

## RESEARCH ARTICLE

# Tracking Residential Real Estate Capital Growth In NSW by Constructing A Price Index from Sales Transactions

Journal of Data Science and Intelligent Systems

yyyy, Vol. XX(XX) 1–5

DOI: [10.47852/bonviewJDSIS32021344](https://doi.org/10.47852/bonviewJDSIS32021344)



Emad Nashed<sup>1</sup> and Michael Baron<sup>1,2,\*</sup>

<sup>1</sup> University of Adelaide, Australia.

<sup>2</sup> Charles Sturt University, Australia.

\*Corresponding author: Michael Baron, Charles Sturt University, Australia. Email: [mbaron@csu.edu.au](mailto:mbaron@csu.edu.au)

**Abstract:** The main objective of this paper is to investigate residential real estate capital growth in the state of NSW (Australia). Tracking capital growth has been instrumental for tracing price variations & value growth of residential real estate. More specifically, residential real estate past sales transactions can be utilized to track capital growth using repeat sales approach. Ordinary least squared method was used to fit a linear model and construct a price index. The results were used to establish a yearly and quarterly price index for the whole state as well as for each respective district. The research shows that while using a mean price as a measure for tracking growth can produce misleading results, the capital growth-tracking can be carried out more accurately by the means of investigating repeat sales. Furthermore, the research reveals trends in the property market as well as areas with the highest and lowest growth. The results produced can help investors, banks, and government organizations to compare capital growth across different districts at different times and drive lending, investment, or development decisions.

**Keywords:** real estate price index, repeat sales, capital growth

## 1. Introduction

Residential real estate industry plays an important role in the Australian economy. The total contribution of that sector was estimated to be over \$150 billion per year in 2016. The growth in residential property building has provided a boost to the national economy and provided tens of thousands of jobs in Australia (Housing Industry Association, 2016). Tracking capital growth in Australian residential real estate properties will help both the supply and demand sides to achieve a healthy market growth.

However, the real estate market is not a homogeneous market (Jones & Trevillion, 2022). Residential properties differ based on many characteristics such as location, size, construction, design, proximity to other facilities, etc. As a result, tracking the performance of the real estate market is a challenge since the heterogenous nature of the market makes it difficult to determine the trend of prices based on individual sale transactions. To address this challenge, residential property price indexes (RPRI) are typically used to track the performance of the market and measure capital growth (Jones & Trevillion, 2022). Additionally, user cost

models for housing incorporates the price index value to estimate the capital growth and the opportunity cost (European Commission et al., 2013). Diewert et al. (2020) argue that capital growth is the most important part in calculating user costs. In this paper we will use the constructed price index to estimate the final sale price of properties at the end of each period which could be used to obtain the capital gain component in the user cost model.

In this research, several methods for establishing price indexes were considered. Given the nature of the available datasets, the repeat sales method was chosen and used to construct the price index. The method was used to build a linear model to predict house prices. The results are evidently more accurate than one's achieved by the means of deployment of a naïve model that uses the mean sale price to track the growth. The results revealed the current trend in the market, as well as areas with the highest and lowest capital growth. This could be beneficial for many stakeholders, such as investors, governments, lenders, and land developers to assist in future development or investment decisions.

## 2. Literature Review

There are different ways to construct a price index for real estate, and not all of them achieve the same purpose (Diewert et al. 2020). In general, there are two main groups of real estate price indexes: appraisal-based, and transaction-based indexes. In appraisal-based indexes, a sample of properties are valued on a regular basis, the valuation results are collected for analysis to measure the performance (Jones & Trevillion, 2022). This is common in commercial properties where past sales transactions are not required. Alternatively, for residential properties, transaction-based indexes are used (Jones & Trevillion, 2022).

There are two main approaches that belong to transaction-based indexes. The first takes into consideration all the characteristics of the property and builds a model to estimate the price from those characteristics. An example of this is the hedonic regression model (Rosen, 1974; Hill, 2021). The shortcoming of this approach is that a bias is introduced to the model if some of the attributes are not accounted for. The second approach is the repeat sales method. As its name implies, it only considers properties that are sold multiple times. This method eliminates the need to account for different property characteristics, assuming that there is no major re-construction or renovation applied between the multiple sales transactions (European Commission et al., 2013).

The original BMN repeat sales method was first developed by Bailey (1963). The main idea is that the price of properties changes over time by the same change in price index. The method uses properties that were sold repeatedly and fits a linear model using least squared method to predict the log of the ratio of first and second sale prices.

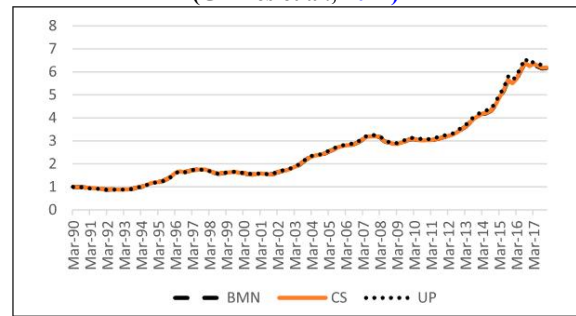
Multiple enhancements were applied to the repeat sales method. Case and Shiller (1987) criticized the assumption of the homoscedasticity of the original method. by arguing that if the time difference between two sales transactions is higher, the variance is different because of the uncertainty about the condition of the property. This means if the period between the two sales of the house is long, the house could have been renovated, or could have deteriorated. They introduced the Weighted Least Squared method (WLS) which applies a random term representing the level of uncertainty depending on the interval between the consecutive sales.

