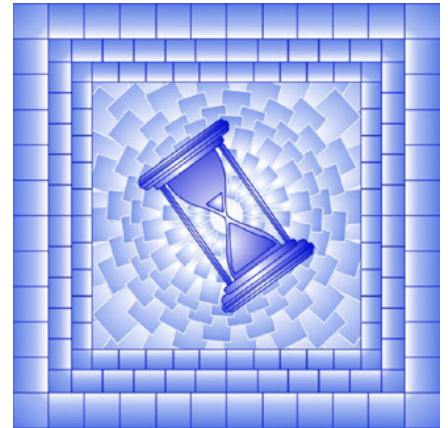


Catalogue no. 62F0014M  
ISSN 1706-7723  
ISBN 978-0-660-31806-6

# Methodology of the Residential Property Price Index (RPPI)

by Steve Martin

Release date: November 14, 2019



Statistics  
Canada

Statistique  
Canada

Canada

---

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, [www.statcan.gc.ca](http://www.statcan.gc.ca).

You can also contact us by

Email at [STATCAN.infostats-infostats.STATCAN@canada.ca](mailto:STATCAN.infostats-infostats.STATCAN@canada.ca)

**Telephone**, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

### Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on [www.statcan.gc.ca](http://www.statcan.gc.ca) under "Contact us" > "[Standards of service to the public](#)."

## Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2019

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

*Cette publication est aussi disponible en français.*

---

## Table of contents

<b>Introduction</b> .....	<b>4</b>
<b>1 The New Housing Price Index (NHPI)</b> .....	<b>4</b>
1.1 Concepts and definitions.....	5
1.2 Data .....	5
1.3 Index calculation.....	5
1.3.1 Model replacement .....	6
<b>2 The New Condominium Apartment Price Index (NCAPI)</b> .....	<b>7</b>
2.1 Concepts and definitions.....	7
2.2 Data .....	7
2.2.1 Sampling .....	7
2.2.2 Cleaning and filtering .....	8
2.3 Index calculation.....	8
<b>3 The Resale Residential Property Price Index (RRPPI)</b> .....	<b>9</b>
3.1 Concepts and definitions.....	9
3.2 Data .....	9
3.2.1 Data sources .....	9
3.2.2 Collection delay.....	10
3.2.3 Cleaning and filtering .....	10
3.3 Index calculation.....	11
3.3.1 The repeat-sales method .....	11
3.3.2 The GRS and ARS indices .....	12
3.3.3 Worked example of the ARS index .....	14
3.3.4 Representativeness of the target population .....	16
3.3.5 Inverse-variance weights .....	16
3.4 Revision .....	17
3.4.1 Accounting for revision in the repeat sales model .....	17
3.4.2 Accounting for revision due to collection delay .....	17
<b>4 The Residential Property Price Index (RPPI)</b> .....	<b>18</b>
<b>References</b> .....	<b>19</b>

# Methodology of the Residential Property Price Index (RPPI)

by **Steve Martin**

## 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 is a quarterly index covering prices for new and resale residential housing in the Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria census metropolitan areas (CMA), as well as a 6 CMA composite, starting in quarter 1 2017. The index is a composite of three separate indices produced at Statistics Canada. New housing is covered by the New Housing Price Index (NHPI) and the New Condominium Apartment Price Index (NCAPI), and resale housing is covered by the Resale Residential Property Price Index (RRPPI). These three indices are aggregated together to form the RPPI.

This document outlines the methodological details behind the NHPI, the NCAPI, and the RRPPI, as well as how these three indices are aggregated to form the RPPI. Sections 1 and 2 cover the NHPI and NCAPI, which are standard survey-based price indices; Section 3 outlines the RRPPI, which uses a more complex repeat-sales methodology. As housing is a fairly heterogeneous good, constructing a price index with a constant-quality interpretation is an important methodological consideration for all three indices. Section 4 details how the NHPI, NCAPI, and RRPPI are aggregated to form the RPPI.

## 1 The New Housing Price Index (NHPI)

The NHPI measures the change over time in builders' selling prices of newly built houses (single/semi-detached and row) in 27 CMAs. For the purpose of constructing the RPPI, the NHPI covers new housing in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria. The NHPI is a monthly index, starting in January 1981. To produce a constant-quality price index, the NHPI uses a matched-model approach—wherein prices for the same house models are compared over time—along with explicit quality adjustments. Data are collected monthly from builders as part of a survey using an electronic questionnaire.

