Residential Property Price Index: Sources and Methods

Version 1
April 2023

Statistics South Africa

Preface

Three of the principal price indices in the system of economic statistics − the consumer price index (CPI), the producer price index (PPI), and the export and import unit value indices (XMUVI) − are well known and closely watched indicators of macroeconomic performance. The residential property price index (RPPI) is a new addition to this family of indices. Statistics South Africa has developed the RPPI in partnership with the South African Reserve Bank (SARB) and with the support of the International Monetary Fund (IMF). The data are currently released as experimental statistics in the RPPI discussion document (D0160) for the purposes of stakeholder evaluation and feedback. This sources and methods document provides an overview of the methodology used to compile the South African RPPI.

Contents

1. Introduction to the South African residential property price index (RPPI)

1.1 Defining the RPPI

An RPPI aims to measure "the rate at which the prices of residential properties (flats, detached houses, terraced houses, etc.) purchased by households are changing over time. Both new and existing dwellings are covered if available, independently of their final use and their previous owners. Only market prices are considered. They include the price of the land on which residential buildings are located" (OECD 2022).

1.2 Uses of the South African RPPI

RPPIs' importance increased worldwide following the global financial crisis and they are typically used:
- as a macro-economic indicator of economic growth;
- in monetary policy and inflation targeting;
- as an input into estimating the value of housing as a component of wealth;
- as a financial stability or soundness indicator to measure risk exposure;
- as a deflator in the national accounts;
- as an input into an individual's decision making on whether to buy (or sell) a residential property; and
- to make inter-area and international comparisons (IMF, 2020).

1.3 Development of the South African RPPI

In 2018, an International Monetary Fund (IMF) team assessed the feasibility of compiling an RPPI for South Africa and concluded the most suitable source of data would be the property transaction records maintained by the Office of the Chief Registrar of Deeds (Deeds office). However, conflicting legislation and the high cost of obtaining the complete dataset prevented direct access to this information. In 2020, the South African Reserve Bank (SARB) entered into a contract with a third party provider to secure access to the Deeds office data for Statistics South Africa (Stats SA) to develop the RPPI.

From 2021, Stats SA, with the support of the IMF and the SARB, worked through the development stages of the RPPI from assessing the quality of the data, to refining editing rules, constructing an appropriate index estimation model, and finalising weighting structures and dissemination formats.

During September and October 2022, the proposed methodology and preliminary results were shared with different stakeholder groups. A number of the suggestions made during the meetings were incorporated into the current version of the indices which are published as a discussion document (D0160) on the Stats SA website.

1.4 Alignment with international best practice

In compiling the South African RPPI, Stats SA largely follows the methodology guidelines in the *Residential Property Price Index: Practical Compilation Guide* (2020) − the RPPI guide − published by the IMF and the Eurostat *Handbook on Residential Property Prices* (2013) − the RPPI handbook. These are the main references for RPPI concepts and definitions, providing the theory, conceptual framework and practical methodological guidelines for the compilation of RPPIs. The sources and methods documents of other statistical agencies are also used as reference material.

RPPI guide: https://www.imf.org/en/Data/Statistics/RPPI-guide
RPPI handbook: https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF

Stats SA also benefitted from the training and technical assistance provided by the IMF during the development of the RPPI.

2.   Data sources and cleaning process

The RPPI handbook lists a number of potential data sources for the RPPI, available at different points in the process of selling a property:
- advertised or asking prices from newspapers, real estate agents or online property portals;
- mortgage applications from mortgage lenders;
- mortgage approvals from mortgage lenders;
- signed contracts from lawyers or notaries; and
- registered transactions from land registries or tax authorities.

The primary source of data used in the compilation of the South African RPPI is the records of property transactions registered with the Deeds office. All transactions that result in a change of ownership of land are regulated by the Deeds Registries Act (1937) which requires the registration of title deeds with the Deeds office. The data are purchased by the SARB from a third party that does some quality assurance of the data and adds a limited number of variables. The continued production of the RPPI currently depends on this SARB sponsorship.

