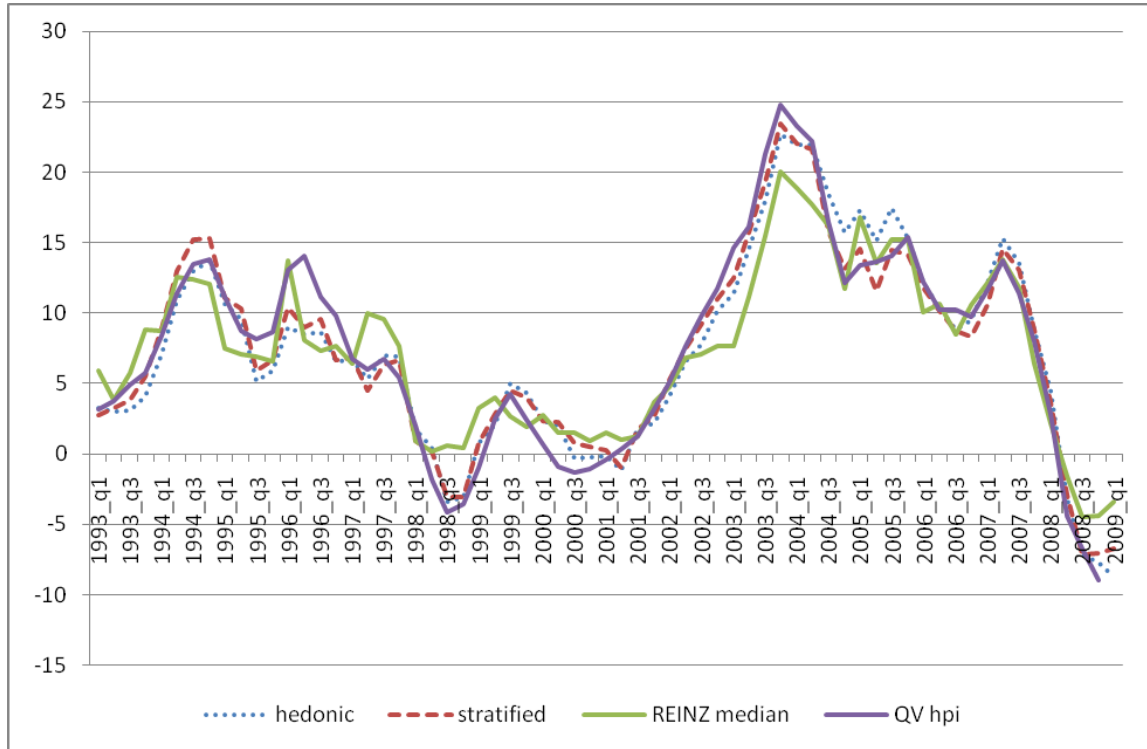


Source: McDonald and Smith (2009)

As part of Statistics NZ's feedback on the stratified method, we calculated a 'semi-hedonic' index using dummy variables for the 1,852 suburbs as the independent variables in the hedonic model. Figure 14 shows that this gave a very similar result to the stratification method, in terms of the annual percentage changes. Figure 15 compares the four measures as indexes. This implies that the stratification approach uses the suburb-level information in the data well. Note that if there is systematic change within suburbs then this semi-hedonic model will not control for that. To do that would require individual-level house sale data.

