CoreLogic

CoreLogic Australia Residential Property Index Series
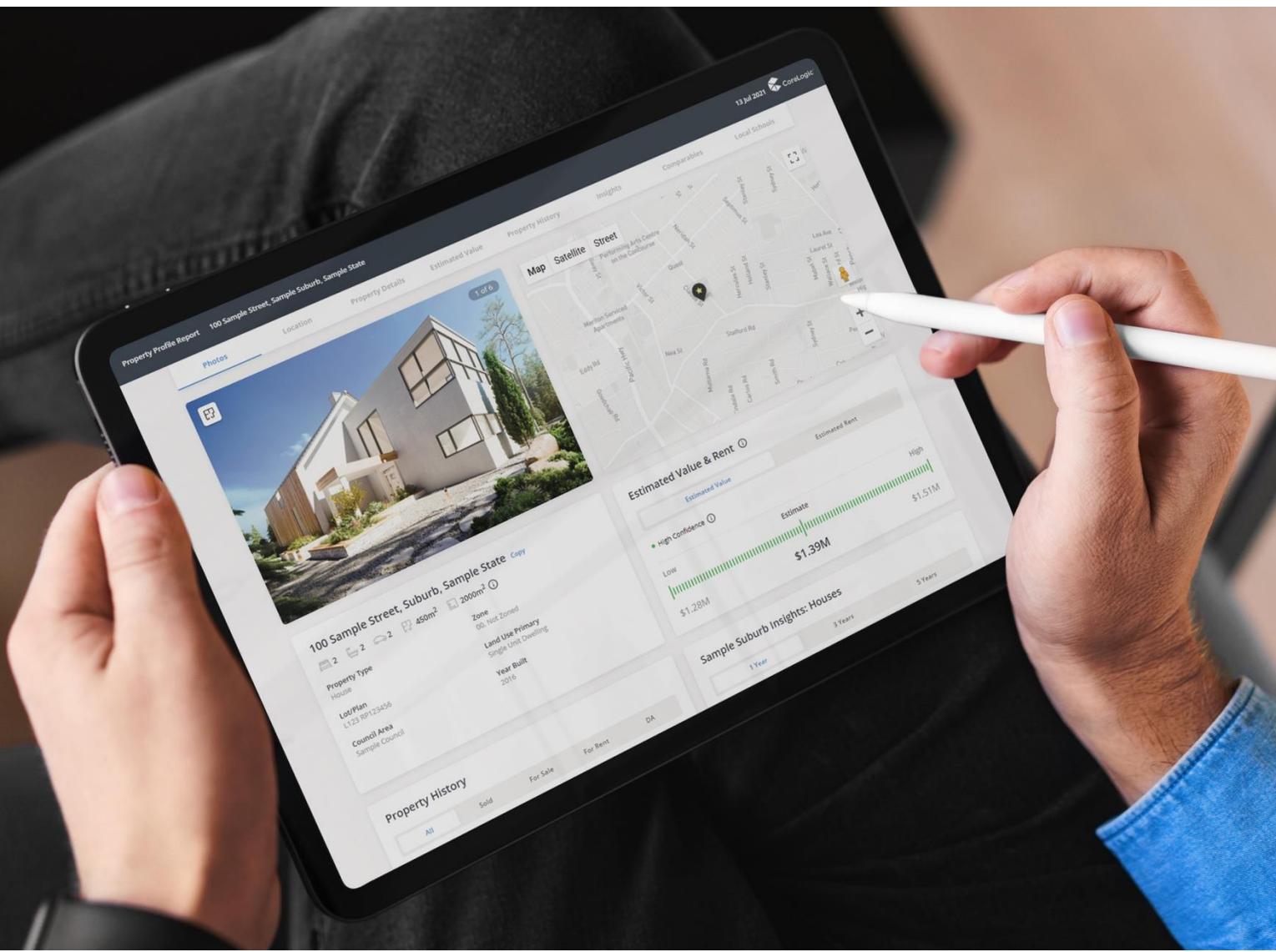
OCTOBER 2023

Table of Contents

Overview

The CoreLogic Australia Residential Property Index Series is designed to provide a reliable and consistent set of benchmarks for the Australian residential real estate market. The market includes all properties that are defined as residential units (apartments/townhomes etc) and houses.

The hedonic imputation approach (based on hedonic regression) to calculate residential property value indices is the preferred method for compiling a constant quality dwelling value index in Australia. Hedonic regression is a statistical technique that measures the relationship between values of residential real estate and the observed values of its characteristics, for example attributes that encompass its geographic location and property features, such as number of bedrooms and bathrooms etc. [1]. The hedonic imputation index model essentially estimates the value of every property each day, using the coefficients derived using hedonic regression, and describes the daily movements in the value of the residential real estate portfolio.