Additional improvements were applied to the repeat sales method by Gao and Wang (2007). Instead of using pair transactions used by Bailey (1963), and Case and Shiller (1987) to compare price changes between two periods, they created a multiple transactions model (UP Model) without breaking them into pairs of transaction.

A study was done in 2021 to compare the three main repeat sales methods in New Zealand by using a dataset from CoreLogic NZ (Grimes et al., 2021). The study compared the original repeat sales method by Bailey

(BMN), the Case and Shiller method (CS) and the Gao and Wang (2007) unbalanced panel approach (UP). Figure 2 shows a sample of Grimes results showing the comparison between BMN, CS and UP prices index. The study concluded that the three indexes produced similar results across different markets.

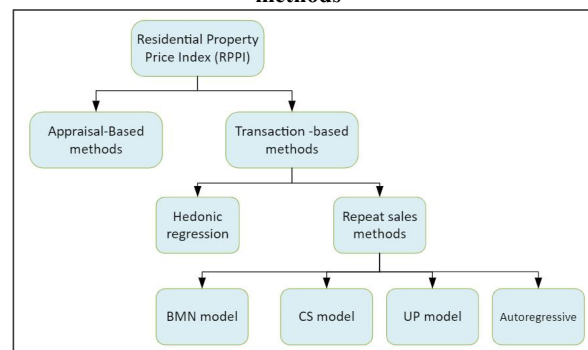
**Figure 2**  
Comparison between BMN, CS and UP price indexes. (Grimes et al., 2021)



One disadvantage of repeat sales method is that it does not use the whole dataset. An alternative approach is to use both single sales and repeat sales in one model. An autoregressive approach was developed that utilizes the whole dataset in addition to using the Zip code (Nagaraja et al., 2011).

Figure 1 shows hierarchy of the surveyed methods for this research. It is not feasible to obtain accurate datasets that contain all the attributes of sold properties, and hence the repeat sales method is the most suitable method for the purpose of this research given the nature of the currently available public datasets.

**Figure 1**  
Hierarchy of surveyed methods



### 2.1 Theoretical Framework

It was shown that the best two methods to establish a price index were BMN and autoregressive. BMN is favored for its simplicity, and autoregressive approach for its accuracy from utilizing the Zip code which was found to be a very significant predictor (Nagaraja et al., 2014). For this

research we chose the original BMN method. Additionally, to make use of the significance of the location, the method was used to build a separate model for each LGA (local government area).

To summarize the BMN method model, if a specific property was sold for a price  $P_t^t$  at time  $t$  and sold again for a price  $P_t^{t'}$  at time  $t'$ , the ratio between the two prices can be expressed as:

$$\log\left(\frac{P_t^{t'}}{P_t^t}\right) = \log\left(\frac{\beta^{t'}}{\beta^t}\right) + \epsilon$$

Where  $\beta$  is the price index at time  $t$ , and  $\beta^{t'}$  is the price index at time  $t'$ , and  $\epsilon$  is the residual error. The model is changed to the logarithmic scale as the following:

$$r_{itt'} = -b_t + b_{t'} + \epsilon$$

Where  $r_{itt'}$  is the logarithm of the ratio between the two prices,  $b_t$  is the logarithm of the price index at period  $t$ , and  $b_{t'}$  is the logarithm of the price index at time  $t'$ . Using the ordinary least squared method (OLS), the model can solve the equation for the price indexes at each period. This model assumes that the error term has a mean of zero, follows the normal distribution and has a constant variance (homoscedasticity assumption).

So, despite these shortcomings of the original BMN method, the empirical results are still similar when compared against other methods. The simplicity of the BMN model and the empirical results show that it is a valid and suitable data analysis method for this task. This research demonstrates how to use NSW public data to construct a price index. The results are separated by property type, used to track capital growth, and compared in different districts.

### 3. Research Methodology

#### 3.1. Dataset and Preprocessing

The dataset is a secondary data source collected from various publicly available sources and published by the NSW government website on a weekly basis (Valuer General NSW Valuation Portal, 2023). The files contain sale transactions data from 2001. Historic data from 1990 till 2001 are also available in a different format. The dataset is available as part of NSW Government open data policy and subject to the Creative Commons attribution 3.0 Australia license.

The original dataset contained a total of 134,500 files having 4,737,735 transactions. Only transactions with settlement date within the range (1/1/2001 - 30/6/2023) were considered. Entries with missing information, such as prices or legal description were removed. Transactions of partial interest sales were removed; this is when a part of the property is sold. Non-residential properties were removed such as: shops, offices, farms, hotels, etc. Some

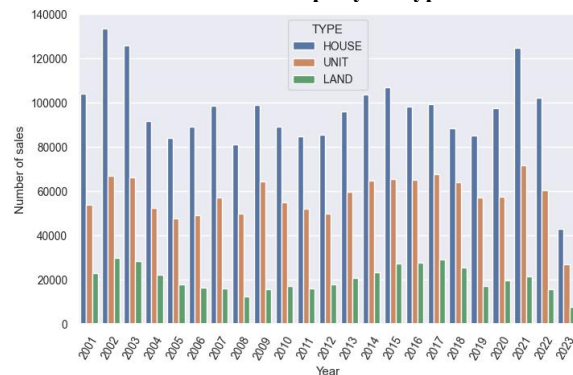
duplicate transactions were also identified and removed using the legal description and either contract or settlement dates.

