Current research and practices for estimating house price indices at Statistics Canada: A note

Paper (preliminary) presented at the Ottawa Group 2007 meetings held in Ottawa, Canada (October 10-12)

Serra Erdur and Marc Prud'Homme

(serra.erdur@statcan.ca or marc.prudhomme@statcan.ca)

Shelter is a significant component of most Consumer Price Indices (CPI), often accounting for well over 20 percent of total expenditures and for some countries over 25 percent. As is well known however, shelter or more precisely for the purpose of this note owner-occupied-housing (OOH), is without doubt one of the most challenging (may we dare add interesting) domains faced by compilers of the index. First, there are the conceptual issues that must be dealt with such as what exactly is the target price movement that we are attempting to estimate. Even if this complicated issue had a simple solution, there still remains the thorny but practical question of getting the price observations. And finally, let's assume just for a moment that the compiler has overcome (or put them aside would be more likely) both of the aforementioned complications, one more significant hurdle still needs addressing and that is the dynamic nature of the market where the sample of units transacted every period changes; thus complicating even more the construction of constant-quality house price indices by not being able to find identical units to compare over adjacent periods.

The problems can easily be avoided by simply omitting OOH from the CPI, an option that some countries have opted for. It does not however appear reasonable for an expenditure such as OOH, for which households dedicate a large portion of their savings for the initial purchase followed by significant amounts of their income to maintain, should be simply left out of a price index of consumer expenditures.

This is essentially Statistics Canada's position, adopted many decades ago when it was decided to include OOH in the CPI. Consequently, OOH is an important feature of the Canadian CPI (CCPI), which accounts for 16.48% of total national household expenditures, consisting of Mortgage interest cost (5.66%), Replacement cost (3.27%), Property taxes (3.31%), Insurance (1.15%), Maintenance and repairs (1.51%), and Other expenses (1.58%). Because the CCPI estimates recurring costs to owner-occupiers and its inclusion of depreciation of the dwelling, the official CPI's treatment to shelter has often been described as a variant of the user cost approach.

The OOH index in the CCPI, is quite data intensive. Good data on house prices (with and without land) are particularly important because they play a role in five different components of OOH: mortgage interest cost, replacement cost (without land), insurance, realtors commissions, and legal fees; the latter two are found within the Other expenses categories.

Research on house prices at Statistics Canada has been motivated by the importance of OOH in the CCPI and the importance of house prices in calculating for the OOH index. Furthermore, good house price indicators have many other uses outside the CPI and have long been identified as an important data gap that needs to be rectified by national statistical agencies. In fact, it is the Governor of the Bank of Canada, David Dodge, in a speech given at the Conference of European Statisticians in 2003, who articulated this need quite clearly:

> "*Given that investment in housing represents a big chunk of household spending, and that for most people their homes represent their most valuable asset, it is surprising that, in many countries, there are no comprehensive quality-adjusted data on housing prices and rents… … There is a need to expand the current limited international experience in constructing standardized housing price and rent indexes.*"

Since then, various international organizations have held two forums on the subject: A joint conference by the BIS and IMF in 2003, and the OECD in 2006.

The current note will briefly discuss some of the research that is currently being conducted at Statistics Canada on the question of house prices. It is divided into three parts: 1) The alternative measures of shelter cost and the CCPI; 2) the hedonic and repeat sales house price indexes; and 3) the re-engineering of the New House Price Index (NHPI)

PART 1: ALTERNATIVE MEASURES OF SHELTER COST AND THE CPI

At the last Ottawa Group meeting in London, a paper was presented on alternative measures of shelter cost and the CPI. Eventually, the indexes produced here, once the project complete, will be made publicly available under the heading of "experimental" or "analytical" indexes and will complement the current "headline" CPI. The project is motivated by the recognition that no single CPI can satisfy the various needs of its users. If ever a component of the CPI made this clearly (or painfully) obvious it is the treatment of shelter.

For example, if one is in need of a deflator for personal consumption expenditures in the national accounts, then the rental equivalence approach is probably the best candidate for this purpose. In contrast, if measuring the change in retail prices is the aim of the CPI, then current movements in house prices should probably be used. But if the CPI is to be used to track the behaviour of the purchasing power of the dollar, then the previous methods are not well suited for this task; the rental equivalence approach is a notional value that does not reflect out-of-pocket disbursements, while limiting OOH to transaction prices neglects some of the multi-period costs faced by home-owners and does account for the investment dimension which is inherent for OOH.

The alternative measures project examines various treatments of shelter and their incidence on the CPI. It is meant as way of fulfilling the needs of users which the official CPI cannot meet on its own. At the previous Ottawa Group meeting, we presented the current state of the project at that time. The results centred around four concepts: the official concept (described above), the rental equivalence approach, the money outlays approach and the net payments approach. The last two also each had two additional variants. Version A of the money outlays approach included equity payments while version B did not; version A of the money outlays approach is based on the purchase price while version B is based on the amount of the down payment plus the discounted value of future equity payments.[1]

[1] A more detailed discussion of these various approaches can be found in the version of the paper presented at the Ottawa Group meetings in London. http://www.statistics.gov.uk/events/ottawa_group/about.asp

Since the last Ottawa Group meetings in London (U.K.) the work on this project has progressed and expanded and now new alternative treatments of OOH have been added to the existing ones. These new versions are:

a) No OOH. For some inexplicable reason this option has always been left out of the these analytical series but its addition will help for international comparability with those countries that do not have OOH in their CPI. This version will also answer the question as to the sensitivity of the CPI when OOH is excluded from its calculation.

b) "Truer" User Cost estimates. This approach borrows from capital theory and although admittedly generates data series that are more volatile than the others because of the inclusion of a capital gains component, the results are quire interesting in their own right; furthermore, in addition it provides insights as to the behaviour of the CPI when the treatment of OOH approximates the one that would be used in the context of a Cost-of-Living index (See Diewert (2004).)