The RPPI handbook points out that although registration records have the advantage of covering all transactions and reflecting the final price, the data are primarily recorded as a step in the administrative process and not as an input into a statistical system, and this needs to be taken into account when using the data. Stats SA takes the data through an eight-step process to ensure that this administrative data set is suited for statistical processing. Figure 1 illustrates the process followed.

Figure 1: Data cleaning process
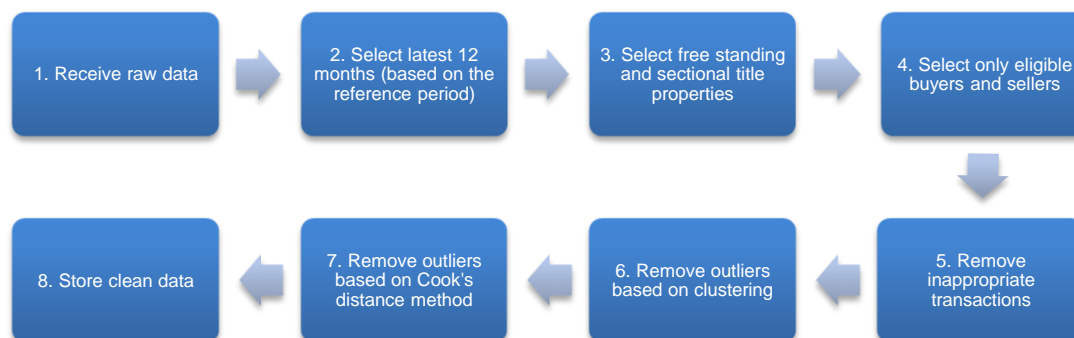
1. Receive raw data
Stats SA typically receives the updated dataset in the third week of each month. Among other variables, the data includes information about the price of the property, the type of property, the type of buyer and seller, and a unique property identifier code. On average, there is a four-month lag between the time a transaction is completed and the time it is recorded by the Deeds office. While both dates are included in the data, the transaction date is the one relevant to the RPPI.

2. Select latest 12 months
Every month, the reference month and 11 months of data prior to the reference month are selected to compile the latest indices. The reference month is five months earlier than the publication date.

3. Select free standing and sectional title properties
In order to exclude non-residential properties, only free standing properties (erf) or sectional title properties are selected. These are assigned property type codes E and U in the Deeds office data.

4. Select only eligible buyers and sellers

In order to ensure that only properties intended for residential use are included, only transactions between eligible parties are included. Purchases by private individuals are included, while the following eight seller types are within scope:

- Body Corporate;
- Company;
- Close Corporation;
- Estate;
- Financial institution;
- Foundation;
- Partnership; and
- Private individual.

5. Remove inappropriate transactions

Three broad categories of records are removed from the data set: where necessary data is missing, where there are duplicate transactions and where the sales value or property size is so small or large that their inclusion will distort the results. These thresholds are based on the data distribution. Properties in areas labelled 'industrial' are also removed. Specifically, transactions are removed when:

- the selling price is null or missing;
- there are multiple buyers;
- sale or transfer of a share of the property;
- duplicate transaction;
- the buyer or seller is missing;
- the buyer and seller are the same person;
- the selling price is less than R80 000;
- the selling prices is greater than R15 million;
- the size of the property is greater than 2 500m^2; and
- the size of the property is smaller than 40m^2.

6. Remove outliers based on clustering

Stats SA has developed a clustering approach based on a k-means methodology. It generates clusters every month based on the selling price for each suburb (Munsub) and allows a more accurate comparison of price data for a similar geographic area. Clusters including less than 10 records are removed from the dataset. Clusters are calculated once a year, using data from the most recent 5 years. At this stage the regression model (see 4.4) is run.

7. Remove outliers based on Cook's distance method

The Cook's distance method identifies any transactions having an outsized impact on the regression results. These observations are flagged for investigation and possible removal from the dataset.

8. Store clean data

The final dataset is now ready for generating the results.

3.   Weights and index reference period

3.1   Weights basis and reference period

The RPPI handbook advises that the basis of the weights will depend on whether the aim of the index is to track the change in the value of the housing stock or the price to acquire housing. If it is the former, then stock value shares are more appropriate. The South African RPPI, however, aims to track changes in the sales values of properties, and therefore uses aggregate sales values as weights.