#### 3.2. Feature Engineering

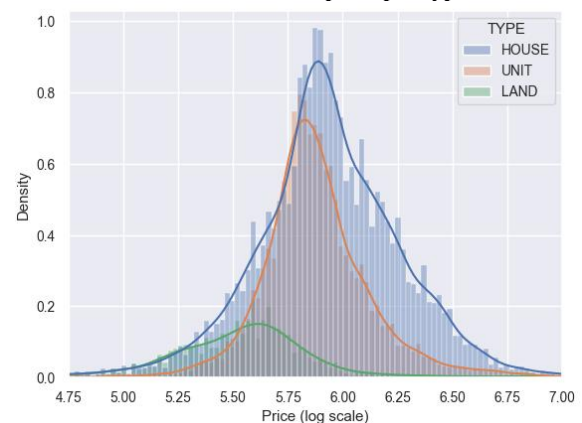
A property type feature was added to the dataset with a value of either House, Unit, or Land. The value was assigned based on the address (whether it contains a unit number), title reference (whether it is a strata title) or based on the purpose of the land (for example townhouses, duplex, and villas were considered as units). Figure 3 and Figure 4 summarize the dataset between Jan 2001 and June 2023. It shows the number of transactions per year per type, and the price distribution per type. It shows that house transactions are the most common with the highest prices, units come second, and land sales are the least common with the lowest prices.

NSW has over 4500 localities, which is too granular. Using the localities as a feature would result in a highly sparse dataset. Alternatively, NSW has only 128 LGA (local government area). LGA name was added to the dataset using the Australian statistical geography standard dataset (Australian Bureau of Statistics, 2021a) to provide a higher-level categorisation of properties.

**Figure 3**  
Number of sales per year/type



**Figure 4**  
Distribution of sale price per type



### 3.3. Transforming the dataset

To use repeat sale method, the dataset was transformed into a pair-transactions for properties that were sold multiple times. The result is a directed graph as the following:

*Initial sale time & price* → *Final sale time & price*

To visualize the distribution of repeat sales, Figure 5 shows a heat map of the number of pair-transactions by the initial year and final sale year. Light regions on the map indicate a higher number of repeat-sales. For example, many properties bought in 2003 were sold again in 2007 and 2009. In general, the light region goes diagonally, begins after two years from initial sale, and starts to fade away after 10 years.

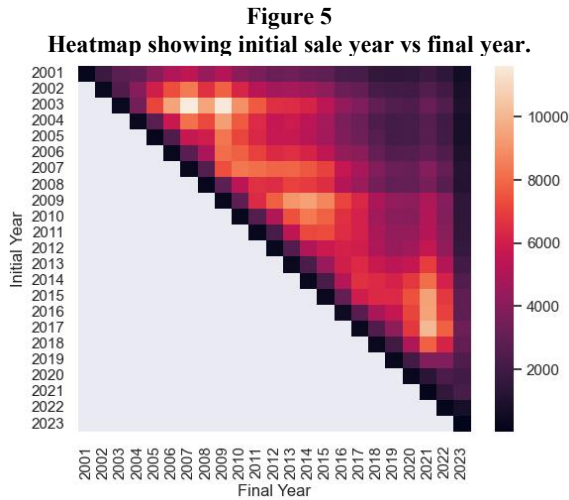
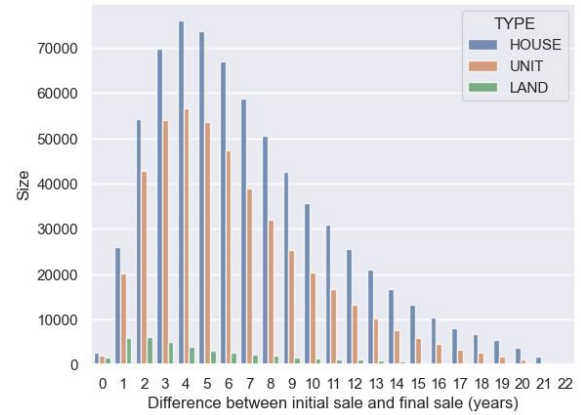


Figure 6 shows the distribution of the duration of owning a property expressed as the difference in years between buying and selling a property. It can be noticed that the peak duration of owning a house is four years, which is the same value for units. However, land resales peak at a shorter period of two years.

**Figure 6**  
Distribution of duration of owning a property in NSW.



### 3.4. Outlier removal

The dataset was inspected and found to have outliers. Table 1 confirms the existence of outliers in the growth rate in pair transactions. It shows that the mean and max values are huge compared to the median and the 75<sup>th</sup> percentile. This could be attributed to major development, or depreciation in the property conditions, as well as mis-reported price values.

