



MONASH
University

Forecast Reconciliation for Land Transfer Duty

Hoang Do

Report for
The Department of Treasury and Finance

29 May 2024

MONASH
BUSINESS
SCHOOL

**Department of
Econometrics &
Business Statistics**

☎ (03) 9903 4416
✉ vdoo0002@student.monash.edu

ABN: 12 377 614 012



1 Abstract

- Overview: Briefly summarize the purpose, methodology, key findings, and implications of the forecast.
- Key Recommendations: High-level actionable recommendations based on the forecast results.

2 Introduction and background

The property sector plays a pivotal role in Australia's economy, accounting for 1 in 4 jobs indirectly and contributing around 13% of Gross Domestic Product (GDP). In the 2021 financial year, property sales totaled approximately \$350 billion (**Real Estate Institute of Australia, 2021**). Land transfer duty, previously known as stamp duty, significantly impacts property transactions and the sector as a whole. A literature published by the New South Wales Treasury found that a 100 basis point (1%) cut in land transfer duty could boost property transactions by 10% (**Malakellis & Warlters, 2021**).

Land transfer duty is a tax applied to the “dutable value” of a property being purchased or acquired, whether it is a first home or an investment property. The dutiable value is determined as either the property's purchase price or its market value, whichever is greater. Several factors influence the amount of duty paid, including the buyer's intended use of the property, foreign purchaser status, and eligibility for exemptions.

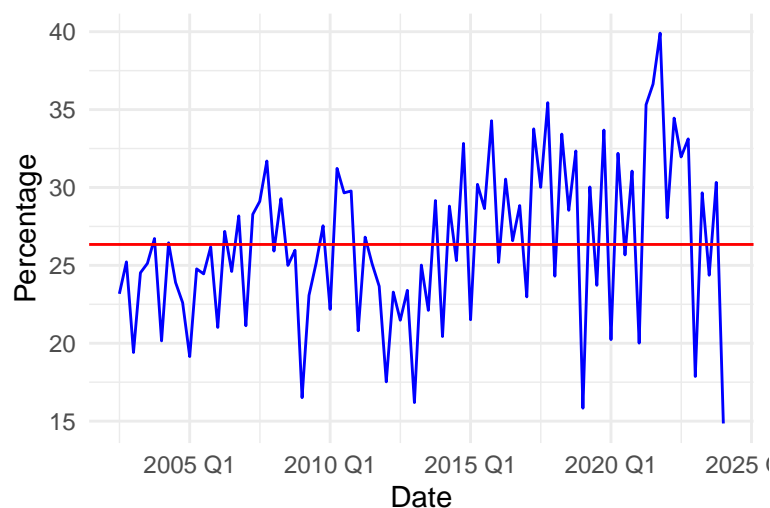


Figure 1: *Dependence of Victoria's Tax Revenue on Land Transfer Duty*

Additionally, Victoria's tax revenue heavily relies on land transfer duty. As can be seen from Figure 1, over the past 20 years, on average, land transfer duty accounts for 27% of Victoria's tax revenue. Despite its perceived inequity and various exemptions designed to aid homebuyers,

abolishing this duty remains challenging due to the need for equivalent revenue replacement. If this duty is removed, the government will need to introduce one or more new taxes to generate equivalent revenue.

To balance the need for sufficient tax income without discouraging property transactions, *the Department of Treasury and Finance (DTF)* has been closely working with the government on policy adjustments, subsidies, exemptions, and restrictions. Accurate forecasts of land transfer duty, both short-term (*1 to 3 months*) and long-term (*12 months*), are essential for these decisions.

This report introduces a new forecasting methodology aimed at improving the accuracy of these predictions. By employing forecast reconciliation and combining cross-sectional and temporal hierarchies, we aim to provide the *Department of Treasury and Finance* with more reliable forecasts to inform policy-making.

3 Data Description

3.1 Data Source

Detail the sources of the LTD data, including collection methods, frequency, and historical range.

3.2 Variables Description

Describe each variable used in the analysis, including dependent and independent variables.