Unlike the NCAPI and the RRPPI, the NHPI is its own index series that is distinct from the RPPI.<sup>1</sup> The RPPI simply uses the city-level NHPI values to capture price changes for new housing. Consequently, this section focuses on the methodological details of the NHPI as they pertain to the RPPI.

---

1. See the [IMDB](#) for more information about the NHPI.

## 1.1 Concepts and definitions

Table 1.1 defines key concepts used for constructing the NHPI, at least for its use in the RPPI.

**Table 1.1**  
**Concepts and definitions for the NHPI**

Concept	Definition
Price	Either the transaction price or the list price for a model of a house as reported by the builder in a given month, exclusive of any sales tax. This is the price received by the builder, and excludes any additional fees paid by the buyer.
Model	The particular floor plan and features of a house.
Sample	See section 1.2.
Target population	All new residential houses (single/semi-detached and row) available for sale or sold in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria in a given month.
Index base period	The period for which the index equals 100. The base period for the NHPI is December 2016 = 100.

## 1.2 Data

Data for the NHPI are collected from a survey of home builders. The sampling frame for the NHPI is Statistics Canada's Building Permits Survey, and the survey uses a multi-stage sample design in which representative models are selected into the sample at each stage. The first stage of sampling involves contacting the top 15% of developers within a CMA, based on the value of their building permits, to determine if they are in scope for the survey. This helps ensure that large tract builders that develop an entire subdivision are included in the sample. Once a builder is identified as in scope, they select the development they are building with the most lots available for sale within a CMA, and up to three of the top selling house models in this development. This helps to ensure that the same models can be followed over time within the same development, and that these models are broadly representative of market activity for new housing.

An electronic questionnaire is used to collect price information for these models each month. If a model does not sell in a particular month, the builder is asked for a list price. The sample is periodically refreshed as developments sell out, and builders enter and exit the market. The data collected from developers are manually reviewed for consistency and completeness, and certain records may be edited or removed based on judgement.

## 1.3 Index calculation

The NHPI is a fairly straightforward matched-model index. Prices are stratified by CMA, builder, and model to produce a price relative for each model that each builder reports in the survey. The value of any promotions or upgrades is subtracted from the price of a model prior to calculating a price relative. Provided that house models do not change over time, this collection of price relatives has a constant-quality interpretation. The price relatives for each model are then aggregated to the CMA level using a Jevons index. Although the NHPI is calculated monthly, the three index values within a quarter are averaged to produce a quarterly index for the RPPI.

To make the index calculation explicit, let  $p_{mbt}$  be the price of model  $m$  by builder  $b$  at time  $t$ . These model prices are used to calculate a price relative between period  $t-1$  and period  $t$ ,  $p_{mbt} / p_{mbt-1}$ , for each model that each builder reports in the survey. To produce a CMA-level index, the price relatives for all models by all builders are aggregated with a Jevons index

$$I_t^{t-1} = \prod_{b=1}^{B_t} \prod_{m=1}^{M_{bt}} \left( \frac{p_{mbt}}{p_{mbt-1}} \right)^{1/\sum_{b=1}^{B_t} M_{bt}},$$

where  $M_{bt}$  is the number of models produced by builder  $b$ , and  $B_t$  is the number of builders. This index is then chained with the previous period's index value  $I_{t-1}$  to produce an index  $I_t = I_t^{t-1} \cdot I_{t-1}$  running from the base period to period  $t$ . Finally, the quarterly CMA-level index is simply the average of the three index values within that quarter. For the quarter starting in month  $q$ , the index is

$$I_q = \frac{1}{3} \sum_{t=q}^{q+2} I_t.$$

The resulting collection of quarterly indices at the CMA level capture the new house side of the RPPI.

### 1.3.1 Model replacement

When a house model is no longer for sale anymore, or no longer representative, and replaced by another model in the sample, a back price for the replacement model is imputed in the first period that it appears in the sample. This allows for a new model to be used in the matched-model index calculation immediately. The imputation is done with a linear regression (hedonic) model that relates house prices to observed characteristics (see de Haan and Diewert (2013, chapter 5) for more details). A separate model is calculated for each of the six cities. No imputation is made when a new model is added to the sample without replacing an old model, nor when a new builder is added to the sample.

