Resale Residential Property Price Index (rRPPI) Methodology

PRODUCER PRICES DIVISION

FINAL DRAFT

Mar 11 2019

Contents

1 Introduction

Following the global financial crisis in 2008, the G-20 identified real estate price indices as an important financial soundness indicator. Linked to these efforts, residential property price indices form a core set of data necessary for financial stability analysis under a new tier of the IMF's Special Data Dissemination Standard, known as SDDS Plus. In order to meet these new data requirements, and to improve the relevance of housing price statistics, the 2016 Federal Budget mandated Statistics Canada to develop an official Residential Property Price Index (RPPI).

The RPPI covers prices for both resale and new housing, with the resale component captured by the Resale Residential Property Price Index (rRPPI).[1] This is a constant-quality price index that measures the change in transaction prices for resale houses and condominium apartments in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria. The data collection, ingestion, editing, and calculation is done in partnership with Teranet and National Bank.

The rRPPI is a quarterly index, composed of 12 sub-indices—two for each property type (house and condo) in each of the six cities—and a set of weights for aggregating these sub-indices to produce a national property price index. Each sub-index is computed using the repeat-sales method, an internationally accepted method for constructing a constant-quality price index as outlined in Eurostat's *Handbook on Residential Property Prices Indices* (IMF, 2015). The purpose of this document is to provide the details of the repeat-sales methodology used for the rRPPI, and how this index is compiled to a national total.

The remainder of the documents proceeds as follows: section 2 summarizes the key concepts and definitions used for the rRPPI; section 3 outlines the data sources for the rRPPI and how these data are prepared; section 4 details the repeat-sales method and how this is used to construct a constant-quality price index; section 5 discusses aggregation of the city-level sub-indices into a national property price index; and section 6 outlines revisions to the rRPPI.

2 Concepts and definitions

Table 1 defines key concepts used for constructing the rRPPI. Note that the concept for the date of sale of a property is the closing date, at which time the property is transferred from the seller to the buyer and subsequently recorded in the land registry. The closing date is later than the date at which a buyer and seller agree on a transaction price for the property.

---

[1] Statistics Canada's New House Price Index (NHPI) and New Condominium Apartment Price Index (NCAPI) capture price changes for new housing. These indices, along with the rRPPI, are aggregated to create the RPPI.

Table 1
Concepts and definitions for the rRPPI

Concept	Definition
Price	Final transaction price at the closing date for the sale of a property and recorded in the provincial land registry.
Sales date	The closing date for the sale of a property.
Sales pair	Prices and sales dates for consecutive sales for the same physical property.
CMA	Census metropolitan areas for Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria, according to Statistics Canada’s Standard Geographic Classification (2016).
Sample	All single/semi-detached houses, row houses, and apartment condos in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria that sold at least twice since January 1, 1998 and appear in the land registry databases.
Target population	All single/semi-detached houses, row houses, and apartment condos in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria, eligible for resale, that actually sold between January 1, 1998 and the current period.
Sales weights	Value shares for all repeat-sale properties by property type (house and condo) and city, based on transaction prices, and used to aggregate city-by-property type sub-indices.
Weight reference period	The three years prior to the quarter 1 index in a given year.
Index base period	The period for which the index equals 100. The base period for the rRPPI is 2018 = 100.

3 Data

3.1 Data sources

Property transaction data for the rRPPI come from the provincial land registry offices in Alberta, British Columbia, Ontario, and Quebec, from 1998 to the current period. As each property sale in Canada is registered in its respective provincial land registry office, these data capture all property transactions over this period. The rRPPI includes only transactions for single/semi-detached houses, row houses, and condominium apartments in the Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria CMAs (according to Statistics Canada’s Standard Geographic Classification for 2016). These data are collected and processed by Teranet-National Bank.

Transaction data from each provincial land registry is provided on a monthly basis. These transactions are then matched to Teranet’s property database to create a sales history for each property. Sales pairs are created for each property that has sold twice, capturing the transaction prices and the closing dates for both sales for that property; sales pairs are created for consecutive sales for properties that have sold three or more times. Properties that have sold only once (e.g., newly built properties) are excluded. Table 2 gives a fictitious example of the resulting sales-pair data.