4 Initial Data Analysis (IDA)

4.1 Data Cleaning and Manipulation

4.2 Descriptive Statistics

Based on summary statistics results from Table 1,

1. Land Transfer Duty Overview:

- The total land transfer duty (**LTD**) reported has varied considerably over the past decade, ranging from approximately 286.2 million to over 1.013 billion over past . This variability indicates significant fluctuations in the real estate market activity and property value changes over the years.
- Residential transactions dominate the LTD collection, contributing between 238.1 million and 761.4 million, representing the largest portion of land transfer duty. This suggests

that residential real estate remains a vital part of the market, possibly reflecting trends in housing demand and price movements.

- **Non-residential transactions**, though smaller than residential, still make a substantial contribution, with LTD from this sector ranging from about 42.9 million to 357.2 million. Within this, the commercial sector is the most significant, followed by industrial and other sectors (such as agricultural).

2. Detailed Sector Analysis:

- **Commercial Sector:** With LTD ranging from 22.1 million to 132.4 million, this sector shows substantial activity, possibly reflecting economic growth, investment trends, and business expansions.
- **Industrial Sector:** This sector has collected LTD between 5.2 million and 214.4 million, highlighting some large-scale industrial transactions or developments during certain periods.
- **Other Sectors:** Including smaller sectors like agricultural, contributions range from 3.3 million to 74.5 million, indicating sporadic activity that may correspond with specific market or economic conditions.

3. Economic Indicators:

- **Sales:** The volume of sales has fluctuated from 4,629 to 15,177, indicating periods of both high and low market activity.
- **Home Value Index (HVI):** The HVI has ranged from 106.2 to 192.4 point, which is indicative of significant changes in home values. A higher HVI suggests increasing property values, which could contribute to higher LTD collections.

5 Exploratory Data Analysis (EDA)

5.1 Time series analysis

Total Land Transfer Duty

Figure 2 demonstrates an upward, non-linear trend, showing that the value of land transfer duty has generally increased over the ten-year period. Additionally, there is a noticeable increase in variability, particularly in recent years, indicating more pronounced fluctuations in the values. This increasing volatility suggests that a transformation of the data may be necessary to better analyze and model these trends.

