



# Regional Housing Market Dynamics in Australia

**Yiran Yao**  
Master of Business Analytics

MONASH  
BUSINESS  
SCHOOL

**Department of  
Econometrics &  
Business Statistics**

📞 (03) 9905 2478  
✉️ BusEco-Econometrics@monash.edu

ABN: 12 377 614 012

Report for  
Neoval

3 November 2025

## 1 Introduction

Housing markets are influenced by a complex interplay of regional, social, and macroeconomic factors that interact over time and across space. Prior research has shown that these dynamics often co-move through common systematic forces, such as macroeconomic cycles and financial conditions, motivating the use of factor-based models to capture shared market behaviour (Case, Cotter & Gabriel 2010). However, while approaches such as region-specific time series models, multivariate regressions, or machine learning methods can account for these dependencies, they are often computationally intensive and may obscure the underlying drivers of price variation.

To address these challenges, Sijp, Panagiotelis, and Francke (2025) (Sijp, Panagiotelis & Francke 2025) propose an analytical framework that applies Principal Component Analysis (PCA) to regional housing price indexes. This data-driven approach systematically extracts the principal sources of variation in housing prices, providing a more interpretable and efficient means of identifying the structural forces shaping regional housing markets.

The PCA results show that the first three principal components capture much of the behavior of local price indexes, as a linear combination of their time series explains most of the variance in the housing index. Nevertheless, the PCA-based linear model lacks robustness across different time windows, and its coefficient estimates must be derived directly from the PCA procedure.

Our exploratory research extends the PCA approach by addressing these limitations through a factor linear model, which uses time series data directly, with independent variables serving as proxies for the PCA-derived principal components. We model the factor proxies as time series using forecasting techniques to obtain their projected values. The forecasted factor series are then integrated into a linear modelling framework to produce projected housing price indexes. An ARIMA specification is employed to account for residual autocorrelation, while time-varying coefficients are introduced to reflect changes in regional dynamics over time.

The resulting factor model can be applied for a variety of purposes, but our central aims is to **assess the regional housing market sensitivity to national and sectoral trends**. By quantifying how each region responds to national, mining, and lifestyle factors, the model enables identification of regions most exposed to structural trends, such as shifts in mining investment or lifestyle migration. These outputs can inform policy analysis, highlighting areas vulnerable to affordability pressures or cyclical volatility, and support investment decisions by revealing which regions are likely to outperform or lag under different macroeconomic conditions.

## 2 Methodologies and Datasets

### 2.1 Methodologies

#### Principal Component Analysis

Principal Component Analysis (PCA) was initially applied to the original housing price indexes at the Statistical Area Level Two (SA2) regional level, after which the results were aggregated to broader geographic scales as required. It is useful to briefly revisit PCA and how it allowed us to identify specific factors. The original analysis was conducted by Will Sijp, the supervisor of this project, with comprehensive details provided in the referenced works. The following section presents a brief summary of his methodology.

PCA is a statistical method that reduces complex datasets into a smaller number of uncorrelated components, each capturing a dominant source of variation. In our project, the PC1 explains 96% of the variance, while PC2 and PC3 account for 2.5% and 0.7%, respectively. Together, these three components capture nearly all the variability in the housing indexes. Figure 1 illustrates the temporal evolution of the first three principal components derived from the PCA of regional housing price indexes, notably, only PC1 exhibits a clear upward trend, while PC2 and PC3 display mean-reverting dynamics.

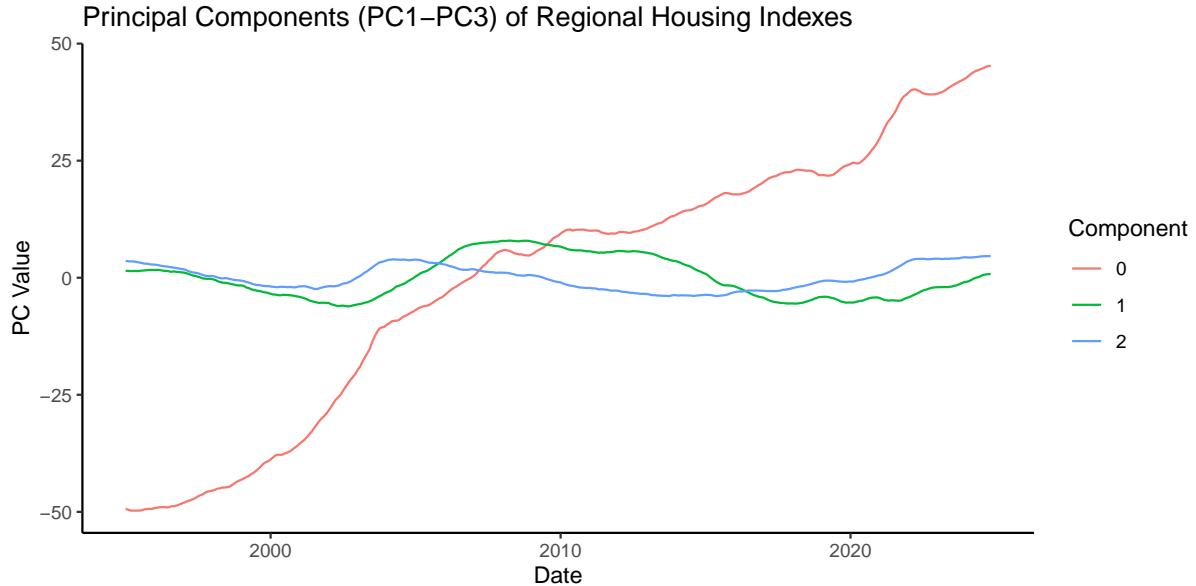
The loadings from the PCA highlight which regions contribute most strongly to each component:

The first principal component (PC1) exhibits uniformly positive loadings across almost all regions, capturing the dominant national housing trend. It reflects broad macroeconomic influences such as monetary policy, interest rates, and aggregate demand that drive housing prices nationwide. As this component explains the majority of total variance in the dataset, it can be interpreted as the “market” factor, representing the shared cyclical movement of housing prices across Australia’s regions.

The second principal component (PC2) displays strong positive loadings in mining-intensive regions such as Western Australia, Queensland, and the Northern Territory, and negative loadings in service-based urban centres like Sydney and Melbourne. This spatial pattern suggests that PC2 represents housing market behaviour linked to the mining industry, characterised by rapid price growth during resource booms and extended periods of decline afterward. It highlights the cyclical and volatile behaviour of regions reliant on mining compared with those driven by other sectors.

The third principal component (PC3) shows strong positive loadings in coastal and amenity-rich areas such as the Central Coast, Noosa, and the South Coast of New South Wales, and negative loadings in inner-city regions. This spatial distribution reflects migration from metropolitan centres toward lifestyle-oriented destinations, particularly during periods of high urban housing costs or increased

remote working. PC3 therefore captures the lifestyle-driven movement toward regional housing, influenced by affordability, natural environment, and quality-of-life factors.



**Figure 1:** The first three principal components summarise the dominant sources of systematic variation in regional housing price dynamics.

### Factor-Based Substitutes for Principal Components

The PCA-derived components provide a compact statistical summary of regional housing dynamics, but their direct use in forecasting can be challenging, as principal components may shift when the estimation window changes. To address this, we construct a set of factor proxies that reproduce the behaviour of the PCA components using simple and economically interpretable index combinations.

Comparable efforts to model spatial and temporal dependencies in regional housing prices include the work of Valentini, Ippoliti, and Fontanella (2013) (Valentini, Ippoliti & Fontanella 2013), who employed a spatial dynamic structural equation model to extract latent factors underlying U.S. housing markets. Although their approach is based on dynamic latent factors rather than Principal Component Analysis (PCA), both methodologies share the aim of dimensionality reduction and identifying the dominant sources of variation across regions.

Each regional index  $\mu_r$  is approximated as a linear combination of these factors and an idiosyncratic residual:

$$\mu_r = \beta_r U + \lambda_r \delta_{PS} + \gamma_r \delta_L + \epsilon_r \quad (1)$$

Equation 1  $U$  denotes the market (national) factor,  $\delta_{PS}$  the Perth–Sydney spread,  $\delta_L$  the lifestyle spread, and  $\epsilon_r$  the residual term. As lifestyle's influence is more subtle and region-specific, it is

generally treated together with the residual term, and only highlighted in areas where lifestyle dynamics play a significant role. This formulation replaces abstract PCA components with observable spreads between city or regional indexes that capture similar variation patterns. Three factors are defined:

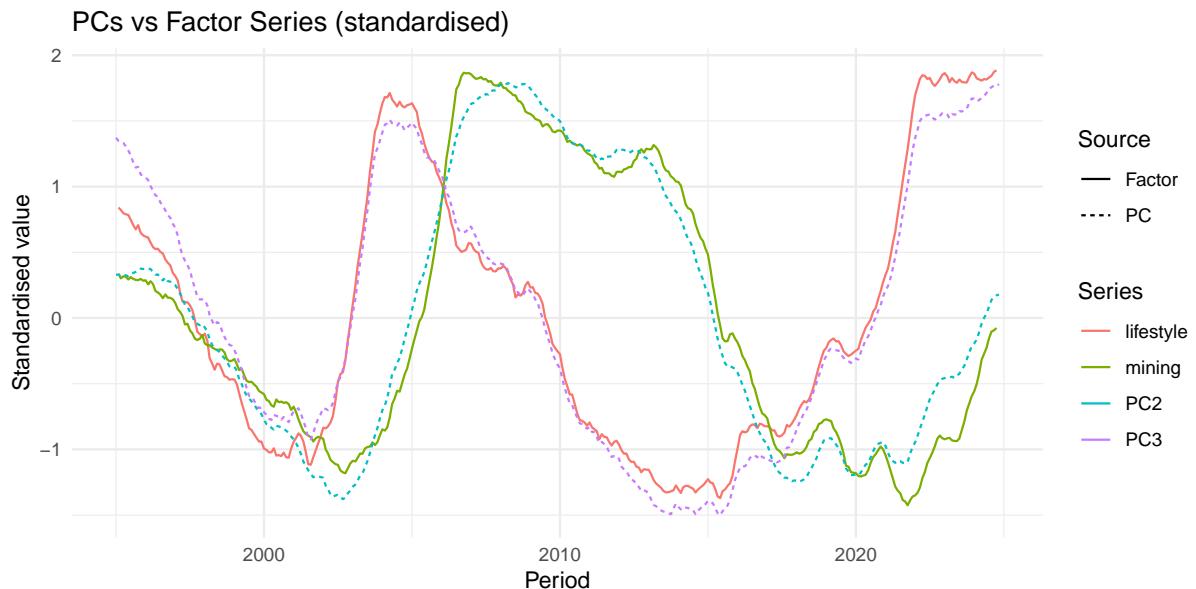
**Market factor** (aligned with PC1): represented by the national housing index  $U$ , capturing the broad macroeconomic trend common to all regions.