The hedonic imputation model operated by CoreLogic utilises comprehensive information on the attributes and characteristics of residential properties to measure quality adjusted changes in property values over time and also to impute the value of all dwellings based on observed characteristics. The hedonic imputation method has the advantage of controlling for compositional bias that may exist in other indices (for example in a median sales index) [2]. The hedonic index model is trained using an exponential weighting scheme with a 360 day decay period. Weighting the regression allows the model to impute more timely valuation estimates. The index is further enhanced by revising the most recent 12 months of growth in a rolling window as new sales and listings data becomes available. This improves the historical accuracy of the index back series and reduces index drift.

In addition to the hedonic index series a set of additional indices are produced, including stratified median and repeat sales indices. Accompanying metrics such as rolling mean and median of various periods are also available.

In order to establish a summary of best practices for residential property price index calculation, the Statistical Office of the European Union (Eurostat) coordinated in 2013 the development of the Residential Property Price Indices Handbook [1]. The handbook proposes a comprehensive presentation and analysis of the methodologies to date and has been widely used as a reference document by financial institutions and legislators [3, 4, 5]. While more recent reviews expand on index construction [6], they do not revolutionise best practices and so the Eurostat paper remains the most comprehensive overview of index methodologies. CoreLogic Australia has adhered to its guidelines when implementing indices methods appropriate to the Australian market while introducing improvements wherever possible.

Index Series

The CoreLogic index series are produced at a variety of geographical regions within Australia. The regions have been defined using the Australian Statistical Geography Standard (ASGS) 2021 standard [7]. Each index can be subcategorised according to any region and property attribute combination and additional subindices can be generated on request.

A broad suite of indices is produced for listings, sales, and rents using different methodologies such as simple tiered median, repeat observation and hedonic imputation. A summary of the indices that are produced are shown in Table 1.

CoreLogic

Table 1. Index Series

Index Method	Index Name	Dataset			
		Sales	Rents	Listings	Properties*
Hedonic	Hedonic Index (sales)				✓
	Hedonic Index (asking price)				✓
	Hedonic Index (rents)				✓
	Tiered High/Mid/Low Hedonic Index (sales)				✓
	Rental Yield (rent-to-sales)				✓
	Rental Yield (rent-to-asking price)				✓
	Total Return (sales plus rents)				✓
	Deciles Hedonic Index (sales)				✓
Stratified Median	Stratified Median Index	✓	✓	✓	
Repeat Observation	Repeat Sales Index	✓			
	Repeat Rents Index		✓		
Median	1 Month Median Index	✓	✓	✓	
	3 Month Rolling Median Index	✓✓	✓	✓	
	6 Month Rolling Median Index	✓✓	✓	✓	
	12 Month Rolling Median Index	✓✓	✓	✓	
Mean	1 Month Mean Index	✓	✓	✓	
	3 Month Rolling Mean Index	✓✓	✓	✓	
	6 Month Rolling Mean Index	✓✓	✓	✓	
	12 Month Rolling Mean Index	✓✓	✓	✓	

* Properties here meaning that the indices are based on value imputations of the properties

Index Maintenance

Daily Valuation

Residential property value indices and accumulation indices based on the hedonic imputation method are calculated daily. The property value indices measure the daily movement in imputed values. The accumulation index series measures the daily movements in residential property values with the full reinvestment of rental income (which is also calculated based on the hedonic imputation method). Daily results are never directly revised, however when the month end results are revised, daily results are normalised to keep the daily and monthly indices inline. This process involves distributing the delta between the original month end result and the revised result evenly across each day of the month. In this way intra-month trends are broadly preserved while maintaining uniformity of all reported indices.

Selection Rules

All residential properties defined as a house or unit and that are known to exist on the index date will be eligible to be included in the index from the initial month after the property was added to the CoreLogic database. From CoreLogic's perspective, a house is any property that is on a Torrens Title, where the title holder claims ownership upon the land which the property resides on. This includes property types such as semidetached dwellings, terraces and duplexes. A unit is any property that is on a strata title, where the title holders own a shared claim to common land that multiple properties may reside on. This includes property types such as villas and townhouses. Both of these only refer to properties that have been deemed by regulatory bodies to be for residential purposes only. Properties that no longer meet eligibility criteria will be removed from the index once the record is removed, which may occur if it is known that a property has been demolished.