Table 2
Example of sales pair data

Address	Property Type	Sales Date	Sales Price	Previous Sales Date	Previous Sales Price
123 Fake St.	Condo	2018-01-08	250,000	2014-02-01	200,000
321 False Dr.	House	2018-01-18	500,000	2005-06-04	400,000
321 False Dr.	House	2005-06-04	400,000	1999-12-15	350,000

3.2 Collection delay

Although the land registry data are received from the provincial land registries every month, there is a delay between when sales are recorded in the land registries and when these data are received by Teranet-National Bank. This delay is particularly severe for British Columbia. Table 3 gives an example of the cumulative proportion of sales received per province at the end of each month, for a fixed month *M*. Due to this collection delay, the rRPPI has a revision of one quarter to ensure that sufficient data are collected to produce reliable index values.

Table 3
Average portion of sales received per province each month

Province	Period M	Period M+1	Period M+2	Period M+3	Period M+4	Period M+5
Alberta	92%	100%	100%	100%	100%	100%
British Columbia	43%	94%	97%	99%	99%	100%
Ontario	90%	95%	97%	100%	100%	100%
Quebec	83%	83%	83%	83%	83%	88%

3.3 Cleaning and filtering

Although data for the rRPPI come from administrative sources—and are therefore fairly clean—some filtering is required to remove property transactions that are not appropriate for constructing the rRPPI, as well as outliers that can have a large influence on the index. This includes removing sales pairs for which one of the transactions may not be at arm's length (e.g., a bequest) or may be a distress sale, or for which the price movement between sales is so extreme as to suggest that the property may have undergone a quality change. These filters are applied to each CMA and property type separately, and are summarized by the order in which they are applied in table 4.

Table 4
Data filters for sales pairs in the rRPPI

Filter	Rationale
Transaction price less than or equal to 10,000 dollars.	These transactions may not be arm's-length transactions (e.g., bequest).
Holding period less than 6 months.	These transactions can be distress sales or speculative transactions (de Haan and Diewert, 2013, section 6.11), or flipped properties for which there is a large change in the quality of the property (e.g., Jansen et al., 2008; S&P Dow Jones, 2018).
Annualized return greater than or equal to 3 median absolute deviations from the median.	There may be a change in the quality of a property that gives rise to an unusually large price change between transactions, or a data entry error for one of the transaction prices. As the rRPPI is based on average transaction prices, this also removes outliers that can exert a large influence on averages (e.g., Rousseeuw and Hubert, 2011).

Prior to these filters being applied, a series of filters are used to remove transactions that may be part of a builder split or a developer block transaction, as these types of transactions fall outside the scope of the rRPPI. Groups of five or more properties within the same Forward Sortation Area (first three digits of a property's postal code), sold on the same date, and for the same price are treated as a block/split transaction. The transaction for each property in the group is removed when this is the most recent transaction for each property.

Transactions in a group can return if the subsequent sale price for at least 75% of the properties in the group is at least 75% of the price of the block/split transaction price, and, for each subsequent sale for each property in the group, there is at most one other property in the same Forward Sortation Area that sold for the same price on the same date. This allows block/split transactions to be used if the price for these transactions is close to the subsequent selling price for most of the properties in the block/split transaction.

4 Index modelling

The repeat-sales method offers a means to construct a constant-quality price index, exploiting multiple sales for the same property over time to control for time-invariant differences in quality between properties. Other approaches for constructing a constant-quality index (e.g., hedonics or stratification) require property characteristics, such as the age of the property, and these are not available in the land registry data. See Hansen (2009) for a comparison of the different approaches for constructing a property price index.

In practice there are a number of methodological choices to make when implementing a repeat-sales index. This section outlines the repeat-sales method and highlights the particular flavour of the repeat-sales index used to construct the rRPPI. See Wang and Zorn (1997) and de Haan and Diewert (2013, chapter 6) for an overview of the repeat-sales method, and Jansen et al. (2008) for an application of the repeat-sales approach.