Excluding OOH from the CPI is not conceptually difficult and is presented here for completeness. No discussion beyond the results is required.

The calculation and inclusion of a "true" user cost formula is another matter and does raise some interesting issues. It will be discussed here in more detail. A subsequent paper will present alternative results based on different views of the user cost formula.

The basic user cost formula in the context of OOH is based on the following relationship:

$$1.\ c_t = (d+i)P_t - (P_t - P_{t-1}) + RC$$

Where:

c_t: User cost at time t

d: Depreciation

i: Interest rate for housing investment

P_t: Price of the house

$P_t - P_{t-1}$: Is the capital gain (or loss) from holding the house for one period.

RC: Recurring costs

It has been modified somewhat in the CPI manual and is presented in the following way:

$$2.\ UC = rM + iE + D + RC - K$$

UC : User cost

M: Mortgage debt

E: Equity in the home

r : Mortgage interest rates

i: Rate of return on alternative assets

D: Depreciation

RC: Recurring costs

K: Capital gains (or losses)

Equations 1 and 2 differ in respect to the treatment of interest rates. Equation 2 recognizes explicitly that homeowners' face two interest rates: the one used to compute the cost of borrowed funds and the rate of return they are forgoing by having money tied up in their house (the equity) and not invested elsewhere. Note that the homeowner with a house that is mortgage free ($M = 0$) will still face the cost of the foregone interest income, iE, of not having the money invested elsewhere. It is not as clear in equation 1 how such a situation would be handled.

Equation 2 requires the calculation of the weights and the price indices for its components. RC is not conceptually difficult and the data are available because they are already included in the current treatment of OOH in the Canadian CPI; some of the components for RC are for instance: legal fees, real estate commissions, and maintenance and repairs. Obtaining the price and weight data for these components is relatively straightforward. The weight for the depreciation component comes from the value of the housing stock for the reference year as determined by the survey of household spending (SHS) multiplied by the rate of depreciation that has been set at 1.5%. The monthly index for depreciation is the rate of depreciation multiplied by the increase (or decrease) in the value of the housing stock as measured by the New House Price Index (NHPI). The value of depreciation used in the UC index is the same as the one used in the official CPI. The amount of mortgage debt outstanding for the reference year is derived from the SHS; movements in this index result from changes in the average rate of interest for mortgages in a given month. The *rM* index

itself used in the UC equation is borrowed directly from the mortgage interest cost index used in the official CPI.

Owner's equity is the difference between the current price of a dwelling and the amount owing on the dwelling. We wish to determine how much an owner would have in his pocket if he had not bought a house but instead invested in another (financial) asset. It can be interpreted as the down-payments opportunity cost because that money could have been allocated to another financially productive asset. Unfortunately this variable is not directly measurable and must be modelled. The SHS provides information for the end of the reference year on the value of a dwelling, the total balance of the mortgage, and the change in the principal owing. There are no questions with regards to the total value of equity tied to the sampled house. For those that have bought the unit during the reference year, getting to their value of equity is as simple as assuming that their down-payment must approximate the equity quite closely. Quite naturally, we would expect that value of the equity to be larger than the down-payment for those who have owned their house for a period longer than the reference year.

For the purpose of this model it is assumed that the value of equity is equal to the difference between the reported value of the dwelling on December 31 of the reference year and the outstanding mortgage balance at that time. However, this only provides a year-end value, which then is deflated by the NHPI to get to the equity amount on Jan 1 of that same year.

Another key point at this stage is that it is assumed that every household paid for their dwelling on Jan 1st and then sold it on Dec 31st of the same year. All lump sum payments as well as the principal paid with the regular mortgage payment during that year are also used to derive the equity position of the household and are assumed to fall on Jan 1st.

Once a total equity estimate is arrived at for Canada, then an interest rate is needed to construct the base weight and then the subsequent monthly movements. To this end, the government 5-year yield bond rate was used for the rate of return on alternative assets, hence representing the opportunity cost of having funds invested in the house.