The weights of the RPPI represent the proportion of sales value per area in a specific calendar year based on the final dataset used to calculate the RPPI. Each metropolitan municipality and each large municipality selected in the RPPI has a weight attached to it, which reflects its relative importance in the overall index. These municipal areas form the elementary indices and are weighted in the process of aggregating the nine provincial, all metropolitan areas and national indices.

The weights reference period should reflect economic conditions that are reasonably normal or stable. Weights are calculated each year using the total sales value of two years prior. The reference period for the 2022 RPPI is 2020. The weights for each area are shown in Appendix A.

3.2   Index reference period

All indices are referenced to 100 in December 2020.

4. Index compilation method

4.1 Introduction

The challenge in compiling any price index is to capture only pure price changes – that is, to separate changes in price due to market dynamics from changes in price caused by shifts in the characteristics of the products being transacted over time. Typically, in measures of inflation such as the consumer and producer price indices, this is achieved by comparing the prices of exactly the same products using a matched model method.

In the case of residential properties, however, price is determined by many different characteristics such as the location, the size of the property, the land it occupies and the type of building, with the result that no two properties are exactly identical. In addition, only a small proportion of the housing stock will be transacted in any given period. The combination of these factors means that the matched model approach cannot easily be used in the case of residential property.

Where possible, RPPIs are computed using a hedonic regression method which estimates shadow prices based on the characteristics of properties (such as size and location) to ensure that quality is held fixed when measuring price changes from one period to the next. The result is an index of pure price change for a consistent set of characteristics.

4.2 The South African model

The South African RPPI uses a form of the hedonic regression model – a rolling window time dummy hedonic regression – to compute the index. The model includes the data of twelve months (the window), with time dummies included for each month. The twelve-month window has the advantage of keeping the coefficients relatively up-to-date and so takes account of changes in price determinants as market conditions change.

The model controls for location, type of property (free standing or sectional title) and the size of the property, which are all available in the data source. Suburbs within each municipality provide the location variable. In cases where there are multiple suburbs with the same name (e.g. Bryanston Ext 1, Ext 2, etc.), these are grouped together. The model includes an interaction term between size and type of dwelling.

Price indices are derived from the estimated coefficients on the time dummies. Equation 1 is the notation of the regression model. Each month the model is moved forward one period and re-estimated. The overall price index is constructed by chaining together the price indices from these rolling windows.

The resulting indices typically show a high level of volatility. In order to provide a more meaningful interpretation of changes in the index, a smoothing algorithm (a Henderson filter) is applied.

Elementary indices are calculated for each of the eight metropolitan and fourteen large non-metropolitan municipal areas.

Equation 1: Time dummy hedonic model

$$\ln p_n^t = \beta_0 + \sum_{t=1}^{T} \delta^t D_n^t + \sum_{k=1}^{K} \beta_k Z_{nk}^t + \gamma^t X_n^t + \varepsilon_n^t$$

where
t – period
n – number of dwellings in period t
k – characteristics
$\ln p_n^t$ – price logarithm
β_0 – intercept
δ^t – coefficient of the time dummy variable that will generate the index
D_n^t – time dummy variables

$\beta_k{}^t$ – 'shadow' price of characteristic k in period t

$Z^t{}_{nk}$ – quantity of characteristic k in period t and dwelling n

γ^t – coefficient of interaction term between property type and property size in period t

X_n^t – interaction term between property type and property size in period t and dwelling n

$\varepsilon_n{}^t$ – error term.

The index for the current period (t) is then derived as follows:

Equation 2: Index derivation

$$I_t = \exp(\hat{\delta}_t) * 100$$

When data for a new period is added to the model, the parameters of the previous periods' dummy variables are affected. In order to ensure an adequate quantity of data for the model and to keep to the coefficients current, a rolling 12-month window is used. This means that data from the last 12 months are pooled together every time the index is calculated for a new period. The indices from the new 12-month window and the previous 12-month window are chained by using the last overlap period between the two windows.