**Mining factor** (associated with PC2): proxied by the Perth–Sydney spread, defined as

$$\delta_{PS} = \mu_P - \alpha\mu_S \quad (2)$$

Equation 2  $\mu_P$  and  $\mu_S$  are the log indexes for Greater Perth and Greater Sydney respectively.  $\alpha$  is a scaling coefficient chosen to remove the shared national trend. It is estimated as  $\alpha = \frac{\text{cov}(\mu_P, U)}{\text{cov}(\mu_S, U)}$ , yielding a value of 0.94. This adjustment isolates the cyclical movements of mining-driven regions relative to service-based economies.

**Lifestyle factor** (corresponding to PC3): constructed by contrasting the average of high-amenity coastal and regional areas with that of major metropolitan centres, thereby capturing migration and demand shifts toward lifestyle-oriented markets. The scaling coefficient is established to be 0.82 using the same covariance-based approach as above.



**Figure 2:** Comparison of PCs 2 and 3 with their corresponding factor series reveals highly similar dynamics, with correlations exceeding 98%. PC1 and factor  $U$  are identical by construction.

These proxies achieve correlations above 98% with their corresponding principal components and display closely aligned patterns Figure 2, indicating that they preserve the explanatory strength of

the PCA while enhancing interpretability and temporal stability for use in regression and forecasting.

### Forecasting of Factors

Once the factor series were defined, we modelled each factor series separately using time series forecasting methods. The process included exploratory analysis, stationarity checks, and STL decomposition to characterise the data. Guided by these diagnostics, we selected suitable ARIMA-type models, enhancing them with Fourier terms to capture seasonality and intervention dummies to address structural breaks.

### Integration into a Factor Modelling Framework

The linear model regresses each regional housing index on the factor proxies, yielding coefficients that represent the sensitivity of local markets to the underlying factors, the coefficients are estimated for each region  $r$ . These coefficients help to capture the response of the local market to each factor. Model performance was evaluated using  $R^2$  and residual diagnostics to check that key assumptions held. Forecasts of the factor proxies were then substituted into the model to obtain projected housing indexes.

### Factor Model Enhancement

Residual diagnostics from the static factor model indicated autocorrelation, suggesting that regional price dynamics were not fully captured by contemporaneous factors alone. To address this, the model was extended using an ARIMAX framework, retaining the same factor proxies as exogenous variables while incorporating autoregressive and moving-average terms to account for temporal dependence. An expanding-window approach was adopted to re-estimate models as new data became available, enabling coefficients to adapt to evolving market conditions and improving both model stability and forecast accuracy.

## 2.2 Use Case

The proposed model establishes a robust analytical framework for examining and forecasting regional housing market dynamics across Australia. By quantifying the sensitivity of each region to national, mining, and lifestyle factors, it facilitates the identification of areas most exposed to long-term structural forces. Furthermore, the model enables an assessment of spatial divergence within the national housing system by evaluating the degree of co-movement between regional and aggregate market trends, thereby elucidating whether regional growth patterns exhibit systematic and predictable behaviour over time.

## 2.3 Datasets

The datasets used in this study are provided by Neoval Pty Ltd. and consist of housing price indexes across Australian regions. These indexes are derived from state valuer general transaction records covering about three million detached house sales between January 1995 and November 2024, including sale prices, dates, and property coordinates. Rather than analysing the raw transactions directly, local repeat-sales indexes are constructed and examined using a three-factor model, with both regional indexes and factors represented as time-series. Owing to confidentiality agreements, further details of the data collection process cannot be disclosed.

All supplementary datasets are derived directly from these housing price indexes and underpin the principal analyses involving PCA and factor construction. Spatial boundary data from the Australian Bureau of Statistics (ABS) are additionally incorporated to support the geographic visualisation of results.

All datasets are organised and available in the data folder:

- Monthly log-transformed changes in housing price indexes at both the city and SA4 levels, named `city_indexes` and `sa4_indexes`
- PCA results including PC scores and loadings, named `df_pcs`, `df_eofs_city`, and `df_eofs_sa4`
- Mean-adjusted factor series derived from the PCA results, named `df_factor_trends`
- Regression coefficients estimated from the factor proxies, named `df_reg_coefs_new`
- Geographic data for SA4 regions, named `SA4_2021_AUST_SHP_GDA2020.shp`

## 2.4 Acknowledgements

This project was conducted under the supervision of Will Sijp from Neoval Pty Ltd, whose expertise in econometrics and housing market analysis was instrumental in shaping the research. Sijp developed the Principal Component Analysis (PCA) framework and factor construction methodology, introduced the underlying formulas, and reviewed the implementation of the factor series. He also provided guidance on the time-series modelling strategy, including the selection of the ARIMAX model and the expanding-window forecasting approach, and continuously advised on the direction and objectives of the research.

The analysis and report preparation were carried out by Yiran Yao, including all R coding, exploratory analysis of the factor series, implementation, selection, and evaluation of time-series forecasting models, static factor and ARIMAX model estimation, and visualisation of results. The written report was authored solely by Yiran Yao.

We also acknowledge Monash University for providing computational resources and academic support throughout the research process.

Finally, parts of the report text were refined using OpenAI's ChatGPT (GPT-5, 2025) to improve clarity and structure [openai2025chatgpt], and minor coding issues were resolved with its assistance.

## 2.5 Project Limitation

This project faces several dataset-related limitations, including but not limited to the following:

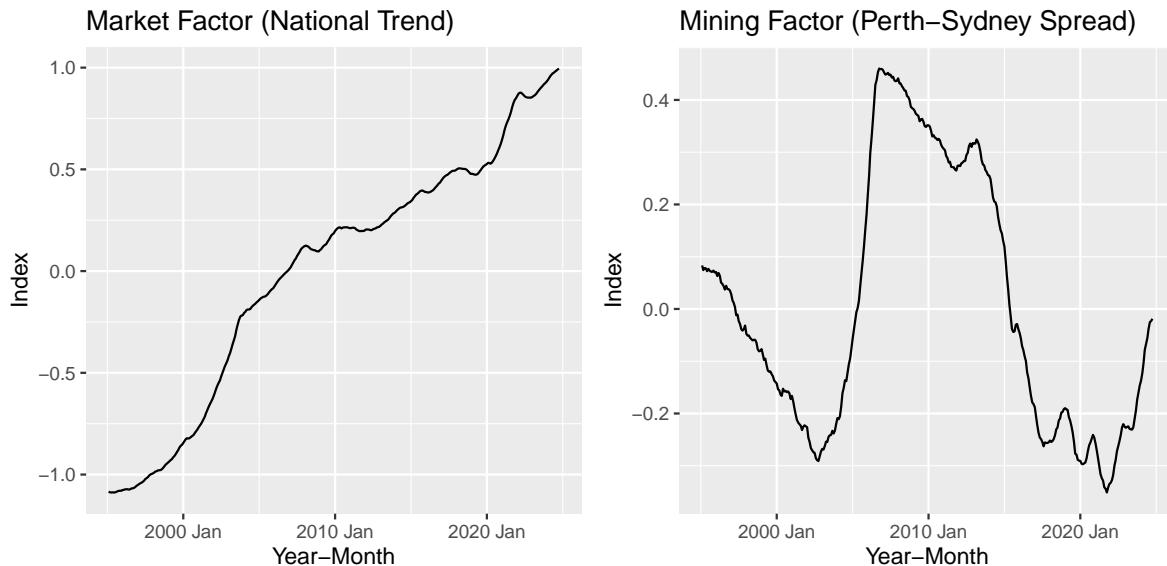
- Historical Scope: PCA results and coefficients are based on 29 years of data (1995–2024), which may not fully reflect future patterns.
- Measurement Bias: Indexes rely on hedonic and repeat-sales methods, which involve assumptions that may introduce bias.
- Proxy Validity: Factor proxies are heuristic choices and may not perfectly represent the underlying principal components.
- Data Sensitivity: Outcomes are highly data-driven and depend strongly on the chosen datasets and methods.
- Structural Shocks: Events like the mining boom, COVID-19, or policy shifts may not be well captured by static models or PCA.
- Macroeconomic Variables: The model does not explicitly incorporate broader economic indicators, which could enhance explanatory power.

## 3 Exploratory Analysis

The exploratory analysis employs a series of visualisations to examine the dynamics of the national and mining trends, which constitute the primary focus of this project. Autoplot functions are first used to illustrate the overall temporal evolution of these factors, followed by seasonal and sub-seasonal plots to identify potential cyclical or recurrent patterns. Together, these visual analyses provide insights into the structural characteristics of the series, forming a basis for subsequent forecasting model design and selection.

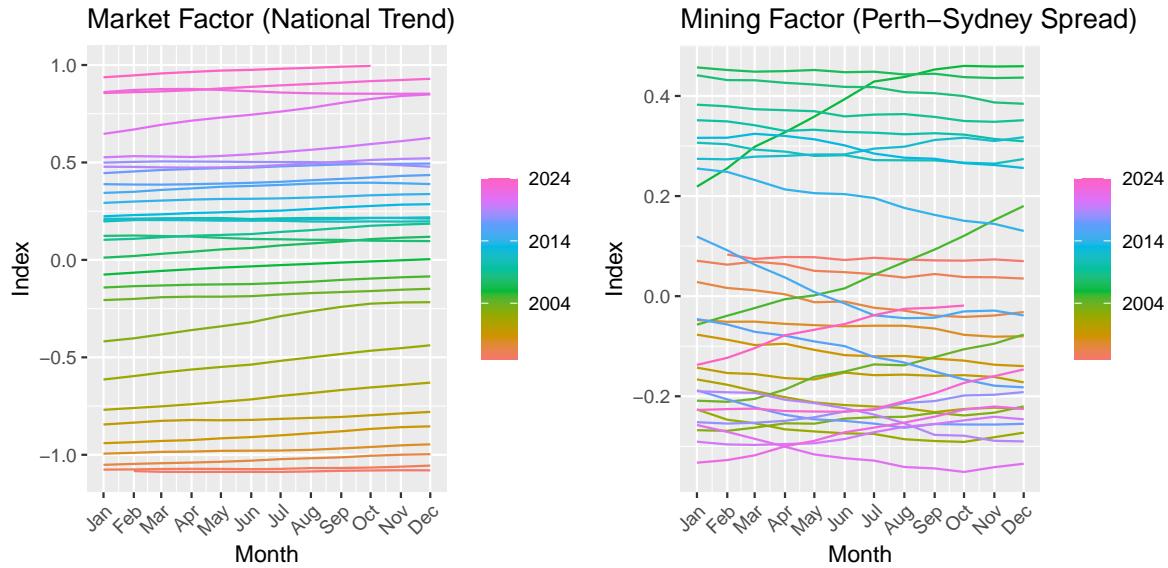
Figure 3 illustrates that the market factor exhibits a sustained upward trajectory from 2003 to 2012, followed by a moderated growth phase extending to 2019. A pronounced regime shift occurs during 2020–2022, coinciding with the COVID-19 period, after which the series remains at elevated yet

volatile levels, with growth rates continuing to vary over time. In contrast, the mining factor displays a more uneven and protracted increase from 2003 to 2012, succeeded by a sharp structural adjustment and subsequent COVID-related fluctuations, reflecting continued time-varying behaviour in both level and volatility.