CoreLogic

Base Date

All of CoreLogic's Hedonic Indices have a baseline value of 100 as at 31/12/2009. The starting date for all hedonic indices is nominally set at 31/01/1980, however this will vary greatly for different series and geographies based on the availability of data.

Index Rebalancing

The portfolio is rebalanced monthly to account for the addition or removal of stock from the residential property portfolio over the preceding month. Changes in property attributes (for example a change in the recorded number of bedrooms for a property) are handled by fixing attributes as those most recently observed at that index date when calculating growth between two periods. This ensures that value estimates imputed at each time are treated as having the same attributes, which eliminates bias in growth estimates due to changes in the underlying properties.

In cases where a property cannot be valued using the hedonic imputation method, for example if there are no relevant observed sales during the period, it will be removed from the index until a value can be imputed from the arrival of new sales data.

Eligibility Factors

CoreLogic maintains comprehensive property and attributes datasets, which are populated via sales and listings timeseries data. The majority of sales are reported through an agents' advice pipeline, which in the 12 months to August 2023 recorded 68.3% of sales prior to official notification of sales by the relevant Valuer General. In combination, CoreLogic is able to support an accurate property dataset backed up by timely and comprehensive attributes, sales, and listings histories. Only due to these rich datasets is CoreLogic able to use hedonic regression, a data intensive method, to produce a compositionally adjusted hedonic index that runs in real time. From within these datasets, all Australian residential properties that are defined as a house or unit are eligible to be included in the index.

The entire known residential property (classified as units and houses) dataset is used as the population for the index model for a given index run date. Similarly, all sales and listings are used. Appropriate filters are in place for incomplete records, missing values and extreme values.

A sale record is excluded if the transfer code indicates that the sale was a non-arm's length transaction, for example a transfer due to a death, divorce, bankruptcy etc.

In addition to transfer exclusions, the land use code at the time of sale is used to determine eligibility for inclusion. The land use code is provided by the Valuer General at the time of sale.

Data Assumptions and Limitations

There are a number of limitations with the data available for the production of a hedonic index

1. It can be reasonably expected that data quality issues will be present in the property databases as the data is derived from a variety of sources and in different formats.
2. Whilst sophisticated techniques exist to ensure that addresses can be matched across multiple data sources, it is recognised that errors can arise.
3. Latitude and longitude are required for matching properties to a statistical region. Latitude and Longitude are present for approximately 98% of the properties and the coverage is continually improving due to data quality initiatives.
4. There is no completely reliable point-in-time property type identifier and therefore there may be instances where it is not possible to tell when a property identifier changed, for example when a house was built on vacant land. This is sometimes possible through the known sale data obtained from the Valuer General but is not always reliable.

5. It is expected that the population of properties and their attributes can change from one day to the next, either due to data quality improvements or due to the arrival of new data on sales and listings.

Index Revision

Despite the timeliness of the datasets, it is inevitable that transactions will be ingested after the monthly index growth rates are published. As a result, the index can be retrospectively improved by recalculating index growth once coverage of transactions has increased. To capture this increased breadth of historical data, the last 12 months of the index will be routinely revised. This ensures that the index can be reported in real time while maintaining the most accurate possible back series. A revisionary index process also means that data issues can be addressed in the underlying dataset and erased from the index as they are identified.

Index Construction

The CoreLogic hedonic index is specifically designed to track daily value changes in the residential property market [2]. It is a constant quality value index for the stock of residential housing at a particular time.

Repeat observation and stratified median methodologies are also available as part of the indices suite. Further comparable metrics are also available, such as several simple median and mean metrics which complement the more complex methodologies.

The steps taken in calculating the hedonic index are:

1. Obtain historical sales over the preceding 360 days;
2. Apply filters and obtain attributes and variables to estimate value function;
3. Estimate value function;
4. Apply value function estimate to population of properties in the valuation portfolio;
5. Identify the composition adjusted portfolio;
6. Calculate the index.

The hedonic regression coefficients are calculated at the Statistical Area 3 level, defined by the Australian Bureau of Statistics, although all other index metrics are produced at all levels of geography.

Data Cleaning and Standardisation

It is often the case that historical sales data have missing values for some of the hedonic attributes, especially for earlier periods. In these cases, the earliest known non-missing variable is used.

There have been cases where a large number of sales have been recorded where the contract date, sale amount and Statistical Area 1 are the same. The best example of where these 'multi-sales' typically occur is where an apartment block is sold as one transaction and the observed sale price of the entire block is attributed to each apartment. Multi-sales are infrequent however, it is prudent to remove any unusual cases of multi-unit sales that have not been identified correctly. Any instances where 5 or more sales have occurred with the same contract price and date within the same Statistical Area 1 are removed.