Letting  $p_{mbt}$  be the price of model  $m$  by builder  $b$  in period  $t$ , the regression model is based on a structural model for house prices

$$\log(p_{mbt}) = \alpha + x_{mbt}\beta + z_{mbt}\gamma + d_b + d_t + \log(\epsilon_{mbt}),$$

where  $x_{mbt}$  is a (row) vector of model characteristics,  $z_{mbt}$  is a vector of location characteristics,  $d_b$  and  $d_t$  are builder and time specific intercepts, respectively, and  $\epsilon_{mbt}$  is an error term. Housing characteristics include the log of lot size and house size (square footage), and dummies for the number of garages, number of bathrooms, and number of bedrooms. Location characteristics include dummies for the property's forward sortation area (first three digits of the postal code). These characteristic data are collected from builders during the sampling process.

The regression model is estimated using a five year rolling window of data collected for the NHPI. Estimation is done with a robust M-estimator, using the bi-square loss function (see Amemiya (1985, section 2.3) or Wooldridge (2010, chapter 12) for more detail about M-estimation). Under the assumptions of the classical linear regression model, this approach to estimation is more robust to outlying price observations than the usual OLS estimator.

When a new house model is introduced into the sample, the characteristics for the new model and the characteristics for the old model are used to calculate a pair of fitted prices from the regression model. The fitted price for the new model is then subtracted from the fitted price for the old model, and this difference is added to the price for the old model to impute the back price for the new model. This effectively accounts for the difference between the characteristics of the old model and the new model, giving an imputation for what the price of the new model would have been in the previous period. That is, plugging the characteristics for a new

model  $n$  into the hedonic model produces a fitted price  $\widehat{\log(p_n)}$ , and plugging the characteristics of the old model  $o$  into the hedonic model produces a fitted price  $\widehat{\log(p_o)}$ . The difference between these fitted prices

$\widehat{\log(p_n)} - \widehat{\log(p_o)}$  is then added to the price for the old model  $\log(p_o)$  to produce a back price for the new

model  $\exp\left(\log(p_o) + \widehat{\log(p_n)} - \widehat{\log(p_o)}\right)$ . The imputed price relative for the new model is then simply

$$\frac{P_n}{\exp\left(\log(p_0) + \log(p_n) - \log(p_0)\right)},$$

and this is used directly in the index calculation.

## 2 The New Condominium Apartment Price Index (NCAPI)

The NCAPI measures changes over time in builders' selling prices of newly built, apartment-style units in condominium buildings in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria. This is a quarterly index, starting in quarter 1 2017, composed of 6 sub-indices (one for each city). Each sub-index is computed using a unit-value approach, wherein the price of a unit is standardized by its square-footage to give a price per square foot. Explicit quality adjustments are made prior to calculating these unit prices in order to produce a constant-quality index. Data for the NCAPI are collected monthly from a survey of builders using an electronic questionnaire.

### 2.1 Concepts and definitions

Table 2.1 defines key concepts used for constructing the NCAPI. An important aspect of the new condo market is that condo units often sell during the presale phase of a building, prior to construction beginning. Prices during the presale phase give an indicator of prices for new condo units, but may not reflect a transfer from the buyer to the seller if, for example, the builder is not able to sell enough units to finance construction of the building.

**Table 2.1**  
**Concepts and definitions for the NCAPI**

Concept	Definition
<b>Price</b>	Either the transaction price or the list price for a unit as reported by the builder in a given month, exclusive of any sales tax. This is the price received by the builder, and excludes any additional fees paid by the buyer.
<b>Unit value</b>	The price of a unit standardized by its square footage, giving a price per square foot.
<b>Unit type</b>	The number of bedrooms in an apartment, with or without a den, in one of the following categories: one bedroom, one bedroom+den, two bedroom, two bedroom+den and three bedroom.
<b>Presale</b>	The period in which units can be purchased prior to construction beginning.
<b>Sample</b>	See section 2.2.1.
<b>Target population</b>	All new residential low rise/high rise apartment condo units available for sale or sold in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria in a given month.
<b>Index base period</b>	The period for which the index equals 100. The base period for the NCAPI is 2017 = 100.

## 2.2 Data

### 2.2.1 Sampling

Data for the NCAPI are collected from a survey of condo builders. The frame is compiled from multiple sources including zoning and planning applications received from municipalities, building permits, builder associations, new home buyer insurance companies and governmental/non-profit home buyer protection services, advertisements, and various internet sources that provide information on upcoming buildings.

