MONASH University

Regional Dynamics of Australian Housing Indexes

Yiran Yao

19 October 2025

Background

In real estate, key questions shape both professional strategy and academic research, guiding long-term policy and investment decisions. These include:

- How can we predict housing prices at the local level?
- What drives price fluctuations over time?
- Which regions will gain desirability or investment potential in the coming decade?

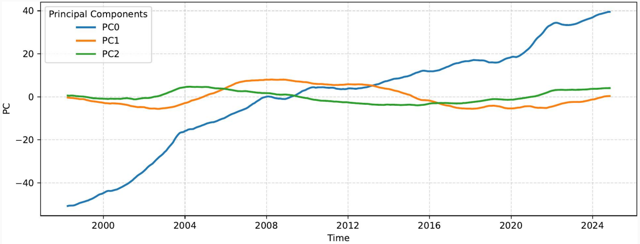
Problem

Existing approaches are effective but depend on diverse data sources that require extensive preprocessing and alignment with housing indexes.

- Time-series models: Region-specific forecasts that work well short-term but are labor-intensive and don't scale easily.
- Econometric models: Capture causal links with economic variables, but forecasts suffer if inputs are uncertain.
- Machine learning: High predictive accuracy (e.g., random forest, XGBoost), though interpretability is limited.

Methodology

W. Silp proposed a PCA-based framework for housing index analysis. We extend it with time series forecasting and multivariate regression using only: Logged monthly housing index growth (1995–2024). The PCA was conducted at the SA2 level, with results subsequently aggregated to SA4 and major city levels to enhance interpretability.



PCA Summary

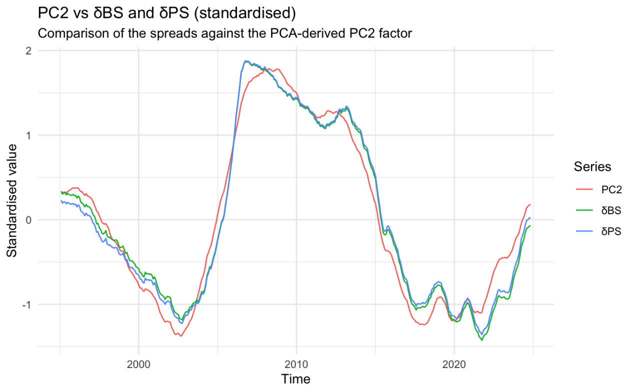
The first three principal components explain 96.5% of variation in housing indexes, reflecting **market, mining, and lifestyle** influences based on regional loadings.

major_city	loading_value
REST OF WA	0.3467
GREATER PERTH	0.2101
GREATER DARWIN	0.1495
REST OF NSW	-0.0468
GREATER MELBOURNE	-0.0749
GREATER SYDNEY	-0.2040

From PC to Factor Series

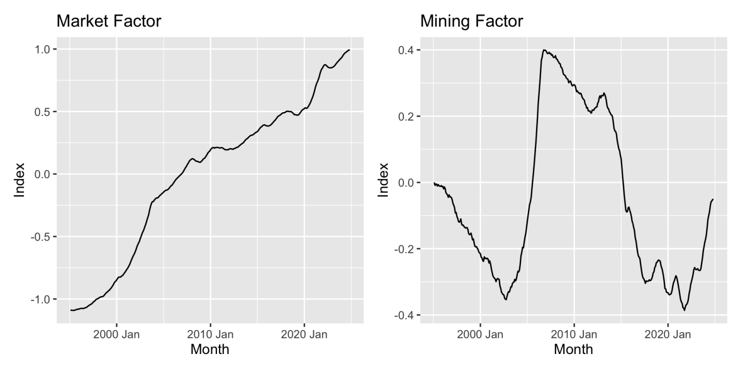
PCs were approximated as weighted averages of regional indexes for greater stability, interpretability, and modeling utility, while retaining **>96% of the original variance**.

$$\delta_{PS}(t) = \mu_P(t) - \frac{\mathrm{cov}(\mu_P, U)}{\mathrm{cov}(\mu_S, U)}\,\mu_S(t)$$