Extreme values may be present in some cases for property attributes, sales or rental values. These cases may represent instances of incorrect data entry or invalid sales, particularly if the event related to a non-arm's length transaction, for example transmission of ownership by death, and the event has not been allocated the correct transfer code. These transfers do not represent market value.

Observed sales that fall outside thresholds determined through expert assessment are also removed (for example less than $1,000 and greater than $100,000,000 for sales).

In the case of sales or listings the following steps are performed:

1. Partition the sales or listings time series by Statistical Area 2 region (or Statistical Area 3 if there are not enough sales at the lower region) [7];
2. Create a date schedule of 30 day time periods ending on the current index calculation date. Note that the first period is likely to be shorter than 30 days and this represents a 'short first' convention;
3. Discard observed sales or listing prices that are identified as extreme values within each 30 day period.

In the case of property attributes the following steps are performed:

1. Partition the properties data set by Statistical Area 3 region and property type [7];
2. Winsorize the data at the 5th and 95th percentiles to reduce the effect of extreme outliers.

Every record is processed programmatically to ensure that missing or extreme variables are managed appropriately. In some cases, records are removed if there are null values for some variables, however for others the variables are imputed.

Outlier Detection

Outliers are detected dynamically on every index run date in order to exclude them from the regression functions. A daily outlier detection routine was chosen so as to ensure that the index did not move suddenly as a result of a change in the distribution of properties due to less frequent outlier filtering.

The outlier detection method is consistent for transfers, sales listings, and rental listings and is determined at the largest Statistical Area below that at which the hedonic regression is trained. If there is sufficient data, the hedonic model is trained at the Statistical Area 3 level, and therefore outlier detection is performed at the Statistical Area 2 level. In the case where there is insufficient data the regression model switches to run at the Statistical Area 4 level, and outlier detection therefore increases to run per Statistical area 3 as well. The outlier detection process is split into two stages (true outliers and price filters) because it was found during testing that outlier detection and price filtering methods are affected by the outliers themselves. The outlier detection stage is described in this section and the price filtering stage is described in the following section.

The outlier detection stage is designed to remove records which are obvious errors in the data (true outliers). The time series is split into periods of 30 days. For the start of each 30 day period, all records are selected from the period of 30 days before and after. If there are less than the required number of records the window is extended by 30 days on either side until there is sufficient data. The threshold is 50 when running for Statistical Area 2, and 100 when running for Statistical Area 3. Additionally, a constant minimum window length of 30 days was used in order to identify monthly outlier levels as much of the indices reporting is based on monthly data. The 2.5th and 97.5th percentiles are identified and any records that are less than 80% of the 2.5th percentile or more than 120% the 97.5th percentile are discarded. Any observations that are less than $1,000 or greater than $100,000,000 are also discarded. This step is required so as not to skew the results of the continuous Gaussian kernel smoother in the subsequent price filtering stage.

Price Filtering

The price filtering stage fits the upper and lower price filter bands around the time series. The time series is again split into periods of 30 days. For the start of each period, all records (which now exclude outliers) are selected from the period of 30 days before and after. If there are less than 50 or 100 records for Statistical Area 2 or Statistical Area 3 respectively, the window is extended by 30 days on either side until there are sufficient records. The 2.5th and 97.5th percentiles are identified. It is worth highlighting that for regions with low sales volumes the same sale could be included in the estimation for the price filtering levels for consecutive sales periods. If less than 50/100 sales are identified in total, no price filter level can be detected for that region. As with outlier detection, this process operates at the Statistical Area level immediately lower than that which is used for hedonic regression training.

A continuous Gaussian kernel smoother is generated and fitted to the 2.5th and 97.5th percentiles. In an effort to ensure that a smooth price filter curve was generated by the Gaussian kernel function regardless of transaction volumes, a bandwidth of 50 was chosen. Through empirical testing, it was found that this was the optimum

bandwidth parameter in order to prevent too many observations from being excluded from the hedonic regression.

Finally, any observations that are outside of the upper and lower bounds specified by the fitted curve are excluded by the regression functions.

Hedonic Imputation Index

There is no uniformity in the application of hedonic regression and it can suffer from omitted variable bias if an important hedonic attribute is missing. As long as there are sufficient data available for all hedonic attributes, the hedonic imputation method is generally the best technique for constructing a constant quality residential property value index.

