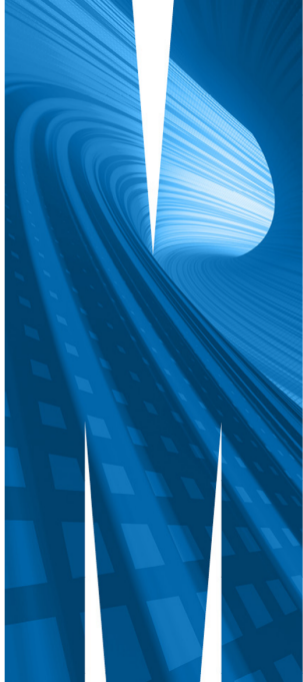


# **Regional Dynamics of Australian Housing Indexes**

Yiran Yao

23 October 2025



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

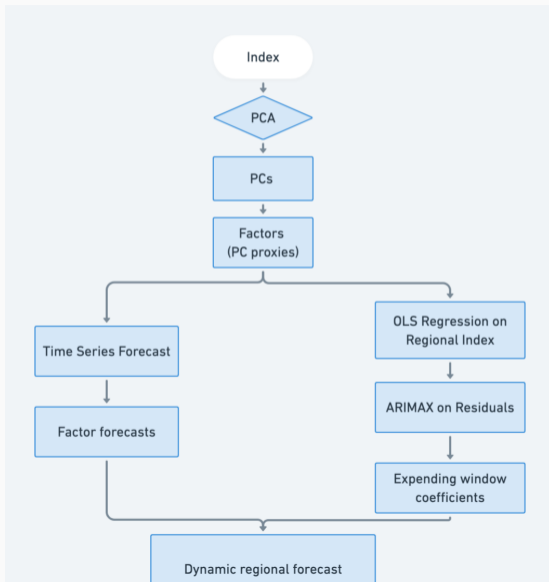
- How can we project housing prices at the local level?
- What drives price fluctuations over time?
- Which regions will gain desirability or investment potential in the coming decade?

# Problem

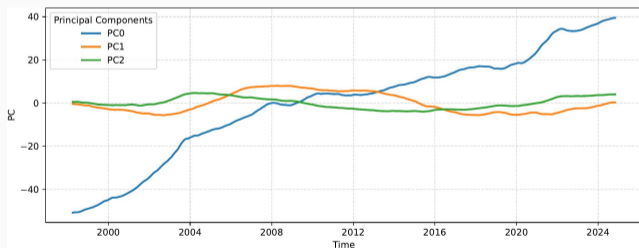
Existing approaches are effective but depend on diverse data sources that require extensive preprocessing and alignment with housing indexes.

- 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.

# Project Framework



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 first three principal components explain 96.5% of variation in housing indexes.



**Figure 2:** Top 3 principal components of regional housing indexes

If we look at those top 3 PCs more closely, we found some shared patterns across regions. We named them: **Market, Mining, and Lifestyle** factors.

**Table 1:** Top and bottom loadings for the first three principal components (PC1–PC3)

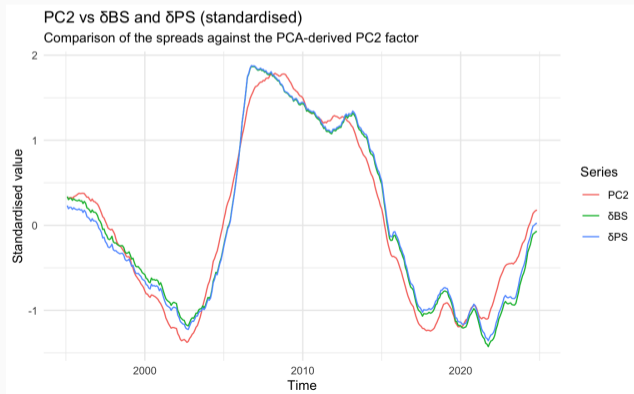
Major city	Mode	Loading	Major city	Mode	Loading	Major city	Mode	Loading
AUSTRALIAN CAPITAL TERRITORY	0	0.9958	REST OF WA	1	0.3467	REST OF TAS.	2	0.1295
GREATER ADELAIDE	0	0.9956	GREATER PERTH	1	0.2101	GREATER HOBART	2	0.1012
REST OF VIC.	0	0.9925	GREATER DARWIN	1	0.1495	REST OF NSW	2	0.0928
REST OF NT	0	0.9506	REST OF NSW	1	-0.0468	GREATER MELBOURNE	2	-0.0907
GREATER DARWIN	0	0.9465	GREATER MELBOURNE	1	-0.0749	GREATER DARWIN	2	-0.2104
REST OF WA	0	0.9055	GREATER SYDNEY	1	-0.2040	REST OF NT	2	-0.2280