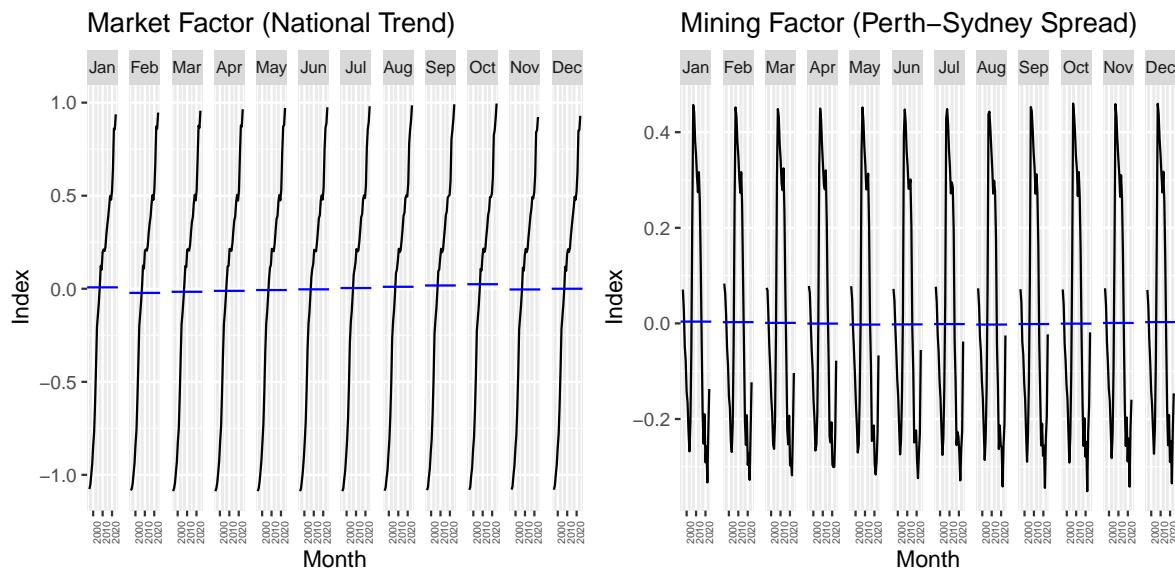
**Figure 3:** Temporal evolution of the market and mining factors. Both series display distinct patterns and structural shifts.

Figure 4 illustrates that the market factor displays pronounced within-year gradients primarily driven by a persistent long-term upward trajectory, with no systematic month-of-year seasonality evident. In contrast, the mining factor exhibits minimal intra-annual variation, showing no consistent pattern across months. However, its overall level fluctuates notably between years, indicating that longer-term structural influences rather than seasonal effects govern its dynamics.



**Figure 4:** Seasonal patterns of the market and mining factors, showing no consistent month-of-year effects and highlighting trend-dominated behaviour.

Figure 5 demonstrates that each monthly panel of the market factor exhibits a consistent upward trajectory, with mean values across months remaining largely stable. This pattern indicates that the series is primarily trend-driven rather than influenced by seasonal effects. Similarly, the mining subseries plot reveals no pronounced seasonal structure, as the monthly profiles maintain comparable shapes and the corresponding means remain relatively flat and centred near zero, underscoring the absence of systematic month-specific variation.



**Figure 5:** Subseries plots of the market and mining factors, indicating stable monthly profiles and the absence of systematic seasonal variation.

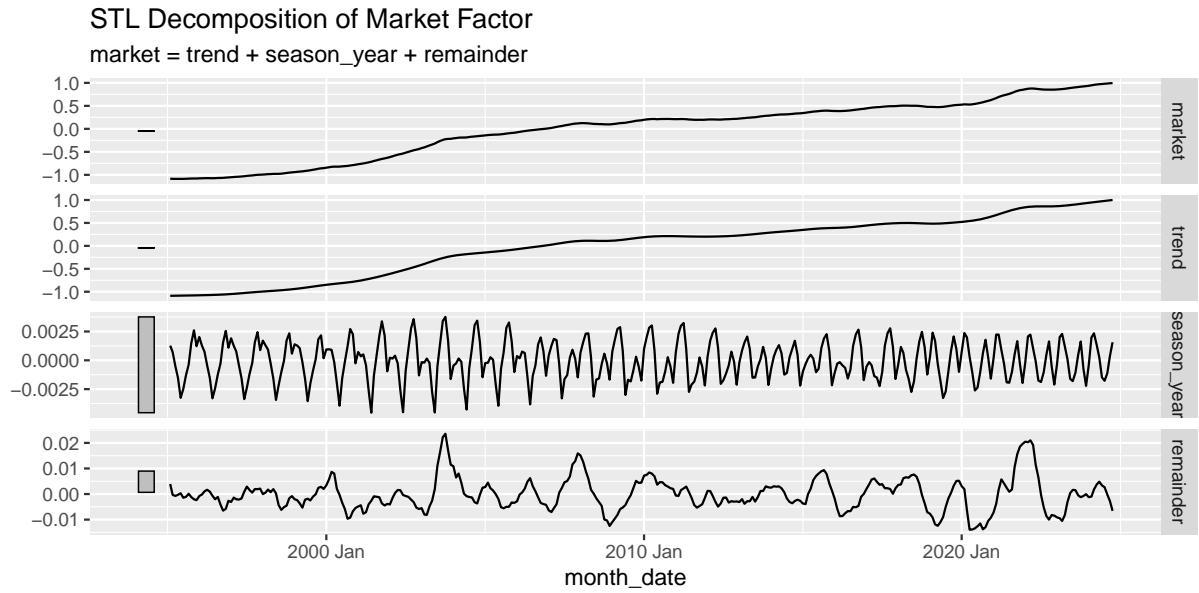
## 4 Time Series Forecasting

Following the exploratory analysis, this section develops time series forecasts for the extracted factors, which are later used as explanatory inputs in the regional factor model. Our approach consists of three main steps:

- Visual inspection and preliminary diagnostics: We begin with graphical and statistical exploration of each factor using STL decomposition, autocorrelation (ACF), and partial autocorrelation (PACF) plots. These tools help identify underlying trend and seasonal components, potential structural breaks, and the degree of persistence in each series.
- Model estimation and comparison: Several competing forecasting models are fitted to each factor, including benchmark methods and parametric time series models. Model selection is guided by information criteria and diagnostic checks on model adequacy.
- Residual analysis and forecast evaluation: The residuals from each fitted model are examined to verify the absence of autocorrelation and non-stationary patterns. Forecast performance is then assessed using the holdout sample (one year of observations), ensuring that the chosen models generalise well beyond the training period.

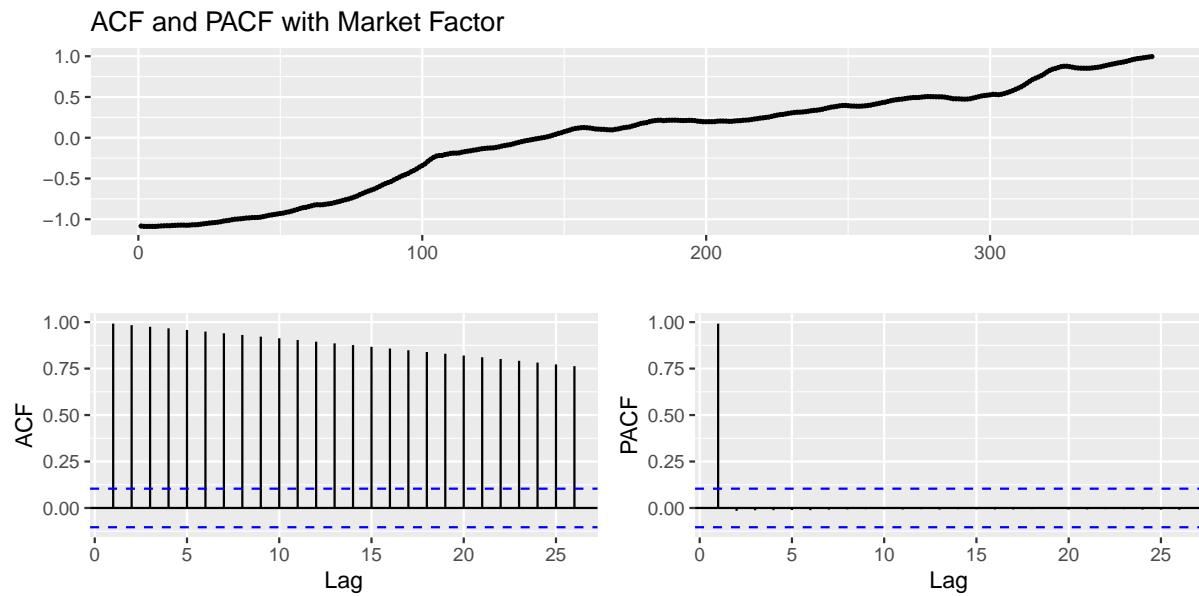
### 4.1 Forecasting the Market Factor

The STL decomposition of the market factor Figure 6 reveals a smooth and persistent upward trajectory in the trend component, while the seasonal fluctuations are of relatively minor magnitude. The remainder term is dominated by low-level noise with occasional shocks, aligning with major economic events such as the Global Financial Crisis (2008), the end of the mining investment boom (2012), and the COVID-19 pandemic (2020). These observations suggest that the market factor is primarily trend-driven, with limited regular seasonality but distinct structural shifts.



**Figure 6:** *STL decomposition of the market factor, highlighting a dominant upward trend with minor seasonal fluctuations and occasional shocks.*

The ACF and PACF plots of the national market factor (Figure 7) reveal strong persistence and a slow decay in autocorrelation, characteristic of a highly non-stationary process dominated by long-term trends rather than cyclical variation. The ACF remains positive and close to unity across multiple lags, while the PACF displays a sharp initial spike followed by gradual tapering, confirming the need for differencing to achieve stationarity. Accordingly, the differencing order is fixed at  $d = 1$ , both to remove the stochastic trend statistically and to remain consistent with standard economic interpretations, where housing price levels are typically modelled as integrated of order one  $I(1)$  processes reflecting cumulative growth over time as discussed by Hyndman and Athanasopoulos (2021) (Hyndman & Athanasopoulos 2021).