The hedonic index requires that the actual prices of all properties are known in order to calculate the movement in values. However, in each period not all properties are sold, therefore their values must be imputed based on sales that were observed for a given Statistical Area, property type and period. The hedonic imputation method is used to estimate the values (or listed values in the case of sales listing and rental indices) of all properties for each period. If the value of property i cannot be imputed, for example if there aren't enough sales in the Statistical Area for that period, the value of property i is imputed from the sample based on the higher level Statistical Area wherever possible. In such a case, transaction data from the Statistical Area 4 would be used to impute the values of properties that could not be established using only Statistical area 3 transaction data.

The hedonic regression formula is shown in Equation (1). It gives the least squares estimate of the log price (or asking rental price in the case of the rental index) of a property at a particular time conditional on the information available and controlling for its most statistically significant observable attributes [8]. Listings data are used in the calculation of equivalent rental and accumulation indices, but not in sale price indices. In the case of the rental index, sales are substituted by rental asking prices found in the listings data. The hedonic variables and their transformations are shown in Table 2.

$$\ln P_i^t = C_0^t + \sum_{k=1}^{K} C_k^t X_{ik}^t + \varepsilon_i^t$$

where:

P_i^t = sale price (or listing price) of property i $(i = 1\ to\ N^t)$ in period t

X_{ik}^t = transformation of the hedonic variable k for property i in period t

$C_{0\ to\ k}^t$ = numerical coefficients to be estimated

ε_i^t = (zero mean) residual error term for property i in period t

The calculation of the hedonic index value is shown in Equation (2).The hedonic index is market value weighted and adjusts as new properties enter or exit the portfolio, reflecting when properties were constructed, demolished or added to the residential database. In the case of a daily index, the rebalancing is performed daily to reflect changes in the level of property stock. All properties are revalued for the previous index date. This technique ensures that the population of properties (and their attributes) of properties are consistent between two consecutive index dates and that the effect of time only is captured in the index. When valuing both the current and prior month, the attributes are fixed as those observed for the more recent month; this formulation is known as a Paasche index. This index is revised every month for the 12 months following release in a routine manner.

$$I_{IMP}^t = I_{IMP}^{\acute{t}} \times \frac{\sum_{i=1}^{N^t} \widehat{P}_i^t}{\sum_{i=1}^{N^t} \widehat{P}_i^{\acute{t}}}$$

where:

$I_{IMP}^t =$ sale price (or listing price) imputation index at period t

$I_{IMP}^{\acute{t}} =$ sale price (or listing price) imputation index at the index run date prior to period t

$\widehat{P}_i^t =$ imputed price (or listing price) of property i at period t

$\widehat{P}_i^{\acute{t}} =$ imputed price (or listing price) of property i at the index run date prior to period t

$N^t =$ number of properties at period t with valuations possible for period t and the prior period

CoreLogic

Table 2. Hedonic variables and their transformations		
Hedonic Variable	Description	Transformation
Statistical Area 2	The Statistical Area 2 that each property belongs to.	This variable is not transformed.
Land area	Total size of the parcel in square metres.	Transformed to a log value. The land size is reserved for houses, not units. It is winsorized at the 5th and 95th percentile for land area in the property's Statistical Area 3 region.
Street classification	Highway, main road, etc..	This variable is transformed as shown in the first column in Table 3.
Property type	Full category for properties, such as house, unit with additional information such as multi storey, duplex, one storey/lowset etc.. The primary land use code which is supplied by the Valuer General's office indicates what usage the government permits for the land (for example, residential, commercial etc.). There are currently around 940 land use types across Australia and 19 residential property types are derived from them.	This variable is not transformed.
Bedrooms	The most recent recorded bedroom count at the time of the recorded sale or listing.	This variable is winsorized at the 5th and 95th percentiles for bedrooms in the property's Statistical Area 3 region and is represented as a factor.
Bathrooms	The most recent recorded bathroom count (inclusive of ensuites) at the time of the recorded sale or listing.	This variable is winsorized at the 5th and 95th percentiles for bathrooms (including ensuites) in the property's Statistical Area 3 region and is represented as a factor.
Bedrooms/ Bathrooms	The ratio of the number of bedrooms to bathrooms.	This variable is binary and indicates whether there are more bathrooms than bedrooms after winsorization.
Car spaces	The most recent recorded car space count at the time of the recorded sale or listing, or the number of garages if the number of car spaces is zero.	This variable is winsorized at the 5th and 95th percentiles for car spaces in the property's Statistical Area 3 region and is represented as a factor.
Pool	A swimming pool is present.	This is either Y or N.
Air conditioning	The property has air conditioning.	This is either Y or N.
View	Property has a scenic view flag.	This is either Y or N.