**Table 1**  
**Annual growth statistics**

Mean	7.765608e+21
Standard deviation	9.266820e+24
Min	2.728775e-24
25 <sup>th</sup> percentile	1.025117e+00
50 <sup>th</sup> percentile	1.055411e+00
75 <sup>th</sup> percentile	1.095935e+00
Max	1.105827e+28

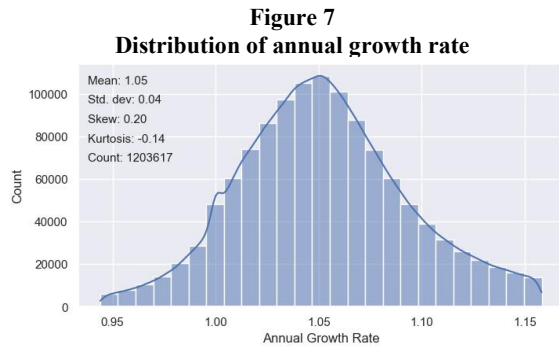
To inspect outliers in the pair transactions dataset,  $g$  is defined as the annual growth as the following:

$$g = \left( \frac{P_1}{P_0} \right)^{\frac{1}{y_1 - y_0}}$$

where:

- $g$  :annual growth
- $P_0$  :Initial price
- $P_1$  :Final price
- $y_0$  :Initial sale year
- $y_1$  :Final sale year

The interquartile range method was used iteratively to remove outliers outside the range (Feasel, 2022). The method simply calculates the interquartile range (IRQ) as the difference between Q3 (75<sup>th</sup> percentile) and Q1 (25<sup>th</sup> percentile) and excludes observations outside the range from  $[Q1 - 1.5 \text{ IQR}, Q3 + 1.5 \text{ IQR}]$ . The result was removing around 10% of pair transactions. Figure 7 shows the distribution of the annual growth rate after removing outliers. It shows that the mean annual growth rate is around 5% for properties in NSW.



### 3.5. Performance metrics

The pair transactions dataset was split into a training set of 80% and test set of 20% and stratified by the final sale year to allow for balanced coverage over the years between the training and test sets. The Ordinary least squared model was fitted (Grimes et al., 2021) using the training set. Using the test set, the model can predict  $\hat{y}$ , the log of the ratio between the final price and the initial price as the following:

$$\hat{y} = \log\left(\frac{P_1}{P_0}\right)$$

To test the model, the result was transformed to calculate the final price ( $P_1$ ) as the following:

**Equation 1**  
Predicting final sale price.

$$P_1 = e^{\hat{y}} \cdot P_0$$

where:

- $P_0$  :Initial price
- $P_1$  :Final price
- $\hat{y}$  :OLS model prediction

The errors in predicted final price were evaluated using Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and  $R^2$ . To validate the model results, a naïve model was built that predicted the growth in price as the average sale price in the final year divided by the average price sale in the initial year. The OLS performance metrics were compared against the naïve model.

## 4. Results

Figure 8 shows the results after applying the ordinary least squares method to solve for yearly houses coefficients and transform back to the price index values for each property type.

Based on the analysis, it has been established that on average, every \$1 invested in houses in 2001 turned into \$4.08 by 2023. While every \$1 invested in a ‘unit’ turned

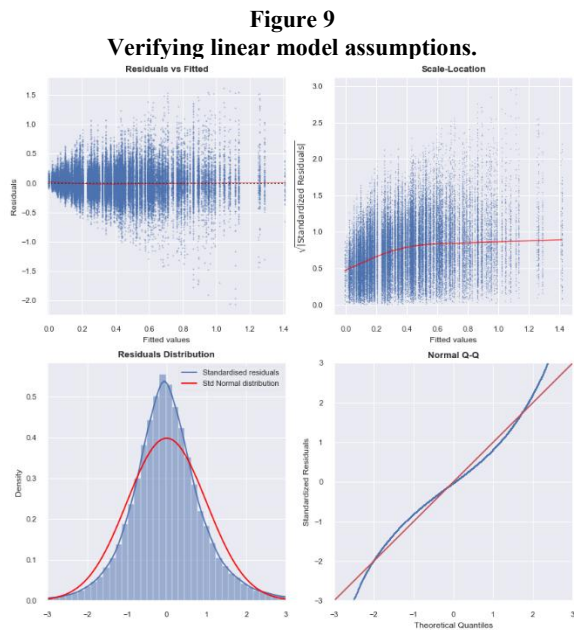
into only \$2.601 by 2023. Land showed the highest capital growth where every \$1 spent turned into \$4.601 by 2023. The results also show that the price index has been increasing since 2001 and peaks in 2022 for all real-estate properties. The only years that suffered a negative growth in prices were 2019 and 2023 (Note that 2023 results are partial and only up to June 2023). Land price indexes seem to have a different pattern than units and houses and only had a negative growth in 2009.