**Figure 7:** ACF and PACF plots of the market factor, indicating strong persistence and non-stationarity dominated by long-term trends.

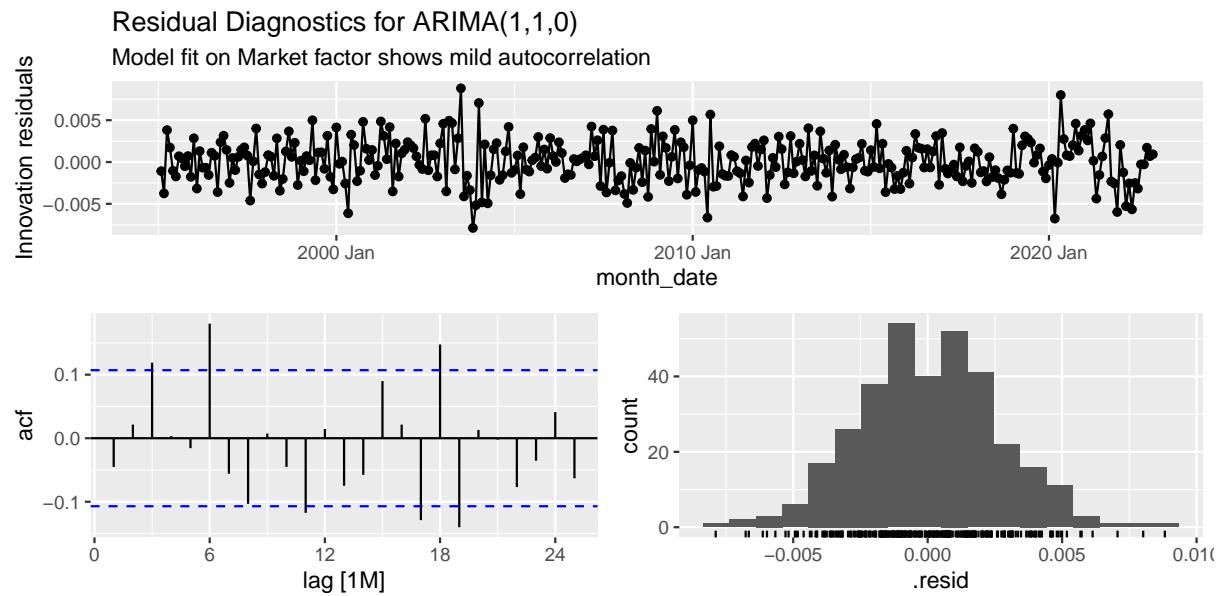
In summary, the market factor exhibits non-stationary behaviour, dominated by a sustained upward trajectory and intermittent fluctuations across the observation period. To address the stochastic trend, the differencing order is set to one. A random walk with drift is employed as the benchmark model, with additional ARIMA variants-incorporating seasonal and piecewise structures-evaluated for comparison. Fourier terms are included to account for irregular yet recurring variations that lack a consistent seasonal pattern.

**Table 1:** Candidate forecasting models for the market factor with fit statistics and residual diagnostics.

Model	AICc	Ljung.Box.p.value	Consideration
autoarima	-3001.47	NA	Automatic ARIMA selection benchmark.
rw_drift	-2445.04	0.0e+00	Baseline random walk with drift capturing persistent trend.
arima110	-3005.60	3.0e-04	ARIMA(1,1,0) with Fourier ( $K = 2$ ) for trend and weak high-frequency variation.
arima111	-3004.12	2.0e-04	ARIMA(1,1,1) with Fourier ( $K = 2$ ) capturing moderate short-term dependence.
arima211	-3002.03	6.0e-05	ARIMA(2,1,1) with Fourier ( $K = 2$ ) modelling complex short-term dynamics.
sarima211	-2985.13	6.7e-03	Seasonal ARIMA extension with yearly persistence term.
piecewise	-2981.34	3.0e-04	Piecewise specification capturing structural breaks (2008, 2012, 2020).
step	-3011.58	1.3e-06	ARIMA(2,1,1) with Fourier and COVID-19 step/pulse interventions.

The candidate model specifications and their corresponding fit diagnostics are summarised in Table 1. Among these, two models demonstrated notably strong performance. The first is an ARIMA(1,1,0) model with Fourier terms ( $K = 2$ ) and a drift component, while the second combines Fourier terms ( $K = 2$ ) with step and pulse intervention dummies within an ARIMA(2,1,1) structure.

After accounting for model-specific degrees of freedom, all models exhibit some degree of residual autocorrelation according to the Ljung–Box test. This outcome is expected, as the models are primarily designed to capture broad structural and cyclical patterns rather than every short-term fluctuation. Attempting to eliminate all autocorrelation would risk overfitting and reduce the models' generalisability. Moreover, any remaining autocorrelation can be more effectively addressed in downstream modelling stages, such as ARIMAX or multivariate regression frameworks.

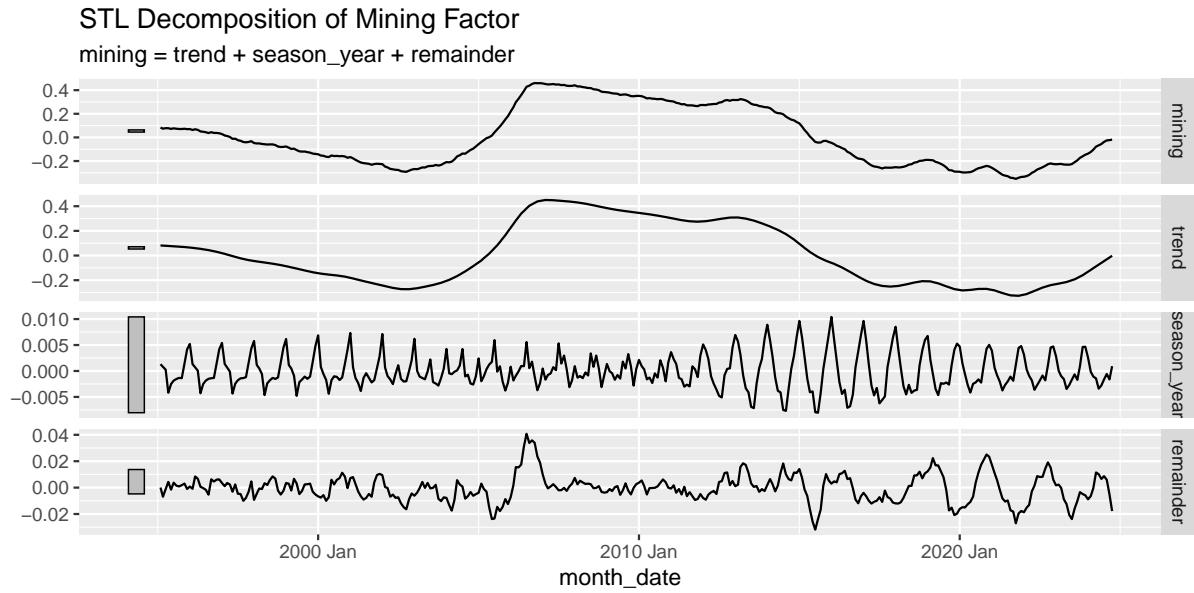


**Figure 8:** Residual diagnostics for the ARIMA(1,1,0) model fitted to the market factor, showing no significant autocorrelation and a symmetric distribution.

The ARIMA(1,1,0) with Fourier terms and drift is ultimately selected as the preferred specification. Among the models without intervention dummies, it achieves the lowest AICc while maintaining a parsimonious structure, thereby enhancing interpretability. The exclusion of explicit seasonal terms aligns with the earlier conclusion that seasonality is weak and not systematically identifiable in this context. Residual diagnostics Figure 8 further support this choice, as the residual series displays no significant or persistent autocorrelation, with most spikes within the significance bounds and a roughly symmetric, zero-centred distribution.

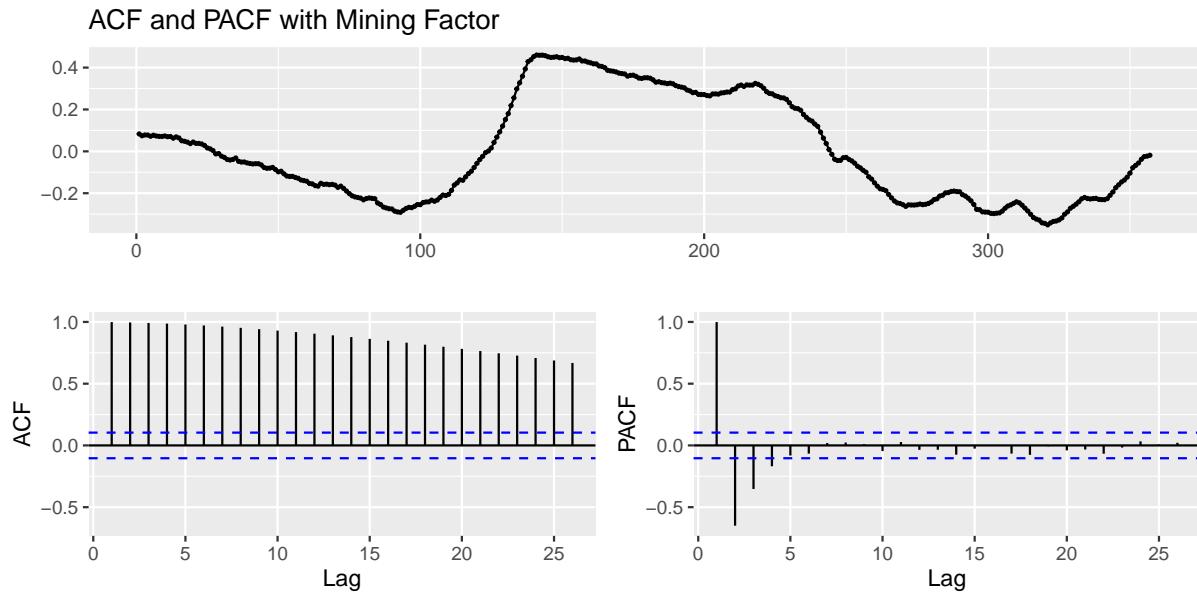
## 4.2 Forecasting the Mining Factor

The mining factor fluctuates around zero and exhibits mean-reverting behaviour, indicating the absence of a persistent long-term trend. As shown in Figure 9, seasonal effects are present but remain relatively consistent across years, while broader multi-year fluctuations highlight the need for a more flexible seasonal specification. The series reflects the dynamics of the resource boom, which can be interpreted as a one-off structural adjustment rather than a permanent driver—characterised by a sharp surge during the 2000s mining expansion, a collapse around 2012, and subsequent stagnation in the following years.