Recently updated	A flag to indicate that the property has been built or refurbished within the previous 10 years.	This is either Y or N.
Density	A metric which describes the density of the property's mesh block (as defined by the ASGS 2021 standard [7]). It is measured in the number of properties per square kilometre and transformed to a log value.	The density is the total number of properties in the property's mesh block region divided by the total area of the mesh block (in square kilometres), transformed to a log value. The mesh blocks are defined by the Australian Bureau of Statistics [7]. (1 is added to the number of properties to safeguard against any errors arising from zero observed properties in a mesh block.)
Spatial Dependency	A variable that captures the spatial dependency of the property.	The spatial dependency is represented by a 2-D cubic spline function of the X and Y Cartesian coordinates of the property [8]; a thin plate regression spline of the coordinates in the GAM model is used [9].
Time Fraction	Period fraction indicating when the property was sold (or listed in the case of the listing or rent indices) within period t.	The period fraction is shown in the following formula: $$F_i^t = \frac{date(P_i^t) - date(t_{start})}{date(t_{end}) - date(t_{start})}$$ $date(P_i^t)$ = date of sale (or listing) of property i in period t $date(t_{start})$ = start date of period t $date(t_{end})$ = end date of period t
Prior Price	Prior sale (or listing) price for the property.	Transformed to a log value. When the prior price is missing then this field is imputed as zero.

| Prior Price Days | Number of days since the previous sale. | This variable is not transformed and is calculated as:

$$days(P_i^t) = date(P_i^t) - date\_prior(P_i^t)$$

$date(P_i^t)$ = date of sale (or listing) of property i in period t

$date\_prior(P_i^t)$ = previous date of sale (or listing) of property i in period t

(An interaction is specified between Prior Price and Prior Price Days). |
| Prior Price Missing | A dummy variable that indicates whether there is an observed prior sale (or listing) for the property. | This variable is not transformed. |

Table 3. Street classifications

		State / Territory								
		ACT	NSW	NT	OT	QLD	SA	TAS	VIC	WA
A	VEHICLE TRACK	✓	✓	✓	✓	✓	✓		✓	✓
A	ACCESS ROAD	✓	✓	✓	✓	✓	✓	✓	✓	✓
A	PROPOSED ROAD		✓	✓		✓	✓	✓	✓	✓
B	HIGHWAY	✓	✓	✓	✓	✓	✓	✓	✓	✓
B	RESTRICTED ACCESS ROAD		✓	✓	✓	✓	✓	✓	✓	✓
C	FREEWAY	✓	✓			✓	✓		✓	✓
C	UNKNOWN	✓	✓	✓	✓	✓	✓	✓	✓	✓
D	CONNECTOR ROAD TUNNEL	✓	✓							
D	LOCAL ROAD	✓	✓	✓	✓	✓	✓	✓	✓	✓
D	SUB-ARTERIAL ROAD	✓	✓	✓		✓	✓	✓	✓	
E	ARTERIAL ROAD	✓	✓	✓		✓	✓	✓	✓	✓
F	CONNECTOR ROAD	✓	✓	✓		✓	✓	✓	✓	✓
F	PATH	✓	✓		✓	✓	✓	✓	✓	✓
F	PRIVATE ROAD	✓	✓	✓	✓	✓	✓	✓	✓	✓
OTHER	Not yet observed based on transfers									

Accumulation Index

The accumulation index assumes that daily rental income is reinvested into the hedonic index daily [10]. Transaction, taxation and ownership costs associated with purchasing and owning residential property are not considered (along with individual tax concessions such as negative gearing). A gross rental yield is calculated. It is assumed that all properties are rented; therefore, the weekly rental income can be estimated from the

hedonic imputation rental model which is based on advertised weekly rents. If the rent of property i cannot be imputed using the hedonic imputation model, for example if there are no listings in period t, the rent of property i is imputed from the hedonic index for the region above. The imputed weekly rents are converted to daily values and reinvested into the daily property value index (assuming an actual/365 day count convention). In line with the hedonic index it is necessary to ensure that the population of properties (and their attributes) is consistent between consecutive index dates. The accumulation index is shown in Equation (3).

$$I_{TR}^t = I_{TR}^{\acute{t}} \times \sum_{i=1}^{N^t} \frac{\widehat{P}_i^t + \widehat{R}_i^t}{\widehat{P}_i^{\acute{t}}}$$

where:

$I_{TR}^t =$ sale price total return index at period t

$I_{TR}^{\acute{t}} =$ sale price total return index at the index run date prior to period t

$\widehat{P}_i^t =$ imputed price (or listing price) of property i at period t

$\widehat{P}_i^{\acute{t}} =$ imputed price (or listing price) of property i at the index run date prior to period t

$\widehat{R}_i^t =$ imputed daily rent of property i at period t

$N^t =$ number of properties at period t with valuations possible for period t and the prior period

Repeat Sales Index

In line with the hedonic indices described previously, the purpose of the repeat sales index is to measure the average change in home price in a particular statistical region. The repeat sales method was originally introduced in 1963 by Bailey et al., and there exist a different number of implementations of the methodology, with the S&P CoreLogic Case-Shiller Home Price Index being the best known.