Due to the smaller number of transactions for condos, the condo sub-index is calculated for each quarter. For houses the index is calculated monthly, with the resulting index values averaged over each quarter to produce a quarterly index.[2]

4.1 The repeat-sales method

There are two broad classes of repeat-sales price indices—the Jevons-like geometric repeat-sales index (GRS index) proposed by Bailey et al. (1963) and the Laspeyres-like arithmetic repeat-sales index (ARS index) proposed by Shiller (1991).[3] The GRS and ARS indices often show similar price movements over time (e.g., Shiller, 1991). The rRPPI uses the arithmetic repeat-sales index outlined in Shiller (1991, section II), similar to that used by S&P Dow Jones (2018).

In addition to the geometric and arithmetic versions of the repeat-sales index, there are various weighting schemes that can be used to weight the price relatives in the index calculation (e.g., Case and Shiller, 1987, 1989; Abraham and Schauman, 1991; Calhoun, 1996). These are inverse-variance weights designed to correct for differences in the variance in transactions prices for properties with different holding periods that can complicate constructing confidence intervals for the index. While weights directly affect the index values, in practice these weights can have an at most marginal impact on the index (e.g., Goetzmann, 1992; Hansen, 2009), especially with large samples. The weighted indices, however, rely on more assumptions than their unweighted counterparts, and cannot be computed if the weights cannot be calculated. Previous studies have also found that the unweighted indices are not inferior to the weighted indices (de Haan and Diewert, 2013, section 6.14). Consequently, as confidence intervals are not reported for the rRPPI, inverse-variance weights are not used to compute the rRPPI.

4.2 The GRS and ARS indices

Historically the GRS index came before the ARS index, starting with the seminal paper by Bailey et al. (1963), and it is easier to understand the ARS index by first developing the GRS index. Letting time periods be indexed by $t \in \{0,1,\ldots,T\}$ and properties be indexed by $i \in \{1,2,\ldots,N\}$, the starting point for the GRS index is a structural (hedonic) model of property prices

$$\log(p_{it}) = \log(P_t) + x_{it}\beta + \epsilon_{it},$$

where p_{it} is the transaction price of property i at time t, P_t is a common city-level price reflecting aggregate price movements, x_{it} is a (row) vector of property characteristics (e.g., number of bedrooms for property i at time t), β is a vector of implicit (hedonic) prices, and ϵ_{it} is an error term.[4] This is simply a time-dummy hedonic model in which properties can sell more than once (e.g., de Haan and Diewert, 2013, chapter 5). In the context of this model, the constant-quality (geometric) price index in period τ with base period 0, denoted by I_τ^G, is $P_\tau/P_0 \cdot 100$. Importantly, P_t is not random—it is a parameter that governs the joint distribution of property prices.

---

[2] The published rRPPI is a quarterly index, although there are internal uses for a monthly rRPPI for houses at Statistics Canada.

[3] There are also hybrid repeat-sales indices that can make hedonic quality adjustments using a random-effects model (see de Haan and Diewert, 2013, section 6.25-6.27), although these methods cannot be used with the land registry data as these data lack property characteristics. An exception is the quality adjustment proposed by Goetzmann and Spiegel (1995) for the GRS index.

[4] The structural models in Case and Shiller (1987) and Shiller (1991, section V) are special cases of this model.

Under the assumption that property characteristics do not change over time (i.e., $x_{it} = x_i$, for all t) and that each property sells twice, the first-difference transformation can be used to deliver

$$\log\left(\frac{p_{is(i)}}{p_{if(i)}}\right) = \log\left(\frac{P_{s(i)}}{P_{f(i)}}\right) + \log\left(\frac{\epsilon_{is(i)}}{\epsilon_{if(i)}}\right)$$
$$= \sum_{t=1}^{T} D_{it} \log\left(\frac{P_t}{P_0}\right) + \log\left(\frac{\epsilon_{is(i)}}{\epsilon_{if(i)}}\right),$$

where $s(i)$ gives the time of the second sale for property i, $f(i)$ gives the time of the first sale for property i, and D_{it} is a dummy variable that takes the value 1 if a property sells for the second time in period t (i.e., $s(i) = t$), -1 if the property sells for the first time in period t (i.e., $f(i) = t$), and 0 otherwise. The assumption that property characteristics do not change over time means that the percent change in a property's price follows the aggregate percent change in property prices, up to an additive error. Properties that sell three or more times can be incorporated in the first difference transformation by treating consecutive pairs of sales as distinct properties.