**Figure 8**  
Results of OLS method.

Year	PriceIndex			Growth		
	HOUSE	LAND	UNIT	HOUSE	LAND	UNIT
2001	1.000	1.000	1.000			
2002	1.169	1.187	1.119	↑ 16.95%	↑ 18.69%	↑ 11.86%
2003	1.402	1.526	1.232	↑ 19.88%	↑ 28.58%	↑ 10.15%
2004	1.530	1.727	1.290	↑ 9.11%	↑ 13.17%	↓ 4.66%
2005	1.531	1.816	1.273	⇒ 0.07%	↑ 5.12%	↓ -1.28%
2006	1.559	1.855	1.269	↓ 1.83%	↓ 2.19%	⇒ -0.32%
2007	1.610	1.928	1.295	↓ 3.27%	↓ 3.89%	↓ 2.05%
2008	1.641	1.983	1.321	↓ 1.91%	↓ 2.87%	↓ 2.00%
2009	1.687	1.966	1.371	↓ 2.80%	⇒ -0.83%	↓ 3.78%
2010	1.808	2.046	1.479	↑ 7.18%	↓ 4.05%	↑ 7.91%
2011	1.840	2.064	1.515	↓ 1.77%	⇒ 0.90%	↓ 2.43%
2012	1.875	2.110	1.537	↓ 1.92%	↓ 2.23%	↓ 1.43%
2013	1.967	2.122	1.631	↓ 4.89%	⇒ 0.54%	↑ 6.09%
2014	2.146	2.221	1.796	↑ 9.14%	↓ 4.69%	↑ 10.12%
2015	2.370	2.342	1.994	↑ 10.44%	↑ 5.44%	↑ 11.04%
2016	2.530	2.478	2.126	↑ 6.73%	↑ 5.81%	↑ 6.60%
2017	2.769	2.663	2.290	↑ 9.44%	↑ 7.45%	↑ 7.76%
2018	2.854	2.877	2.307	↓ 3.08%	↑ 8.05%	⇒ 0.74%
2019	2.843	3.008	2.249	⇒ -0.40%	↓ 4.54%	↓ -2.54%
2020	3.039	3.151	2.345	↑ 6.90%	↓ 4.78%	↓ 4.29%
2021	3.617	3.679	2.553	↑ 19.05%	↑ 16.73%	↑ 8.84%
2022	4.115	4.535	2.663	↑ 13.75%	↑ 23.29%	↓ 4.34%
2023	4.080	4.601	2.601	⇒ -0.84%	↓ 1.45%	↓ -2.36%

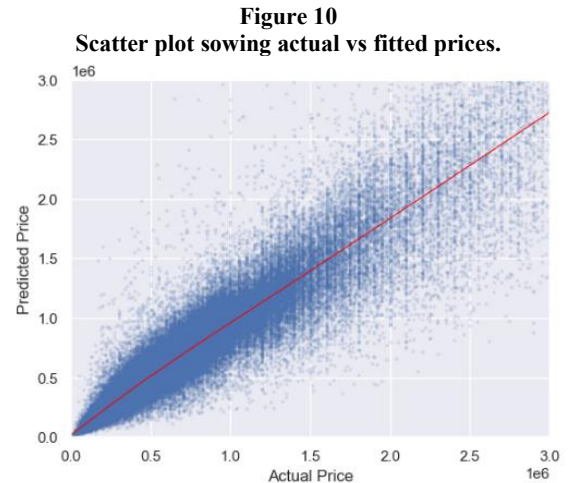
### 4.1. Verifying model assumptions

The OLS method assumes that residuals are normally distributed with the mean of zero and having constant variance (Gupta et al., 2020). Figure 9 shows four plots used to verify the linear model assumptions using a graphical method (Chatterjee & Hadi, 2006). First, the residuals vs the fitted values were plotted. The result is a flat line at 0.0, which verifies the linearity assumption. Secondly, the study used the scale-location plot to check for homoscedasticity assumption (constant variance). The plot shows the square root of absolute standardized residuals versus the fitted values. The results show a slight growth in the line, then it becomes flat. This indicates that the variance of the error is increasing slightly as the fitted values are increasing. The final two plots show the distribution of the residuals against the normal distribution. The distribution and the QQ-plot show that that it is very close to normal distribution, and hence, the normality assumption of the model can be accepted.



model. The performance metrics show that using repeat sale OLS method provides significant improvement than the naïve model, which validates the model. For example, using the naïve model would result in an absolute error of \$276k on average, compared to only \$134k if the OLS model was used.

Figure 10 shows the plot of actual prices vs predicted prices by the model. It shows the values are scattered around the line with slope close to 1.0.



## 4.2. Model Performance

To measure the model performance, final sale price was predicted using Equation 1. Table 2 summarises the results comparing metrics of the model against the naïve

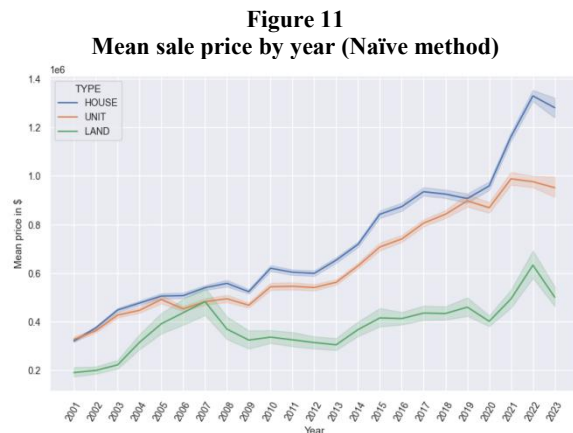
**Table 2**  
**Comparing model performance vs naïve model for yearly price-index.**