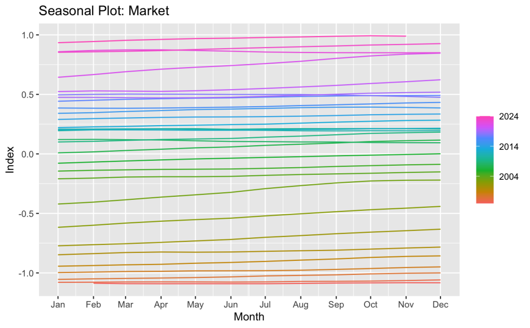
EDA

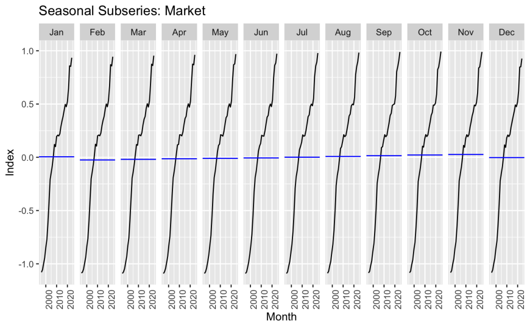
- The market factor shows long-term growth, a structural shift during the pandemic and remains volatile.
- The mining factor rises, drops and fluctuates, reflecting resource-sector shifts.



EDA

- Neither factor shows seasonality; variation is driven by long-term structural changes.
- These non-seasonal trends suggest forecasting should emphasize regime shifts and structural breaks over seasonal adjustments.



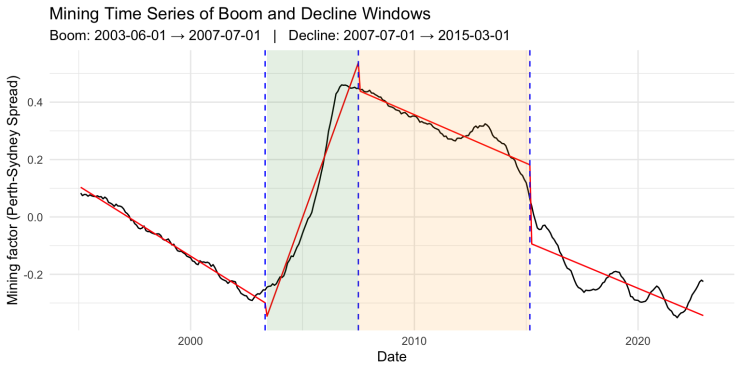


Forecasting Factors: Market

- The market factor is non-stationary and resembles a random walk with drift. ARIMA(0,1,0) was used as a benchmark and compared with neighboring models using AICc and Ljung-Box statistics.
- None passed the Ljung-Box test at 5%, likely due to subtle fluctuations; this was addressed downstream via ARIMAX.
- ARIMA(1,1,0) with a Kourier term of 2 was selected for its simplicity, assumption alignment, and residuals showing no visible trend or seasonality.

Forecasting Factors: Mining

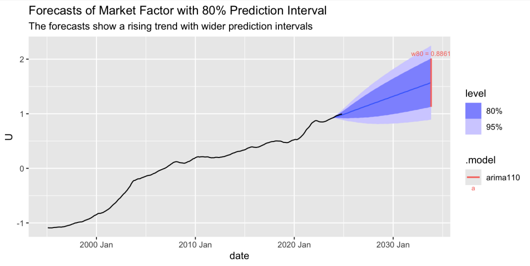
The mining factor exhibits a temporary structural break due to the mining boom. This was addressed using dummy variables as exogenous regressors. Residual dynamics were modeled with ARIMA(2,1,1), which satisfied the Ljung-Box test.

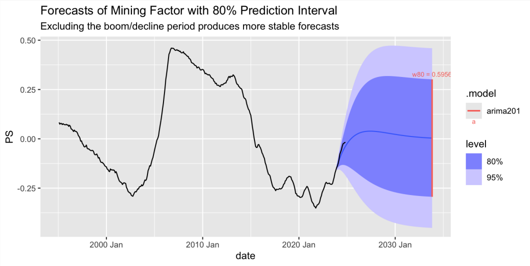


Test Set Performance & Prediction Interval

The market factor yields low forecast error (RMSE = 0.0144, MAE = 0.0131), consistent with its stable dynamics. The mining factor exhibits higher error (RMSE = 0.0214, MAE = 0.0180), reflecting its greater volatility and cyclical nature.

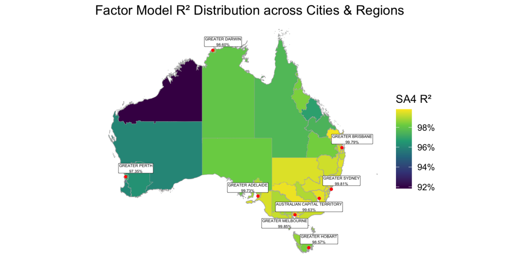
Ten-year forecasts show a rising market trend and a stable mining path, with reduced volatility from excluding boom–bust effects.