Under the assumption that the error terms are strictly exogenous, so that $E[\log(\epsilon_{it})|D_{i1}, D_{i2}, \ldots, D_{iT}] = 0$—a standard assumption in panel-data applications (e.g., Wooldridge, 2010, chapter 10)—the assumption that property characteristics do not change over time allows for the price index to be identified from the linear regression

$$\log\left(\frac{p_{is(i)}}{p_{if(i)}}\right) = \sum_{t=1}^{T} D_{it}\gamma_t + \log\left(\frac{\epsilon_{is(i)}}{\epsilon_{if(i)}}\right),$$

so that $I_\tau^G \equiv P_\tau/P_0 \cdot 100 = \exp(\gamma_\tau) \cdot 100$. The first difference transformation turns a structural model that depends on property characteristics into an estimating equation that depends on only the time when a property sells.[5]

It is instructive to derive the form of the GRS index as an index number to make the link with the ARS index. Letting $N_f(\tau)$ be the set of properties that sell for the first time in period τ, $N_s(\tau)$ be the set of properties that sell for the second time in period τ, and $N(\tau) = |N_f(\tau)| + |N_s(\tau)|$ (the number of properties that sell in period τ), it can be shown that

$$I_\tau^G = \prod_{i \in N_f(\tau)} \left(\frac{p_{i\tau}}{\frac{p_{is(i)}}{I_{s(i)}^G}}\right)^{\frac{1}{N(\tau)}} \prod_{i \in N_s(\tau)} \left(\frac{p_{i\tau}}{\frac{p_{if(i)}}{I_{f(i)}^G}}\right)^{\frac{1}{N(\tau)}}.$$

The GRS index is simply a matched-model Jevons index with a twist. Rather than use only property transactions that occur in period 0 and period τ, the index itself is used to extrapolate prices across time for all properties that sell in period τ by deflating prices for sales that do not occur in the base period using

[5] If property characteristics systematically change over time, then D_{it} is correlated with the error $\log(\epsilon_{is(i)}/\epsilon_{if(i)})$, although the direction of this correlation depends on whether there is net depreciation or net appreciation of properties over time. With net depreciation (net appreciation), the GRS index will tend to overstate (understate) changes in property prices.

that period's index. This allows all properties that sell in period τ to be used in the index calculation, whether that property sells in period 0 or not.

As an alternative to a geometric index, Shiller (1991) proposes the ARS index, denoted by I_τ^A, that simply replaces the geometric averages in the GRS index with arithmetic averages:

$$I_\tau^A = \frac{\sum_{i \in N_f(\tau)} p_{i\tau} + \sum_{i \in N_s(\tau)} p_{i\tau}}{\sum_{i \in N_f(\tau)} \frac{p_{is(i)}}{I_{s(i)}^A} + \sum_{i \in N_s(\tau)} \frac{p_{if(i)}}{I_{f(i)}^A}}.$$

Price relatives are formed in the same way as the GRS index, except now a Laspeyres index is used to combine price relatives, rather than a Jevons index. This is the index used to calculate the rRPPI.

Computing the ARS index requires solving a system of equations to calculate the index in each period. As with the GRS index, the ARS index can be computed as a linear regression, although now with a set of instrumental variables—this provides a convenient way to calculate the index and determine its statistical properties. Letting

$$Y_i = \begin{cases} p_{if(i)} \text{ if } f(i) = 0 \\ 0 \text{ if } f(i) > 0 \end{cases},$$

and

$$X_{it} = \begin{cases} -p_{if(i)} \text{ if } f(i) = t \\ p_{is(i)} \text{ if } s(i) = t \\ 0 \text{ otherwise} \end{cases},$$

the ARS index is the reciprocal of the instrumental variables (IV) estimator for the regression