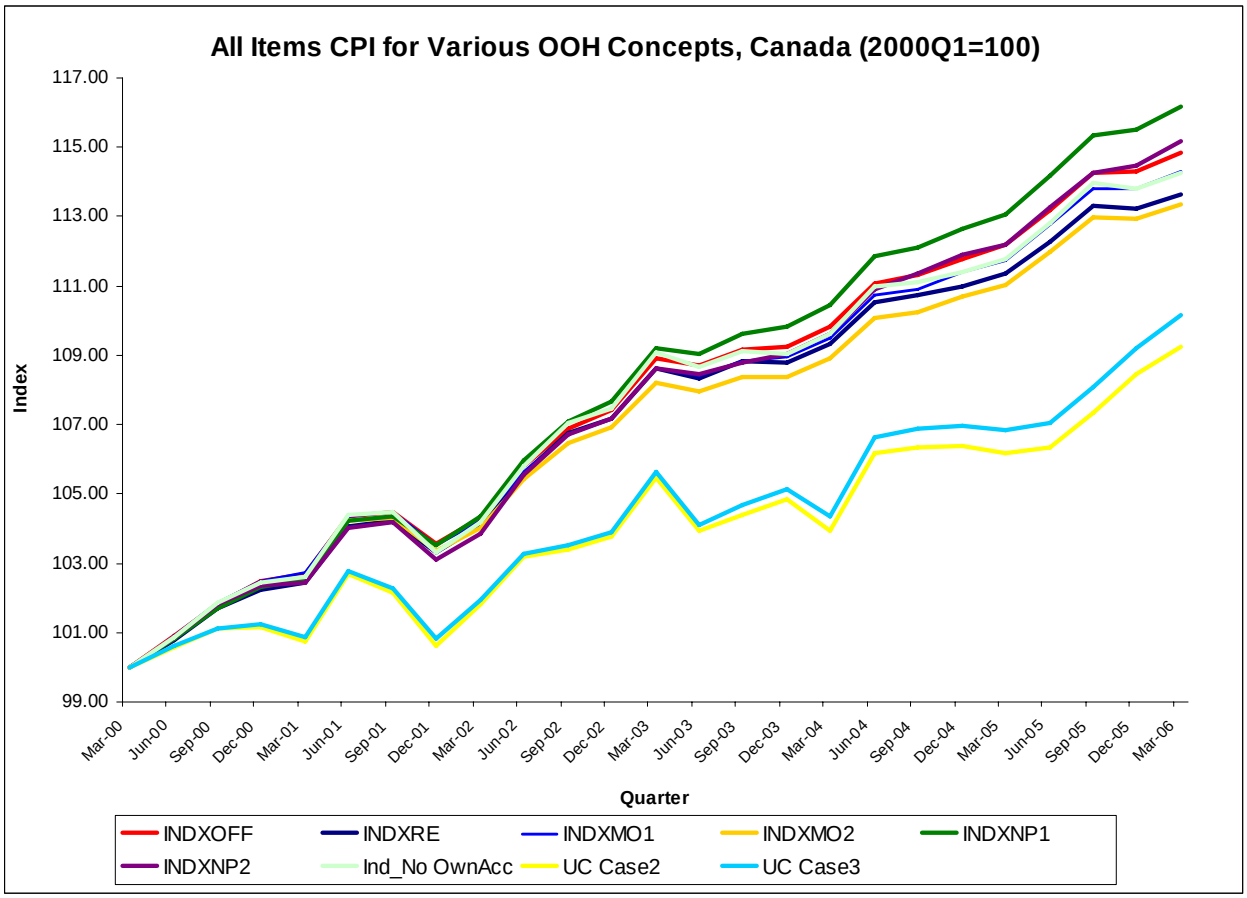
Lastly, the capital gains (or losses) must be computed. Typically, the weight of capital gains would be estimated as the increase in the price of the dwelling over the reference year. The

SHS does not provide the information to be able to calculate with full accuracy the total value of capital gains for a given reference year. For instance, capital gains accrue for any homeowners who sell their house during the reference year. The SHS will ask respondents if they sold their house during the year but will not ask about the purchase price unless it was bought during the reference year. Therefore the information for estimating capital gains is limited. To remedy this data gap, an average purchase price was estimated for January 1 of the reference year and then the movement in the NHPI was used as a proxy for the increase in overall house prices over the year and thus generating an estimate for capital gains for the reference year. The capital gains index is calculated as the change in the NHPI in relation to the base year value for capital gains.

The following table provides the overall growth rates over the 2000 to 2006 period for all the aforementioned approaches at three levels of aggregation from the current analytical series including the two new ones just presented. The accompanying graph illustrates the sensitivity of the results to the various approaches used.

Annual Growth Rates

Commodity group	Official Concept	No Own. Acc	User Cost Case 2	User Cost Case 3	RE Concept	Money Outlays Concept		Net Purchase Concept	
						No equity	Equity	Purchases	Down-payment
Owned Accommodation	2,2%		-2,1%	-1,3%	1,4%	1,8%	1,0%	3,3%	2,3%
Shelter	2,3%	1,1%	-0,6%	-0,1%	1,8%	2,0%	1,6%	2,9%	2,3%
All items	1,9%	1,9%	1,0%	1,1%	1,8%	1,8%	1,7%	2,1%	1,9%


All Items CPI for Various OOH Concepts, Canada (2000Q1=100)
Index
117.00
115.00
113.00
111.00
109.00
107.00
105.00
103.00
101.00
99.00
Mar-00
Jun-00
Sep-00
Dec-00
Mar-01
Jun-01
Sep-01
Dec-01
Mar-02
Jun-02
Sep-02
Dec-02
Mar-03
Jun-03
Sep-03
Dec-03
Mar-04
Jun-04
Sep-04
Dec-04
Mar-05
Jun-05
Sep-05
Dec-05
Mar-06
Quarter
INDXOFF
INDXRE
INDXMO1
INDXMO2
INDXNP1
INDXNP2
Ind_No OwnAcc
UC Case2
UC Case3

PART 2: HEDONIC AND RESALE PRICE INDICES

The treatment of OOH in the official CPI requires house prices for its calculation. Currently the proxy that is used for this purpose is the NHPI. Conceptually there is no reason why the data should be limited to prices of new houses; it is more operational and practical considerations that make it so. In fact, the house prices, which are used in the OOH of the CPI would and should include resale house prices also. There are however no data which are produced at Statistics Canada on resale house prices and therefore the NHPI acts as a proxy in this area. New house prices may behave differently than those of resale houses, motivating interest in exploring the potential of data on resale housing for the CPI.

For some time now, we have had access to a very detailed and rich database of resale house prices; but limited to one geographic area, i.e. Ottawa. The data in question is from Multiple Listing Services® (MLS®) provided to us from the Ottawa Real Estate Board. In spite of the restricted geographical coverage, the data have proven to contain a wealth of information on "transaction" prices and the multitude of features (characteristics) associated with every house that has sold and that was advertised through MLS® in the last 10 years or so. The database is populated with over 100,000 observations and therefore can confidently support a number of empirical studies of various kinds.

As is well known, producing reliable estimates of resale house price indices has its challenges. Using the matched sample approach is not an option because of the impossibility of having an identical house that is sold on the market over adjacent periods. Many housing market analysts and urban economist have used with much success hedonics and repeat sales applied to various data sources to estimate reliable price indices for resale houses. Statistics Canada has followed this lead by using the MLS® database at its disposal to produce various versions of quality-adjusted hedonic price indexes for housing in Ottawa in addition to various versions of the repeat sales method.[2]

This note presents an extension of this analysis with updated data and additional version of the repeat sales method over the 1996 to 2005 period. Three hedonic models are presented here.