**Figure 9:** STL decomposition of the mining factor, revealing mean-reverting behaviour with a significant structural shift during the mining boom period.

The ACF and PACF of the mining factor Figure 10 show slow, oscillatory decay, indicating medium-term persistence and cyclical behaviour rather than a strong trend. The PACF's alternating signs suggest a damped cycle consistent with mean reversion around a stable long-run mean. This supports modelling the series with low differencing and flexible cyclical terms to capture mining-related fluctuations.



**Figure 10:** ACF and PACF plots of the mining factor, indicating potential non-stationarity influenced by structural shifts.

The preceding exploratory plots provide mixed evidence regarding the stationarity of the mining factor, leaving its underlying stochastic properties uncertain. Although the mining factor appears

to exhibit mean-reverting behaviour, the presence of a prolonged structural shift suggests potential non-stationarity. To formally assess this, we applied the Augmented Dickey–Fuller (ADF) and KPSS tests, along with the Zivot–Andrews (ZA) test, to the mining factor.

The ADF and KPSS tests evaluate the presence of a unit root and stationarity around a deterministic level or trend. As shown in Table 2, both tests failed to reject the null hypothesis, indicating that the mining factor is non-stationary. However, the Zivot–Andrews test, which accounts for a single endogenous structural break, rejected the null of a unit root, suggesting that the mining factor becomes stationary once structural breaks are considered.

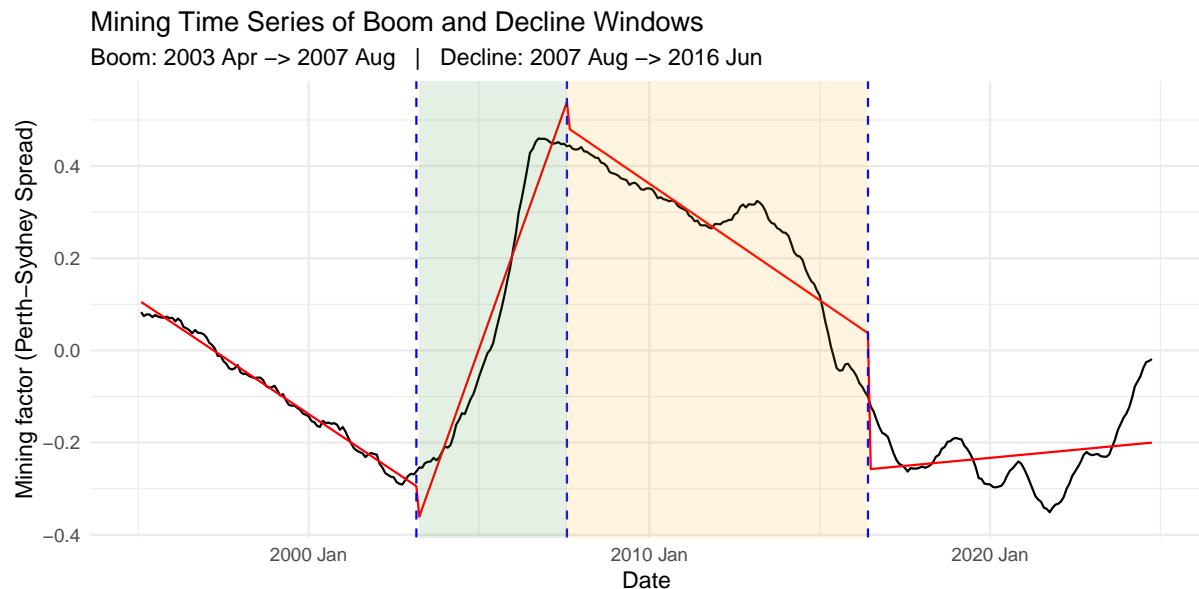
**Table 2:** Results of unit root and stationarity tests for the mining factor. Zivot–Andrew test shows stationarity after accounting for structural breaks.

**Table 2:** Unit root and stationarity test results for the mining factor.

Test	Statistic	p_value	Interpretation
Augmented Dickey–Fuller	-2.0117	NA	Non-stationary (fail to reject H0)
KPSS	1.1128	0.01	Non-stationary (reject H0)
Zivot–Andrews	-4.9751	NA	Stationary with one structural break (reject H0)

To address this, we treat the boom as a temporary structural shock. We introduce a dummy variable (set to 1 during the boom/decline period and 0 otherwise) when modelling the mining factor, and then forecast only the residual stationary component. This approach effectively assumes that another boom of comparable scale is unlikely.

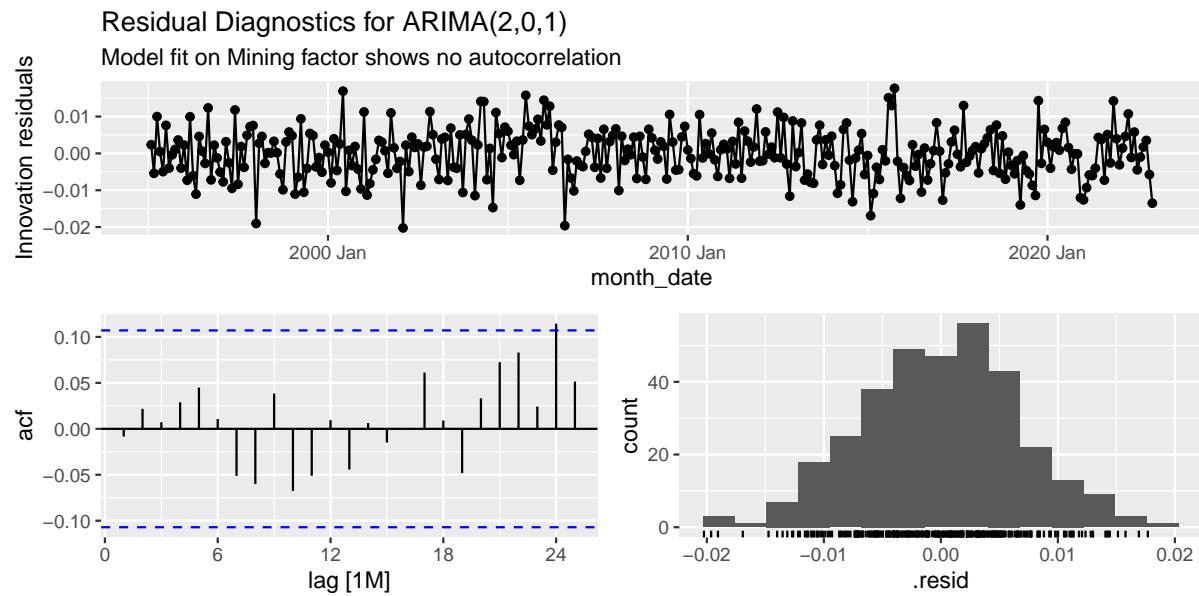
The boom and decline phases were identified using the `breakpoints` function from the `strucchange` package, which detected three structural breaks in the series. The segment containing the global peak was classified as the “boom” period, followed by the “decline” period. These periods were then encoded into dummy variables for use in the modelling process, essentially fitting an ARIMA model to the residuals after accounting for the boom and decline periods.



We estimated multiple ARIMA specifications for the mining factor, incorporating boom and decline dummies as exogenous regressors. Model selection was guided by visual inspection of the ACF and PACF plots, as well as prior knowledge of the factor's cyclical and mean-reverting characteristics.

Among all candidate models, only the ARIMA(2,0,1) specification converged successfully and yielded well-behaved residuals. Simpler autoregressive models, such as AR(1) or AR(2), violated the stationarity condition, while the inclusion of the MA(1) term in ARIMA(2,0,1) stabilised the process and ensured that the AR coefficients remained within the stationary region. The MA(1) component also captured short-term fluctuations not explained by the boom and decline dummies, resulting in residuals that exhibited no significant autocorrelation. This is confirmed by the Ljung–Box test ( $p = 0.700$ ), indicating that the residuals can be regarded as white noise. In contrast, higher-order or alternative specifications either failed to converge or produced serially correlated residuals.

The residuals are well-behaved, fluctuating randomly around zero with no autocorrelation, confirmed by ljustng-box test. Their near-normal distribution suggests the model captures the mining factor's dynamics effectively.



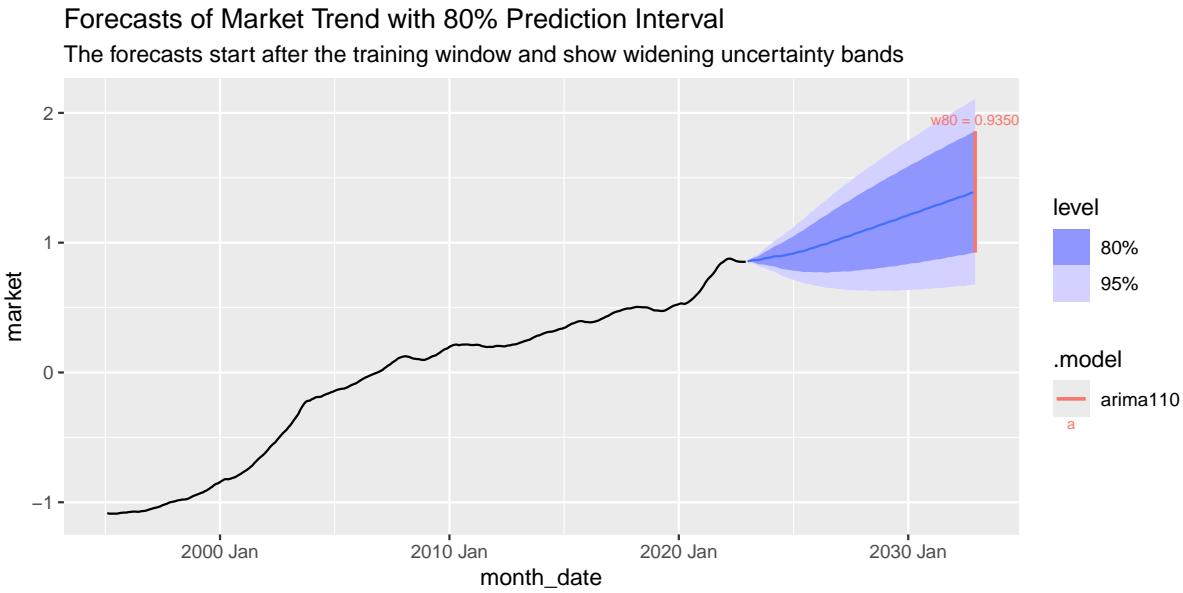
**Figure 11:** Residual diagnostics for the ARIMA(2,0,1) model fitted to the mining factor, showing no significant autocorrelation and a symmetric distribution.