$$Y_i = \sum_{t=1}^{T} X_{it}\, \beta_t + v_i,$$

with D_{it} as an instrument for X_{it}. Letting $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})$ and $D_i = (D_{i1}, D_{i2}, \dots, D_{iT})$, the entire series of ARS indices from period 1 to period T is computed as

$$\left(I_1^A, I_2^A, \dots, I_T^A\right)' = \operatorname{diag}\left(\left[\sum_{i=1}^{N} D_i' X_i\right]^{-1} \sum_{i=1}^{N} D_i' Y_i\right)_{T\times T}^{-1} \begin{bmatrix} 100 \\ 100 \\ \vdots \\ 100 \end{bmatrix}_{T\times 1}.$$

The validity of the IV estimator rests on a statement that the index to calculate is an arithmetic index (Shiller, 1991, p. 115). Given a sample of repeat-sales transactions for properties over T periods, the IV estimator is consistent under fairly weak conditions on the sampling process (e.g., White, 2001, theorem 3.15; Wooldridge, 2010, theorems 5.1 and 8.1), so that the estimator for ARS index converges in probability to the population ARS index (i.e., it is unbiased in large samples).

4.3 Representativeness of the target population

The target population for the rRPPI is all properties that are eligible for resale and have actually sold since January 1, 1998. In practice, sales pair data are only available for properties that sell two or more times over this period; properties that sell only once are missing from the sample. This is a sample selection problem—repeat-sale properties may not be representative of all transacted properties—and the resulting repeat-sales index may not capture the price movement for the target population of all transacted properties. Producing a representative index rests on the assumption that there are no systematic differences in latent selling prices and holding periods between properties that transact only once and those that transact twice or more. (See Wooldridge (2010, theorem 19.1) for precise conditions under which sample selection can be ignored with an IV estimator.) Previous studies have found some evidence to support this assumption (see de Haan and Diewert, 2013, section 6.17).

As the rRPPI focuses on resale properties, properties that sell only once because they are newly built do not contribute to a selected sample. The only divergence between the target population and the sample of transactions available are properties that sold prior to January 1998 and only once since then. These properties are not used to calculate the rRPPI, but fall in the scope of the target population as these properties are both eligible for resale and actually sold after January 1998.

4.4 Inverse-variance weights

Case and Shiller (1987, 1989) argue that the variance of transaction prices for sales pairs increases with the holding period for a property, in which case the error term in the regression for the GRS index can be heteroskedastic.[6] This means that the usual OLS standard errors for the GRS index are inconsistent, and the OLS estimator is no longer minimum variance; the same applies to the IV estimator for the ARS index. If the relationship between holding period and variance in transaction prices is known, the generalized least squares (GLS) and generalized instrumental variables (GIV) estimators, using inverse-variance weights, are more efficient alternatives than their unweighted counterparts, and provide a consistent estimator for their standard errors (White, 2001, theorem 4.62; Wooldridge, 2010, theorem 8.5).

Heteroskedasticity is not particularly problematic for the rRPPI; as with most national indices, standard errors are not reported for the rRPPI, and there is a sufficiently large sample that asymptotic efficiency is not a concern (Wang and Zorn, 1997, section 4.4). Using inverse-variance weights, however, modifies the index values. This is problematic as the GLS and GIV estimators require stronger assumptions than the usual OLS and IV estimators (e.g., the relationship between variance and holding period must be known), and failure of these assumptions can result in the GLS or GIV estimator becoming inconsistent or having poor small-sample properties, as well as failing to produce correct standard errors or give a more efficient estimator (e.g., Angrist and Pischke, 2009, section 3.4.1; Wooldridge, 2010, section 4.2.3). There is also no guarantee that inverse-variance weights can be calculated at any point in time (e.g., Calhoun, 1996), and since the weights affect the index values, the index cannot be calculated if the weights fail. Consequently, the rRPPI does not use inverse-variance weights.[7]

---

[6] Under the assumption of strict exogeneity, $E(\log(p_{is(i)}/p_{if(i)})|D_{i1}, D_{i2}, \ldots, D_{iT})$ is linear, so that $var(\log(p_{is(i)}/p_{if(i)})\,|D_{i1}, D_{i2}, \ldots, D_{iT}) = var(\log(\epsilon_{is(i)}/\epsilon_{if(i)})\,|D_{i1}, D_{i2}, \ldots, D_{iT})$.