Property Type	Training Size	Test Size	Naïve Model Results				OLS Model Results			
			MAE	MAPE	RMSE	R <sup>2</sup>	MAE	MAPE	RMSE	R <sup>2</sup>
House	558,823	139,706	275,894	0.350	532,351	0.597	134,286	0.171	290,263	0.880
Unit	366,784	91,696	179,971	0.269	316,248	0.689	91,605	0.140	189,187	0.888
Land	31,318	7,830	103,607	0.318	332,108	0.674	62,420	0.204	216,652	0.861

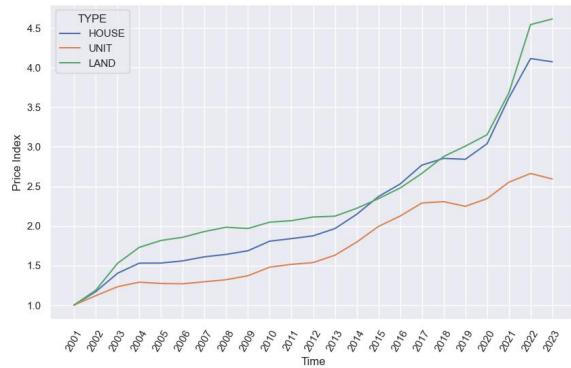
## 5. Discussion

### 5.1. Heterogeneity problem and validity of the research

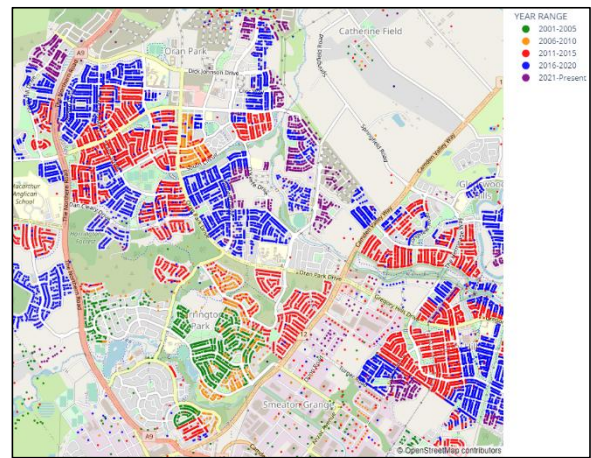
One can argue that mean sale price could be used to measure capital growth given a large enough sample. However, results in Table 4 showed otherwise. Additionally, it could be illustrated visually that the mean price is not suitable to measure growth. Figure 11 shows that considering mean sale price, land growth is less than houses and units. However, results from yearly price index in Figure 12 shows the opposite case where land is achieving a higher growth. This could be explained by the heterogeneity nature of land sales. The available land lots are intrinsically different every year. Figure 13 illustrates this point, by showing a scatter plot of land sales in Camden area, where land sales are progressively moving by time and are not randomly distributed. And hence, the price index method would overcome this shortfall and produce a more accurate result.



**Figure 12**  
Yearly price index



**Figure 13**  
Scatter plot of land sales in Camden, NSW over time.



## 5.2. Comparison with current market data

Australian Bureau of Statistics (ABS) has been publishing residential real estate price index for the eight capital cities in Australia including Greater Sydney in NSW (Australian Bureau of Statistics, 2021b). To compare the results of the model deployed against ABS results, the quarterly price index was constructed from Q4/2011 to Q2/2023 for both houses and units. Figure 14 plots the model's House and Unit price index for all LGAs constituting Greater Sydney area against the ABS price index for the same area.

The results show the model trend is aligned with the ABS price index up to Q4/2021 which validates our model. In addition to that, the model has additional contributions:

- The model can provide up to date insights using the available public dataset. Figure 14 shows the trend up to Q2/2023.
- The model can provide clear separation between houses, units, and lands.

**Figure 14**  
Greater Sydney Price index since 2011



The results in Figure 14 also show that property prices suffered a decline since the second half of 2022. However,

the market has been recovering and property prices are starting to rise again from Q2/2023. This aligns with CoreLogic price index based on their proprietary data that shows that Sydney area had a negative growth year on year, and positive growth in in Q2/2023 (CoreLogic, 2023).

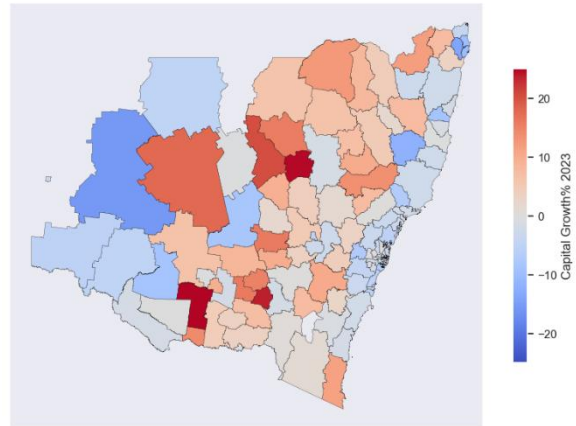
### 5.3. Cash rate impact on property prices

Market experts have traditionally been relating the drop of property prices in the second half of 2022 and 2023 to the rising interest rates (Terzon, 2022). The Reserve Bank of Australia (2023) has raised the cash rate for 12 times for the period from May 2022 to June 2023. Results shown above in Figure 8 and Figure 14 confirm with those market reports. For example, this research shows that houses in NSW lost on average 1.1% of their value in 2023, while units lost 2.7% of their value.