The NCAPI uses a multi-stage sample design in which units are selected into the sample at each stage. The first stage of sampling involves contacting developers in the survey frame to determine if they are in scope for the survey. To ensure that the same building can be followed through time, if a developer is in scope they are asked to report up to four buildings they are developing in which less than 70% of at least one of the target unit types have been sold. The second stage of sampling involves selecting one of these buildings into the sample. An electronic questionnaire is then used to collect price information from developers for up to three units of each type in a building each month. Developers also report any premia applied to a unit (e.g., the value of a parking spot, or a better orientation within the building), and are asked for a list price if no units of a particular type sold that month.

The same premium information is also collected for list prices. The sample is periodically refreshed as buildings sell out and builders enter and exit the market.

## 2.2.2 Cleaning and filtering

The data collected from developers are manually reviewed for consistency and completeness, and certain records may be edited or removed based on judgement. In addition to this manual cleaning, any price relatives (see section 2.3) greater than or equal to 3 absolute deviations from the median are not included in the index calculation. As NCAPI is based on average transaction/list prices, this is a standard filter to remove outliers that can exert a large influence on averages (e.g., Rousseeuw and Hubert, 2011). In order to adequately clean the data, the NCAPI has a one quarter revision. This is due in part to the small sample size in most months.

## 2.3 Index calculation

The index calculation for the NCAPI is fairly straightforward, and is similar in spirit to the NHPI. First, any premia are subtracted from the price of a unit to arrive at a quality-adjusted price for a “no-frills” reference unit. The quality adjusted price is then standardized by the square footage of a unit to arrive at a quality-adjusted unit price. Units are stratified by CMA, building, and unit type, and an unweighted geometric index is calculated for each stratum, giving a price relative for each stratum. The combination of stratification and explicit quality adjustment means that the same type of unit within each building is compared over time, giving these price relatives a constant-quality interpretation.<sup>2</sup> These stratum-specific price relatives are then aggregated to the CMA level using a Jevons index. The NCAPI is calculated monthly, and the three index values within a quarter are averaged to produce a quarterly index.

To make the index calculation explicit, let  $p_{usbt}$  be the price of unit  $u$  of type  $s$  in building  $b$  at time  $t$ , let

$\Delta_{usbt}$  the value of the premia for this unit, and let  $a_{usbt}$  be its square footage. The quality-adjusted unit price is calculated as

$$\rho_{usbt} = \frac{p_{usbt} - \Delta_{usbt}}{a_{usbt}}.$$

These unit prices are used in a geometric index to produce a collection of strata-level indices between period  $t-1$  and period  $t$ ,

$$I_{sbt}^{t-1} = \frac{\prod_{u=1}^{U_{sbt}} (\rho_{usbt})^{1/U_{sbt}}}{\prod_{u=1}^{U_{sbt-1}} (\rho_{usbt-1})^{1/U_{sbt-1}}},$$

where  $U_{sbt}$  is the number of units sold of type  $s$  in building  $b$  at time  $t$ . To produce a CMA-level index, the within-CMA relatives for each unit type in each building are aggregated with a Jevons index

$$I_t^{t-1} = \prod_{b=1}^{B_t} \prod_{s=1}^{S_{bt}} (I_{sbt}^{t-1})^{1/\sum_{b=1}^{B_t} S_{bt}},$$

where  $S_{bt}$  is the number of unit types in building  $b$  and  $B_t$  is the number of buildings. These period-over-period indices are chained with the previous period's index value to give the current-period index value

$$I_t = I_t^{t-1} \cdot I_{t-1},$$

2. Two notable characteristics for which no adjustment is made are number of bathrooms in a unit and outdoor living area (e.g., balcony), as builders are unable to explicitly price these characteristics.



where  $I_{t-1}$  is the index that runs from the base period to period  $t - 1$ . If a new building is introduced into the sample in a period, there is no attempt to impute back prices for the units in this building. This means that a building is not included in the index calculation in the first period that it is introduced into the sample.

Finally, the quarterly CMA-level index is simply the average of the three index values within that quarter. For the quarter starting in month  $q$ , the index is

$$I_q = \frac{1}{3} \sum_{t=q}^{q+2} I_t.$$

The resulting collection of quarterly indices at the CMA level capture the new condo side of the RPPI.

### 3 The Resale Residential Property Price Index (RRPPI)

The RRPPI measures the change in transaction prices over time for resale houses and condominium apartments in Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria. This is a quarterly index, starting in quarter 1 2017, composed of 12 sub-indices—one for each property type (house and condo) in each of the six cities. 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 data collection, ingestion, editing, and calculation is done in partnership with Teranet and National Bank.<sup>3</sup>

#### 3.1 Concepts and definitions

Table 3.1 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.