# From PC to Factor Series

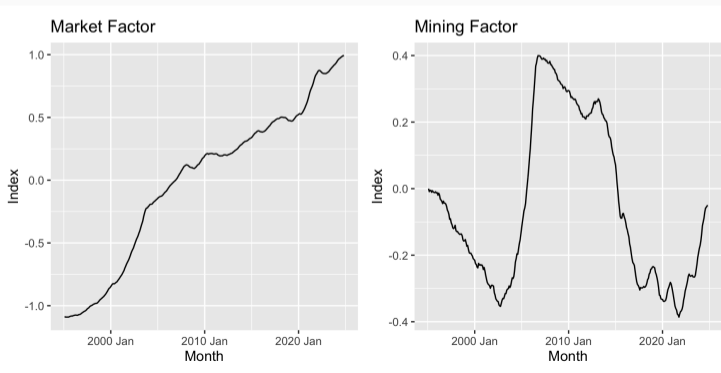
PCs were approximated as weighted averages of regional indexes for greater stability, interpretability, and modeling utility, while retaining **>96% of the original variance**.

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

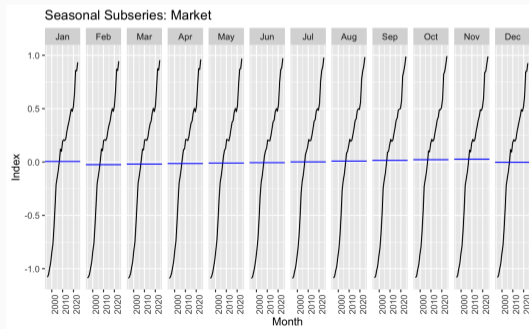
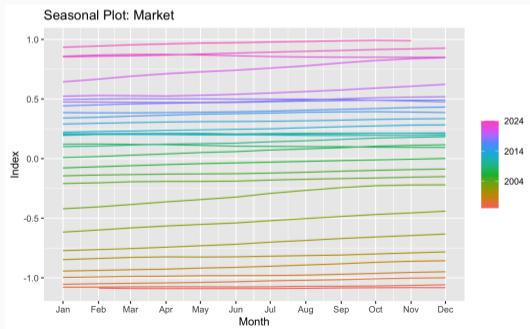
**PC2 proxy (Perth–Sydney adjusted spread):** capturing the difference between Perth and Sydney after removing their shared national market component.



- The market factor shows long-term growth, behaving like a random walk with drift.
- The mining factor rises, drops and fluctuates, reflecting resource-sector shifts.



- 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.

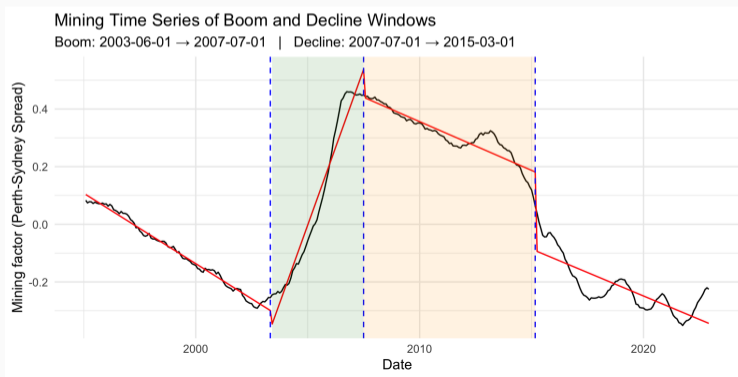


## Forecasting Factors: Market

- 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 Fourier term of 2** was selected for its simplicity, assumption alignment, and residuals showing no visible trend or seasonality. Fourier captured the **quasi-periodic cycles** that are not exactly the same length or amplitude each time.