[2] The results from a previous study were presented at the joint IMF/OECD workshop on real estate price indices held in 2003. http://www.oecd.org/document/47/0,3343,en_2825_495691_37582447_1_1_1_1,00.html

1. The characteristics price index approach where the coefficients are estimated for the base year and then the observations for future periods are applied to the estimated coefficients in order to compute the index.

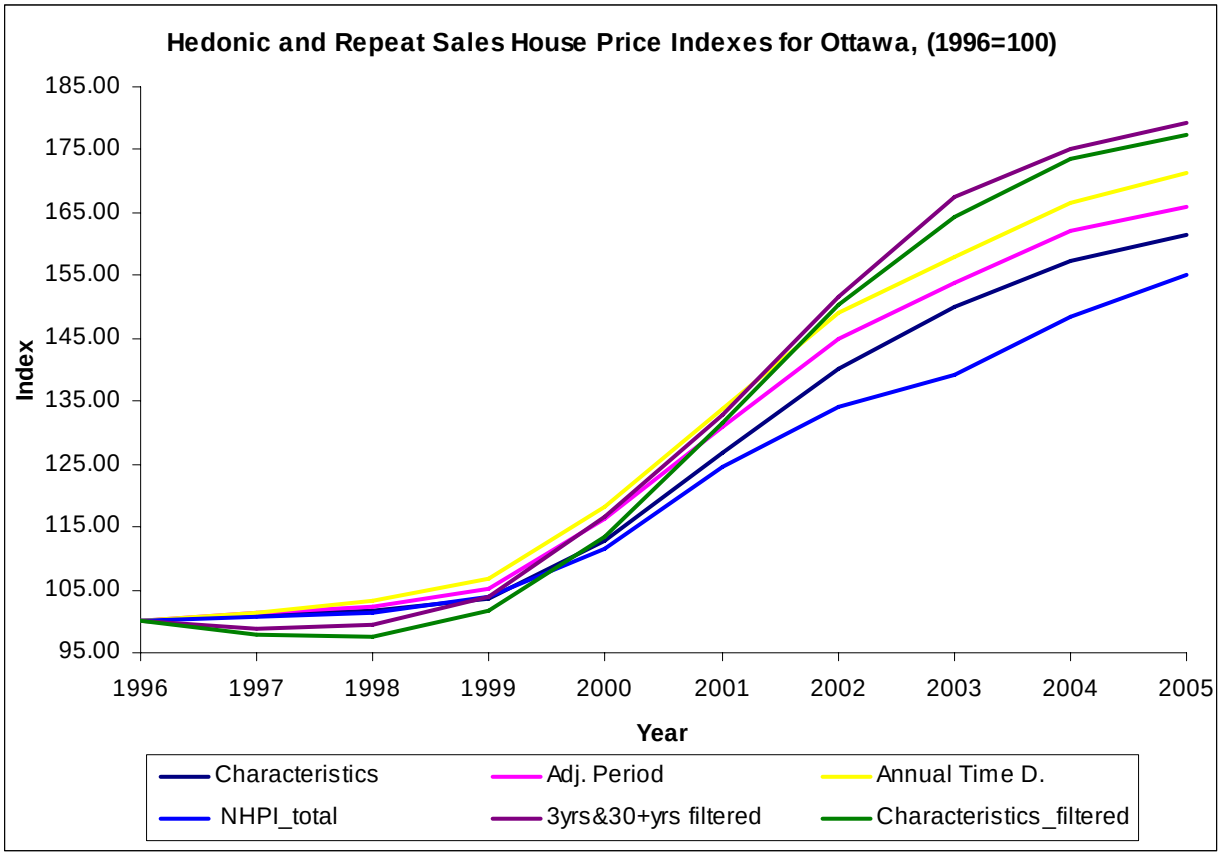
2. The adjacent year dummy price index approach where time dummy coefficients are estimated from two years of pooled data and the indexes from every pair of years are chained together to produce a continuous series.

3. The time-dummy price index approach where all the data are pooled over the entire period for which the observations are available.

In addition to the hedonic indexes, repeat sales indexes have also been computed. This method is presented here because of its widespread use in the literature on house prices and it also a method, which is widely known and referred to in the United States (Case and Schiller). Presented here are two versions of the repeat sales method. The first one (characteristics) matches those house observations that have sold more than once in the database and for which the characteristics have not changed. The other repeat sales index (filtered 3 and 30 years) is a partial attempt at filtering those units that may be problematic for the construction of repeat sales indexes. For instances, units that have been on the market more than once over a period of three years or less may be qualified as "lemons" and thus be the source of "selectivity bias" in the index. Moreover, houses that have sold more than once but are 30 years old or more are also excluded from the calculations because it is thought that these older houses must have gone through some renovations but the information does not show up in the data. As a point of reference, the NHPI for Ottawa is also provided.

The following table shows the average annual growth rates for the various price indexes. The accompanying graph shows the progression of the index over the period.

WRS Nonfiltered	*WRS 3&30 yrs filtered*	*WRS Characteristics filtered*	*Characteristics hedonic*	*Adj. Period*	*Yearly Time Hedonic*	*Monthly Dummy Hedonic*	*Average*	*Median*	*NHPI*
5.9%	5.7%	5.9%	4.7%	5.0%	5.4%	5.8%	6.2%	6.2%	4.5%



It is hoped that in the future the data will be extended to Canada's other major urban centres. Such information could potentially be used in the OOH index of the CPI and also as a stand-alone quality-adjusted price index for houses in Canada thus filling the data gap that was identified by David Dodge. As of the writing of this note, there have been some promising developments on this front. Just recently, the Canadian Real Estate Association has expressed an interest in producing a hedonic version of a price index so that it can better track market developments in this area.