[7] Inverse-variance weights do not offer a correction for quality changes within a property over time. The original weights by Case and Shiller (1987) put less weight on properties with longer holding periods, giving properties with a longer time between sales, and hence more opportunity for quality changes, less weight in the index calculation because price variance increased with holding time. Jansen et al. (2008), however, find a non-monotonic relationship between holding period and variance, so that these weights cannot adjust for quality differences within a property. If the assumptions required for inverse-variance weighting hold, then the weighted and unweighted indices converge in probability to the same value—if weighting offered a correction for quality changes over time, then the weighted index should converge to a different value than the unweighted index.

Standard errors that are robust to unknown heteroskedasticity can easily be computed for the ARS index without changing the index values, and these are proposed by Shiller (1991, section II). It is also straightforward to generalize these standard errors to account for serially-correlated errors over time for properties that sell three or more times (e.g., White, 2001, chapter 6; Wooldridge, 2010, chapter 8).

4.5 Worked example of the ARS index

To work through an example of the ARS index, suppose there are 3 periods—an initial period 0 that serves as the base period, followed by periods 1 and 2—and three houses, labelled as a, b, and c. House a sells for the first time in period 1 and for the second time in period 2; house b sells for the first time in period 0 and for the second time in period 2; and house c sells for the first time in period 0 and for the second time in period 1. Table 5 summarizes these data.

Table 5
Sales pair data

House	Sales Date	Sales Price	Previous Sales Date	Previous Sales Price
a	2	p_{a2}	1	p_{a1}
b	2	p_{b2}	0	p_{b0}
c	1	p_{c1}	0	p_{c0}

With these data, the ARS index proposed by Shiller (1991) is

$$I_1^A = \frac{p_{a1} + p_{c1}}{\frac{p_{a2}}{I_2^A/100} + p_{c0}} \cdot 100$$

$$= \left(\frac{p_{a1}}{\frac{p_{a2}}{I_2^A/100}} \cdot \frac{\frac{p_{a2}}{I_2^A/100}}{\frac{p_{a2}}{I_2^A/100} + p_{c0}} + \frac{p_{c1}}{p_{c0}} \cdot \frac{p_{c0}}{\frac{p_{a2}}{I_2^A/100} + p_{c0}} \right) \cdot 100$$

$$I_2^A = \frac{p_{a2} + p_{b2}}{\frac{p_{a1}}{I_1^A/100} + p_{b0}} \cdot 100$$

$$= \left(\frac{p_{a2}}{\frac{p_{a1}}{I_1^A/100}} \cdot \frac{\frac{p_{a1}}{I_1^A/100}}{\frac{p_{a1}}{I_1^A/100} + p_{b0}} + \frac{p_{b2}}{p_{b0}} \cdot \frac{p_{b0}}{\frac{p_{a1}}{I_1^A/100} + p_{b0}} \right) \cdot 100.$$

This is like a pure matched-model Laspeyres index, except that house a can be included in the index calculation by deflating its price to get a pseudo period 0 price.[8] Doing this, however, means that the index is defined by a system of equations—one for each time period—that must be solved to get the index for a given period. The ARS index is defined simultaneously for each period.

To get a closed-form solution for the ARS index, note that

[8] With a pure matched-model index, the period 1 index is simply $p_{c1}/p_{c0} \cdot 100$ and the period 2 index is $p_{b2}/p_{b0} \cdot 100$.

$$D \equiv \begin{bmatrix} D_a \\ D_b \\ D_c \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix},$$

$$X \equiv \begin{bmatrix} X_a \\ X_b \\ X_c \end{bmatrix} = \begin{bmatrix} -p_{a1} & p_{a2} \\ 0 & p_{b2} \\ p_{c1} & 0 \end{bmatrix},$$

and

$$Y \equiv \begin{bmatrix} Y_a \\ Y_b \\ Y_c \end{bmatrix} = \begin{bmatrix} 0 \\ p_{b0} \\ p_{c0} \end{bmatrix}.$$

The ARS index comes from the IV estimator for the linear regression

$$Y = X\beta + v$$

with D as an instrumental variable.