# Forecasting Factors: Mining

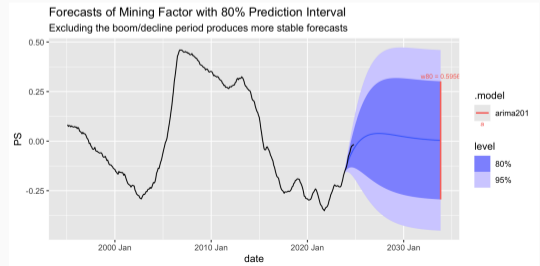
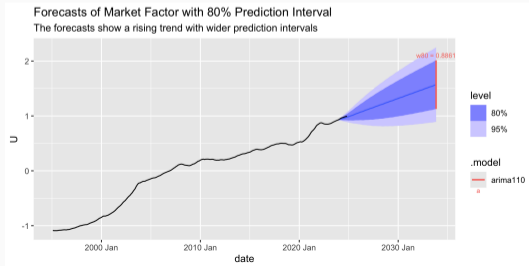
The mining factor **oscillated around zero with abrupt regime changes** during mining booms and busts. This was addressed using **dummy variables as exogenous regressors**. Residual dynamics were modeled with **ARIMA(2,1,1)**, which satisfied the Ljung-Box test.



# 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.

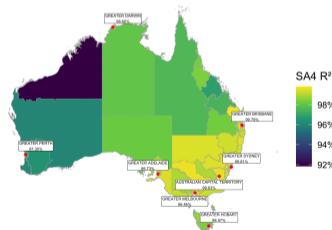


# Factor Regression: Link Regional Indexes to Factors

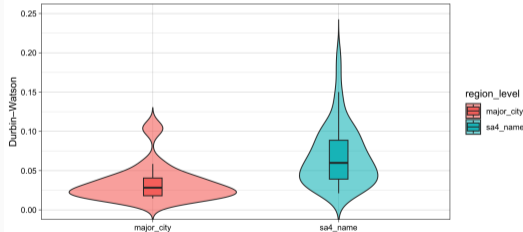
$$\mu_r(t) = \alpha_r + \beta_r M(t) + \lambda_r \delta_{PS}(t) + \gamma_r \delta_L(t) + \varepsilon_r(t)$$

**Static OLS regression:** linking each region's housing index  $\mu_r(t)$  to the national market factor  $M(t)$ , the mining factor  $\delta_{PS}(t)$ , and the lifestyle factor  $\delta_L(t)$ .

Factor Model R<sup>2</sup> Distribution across Cities & Regions

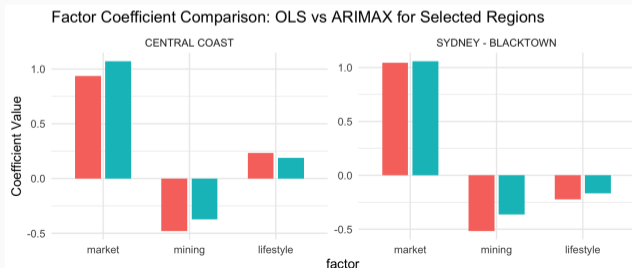


Durbin-Watson by region type

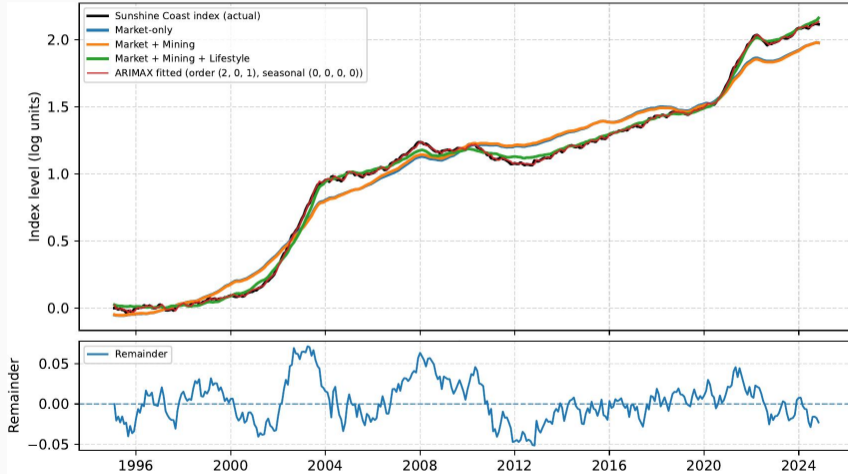


# Model Enhancement: ARIMAX

- 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

# Conclusion

We delivered a scalable, interpretable tool for understanding Australia's housing cycles beyond traditional region-by-region models.

- A unified factor-forecasting framework that links regional housing indexes to interpretable national and mining dynamics through PCA, ARIMA, and ARIMAX modelling.
- Region-specific forecasts with stable, data-driven coefficients that capture both structural trends and time-varying effects.