PART 3: THE "NEW" NEW HOUSE PRICE INDEX: FUTURE DIRECTIONS

Statistics Canada's New House Price index (NHPI) measures changes over time in the contractor's selling prices of new residential houses. Monthly index where the pricing date is the 15th of each month. The questionnaire asks for the total selling price and also collects information about the contractors' estimate of the current market value of the land the unit is built on. It is produced for 21 metropolitan areas and then aggregated to the Canada total.

Its main uses within Statistics Canada are in the CPI where it is used for calculating replacement cost (structure only), insurance (structure only), mortgage interest costs, and real estate commissions. It is also used in the National Accounts for deflating the current dollar value of new residential construction. Additional uses include private consultants and housing market analysts, other individuals or organizations that have an interest in house price trends, and the media.

To construct the NHPI, contractors are selected using purposive sampling and subject matter judgment. The builder must be a "significant" force in the market and can provide prices for comparable homes over time. The sample includes small contractors to large operations with 100s of employees. There are approximately 250 builders across the country who are asked to provide an "transaction" price for a representative house. The sample size is in a constant state of flux but currently stands at about 700 units (a specific model of house with constant features); for example, in one month there were 85 new units introduced in the sample while 51 were discontinued.

Pricing the same model house in the "almost" same location from month-to-month controls for quality. If an identical house is not found, then a close substitute similar house is found and an "explicit" quality adjustment is made based on the "options cost" approach to ensure comparability. If no new house is found in the approximate vicinity as the previous one, then a new model is linked in, often using the "linked-to-show-no-price-change" approach. The incidence of each method varies but it appears that option 2 is the most frequently used.

The NHPI questionnaire includes a list of variables which are: Landscaped (full or partial), Central Vacuum, Dish Washer, Distance from city centre, Number of bedrooms, Number

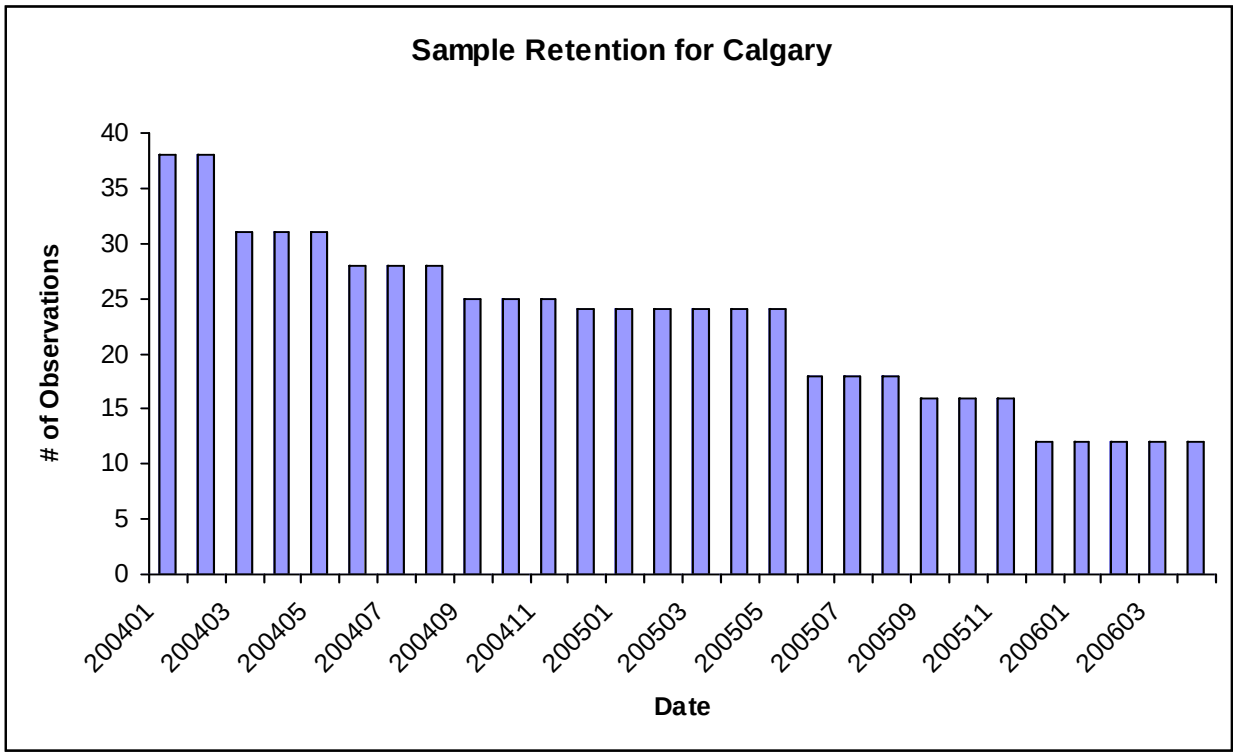
of bathrooms, Lot size, Living area, Number of rooms, Number of fireplaces, Garage, Single - 1 storey, Single - 2 storeys, Split level, Semi.