**Table 3.1**  
**Concepts and definitions for the RRPPI**

Concept	Definition
<b>Price</b>	Final transaction price at the closing date for the sale of a property and recorded in the provincial land registry.
<b>Sales date</b>	The closing date for the sale of a property.
<b>Sales pair</b>	Prices and sales dates for consecutive sales for the same physical property.
<b>Sample</b>	All residential 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.
<b>Target population</b>	All residential 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.
<b>Index base period</b>	The period for which the index equals 100. The base period for the RRPPI is 2017 = 100.

#### 3.2 Data

##### 3.2.1 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 residential single/semi-detached houses, row houses, and condominium apartments

3. The resale residential property price indexes (RRPPI) are produced jointly by Teranet Inc. (Teranet) and National Bank of Canada (NBC) based on a customized methodology developed in collaboration with Statistics Canada. Teranet and NBC have licensed the use of the RRPPI to Statistics Canada for publication and distribution.

in the Calgary, Montreal, Ottawa, Toronto, Vancouver, and Victoria CMAs. These data are collected and processed by Teranet and 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 3.2 gives a fictitious example of the resulting sales-pair data.

**Table 3.2**  
**Example of sales pair data**

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

### 3.2.2 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 and National Bank. This delay is particularly severe for British Columbia. Table 3.3 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 3.3**  
**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
	percent					
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

### 3.2.3 Cleaning and filtering

Data for the RRPPI come from administrative sources—and are therefore fairly clean—although 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 quality of the property may have changed (e.g., due to renovations). These filters are applied to each CMA and property type separately, and are summarized by the order in which they are applied in table 3.4.

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 (i.e., a bundled sale of multiple properties), 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.

**Table 3.4**  
**Data filters for sales pairs in the RRPPI**

Filter	Rationale
<b>Transaction price less than or equal to 10,000 dollars.</b>	These transactions may not be arm's-length transactions (e.g., bequest).  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).
<b>Holding period less than 6 months.</b>	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).
<b>Annualized return greater than or equal to 3 median absolute deviations from the median.</b>	

### 3.3 Index calculation

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.

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.

#### 3.3.1 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).<sup>4</sup> 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; 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. 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.

### 3.3.2 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  $t \in \{0, 1, \dots, T\}$  and properties be indexed by  $i \in \{1, 2, \dots, N\}$ , the starting point for the GRS index is a structural (hedonic) model of property prices

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

$p_{it}$  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  $t$ ),  $\theta$  is a vector of implicit (hedonic) prices, and  $\epsilon_{it}$  is an error term.<sup>5</sup> 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  $\tau$  with base period 0, denoted by  $I_\tau^G$ , is  $I_\tau^G \equiv P_\tau / P_0$ . Importantly,  $P_t$  is not random—it is a parameter that governs the joint distribution of property prices.

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

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

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  $E[\log(\epsilon_{it}) | D_{i1}, D_{i2}, \dots, D_{iT}] = 0$ —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

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

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

so that  $I_{\tau}^G \equiv P_{\tau} / P_0 = \exp(\gamma_{\tau})$ . 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.<sup>6</sup>

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  $\tau$ ,  $N_s(\tau)$  be the set of properties that sell for the second time in period  $\tau$ , and  $N(\tau) = |N_f(\tau)| + |N_s(\tau)|$  (the number of properties that sell in period  $\tau$ ), it can be shown that

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

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  $\tau$ , the index itself is used to extrapolate prices across time for all properties that sell in period  $\tau$  by deflating prices for sales that do not occur in the base period using that period's index. This allows all properties that sell in period  $\tau$  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  $I_{\tau}^A$ , 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

6. If property characteristics systematically change over time, then  $D_{it}$  is correlated with the error  $\log(\epsilon_{it(i)} / \epsilon_{r(i)})$ , although the direction of this correlation depends on whether there is net depreciation or net appreciation of properties over time.

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

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

$$(I_1^A, I_2^A, \dots, I_T^A)' = \text{diag} \left( \left[ \sum_{i=1}^N D_i' X_i \right]^{-1} \sum_{i=1}^N D_i' Y_i \right)^{-1}.$$

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).

### 3.3.3 Worked example of the ARS index

The simplest non-trivial example of a repeat-sales index has 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 3.5 summarizes these data.