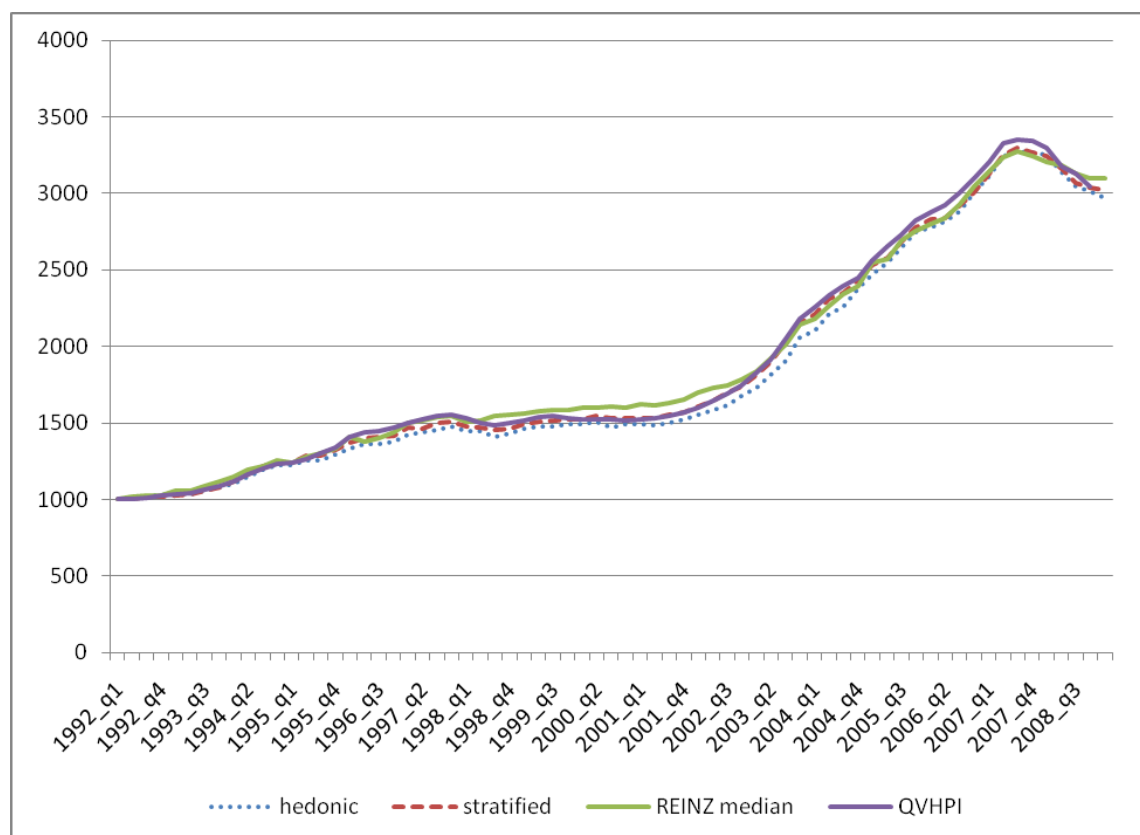
Figure 14

Annual percentage change of house price measures, including a hedonic measure



**Figure 15**

**House price indexes, comparison of different methods**



## 5. Survey data for hedonic estimation

Hedonic models are often developed using information from catalogues on price and characteristics. One of the key limitations of catalogue data for hedonic estimation, however, is that there are generally no quantity weights. If the prices of different specifications of the item being priced (eg computers, rental dwellings, fridges) move differently, then indexes will be biased towards the movements of rarer specifications if appropriate weights are not incorporated into the models.

In all the situations outlined above we were fortunate to have had representative sample data (in the case of used cars and rental dwellings) or population data (in the case of house sales) to work with. So, explicitly or implicitly, there were quantity weights. Rare models of used cars, or unusual houses, would be represented as such in the sample, and therefore the hedonic estimation gives them the appropriate weight in the estimation.

Consideration could be given to collecting survey data with hedonic estimation in mind. Data could be collected using either a simple random sample (ie equally weighted) or a more complex design incorporating stratification and/or clustering. With a hedonic estimation of price movement, less data is required than for a price movement

Use of Hedonics at Statistics New Zealand, by Frances Krsinich  
calculation via the calculation of a total or average price. This, then, would result in  
lower respondent burden and less processing work.

## 6. Hedonic models with dummies for small aggregates

Sometimes individual-level characteristics are not available, but identification of small aggregates are. For example, in house price measurement, we may not have any of the characteristics of the houses being sold, but we might have fine area-level identifiers such as suburbs. This was the case in the REINZ housing price index discussed above. In that case, we can estimate hedonic models where dummy variables representing those small aggregates are included as characteristics in the model. While these hedonic models will be biased by any systematic shift in the composition of the sample below the small-aggregate level, they have two advantages over a stratification approach:

- they will not be affected by 'noise' in the changing composition, as this gets passed to the error term in the hedonic model
- they use the data more efficiently, and require less manual intervention.

## 7. References

Krsinich, F (2009). *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 2009. Available from [www.stats.govt.nz/Publications/Product-Development-and-Publishing/nzae-2009/using-hedonic-regression-to-assess-the-housing-rentals-component-nz-cpi.aspx](http://www.stats.govt.nz/Publications/Product-Development-and-Publishing/nzae-2009/using-hedonic-regression-to-assess-the-housing-rentals-component-nz-cpi.aspx)

McDonald, C, & Smith, M (2009). *Developing stratified house price measures for New Zealand*. (Discussion Paper Series, DP 2009/07). Wellington: Reserve Bank of New Zealand. Available from [www.rbnz.govt.nz](http://www.rbnz.govt.nz).

Statistics New Zealand (2007, July). How used-car price changes are measured in the CPI. *Price Index News*. Available from [www.stats.govt.nz/methods\\_and\\_services/newsletters/price-index-news.aspx](http://www.stats.govt.nz/methods_and_services/newsletters/price-index-news.aspx).