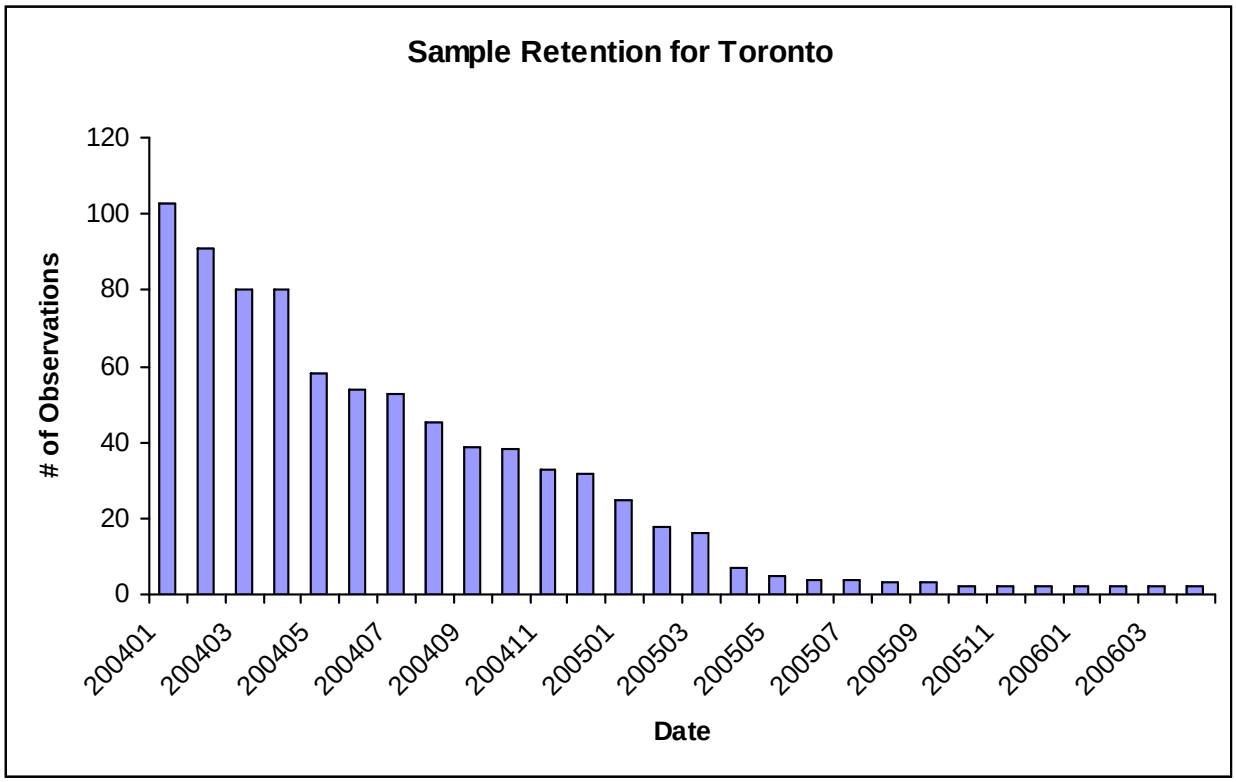
A hedonic regression is suggested as an alternative method for adjusting for quality change. For this purpose, two models were generated: the full model with all the variables from the survey including geographical location dummy variables, and a reduced model with living area, lot size, and the geographical location dummy variables. The reduced model is produced in order to explore the tolerance of the results to a reduction in the number of explanatory variables. By having a smaller number of questions on the survey, respondent burden is reduced and response rates can potentially increase.

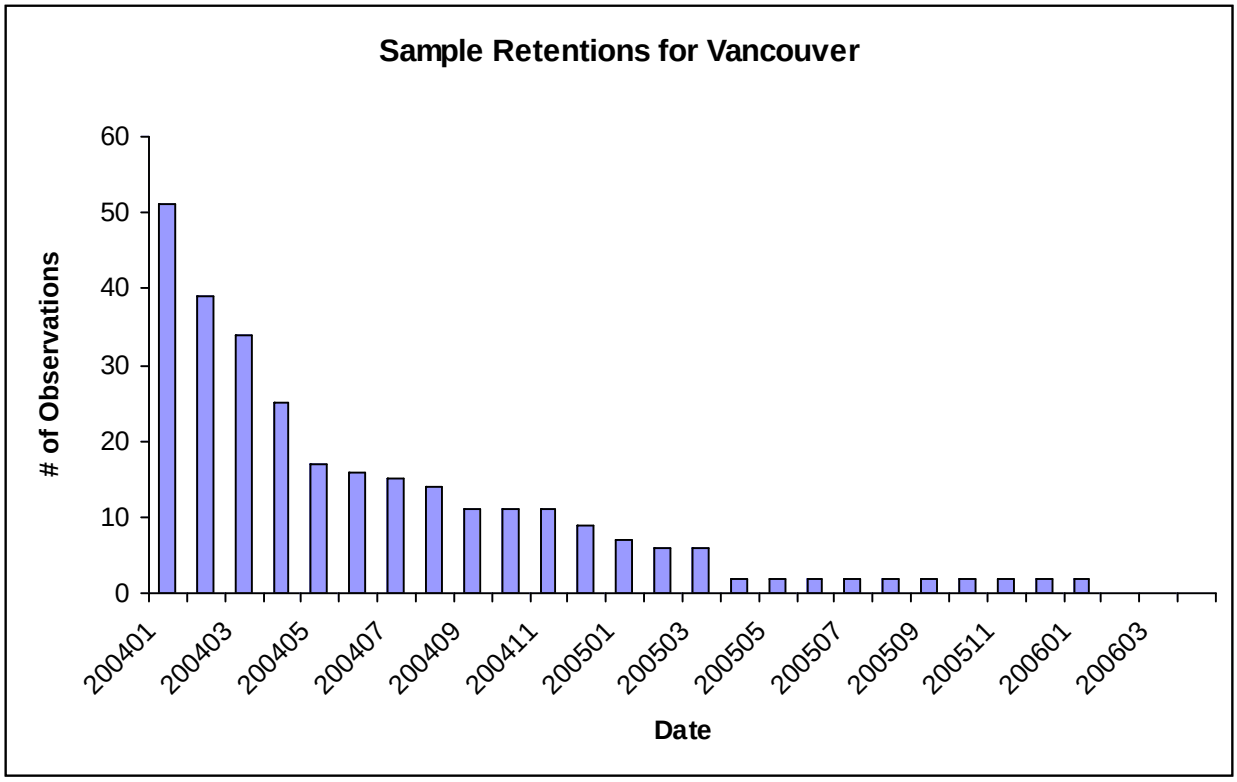
Three types of hedonic regressions are applied to the data: Pooled approach - all periods, Adjacent period approach, and the Hedonic replacement approach. The following table presents the size of the sample for the three cities under study:

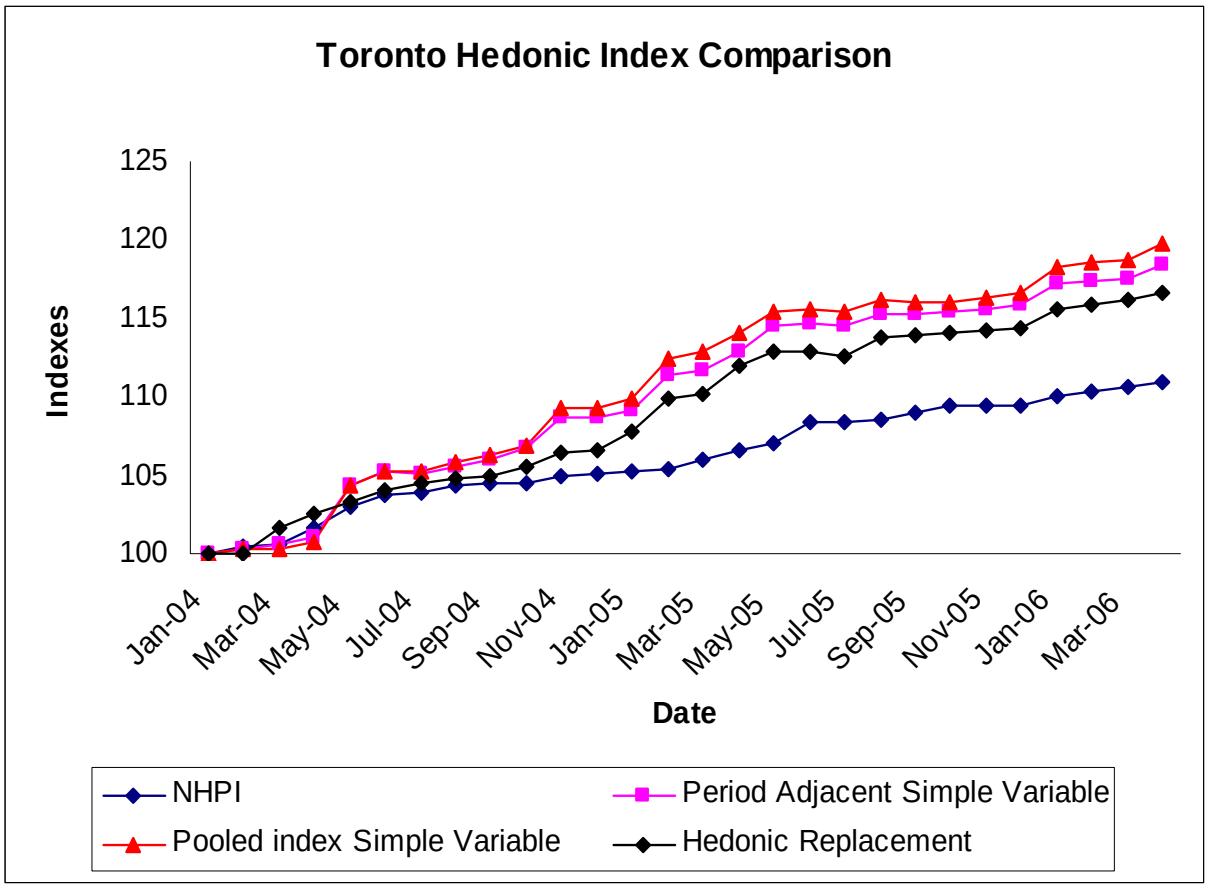
City	Date	Max	Date	Min	Change
Calgary	200403	39	200604	36	-7.69%
Vancouver	200411	51	200405	36	41.67%
Toronto	200604	130	200502	102	27.45%