### 4.3 Forecast Performance

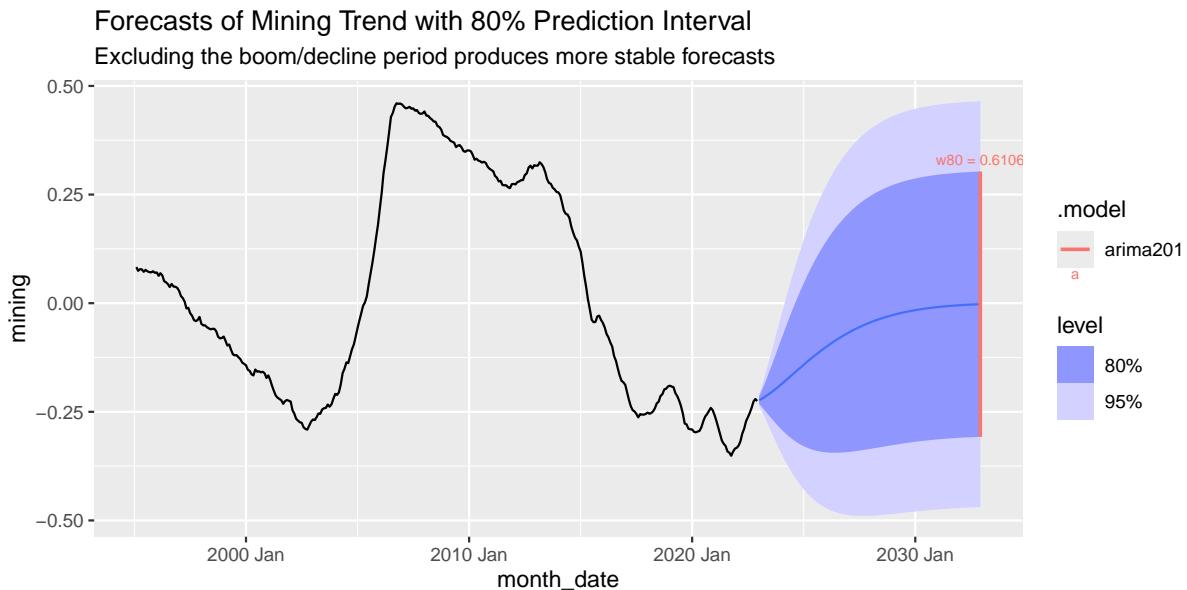
For the forecasting stage, we use the best-performing models (ARIMA(1,1,0) with Fourier terms ( $K = 2$ ) for the markets factor and ARIMA(2,0,1) for the mining factor identified earlier to project values 10 years ahead (120 months). We generate forecasts with prediction intervals at both 80% and 95%. These intervals show the range in which future values are likely to fall, giving us a measure of uncertainty, an 80% interval means there is an 8 in 10 chance that the true value will lie within that range.

For evaluation, we set aside one year of data as a test set. Although a typical split is about 20% of the dataset, our series only spans 29 years, so using 20% would cut off too much valuable information for forecasting. A one-year test period strikes a better balance.

The forecasts indicate that the market trend Figure 12 is expected to rise, with an 80% prediction interval width of 0.9350, reflecting higher uncertainty. In contrast, the mining spread Figure 13 shows a flatter path, with an 80% interval width of 0.6107. This narrower range reflects greater stability, largely because we excluded the one-off boom and decline period, which would otherwise make the series more volatile.



**Figure 12:** Forecasts of the market factor with 80% prediction intervals, showing a sustained upward trajectory accompanied by widening uncertainty bands.



**Figure 13:** Forecasts of the mining factor with 80% prediction intervals, showing stable projections due to exclusion of the boom/decline period.

The test results Table 3 show that the chosen model for the market trend performs well, with low error ( $\text{RMSE} = 0.056$ , small relative to the series scale). For the mining trend, the model with boom and decline dummies yields forecasts with an average error of about 7%, which is reasonable given the mining factor's higher volatility and cyclical nature. Despite minor residual autocorrelation, both models perform robustly overall.

**Table 3:** Forecast accuracy on the test set for market and mining factors, showing low RMSE and MAE values.

series	.model	.type	RMSE	MAE
Market	arima110	Test	0.0541345	0.0454564
Mining	arima201	Test	0.0737446	0.0557371

## 5 Factor Model Application

The linear decomposition of local price indexes into common factors (Equation 1) motivates a corresponding regression framework linking regional price indexes to these factors. This model aims to provide a simple yet informative representation of local housing performance and to quantify each region's sensitivity to national and mining influences. The resulting coefficients offer insight into regional responses to aggregate market dynamics and help evaluate the model's explanatory strength.

The model is estimated in a static form, where each region's sensitivity to the underlying factors is assumed constant across the full sample period. This simplification overlooks potential temporal variation in regional responses but nonetheless provides a valuable baseline for understanding long-term dynamics. The subsequent analysis examines whether this static specification remains stable and sufficiently flexible to approximate local house price movements across three key dimensions:

- Model fit:  $R^2$  measures how much of the variation in regional house prices is explained by the factors, where a higher  $R^2$  indicates that the model captures a larger share of systematic price movements across regions.
- Residual autocorrelation: Examining the residuals helps identify whether the model has adequately accounted for temporal dependencies. If residual autocorrelation is present, then some dynamic structure remains unexplained and should be taken into account.
- Coefficient distribution: The distribution of estimated coefficients across regions reveals how sensitively different areas respond to the common factors. This comparison provides insight into the geographic heterogeneity of market and mining influences.

### 5.1 Model Fit

The average  $R^2$  across all regions is 0.985, indicating that the factor model accounts for the vast majority of variation in local house price movements. Some regional variation is evident, suggesting

that certain areas exhibit distinct dynamics not fully captured by the common factors. As shown in Table 4, the highest  $R^2$  values are concentrated in the major east coast capitals and their surrounding metropolitan belts, whereas lower values tend to occur in smaller or resource-dependent markets.

**Table 4:** Top 5 and Bottom 5 Major City and Area by  $R^2$

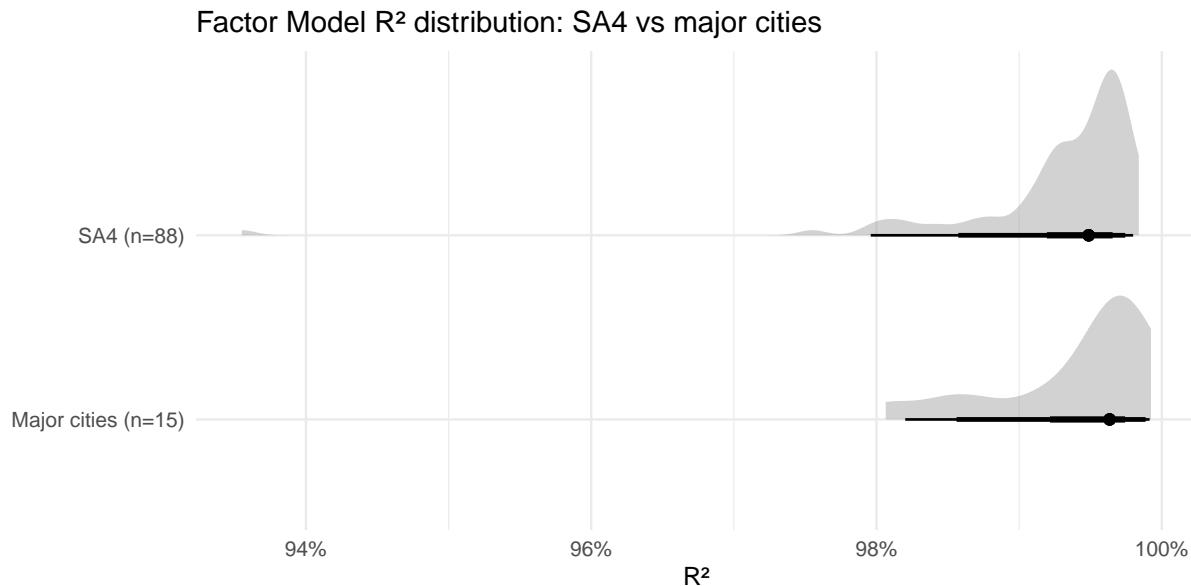
Top 5			Bottom 5		
Rank	Region	$R^2$	Rank	Region	$R^2$
1	rest_of_nsw	0.9992	15	rest_of_nt	0.9806
2	rest_of_qld	0.9990	14	greater_hobart	0.9846
3	greater_melbourne	0.9987	13	rest_of_tas	0.9872
4	greater_brisbane	0.9978	12	rest_of_wa	0.9909
5	greater_adelaide	0.9971	11	greater_darwin	0.9934

Examining the results at the SA4 level Table 5 offers a more granular view of model performance. The pattern remains consistent. Regions with the lowest fit are predominantly located in Western Australia and Queensland, reflecting their greater exposure to mining-related volatility.

**Table 5:** Top 10 and Bottom 10 SA4 Regions by  $R^2$

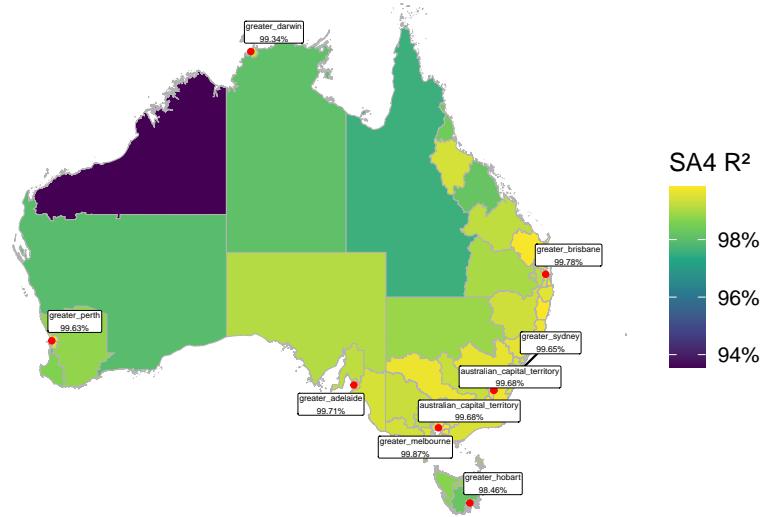
Top 10			Bottom 10		
Rank	Region	$R^2$	Rank	Region	$R^2$
1	melbourne_outer_east	0.9984	88	western_australia_outback_north	0.9355
2	sunshine_coast	0.9982	87	queensland_outback	0.9755
3	coffs_harbour_grafton	0.9980	86	western_australia_outback_south	0.9794
4	newcastle_and_lake_macquarie	0.9979	85	mandurah	0.9805
5	melbourne_north_east	0.9977	84	northern_territory_outback	0.9806
6	wide_bay	0.9977	83	mackay_isaac_whitsunday	0.9820
7	adelaide_west	0.9977	82	south_east	0.9821
8	adelaide_central_and_hills	0.9977	81	cairns	0.9839
9	brisbane_west	0.9975	80	hobart	0.9846
10	brisbane_north	0.9974	79	bunbury	0.9862

Both SA4 regions and major cities exhibit highly concentrated distributions of  $R^2$  values, the mean is nearly identical between the two groups, however, the distribution for major cities shows slightly lower dispersion, implying that urban housing markets behave more uniformly in response to the common factors, whereas regional SA4 markets display marginally greater heterogeneity in their degree of fit Figure 14.