The main variable used for index calculation is the price change (sale pair) which occurred between two arms-length transactions of the same property [11]. Thus, the repeat sales method aims to capture pure price changes across different time periods, but does not control for a change in quality (which can account for a significant part of the variability between two sales of the same property) [12]. A dummy variable regression model is estimated on the pooled data (sales pairs) across the sample period [1]. So, the estimated growth rate in any period is determined not only by the prices of properties sold in that period, but also by the prices of properties for which consecutive transactions occurred on either side of that period [12]. By design, the method is revisionary and the most recent periods are subject to larger revisions, as the properties which changed hands in a period constitute a relatively small proportion of the total number of sale prices that will eventually affect the index in that period.

If we have a sample of properties which transacted more than once over the sample period t = 0, 1, ..., T (with 0 ≤ s < t ≤ T), the sales pairs are pooled together and the model is estimated with the standard repeat sales Equation (4).

$$\ln(P_n^t / P_n^s) = \sum_{t=0}^{T} \gamma^t D_n^t + \mu_n^t$$

where:

P_n^t = Price of property *n* at the end of period *t*

P_n^s = Price of property *n* at the end of period *s*

D_n^t = Dummy variable equals to 1 in the period that the second sale occurs, -1 in the period that the first sale occurs, and 0 otherwise

γ^t = Coefficient for period t

μ_n^t = error term for property *n* during period *t*

Under the classical assumptions that the errors have a zero mean and constant variance, the repeat sales formula shown in Equation (4) can be estimated by the OLS regression. The coefficient for each period can be taken to represent the cumulative growth in property prices since the first period. Therefore, the repeat sales index from period 0 to period t can be obtained by exponentiation of the coefficients γ^t as in Equation (5).

$$P_{RS}^{0t} = exp(\gamma^t)$$

Stratified Median Index

Stratified median price series are also implemented as part of the CoreLogic Australia Residential Property Index Series. Stratification is a process for creating subsets of dwellings which are qualitatively similar. Unique price series are created for these subsets which are then aggregated to estimate strata adjusted price movements in the overall market. The strata definitions used to classify properties into subsets are based on price, geography and interactions of these variables.

The stratified median for units and houses groups statistical regions by their long-term median transaction price within a higher level statistical region. Sales transfers are grouped by property type, Statistical Area 2 (defined by the Australian Bureau of Statistics [7]) and five year median price within each transfer's Statistical Area 4 statistical region [7]. The Statistical Area 2 regions are ranked and grouped by deciles of equal numbers (the first decile may contain more Statistical Area 2 regions if equal number of Statistical Area 2 regions cannot be determined for each decile). The stratified median for a particular region (Statistical Area 4 level or above) is the weighted mean of the log return of each strata within that region (weighted by volume of sales).

Stratified medians are simple to calculate and provide the price of the middle ranking property observed to sell in a given measurement period. Further, rather than selecting the 50th percentile (i.e., the median) from whatever properties are observed to sell over a particular measurement period, other percentiles such as the 25th or 75th percentile can equally be selected using the same method so as to create other percentile series.

The major shortcoming of median (or any other selected percentile) price series is that they do not represent changes in the value of the residential property market portfolio. Median prices do not track the same properties through time. In fact, it is most unlikely that any of the properties observed selling in one period will be the same as the properties selling in the next measurement period when examining quarterly or monthly median price series. By way of example, in one period the median priced property observed may happen to be a three bedroom home and in the next period it may be a four bedroom home. Further, in order to provide useful statistics, one needs to aggregate a sufficiently large number of property sales over what can be long periods of

time. As such, the publication frequency of median price series is often limited to quarterly, although CoreLogic do offer a monthly median series (in addition to many other frequencies).

The fact that median or other percentile based series cannot be used to track changes in value of a market portfolio does not make them wrong: it is simply that they have different applications than hedonic indices. For example, median price series are useful in answering economic policy questions relating to housing affordability [13].