### 5.4. Comparing districts

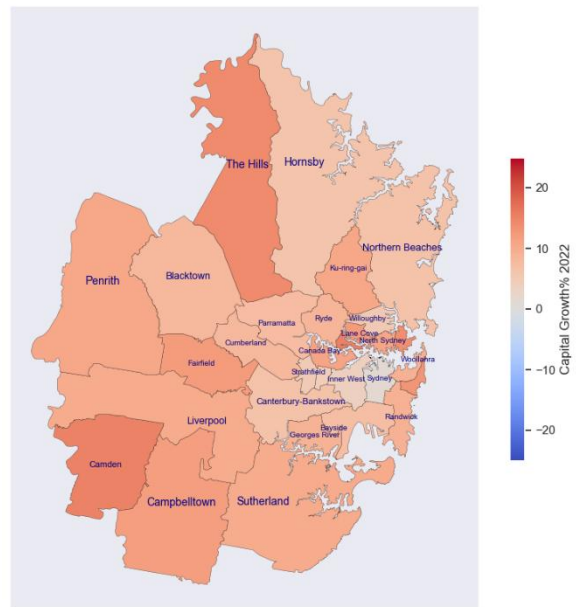
The results of the model can be used to compare areas in terms of annual capital growth. Table 3 & Table 4 show five LGAs with the highest and lowest capital growth for houses in 2023 (up to May 2023). It is noticeable that some areas have significantly higher capital growth than the state average. For example, ‘Murrumbidgee’ has achieved a growth of 29.7% growth. It is out of the scope of this research to explain why some particular areas have significantly higher capital growth than other areas, however, it is evident from Table 4 that some areas such as ‘Lane Cove’ that had very high growth in recent years, suffered from a 10.6% decline in 2023. Figure 15 shows the heatmap for capital growth across NSW in 2023.

**Figure 15**  
Heatmap comparing capital growth in 2023 in NSW.



Additionally, Figure 16 & Figure 17 compare areas in Greater Sydney between 2022, and 2023. It shows that there was a positive overall growth in 2022 and negative growth in 2023. Comparing districts can have many benefits, for example property investors can have more informed decisions on where to invest next or how much are their properties currently worth. On the other hand, mortgage lenders can better assess the risks on loans and mitigate future losses.

**Figure 16**  
Comparing districts of Greater Sydney Area in 2022.



**Table 3**  
Areas with the highest house capital growth in 2023.

LGA	Capital Growth %	
	2022	2023 <sup>1</sup>
Murrumbidgee	14.6%	29.7%
Gilgandra	18.1%	24.7%
Junee	18.7%	23.8%
Warren	10.0%	20.5%
Cobar	3.8%	18.0%

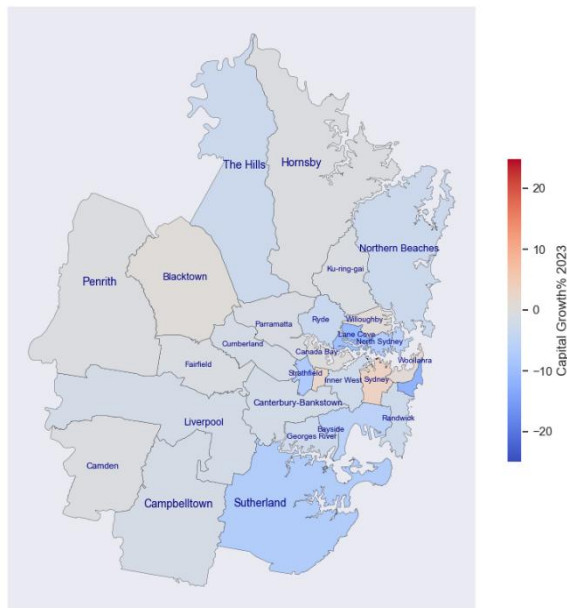
**Table 4**  
Areas with the lowest house capital growth in 2023.

LGA	Capital Growth %	
	2022	2023
Central Darling	19.0%	-15.8%
Lismore	8.7%	-14.0%
Walcha	48.0%	-12.7%
Waverley	13.0%	-11.7%
Lane Cove	12.2%	-10.6%

<sup>1</sup> 2023 results are up to June 2023



**Figure 17**  
**Comparing districts of Greater Sydney Area in 2023.**



### 5.5 Limitations of the study

The study uses a public dataset collected from multiple sources and provided by the NSW Government. The dataset used covers only residential real estate properties sold in NSW between 2001 to 30/06/2023. The NSW government does not guarantee the accuracy nor the completeness of the data, and hence the accuracy of the results is limited by the quality of the dataset.

### 5.6 Future considerations

The results of the study can have implications for many sectors. A frequently updated price index for residential real estate could help property investors make more informed decisions. Banks and lenders could benefit from the geographic comparison of districts and adopt a capital growth risk-based approach. But most importantly, government agencies could use the results in planning land development in rapidly growing districts to better balance supply and demand or introduce more community development programs in slower growing districts.