**Figure 14:** Comparison of  $R^2$  distributions between SA4 regions and major cities, indicating slightly higher uniformity in urban markets.

The choropleth Figure 15 reinforces the findings from the summary table. The east-coast capitals and coastal belts from Melbourne through Sydney into the South-East of Queensland show the strongest fit with most regional New South Wales and Victoria are also high. While fit weakens inland and to the west, notably across Western Australia and along Queensland's mining belts. The dark purple areas mark the lowest  $R^2$  and are typically remote and resource-exposed regions with possible boom-bust timing and sparse transactions that our three factors don't fully capture.

Factor Model  $R^2$  Distribution across Cities & Regions


**Figure 15:** Choropleth of  $R^2$  distribution across SA4 regions, highlighting stronger model fit in east-coast capitals and weaker fit in mining-exposed areas of Western Australia and Queensland.

## 5.2 Residual Autocorrelation

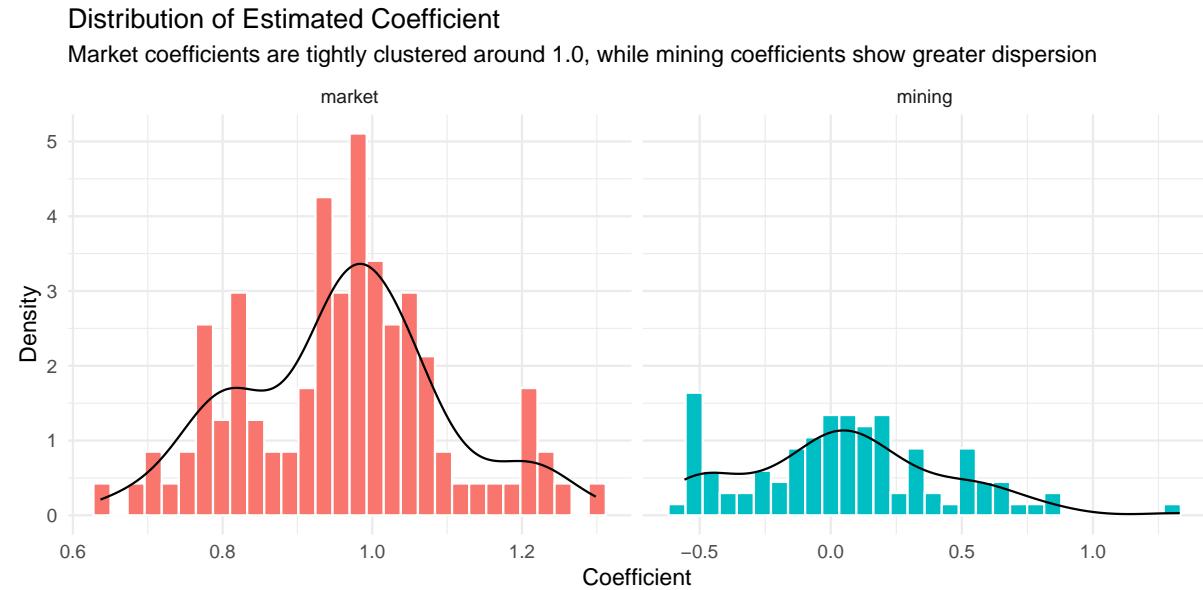
At this stage, the model does not account for temporal dependence in the residuals. Although it explains a substantial proportion of the variance in both city-level and regional indexes, the inherently trending and serially correlated nature of housing prices leads to residuals exhibiting pronounced positive autocorrelation. This is reflected in the Durbin–Watson statistics, which are consistently close to zero across all regions. Major cities cluster between 0.02 and 0.05, while SA4 regions show a slightly wider range extending up to 0.2, still well below the benchmark value of 2 Figure 16. These results highlight the limitations of a static regression framework when time-series dynamics are not explicitly modelled.



**Figure 16:** Violin plot of Durbin–Watson statistics by region type, illustrating pronounced positive autocorrelation in residuals across all areas.

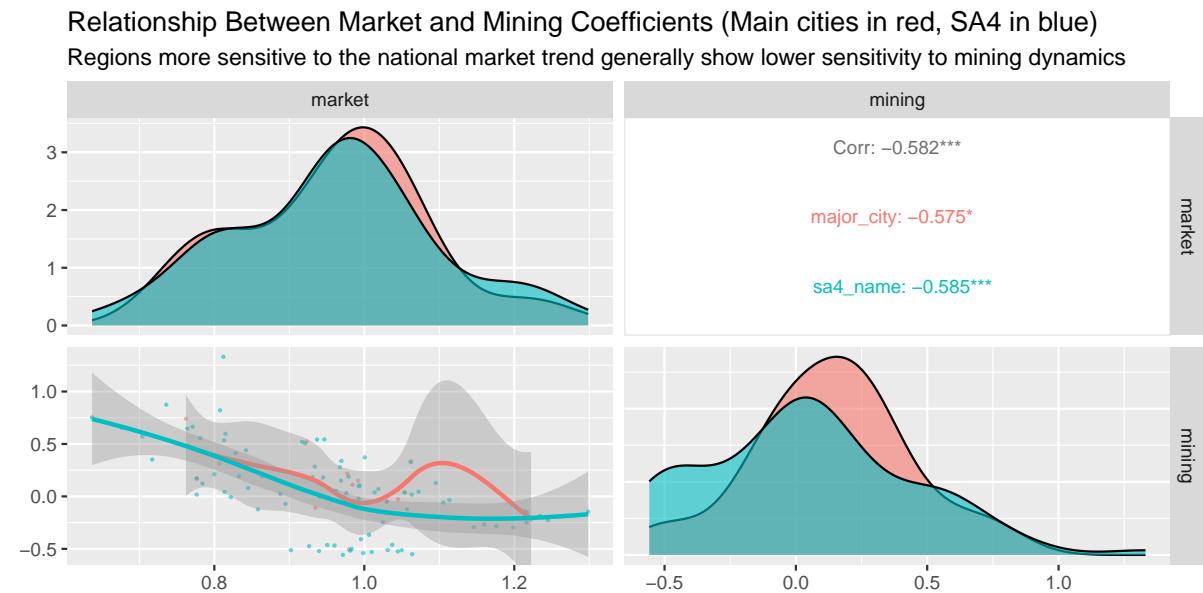
### 5.3 Coefficient Pattern

Examining coefficient patterns revealing whether regional dynamics move uniformly with the national trend or diverge due to mining influences. The distribution of market coefficients is tightly centred around 1.0 with relatively little variation. Therefore most regions respond in a similar way to the national housing trend. In contrast, the mining coefficients are far more dispersed with a long right tail and several extreme outliers, suggesting that exposure to mining-related dynamics differs substantially across regions.



**Figure 17:** Density histogram of estimated coefficients for market and mining factors, highlighting the tight clustering of market coefficients around 1.0 and the greater dispersion of mining coefficients.

Figure 18 shows the relationship between the estimated market and mining coefficients across all regions. The scatterplot in the lower-left panel reveals a clear negative association between the two factors: regions with stronger loadings on the national market factor tend to display weaker, or even negative, sensitivities to the mining factor. The red and blue densities represent the marginal distributions of coefficients for major cities and SA4 regions respectively. Their strong overlap indicates that both spatial levels exhibit similar patterns, with only minor differences in dispersion. The correlation coefficients confirm this consistency. Overall, the figure supports that the inverse trade-off between market and mining sensitivity is a systematic national pattern rather than an artefact of the level of spatial aggregation.



**Figure 18:** Pairwise plot of market and mining coefficients, showing a negative relationship between the two factors across regions, consistent across both major cities and SA4 areas.

## 6 Model Enhancement

The current static factor model exhibits notable residual autocorrelation and spatially uneven fit quality, particularly in mining-exposed regions. To enhance both model performance and adaptability, two methodological refinements were introduced, producing a more responsive specification that accounts for evolving regional relationships while maintaining statistical coherence. These refinements align with methodologies discussed by Stock and Watson (2020) (Stock & Watson 2020), who demonstrate that ARIMAX modelling and expanding-window estimation can effectively address serial correlation and parameter drift in dynamic econometric contexts. The key enhancements are:

- ARIMAX-based residual adjustment, which explicitly models temporal dependence in the errors.
- Time-varying estimation via expanding windows, allowing the regression coefficients to evolve with new data over time.

### 6.1 ARIMAX-Based Residual Control

The ARIMAX framework maintains fixed regression coefficients while addressing serial dependence in the residuals through an ARIMA process. In this context, the differencing order is constrained to zero to ensure that the analysis captures long-term co-movements between the dependent variable and the underlying factor trends, rather than focusing on short-term fluctuations.

Several baseline specifications were tested through a grid search for the orders, and the optimal model was chosen by balancing AICc performance with Ljung–Box test results, ensuring that the residuals approximated white noise.

As a result, Table 6 five ARIMAX models were selected and fitted across regions according to their individual time series dynamics. Among the 88 SA4 regions, 76 passed the Ljung–Box test, indicating that their residuals approximate white noise.

**Table 6:** Summary of ARIMAX model selection across SA4 regions, showing the distribution of best-fitting models and their pass rates for the Ljung–Box test.

best_model	n	ljung_box_pass_rate
ARIMAX(2,0,1) d=0	7	1.0000000
ARIMAX(1,0,1) d=0	20	0.9500000
ARIMAX(1,0,2) d=0	8	0.8750000
ARIMAX(2,0,2) d=0	7	0.8571429
ARIMAX(2,0,0) d=0	31	0.8064516
ARIMAX(1,0,0) d=0	15	0.7333333

The coefficient distributions for the three factors across the two modelling frameworks (static OLS and ARIMAX) Figure 19 reveal broadly consistent patterns. Both models yield strong, tightly clustered positive coefficients for the market factor, with nearly identical medians and narrow interquartile ranges, confirming a uniform national influence. For the mining factor, the medians remain close to zero under both specifications, however, the OLS model exhibits a wider dispersion and more positive outliers, whereas the ARIMAX approach produces more stable estimates once autocorrelation is controlled. The lifestyle factor displays similar distributions across models, though the ARIMAX coefficients are slightly more concentrated, indicating a mild attenuation of the lifestyle signal when temporal dependence is accounted for.