Index Governance

Index audit

In 2017 the consultancy KPMG was engaged to audit the hedonic index. The purpose of the collaboration was to determine that the implementation of the index was in line with the methodology CoreLogic Australia had retained for the daily index.

In addition to the audit of the implementation, the model was subject to an academic review. The purpose of such an academic review was twofold:

i.   Ensure that the new hedonic imputation model was in line with leading research and industry best practices;
ii.   Establish that the variables (features) used in the model were appropriate as intended in the methodology.

These audits were conducted prior to both weighting the hedonic regression, and revising the hedonic index in a rolling 12 month window.

Index Announcements

Announcements with respect to changes in the indices models will be communicated periodically in line with continuous improvement initiatives.

Recalculation Policy

Beyond the standard 12 months of revision, there may be occasions where the indices need to be recalculated. This may occur when material changes to the historic data set have been implemented or where errors are detected in the base level data or outputs. Any errors that arise will be corrected immediately and the changes reflected in the relevant index. Any significant recalculations will be alerted within the Index Announcements.

Index Dissemination

CoreLogic indices are available via subscription and a number of portals. Please contact CoreLogic for further information.

CoreLogic Model Governance

CoreLogic has a formal governance program in place for all analytic models. The objective of this program is to (1) establish model risk management policy and governance program over all CoreLogic models; (2) provide guidance to help CoreLogic business units and mitigate risks arising from customer reliance on computer-based models; and (3) define key model development, testing, validation and monitoring processes and best practices The program is modelled on published Office of the Comptroller of the Currency bulletins, including 'Supervisory Guidance on Model Risk Management" (OCC 2011-12) and "Third-Part Relationships" (OCC 2013-29).

CoreLogic

The hedonic imputation index method has been reviewed by the CoreLogic Model Governance committee and approved for client use. This committee maintains oversight on all models approved for client use and requires regular reviews of these models to ensure continued compliance.

References

[1]   Eurostat, "Handbook on Residential Property Price Indices (RPPIs)," Eurostat Methdologies & Working Papers, Luxembourg, 2013.

[2]   CoreLogic, "CoreLogic RP Data Daily Index," 12 2011. [Online]. Available: https://www.corelogic.com.au/our-data/corelogic-indices#daily-indices. [Accessed 28 07 2023].

[3]   F. S. B. International Monetary Fund, "The Financial Crisis and Information Gap," September 2016. [Online]. Available: https://www.imf.org/external/np/g20/pdf/2016/090216.pdf.

[4]   B. F. I. Settlements, "Analysis of residential property price developments," February 2017. [Online]. Available: http://www.bis.org/statistics/pp_analysis.htm.

[5]   T. E. P. a. Council, "Regulation on harmonised indices of consumer prices and the house price index," May 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0792.

[6]   R. J. Hill and A. N. Rambaldi, "Hedonic Models and House Price Index Numbers," Advances in Economic Measurement, pp. 413-444, 2022.

[7]   Australian Bureau of Statistics, "Australian Statistical Geography Standard (ASGS)," [Online]. Available: https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/latest-release. [Accessed 25 07 2023].

[8]   M. Hardman, "Calculating High Frequency Australian Residential Property Price Indices," 12 2011. [Online]. Available: http://www.corelogic.com.au/resources/pdf/rismark/technical_method_paper.pdf. [Accessed 07 12 2015].

[9]   S. Wood, "Package 'mgcv'," 8 February 2017. [Online]. Available: https://cran.r-project.org/web/packages/mgcv/mgcv.pdf.

[10]   S&P Dow Jones Indices, "Index Mathematics Methodology," S&P Global, 2016.

[11]   "S&P CoreLogic Case-Shiller Home Price Indices," [Online]. Available: http://us.spindices.com/documents/methodologies/methodology-sp-cs-home-price-indices.pdf.

[12]   RBA, "Reserve Bank of Australia Bulletin," July 2004. [Online]. Available: http://www.rba.gov.au/publications/bulletin/2004/jul/pdf/bu-0704-1.pdf.

[13]   P. o. Australia, "Out of reach? The Australian housing affordability challenge," 8 May 2015. [Online]. Available: http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Economics/Affordable_housing_2013/Report.

[14]   CoreLogic, "Model Governance," 2014. [Online]. Available: http://sites.corelogic.net/sites/MG/Policies%20and%20Guidelines/Model%20Governance%20Policy%20-%20Approved%2011.18.15.pdf. [Accessed 1 June 2016].

[15]   CoreLogic, "CoreLogic Daily Index - Type of Indices," 17 March 2017. [Online]. Available: https://www.corelogic.com.au/research/types-of-indices.html. [Accessed 17 March 2017].