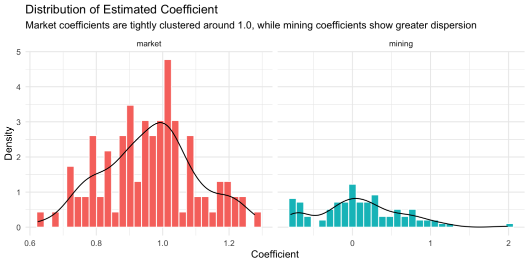




Static OLS Model

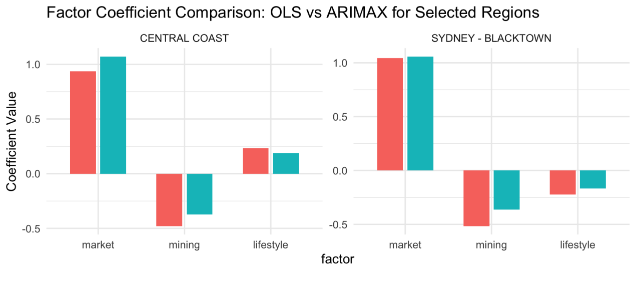
- Residuals show strong autocorrelation, reflecting the persistent nature of housing price movements.
- Model fit varies spatially, with mining-exposed regions showing weaker R^2 and more dispersed coefficients.
- Static coefficients limit temporal adaptability, failing to capture evolving links between regional and national factors.



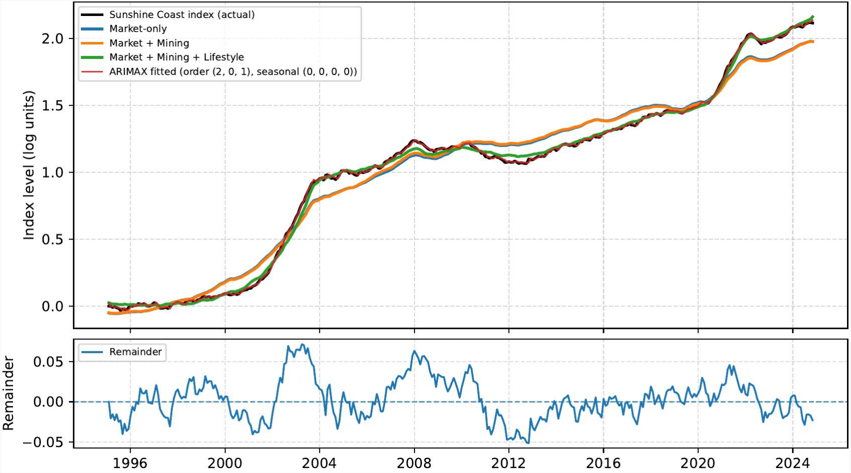


Model Enhancement: ARIMAX

- ARIMAX extends ARIMA by adding exogenous regressors and modeling serial dependence, with orders chosen via grid search. Five models were fitted; 75 of 88 regions passed diagnostic checks.
- Under ARIMAX, lifestyle effects weaken slightly in lifestyle regions, mining effects soften in mining regions, while market coefficients remain largely stable.



Model Enhancement: ARIMAX



Model Enhancement: Moving Window

- We extended the sample to test stability in factor–housing relationships as new data was added.
- OLS was refitted using data up to Nov 2022 and Nov 2024 to assess coefficient changes.
- Intercept and market effects shifted slightly; mining and lifestyle remained stable. A moving window could improve temporal robustness.

factor	mean_change	sd_change
intercept	0.0643	0.0103
lifestyle	0.0014	0.0083
market	0.0319	0.0054
mining	-0.0005	0.0026

Summary of Process

Pipeline

1. Input $\rightarrow$ Regional price indexes $\rightarrow$ Factors
2. PCA $\rightarrow$ Extract latent factors $\rightarrow$
   - ARIMA $\rightarrow$ Factor-level ARIMA $\rightarrow$ Univariate forecasts
   - ARIMAX $\rightarrow$ Static OLS $\rightarrow$ Expanding-window ARIMAX $\rightarrow$ Dynamic forecasts
3. Output $\rightarrow$ Combine forecasts $\rightarrow$ Region-specific, time-varying predictions

Moving Onward

- **Limitations**: The current model effectively captures market and mining trends but excludes broader macroeconomic and social drivers such as interest rates and migration. It remains highly data-driven, with limited economic interpretation, and the 29-year time span restricts our ability to capture longer housing cycles or rare structural shifts.
- **Next steps**: We aim to refine the lifestyle factor to complete the multi-factor framework and incorporate key macroeconomic indicators, such as GDP, exchange rates, and interest rates to strengthen both forecast accuracy and interpretability.