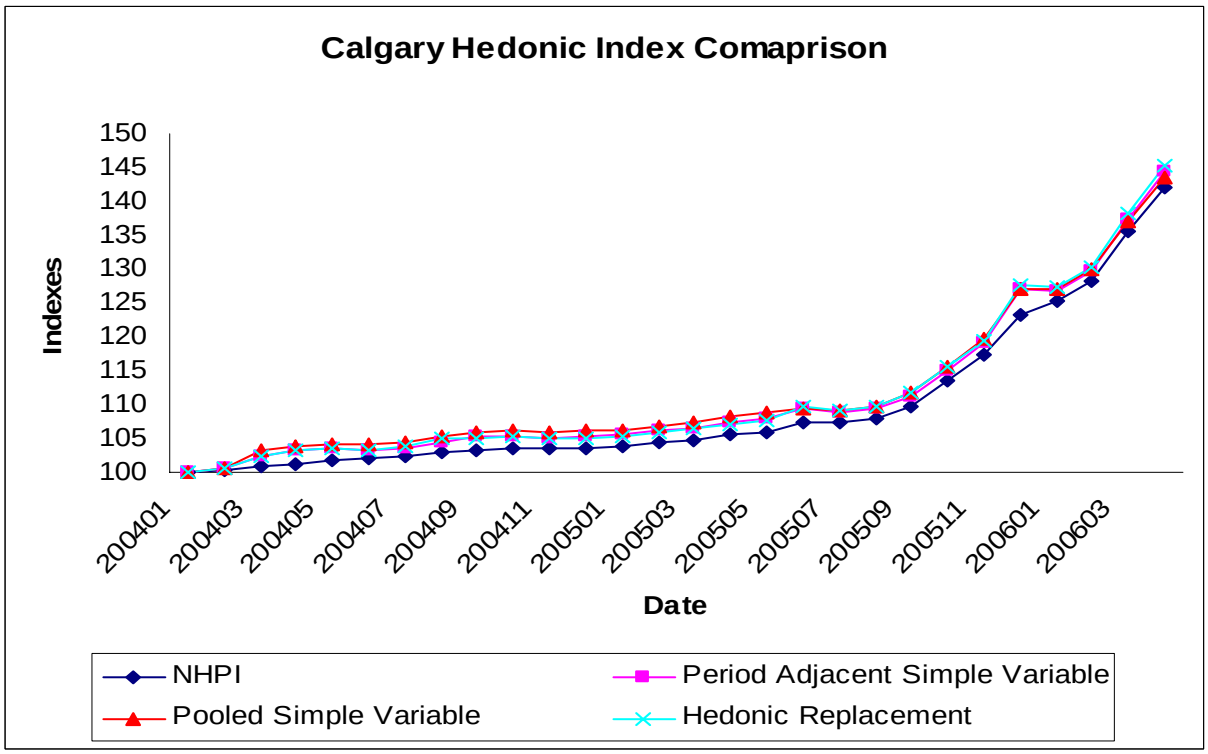
The following graphs show what we call the sample decay over the period January 2004 to March 2006. It is an illustration of the length of time a similar unit stays in the sample. A rapidly falling value means a higher incidence of replacements and potentially biased index as a result of the method which is currently used to adjust for quality differences when a new model is introduced into the sample. Subsequent graphs show the differences among the various hedonic regressions which are used compared with the NHPI.


Sample Retention for Calgary
# of Observations
40
35
30
25
20
15
10
5
0
200401
200403
200405
200407
200409
200411
200501
200503
200505
200507
200509
200511
200601
200603
Date


Sample Retention for Toronto
# of Observations
120
100
80
60
40
20
0
200401
200403
200405
200407
200409
200411
200501
200503
200505
200507
200509
200511
200601
200603
Date


Sample Retentions for Vancouver
# of Observations
60
50
40
30
20
10
0
200401
200403
200405
200407
200409
200411
200501
200503
200505
200507
200509
200511
200601
200603
Date


Toronto Hedonic Index Comparison
Indexes
125
120
115
110
105
100
Jan-04
Mar-04
May-04
Jul-04
Sep-04
Nov-04
Jan-05
Mar-05
May-05
Jul-05
Sep-05
Nov-05
Jan-06
Mar-06
Date
NHPI
Period Adjacent Simple Variable
Pooled index Simple Variable
Hedonic Replacement


Calgary Hedonic Index Comaprison
Indexes
150
145
140
135
130
125
120
115
110
105
100
200401
200403
200405
200407
200409
200411
200501
200503
200505
200507
200509
200511
200601
200603
Date
NHPI
Period Adjacent Simple Variable
Pooled Simple Variable
Hedonic Replacement

The appendix includes some of the comments that were provided to use from the panel of experts that reviewed some of the work that was done on this project.

CONCLUDING REMARKS

This brief note has presented some of the on-going projects related to OOH in Prices Division. Although incomplete in many regards, it does provide insights about data availability and the results that can be achieved from them. The alternative OOH project will be a very useful tool for those that that need a CPI that is better tailored to their particular needs. The hedonic research can be of use not only in the context of the CPI but can also potentially serve the growing requirements for more and better indicators for the housing market. Finally, the initiatives to improve the New House Price Index will also serve a multitude of current and future needs for national accountants, the CPI, and the other numerous users of these data.

APPENDIX

A panel of four real estate economists were asked to come and review the work that was produced on the hedonic regressions in December 2006. Some of their comments are summarized here:

Marion Steele

I would at this point choose the hedonic replacement, partly because it represents the least disruption from the current procedure. The only change I would make to the method given for this in the paper is the increase in sample size. Points to consider in the future include investigating the issue of whether house models should be weighted by number of units sold. To increase homogeneity in the sample further, do separate regressions for row and single houses. Doing this should increase robustness of the simple model. Use a slightly beefed up simple model. Add as additional variables costly characteristics which households value: Luxury level of bathrooms, Luxury feature, fireplaces. I'd try two dummies: for one fireplace, and for two or more Possibly include dummy for 2 car garage, dummy for 3 car garage. Replace lot area with lot frontage, or include lot frontage as well as lot area. If lot frontage were included then the garage dummies might not be needed. In a market like Toronto virtually all single-detached houses with a frontage above some threshold would have a double garage.

Tsur Sommerville

"The most important recommendation I have is that any variant of a hedonic index will be an important improvement over the existing method". "Relative to moving to a hedonic approach, the differences among the alternative methods for generating a hedonic based index are in my mind second order. The key is moving to hedonics." Questionnaire : "since builders will often add in extras, low priced financing, or pay the GST instead of lowering the list price, a question on the value of extras or discounts not included in the sales price would be helpful."

François DesRosiers

"Turning to hedonics for revisiting the NHPI methodology is the best decision Statistics Canada could take. Considering though that a drawback of using this approach is the relatively small sample size available - between 130 (Toronto) and 38 (Calgary) new house sales, I would suggest to compute the index for a given month "t" by pooling the samples of the three preceding months while resorting to some kind of moving average; for instance: Index for January 2007 is based on October, November and December 2006 data; Index for February 2007 is based on November and December 2006 and on January 2007."

Paul Anglin

"I think that there should be an emphasis on using a method that is easy to compute two reasons: to make it easy to communicate to an outside audience and to produce robust estimates."