The moment condition for the IV estimator, $\hat{\beta}$, is

$$D'X \cdot \hat{\beta} = D'Y$$

$$\begin{bmatrix} p_{a1} + p_{c1} & -p_{a2} \\ -p_{a1} & p_{a2} + p_{b2} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} p_{c0} \\ p_{b0} \end{bmatrix},$$

the solution to which is

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \frac{1}{(p_{a1} + p_{c1})(p_{a2} + p_{b2}) - p_{a1}p_{a2}} \begin{bmatrix} p_{a2} + p_{b2} & p_{a2} \\ p_{a1} & p_{a1} + p_{c1} \end{bmatrix} \begin{bmatrix} p_{c0} \\ p_{b0} \end{bmatrix}.$$

The ARS index for period t is simply $100/\hat{\beta}_t$, and thus

$$I_1^A = \frac{(p_{a1} + p_{c1})(p_{a2} + p_{b2}) - p_{a1}p_{a2}}{p_{c0}(p_{a2} + p_{b2}) + p_{b0}p_{a2}} \cdot 100$$
$$I_2^A = \frac{(p_{a1} + p_{c1})(p_{a2} + p_{b2}) - p_{a1}p_{a2}}{p_{b0}(p_{a1} + p_{c1}) + p_{c0}p_{a1}} \cdot 100.$$

5 Index Aggregation

The rRPPI is calculated separately for houses (single/semi-detached and row) and condos, for the Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria CMAs. This means there are two indices for each CMA—one for condos and one for houses—for a total of 12 indices. These 12 indices are aggregated to produce a national rRPPI. As the sub-index for houses is calculated monthly, with the sub-index for condos calculated quarterly due to fewer condo transactions, the three index values in a quarter for the house sub-index are averaged each quarter prior to aggregation.

The weights for aggregation in the rRPPI are sales weights that give the value share for sales of each property type in each CMA, using the repeat-sale transactions that occur in a rolling window of the three most recent calendar years. With the assumption that there are no systematic differences in latent selling prices between properties that transact only once and those that transact twice or more—a requirement for the repeat-sales index to be representative of the target population—these weights give the value share of each property type in each CMA for all transacted properties.

Weighting allows for price movements of certain markets to move the national total in a manner reflective of that market's size and activity. However, anomalies in sales markets can occur and a three year window is chosen to form the weight reference period in order to prevent drastic changes or swings in sales from impacting future index movements. The weight reference period is updated every year with the release of the quarter 2 index.[9]

6 Revision

6.1 Accounting for revision in the repeat sales model

A disadvantage of any repeat-sales index is that it is subject to perpetual revision. Computing the index for one period requires computing the index for all periods and, as new data become available, this will change the index values for previous periods.

The rRPPI avoids revision by using a movement splice to update the index when new periods of data become available. With this approach, the price movement of the series computed with the most recent data is chained together with the last index value of the original series, thereby avoiding revision of the original series. This method of successively chaining together indices is used with hedonic price indices to avoid this same type of revision (e.g., de Haan and Diewert, 2013, section 5.18).[10]

To fix notation, let $I_0^S, \ldots, I_T^S$ be a series of repeat-sale price indices running from period 0 to period T, calculated using the first $S \leq T$ periods of data. This series can be updated with a movement splice as follows. First, with $T+1$ periods of data available, calculate the series of indices $I_0^T, \ldots, I_{T+1}^{T+1}$; that is, recalculate the entire series using all available data. To then update the original series of indices that runs until period T, simply calculate the index value in period $T+1$ as $I_T^S \cdot I_{T+1}^{T+1}/I_T^{T+1}$, and append this value the original series. Thus, the original series of indices becomes

$$I_0^S, I_1^S, \ldots, I_T^S, I_T^S \cdot \frac{I_{T+1}^{T+1}}{I_T^{T+1}}.$$

The impact of any drift in the index from this type of splicing can easily be evaluated over time by comparing the index calculated using all of the data to the spliced index, and this is part of the quality assurance work done when producing the rRPPI.