These results suggest that national market forces exert a consistently strong influence on regional price dynamics regardless of model specification. Controlling for serial correlation reduces the apparent volatility of the mining factor—implying a more conservative estimation of its cyclical influence—while lifestyle effects remain modest and less regionally variable once time-dependent patterns are incorporated.



**Figure 19:** Boxplots comparing coefficient distributions for market, mining, and lifestyle factors between OLS and ARIMAX models, highlighting the impact of accounting for autocorrelation on coefficient stability.

To further illustrate this, we compare factor coefficients for Melbourne and Sydney's inner areas, along with Blacktown and the Central Coast Table 7, which are regions influenced by the mining and lifestyle factors. The ARIMAX model provides a more reliable fit by addressing autocorrelation, resulting in more stable and realistic factor sensitivities across regions. Overall, when time dependence is modelled with ARIMAX, the market sensitivity increases across all regions, while the absolute influence of the mining and lifestyle factors weakens:

- **Market Factor:** Each region becomes more tightly linked to the national housing cycle once autocorrelation is controlled for. For example, the Central Coast rises from 0.926 to 0.999, and Sydney–Inner South from 1.04 to 1.07.
- **Mining Factor:** The negative mining effect softens everywhere (e.g., Sydney–Blacktown from -0.511 to -0.364), implying mining trend of the OLS came from persistent shocks rather than true long-run sensitivity.
- **Lifestyle Factor:** The lifestyle effect also becomes less extreme, most notably in the Central Coast, where it drops slightly but remains positive, confirming its coastal lifestyle character, while capital and suburban areas stay negative.

**Table 7:** Comparison of factor coefficients between OLS and ARIMAX models for selected regions, highlighting the effect of autocorrelation adjustment on sensitivity estimates.

**Table 7: Factor Coefficient Comparison: OLS vs ARIMAX for Selected Regions**

region	model_type	market	mining	lifestyle
central_coast	OLS	0.926429	-0.4740493	0.2350783
central_coast	ARIMAX	0.999559	-0.3792453	0.2046319
melbourne_inner	OLS	1.146014	-0.2932346	-0.8432711
melbourne_inner	ARIMAX	1.144550	-0.3255767	-0.8413226
sydney_blacktown	OLS	1.031223	-0.5111639	-0.2252099
sydney_blacktown	ARIMAX	1.036702	-0.3510197	-0.1852012
sydney_city_and_inner_south	OLS	1.044756	-0.5245910	-0.5246441
sydney_city_and_inner_south	ARIMAX	1.070404	-0.5089045	-0.5505255

## 6.2 Time-dependent Coefficients with Expanding Window

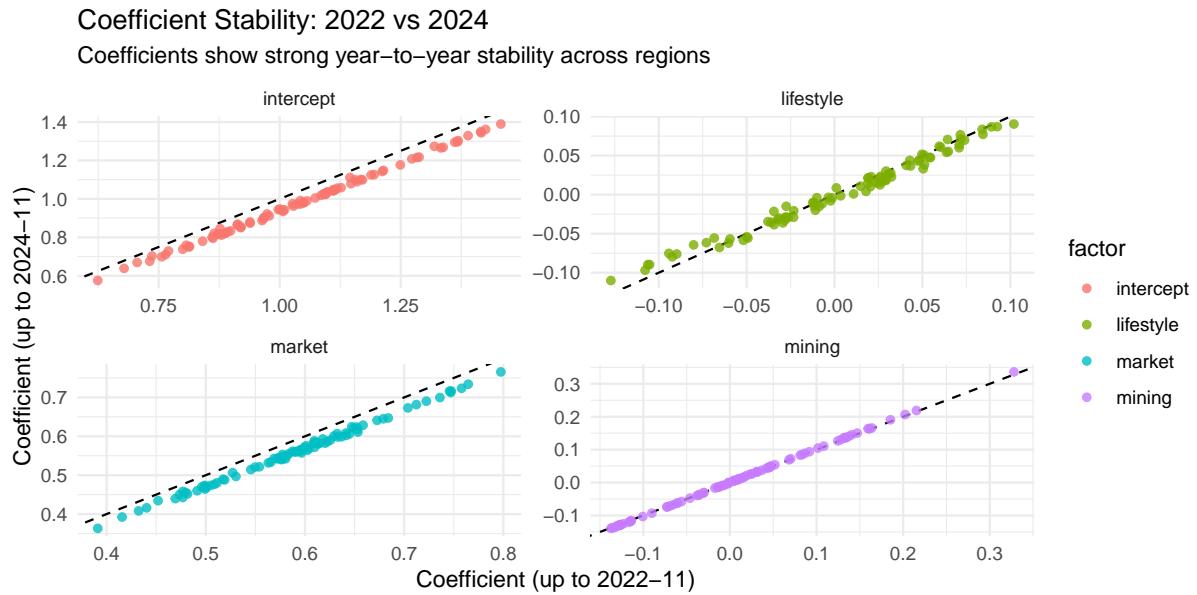
The expanding window approach was used to assess the stability of model coefficients over time. By gradually extending the estimation sample, we tested whether the relationships between regional housing indices and key factors remained consistent as new data of the following years was added. The regression was refitted twice, first using data up to November 2022, and again including data up to November 2024. Comparing these two sets of coefficients allowed us to evaluate how much the estimated sensitivities of each factor changed when the sample expanded by two years.

Table 8 results displays the average yearly change and standard deviation of coefficients across regions. The intercept and market factors show small but positive shifts, while mining and lifestyle remain nearly unchanged. This suggests that the overall influence of these factors on housing prices has been highly stable between 2022 and 2024.

**Table 8:** Summary of coefficient changes between expanding window estimations up to 2022 and 2024, indicating strong stability in factor sensitivities over time.

factor	mean_change	sd_change
intercept	0.0616	0.0098
lifestyle	0.0014	0.0079
market	0.0306	0.0049
mining	-0.0005	0.0022

Figure 20 compares the 2022 and 2024 coefficient estimates for each factor. The points align closely with the 45-degree line, indicating that factor effects remain largely unchanged across years. This stability suggests that incorporating the latest data has not materially affected the underlying relationships, reinforcing the robustness of the factor structure.



**Figure 20:** Scatterplot comparing expanding-window estimates for 2022 and 2024, showing that factor sensitivities have remained remarkably consistent despite new data.

## 7 Conclusion

The analysis extends the factor modelling framework beyond static estimation to encompass dynamic forecasting and time-varying inference. The factor forecasts reveal a clear contrast in behaviour of the market and mining factors. The market factor projects a continuation of the national housing uptrend, with an 80% prediction width of approximately 0.04 - 0.05, indicating that while uncertainty widens modestly with horizon, the aggregate trend remains firmly upward. In contrast, the mining factor displays tighter prediction intervals and a more symmetrical distribution around its mean, consistent with mean-reverting dynamics. This behaviour reflects the transitory nature of resource-driven shocks: Expansions in mining regions tend to unwind rather than compound, producing bounded rather than explosive cycles. The forecasting results therefore delineate two distinct temporal modes: national persistence and regional reversion, under a shared macroeconomic backdrop.

Dynamic estimation confirms that these relationships are structurally stable. ARIMAX residual control and expanding-window estimation both preserve the factor hierarchy observed in the static framework while refining its precision. Across the 2022 to 2024 window, factor coefficients remain highly correlated ( $\rho > 0.95$ ), signifying negligible structural drift despite the inclusion of new data.

The market factor retains a median coefficient near 1.0, implying that a 1 % rise in the national index corresponds to roughly a 1 % increase across most regions. Mining sensitivities remain centred around zero but exhibit a marked contraction in spread once autocorrelation is modelled, indicating a more conservative attribution of cyclical resource effects. Lifestyle coefficients, averaging around 0.2, also tighten slightly under ARIMAX, suggesting that these regional differentials are persistent but limited in amplitude.

Taken together, these findings portray a housing system governed by a persistent national growth engine and self-correcting regional cycles. National shocks transmit broadly and uniformly, while local fluctuations linked to the mining sector revert toward long-term equilibrium rather than diverge. The stability of coefficients across time and specification demonstrates that the principal drivers of the regional housing prices are enduring and interpretable, not artefacts of sample period or model choice. Accounting for temporal dependence thus sharpens inference rather than alters it, yielding a transparent mapping between national and regional market dynamics.

From a broader perspective, the extended model underscores the resilience of Australia's housing structure: Long-run national growth coexists with spatial heterogeneity that is cyclical, not structural. Future research could extend this framework toward state-space or time-varying parameter factor models to examine how these relationships evolve under macroeconomic shocks such as interest-rate adjustments or migration surges. By bridging static factor models and dynamic forecasting, this analysis provides a foundation for understanding how systemic and regional forces interact to shape the geography of housing growth across Australia.

Future work could explore more sophisticated time-varying parameter models to capture potential non-linearities and regime shifts in regional housing dynamics. Additionally, incorporating macroeconomic covariates like interest rates, employment trends, and migration patterns could enhance the explanatory power of the model. Finally, extending the analysis to include more granular spatial units or alternative asset classes may provide further insights into the interplay between national trends and local market conditions.

## 8 Code and Data files

All code and processed data supporting this report are available in the project's GitHub repository:  
<https://github.com/cherylbunny/neoval-project>.

## References

- Case, KE, J Cotter & SA Gabriel (2010). *Housing Risk and Return: Evidence from a Housing Asset-Pricing Model*. Tech. rep. WP201005. Geary Institute, University College Dublin. <http://www.ucd.ie/geary/static/publications/workingpapers/gearywp201005.pdf>.
- Hyndman, RJ & G Athanasopoulos (2021). *Forecasting: Principles and Practice*. 3rd. OTexts. <https://otexts.com/fpp3/>.
- Sijp, WP, A Panagiotelis & MK Francke (2025). Data-sparse price indexes by spatio-temporal regularization and PCA: An application to the Australian housing market. *Real Estate Economics* 53(2), 266–296.
- Stock, JH & MW Watson (2020). *Introduction to Econometrics*. 4th. The Pearson Series in Economics. Discusses ARIMAX models for serially correlated residuals and expanding-window estimation for evolving parameters. Harlow, England: Pearson Education. <https://www.pearson.com/us/subject-catalog/p/introduction-to-econometrics/P200000006421/9780136879787>.
- Valentini, P, L Ippoliti & L Fontanella (2013). Modeling US housing prices by spatial dynamic structural equation models. *The Annals of Applied Statistics* 7(2), 763–798.