**Table 3.5**  
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 is

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

and

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

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.<sup>7</sup> 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

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

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

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 (orthogonality) 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  $1 / \hat{\beta}_1$ , and thus

7. With a pure matched-model index, the period 1 index is simply  $p_{c1} / p_{c0}$  and the period 2 index is  $p_{b2} / p_{b0}$ .



$$I_1^A = \frac{(P_{a1} + P_{c1})(P_{a2} + P_{b2}) - P_{a1}P_{a2}}{P_{c0}(P_{a2} + P_{b2}) + P_{b0}P_{a2}}$$

and

$$I_2^A = \frac{(P_{a1} + P_{c1})(P_{a2} + P_{b2}) - P_{a1}P_{a2}}{P_{b0}(P_{a1} + P_{c1}) + P_{c0}P_{a1}}$$

Despite the conceptual simplicity of the ARS as a matched-model index, it nonetheless has a fairly complex non-linear structure.

### 3.3.4 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 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. This discrepancy between the target population and the sample will disappear over time.

### 3.3.5 Inverse-variance weights

Case and Shiller (1987) 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.<sup>8</sup> 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 price 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).<sup>9</sup> 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 undermine the usefulness of these estimators (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

8. Under the assumption of strict exogeneity,  $E(\log(P_{a(t)}/P_{a(0)})) | D_1, D_2, \dots, D_T)$  is linear, so that  $\text{var}(\log(P_{a(t)}/P_{a(0)}) | D_1, D_2, \dots, D_T) = \text{var}(\log(\epsilon_{a(t)}/\epsilon_{a(0)}) | D_1, D_2, \dots, D_T)$ .

9. 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.



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.<sup>10</sup>

### 3.4 Revision

#### 3.4.1 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).<sup>11</sup>

To fix notation, let  $I_0^S, \dots, I_T^S$  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  $I_0^T, \dots, I_{T+1}^T$ ; 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  $I_T^S \cdot I_{T+1}^T / I_T^T$ , and append this value to the original series. Thus, the original series of indices becomes

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

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. Provided that the historical index series is relatively stable over time, there should be minimal drift from splicing.

#### 3.4.2 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.

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

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

and the revised index is calculated as

10. 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).

11. 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.

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

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.

## 4 The Residential Property Price Index (RPPI)

The RPPI aggregates the CMA-level indices from the NHPI, NCAPI, and RRPPI to produce a price index for residential properties in Calgary, Montreal, Ottawa, Toronto, Vancouver, Victoria, and a 6 CMA composite. The target population for the RPPI is the union of the target populations for each of the three component indices. Each of the four indices (new house, new condo, resale house, resale condo) are aggregated with a Young index, with sales weights capturing the value share of new versus resale properties, and houses versus condo apartments, sold in each CMA. The RPPI is a quarterly index, as both the NCAPI and the RRPPI are quarterly, starting in quarter 1 2017. To keep in line with the NCAPI and RRPPI, the RPPI has a one quarter revision.

The weights for the RPPI are derived from the Canada Mortgage and Housing Corporation's Market Absorption Survey and the inventory of repeat-sales transactions from Teranet and National Bank.<sup>12</sup> Both of these sources capture the value of all new and repeat-sales transactions respectively for residential single/semi-detached houses, row houses, and low rise/high rise apartment condos; consequently, the aggregate values are comparable in order to produce a value share for new versus resale properties, as well as houses versus condo apartments. The weight reference period is the three calendar years prior to the current year of the index, and these weights are updated annually. To avoid overlap with the revision period, the weights are updated in quarter 2 of the year.

12. To help reduce the discrepancy between the target population and the sample, repeat-sales from 1990 to the current period are used to calculate the weights. Although data from 1990 to 1998 are not suitable for calculating an index in all CMAs, this has no impact on the weights as they only use transactions from 2014 onwards. The benefit of this longer window is that properties that sold between 1990 and 1998 and again in the weight reference period are part of the target population and are now used to calculate the weights, whereas using only the repeat-sales data from 1998 onwards omits these properties.

## 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.
- Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.
- 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](http://www.ofheo.gov/Media/Archive/house/hpi_tech.pdf). Office of Federal Housing Enterprise Oversight. Retrieved from [http://www.ofheo.gov/Media/Archive/house/hpi\\_tech.pdf](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.
- 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.
- 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](https://www.imf.org/external/pubs/ft/sdds/guide/plus/2015/sddsplus15.pdf). 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(1): 110-126.
- S&P Dow Jones. (April 2018). [S&P CoreLogic Case-Shiller Home Price Indices Methodology](https://us.spindices.com/index-family/real-estate/sp-corelogic-case-shiller). 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* (2<sup>nd</sup> edition). MIT University Press.