6.2 Accounting for revision due to collection delay

The rRPPI has a one quarter revision to account for the delay of incoming data from the land registries. This revision means that the index is computed twice for each period. For example, when computing the 2018 quarter 1 index, the index is first computed in quarter 2 of 2018 using all of the data received in quarter 1 of 2018, and is then computed again in quarter 3 of 2018 once the majority of the quarter 1 2018 data has been received from the land registries in quarter 2 of 2018.

---

[9] The rRPPI has a one quarter revision period, so the weights are updated in the second quarter to avoid revisions overlapping a change in weights.

[10] With hedonic indices, a rolling window is sometimes used to avoid drift in the index due to high-frequency chaining. This is not sensible with a repeat-sales index, however, as using a rolling window would discard properties that sold for the first time early in the series.

This revision means that the index must be spliced with two different index series. Using the notation above, the preliminary index is calculated as

$$I_0^S, I_1^S, \dots, I_T^S, I_T^S \cdot \frac{I_{T+1}^{T+1}}{I_T^{T+1}},$$

and the revised index is calculated as

$$I_0^S, I_1^S, \dots, I_T^S, I_T^S \cdot \frac{I_{T+1}^{T+2}}{I_T^{T+2}}.$$

This approach to splicing allows for a one quarter revision to the index, so that additional data can be collected from the land registries, while avoiding perpetual revision of the repeat-sales index.

References

Abraham, J. M. and Schauman, W. S. (1991). New evidence on home prices for Freddie Mac repeat sales. *Real Estate Economics*, 19(3): 333-352.

Angrist, J. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Bailey, M., Muth, R., and Nourse, H. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association,* 58(304): 933-942.

Calhoun, C. (1996). *OFHEO House Price Indexes: HPI Technical Description*. Office of Federal Housing Enterprise Oversight. Retrieved from http://www.ofheo.gov/Media/Archive/house/hpi_tech.pdf.

Case, K. and Shiller, R. (1987). Prices of single-family homes since 1970: New indexes for four cities. *New England Economic Review*: 45-56.

Case. K. and Shiller, R. (1989). The efficiency of the market for single-family homes. *American Economic Review*, 79(1): 125-137.

de Haan, J. and Diewert, W. E. (Eds.). (2013). *Handbook on Residential Property Prices Indices (RPPIs)*. Eurostat.

Goetzmann, W. (1992). The accuracy of real estate indices: Repeat sale estimators. *Journal of Real Estate Finance and Economics,* 5(1): 5-53.

Goetzmann, W. and Spiegel, M. (1995). Non-temporal components to residential real estate appreciation. *Review of Economics and Statistics*, 77(1): 199-206.

Hansen, J. (2009). Australian house prices: A comparison of hedonic and repeat-sales measures. *Economic Record*, 85(269): 132-145.

IMF. (2015). *The Special Data Dissemination Standard Plus: Guide for Adherents and Users*. Retrieved from https://www.imf.org/external/pubs/ft/sdds/guide/plus/2015/sddsplus15.pdf.

Jansen, S., de Vries, P., Coolen, H., Lamain, C., and Boelhouwer, P. (2008). Developing a house price index for The Netherlands: A practical application of weighted repeat sales. *Journal of Real Estate Finance and Economics*, 37(2): 163-186.

Rousseeuw, P. J. and Hubert, M. (2011). Robust statistics for outlier detection. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1): 73-79.

Shiller, R. (1991). Arithmetic repeat sales price estimators. *Journal of Housing Economics*, 1: 110-126.

S&P Dow Jones. (April 2018). *S&P CoreLogic Case-Shiller Home Price Indices Methodology*. Retrieved from https://us.spindices.com/index-family/real-estate/sp-corelogic-case-shiller.

Wang, F. and Zorn, P. (1997). Estimating house price growth with repeat sales data: What's the aim of the game? *Journal of Housing Economics*, 6: 93-118.

White. H. (2001). *Asymptotic Theory for Econometricians* (revised edition). Emerald Group Publishing.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd edition). MIT University Press.