## 6. Conclusion

Based on the residential properties sales information for the state of NSW (Australia) the research has led to construction of a price index using the repeat sales method. The price index was used to estimate annual and quarterly capital growth in houses, units, and lands. Results showed that using the price index achieved significantly higher accuracy in estimating growth than using the mean sale price. The results were aligned with the Australian Bureau

of Statistics property index for Greater Sydney area, additionally, it provided more insights about the current market trend. Properties in Greater Sydney areas dropped in value since mid-2022, however this trend is starting to change directions and go up from Q2/2023. The research also covered all areas in NSW; the comparison between different areas revealed areas with the highest or lowest growth. This model can be used with current sales data to provide up-to-date residential real estate price index. The results can be used by investors to have more informed decisions about their future investment or estimate the current value of their properties. Mortgage lenders can also use the results to better assess risks on future loans or estimate equity on current ones. Government agencies could rely on the results for planning future projects, to release to target slower growing districts.

### Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

### References

- Australian Bureau of Statistics. (2021a). *Australian statistical geography standard (ASGS) edition 3*. Retrieved from: <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026>
- Australian Bureau of Statistics. (2021b). *Residential property price indexes: Eight capital cities*. Retrieved from: <https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/residential-property-price-indexes-eight-capital-cities/latest-release>
- Bailey, M. J., Muth, R. F., & Nourse, H. O. (1963). A regression method for real estate price index construction. *Journal of the American Statistical Association*, 58(304), 933-942. <https://doi.org/10.1080/01621459.1963.10480679>
- Case, K. E., & Shiller, R. J. (1987). Prices of single-family homes since 1970: New Indexes for Four Cities. *NBER Working Paper Series*, 1-51. <https://doi.org/10.3386/w2393>
- Chatterjee, S. & Hadi, A. S. (2006). *Regression analysis by example*. USA: John Wiley & Sons. <https://doi.org/10.1002/0470055464>
- CoreLogic. (2023). *CoreLogic daily home value index*. Retrieved from: <https://www.corelogic.com.au/our-data/corelogic-indices>
- Diewert, W. E., Nishimura, K. G., Shimizu, C., & Watanabe, T. (2020). *Property price index theory and practice*. Japan: Springer. <https://doi.org/10.1007/978-4-431-55942-9>
- European Commission, Eurostat, Organisation for Economic Co-operation and Development, & World Bank. (2013). *Handbook on residential property prices indices (RPPIs)*. France: OECD Publishing. <https://data.europa.eu/doi/10.2785/34007>

- Feasel, K. (2022). *Finding ghosts in your data: Anomaly detection techniques with examples in python*. USA: Apress. <https://doi.org/10.1007/978-1-4842-8870-2>
- Gao, A. & Wang, G. (2007). Multiple transactions model: A panel data approach to estimate housing market indices. *Journal of Real Estate Research*, 29(3), 241–266. <https://doi.org/10.1080/10835547.2007.12091202>
- Grimes, A., Sorensen, K., & Young, C. (2021). Repeat sales house price indices: Comparative properties under alternative data generation processes. *New Zealand Economic Papers*, 55(1), 7–18. <https://doi.org/10.1080/00779954.2019.1612937>
- Gupta, B. C., Guttman, I., & Jayalath, K. P. (2020). Simple linear regression analysis. In *Statistics and probability with applications for engineers and scientists using MINITAB, R and JMP* (pp. 622–692). John Wiley & Sons. <https://doi.org/10.1002/9781119516651.ch15>
- Hill, R. (2011). Hedonic price indexes for housing. *OECD Statistics Working Papers*, 2011(01). <https://doi.org/10.1787/5kghzpxt6g6f-en>
- Housing Industry Association. (2016). *Increasing Australia's future prosperity discussion paper – 5-year productivity review*. Retrieved from: [https://www.pc.gov.au/\\_data/assets/pdf\\_file/0005/211649/sub028-productivity-review.pdf](https://www.pc.gov.au/_data/assets/pdf_file/0005/211649/sub028-productivity-review.pdf)
- Jones, C. A. & Trevillion, E. (2022). *Real estate investment: Theory and practice*. UK: Palgrave Macmillan. <https://doi.org/10.1007/978-3-031-00968-6>
- Nagaraja, C. H., Brown, L. D., & Zhao, L. H. (2011). An autoregressive approach to house price modeling. *The Annals of Applied Statistics*, 5(1), 124–149. <https://doi.org/10.1214/10-AOAS380>
- Nagaraja, Brown, L. D., Brownz, L. D., & Wachter, S. M. (2014). Repeat sales house price index methodology. *Journal of Real Estate Literature*, 22(1), 23–46. <https://doi.org/10.1080/10835547.2014.12090375>
- Reserve Bank of Australia. (2023). *Cash rate target*. Retrieved from: <https://www.rba.gov.au/statistics/cash-rate>
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55. <https://doi.org/10.1086/260169>
- Terzon, E. (2022). *Property prices drop nationally again, with Sydney diving 6pc annually and regional Australia following suit*. Retrieved from: <https://www.abc.net.au/news/2022-10-03/property-prices-corelogic-rents-september-2022/101495078>
- Valuer General NSW Valuation Portal. (2023). *Bulk property sales information*. Retrieved from: <https://valuation.property.nsw.gov.au/embed/propertySalesInformation>

## How to Cite

Nashed, E. & Baron, M. (2023). Tracking Residential Real Estate Capital Growth In NSW by Constructing A Price Index from Sales Transactions. *Journal of Data Science and Intelligent Systems*. <https://doi.org/10.47852/bonviewJDSIS32021344>