4.3   Aggregation structure

The elementary indices calculated for the metropolitan and larger municipalities are aggregated firstly to a provincial and then to a national level. The metropolitan areas are aggregated in a separate table to 'All metropolitan areas'. The indices are aggregate using weighted arithmetic means – the same method as used in the consumer price index (CPI) and producer price index (PPI). The weights are proportions of total sales from two years prior (see section 3.1). Appendix A shows the detailed coverage and weights of the national and all metropolitan areas RPPIs.

4.4   Changes in weights – linking the old and new index series

RPPI weights are updated annually with effect from each January. It is important to ensure that the change in weights does not distort the month-on-month movements of the indices between December and January. In order to maintain the continuous time series, the January index using the new weights is linked to the December index (using old weights) by means of a linking factor. The linking factor is the ratio of the December index calculated using the old and new weights. The linking factor is applied during the year, until a new linking factor is calculated to link the next January.

Linking results in non-additivity of the new, linked index. In other words, if components are multiplied by their weights and aggregated to higher levels, the results may not be the same as those published.

4.5   Revisions

The Deeds office data is continuously updated and therefore Stats SA allows for revision of the indices. The RPPIs for the latest 12 months will be considered preliminary and this time frame is aligned with the 12-month rolling window period of the regression model. For example, in the October 2022 RPPI statistical release, the RPPIs for November 2021 to October 2022 are preliminary. When the RPPI for November 2022 is published, only the RPPIs for November 2021 to October 2022 will be revised.

4.6   Further work

The RPPI is labelled experimental with the aim of inviting comment on the methodology and the scope of information provided. As the composition and pricing of the property market is largely determined by location, Stats SA is working on making RPPIs for selected lower-level geographic areas available. Provision of further analytical indices may also be considered, for example for different property types, price bands or new and existing homes. Analysis will be conducted to assess and report the extent of revisions.

Appendix A: RPPI coverage and weights (2020)

Province and municipality	National RPPI	All metropolitan areas RPPI
	Weight	
Western Cape	29,00	
City of Cape Town	23,72	28,21
Drakenstein	1,29	
Overstrand	1,06	
Mossel Bay	1,12	
Saldanha Bay	0,78	
George	1,03	
Eastern Cape	5,23	
Buffalo City	1,75	2,09
Nelson Mandela Bay	3,48	4,13
Northern Cape	0,90	
Sol Plaatje	0,90	
Free State	2,68	
Mangaung	2,27	2,70
Matjhabeng	0,41	
KwaZulu-Natal	11,62	
Ethekwini	8,88	10,56
KwaDukuza	1,72	
Msunduzi	1,02	
North West	2,52	
City of Matlosana	0,77	
Rustenburg	1,75	
Gauteng	44,00	
City of Johannesburg	20,09	23,89
City of Tshwane	13,50	16,05
Ekurhuleni	10,41	12,37
Mpumalanga	2,72	
City of Mbombela	1,55	
Emalahleni	1,17	
Limpopo	1,33	
Polokwane	1,33	

List of references

Eurostat. 2013. Handbook on Residential Property Prices Indices (RPPIs). Source: https://ec.europa.eu/eurostat/documents/3859598/5925925/KS-RA-12-022-EN.PDF

IMF. 2020. Residential Property Price Index Practical Compilation Guide 2020. Source: https://www.imf.org/en/Data/Statistics/RPPI-guide

OECD. 2022. What are Residential Property Price indices?. Source:: https://www.oecd.org/sdd/prices-ppp/Residential-Property-Price-Indices-RPPIs-and-related-housing-indicators-FAQs.htm#Q1

Contact

This report is available on the Stats SA website at https://www.statssa.gov.za/?page_id=2528

The monthly discussion document is available on the Stats SA website at
 https://www.statssa.gov.za/?page_id=1854&PPN=D0160&SCH=73577

For more information, please contact:

Matlhatsi Mogalanyane
012 310 4723
matlhatsir@statssa.gov.za

Patrick Kelly
012 310 8290
patrickke@statssa.gov.za