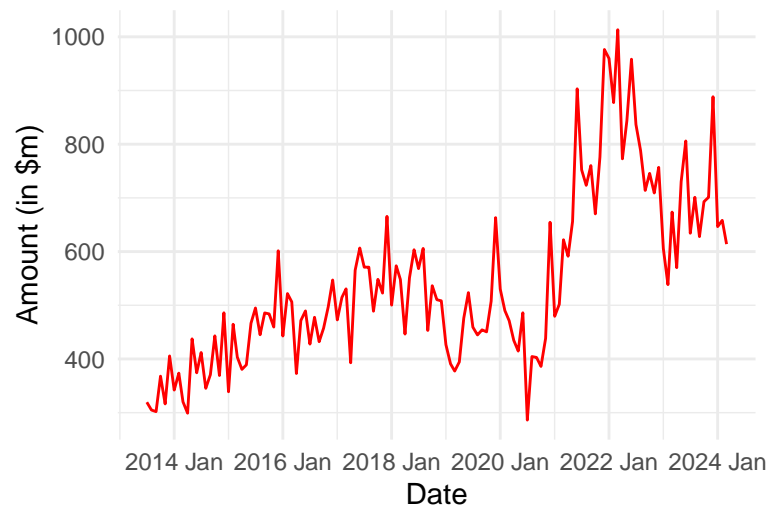


Figure 2: Time plot of Land Transfer Duty in Victoria over time

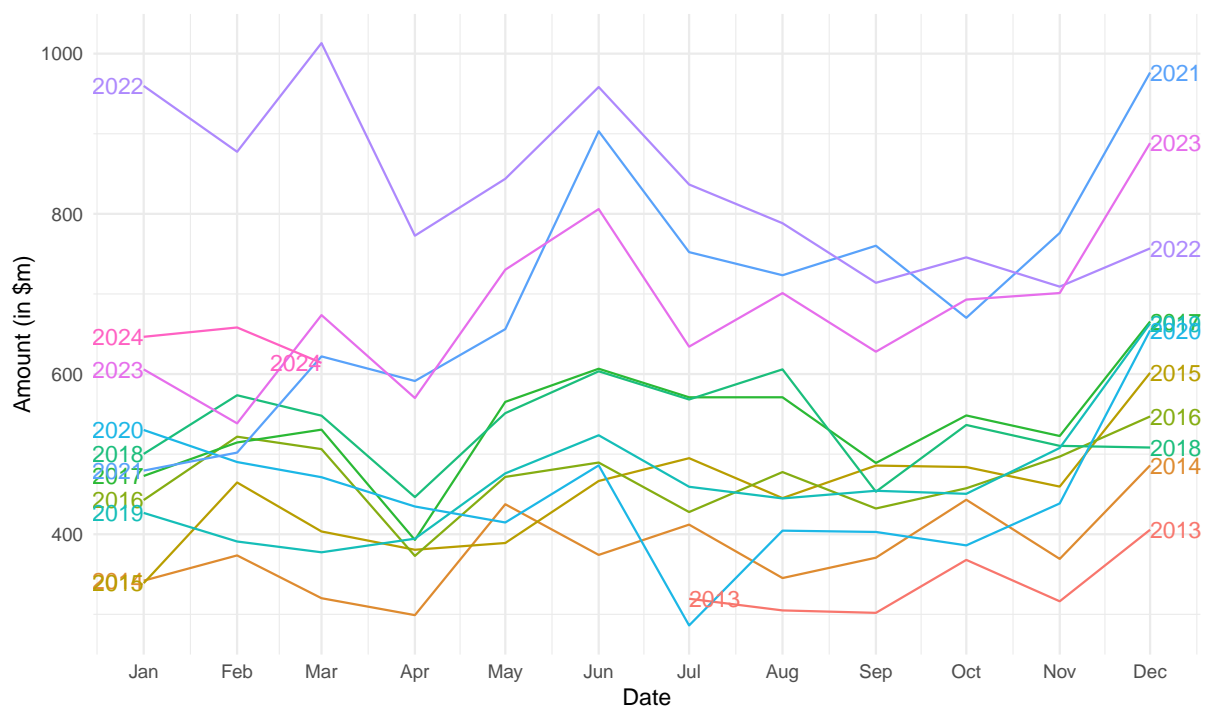


Figure 3: Seasonal trend of Land Transfer Duty in Victoria

Considering the seasonal patterns of land transfer duty, Figure 3 reveals that, apart from slight increases in June and December, there is no clear indication of a prominent seasonal pattern.

Figure 4 shows that there is a seasonal pattern for log transformed land transfer duty but the large grey bar in the seasonal panel shows that the variation in the seasonal component is smallest compared to the variation in the data.

Moreover, Figure 4 demonstrates that after applying a log transformation, the variability becomes

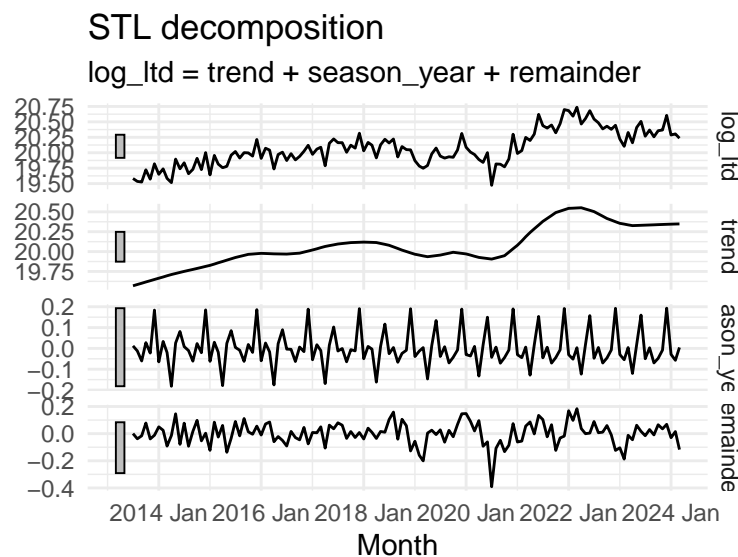


Figure 4: *Decomposition of log of Land Transfer Duty in Victoria*

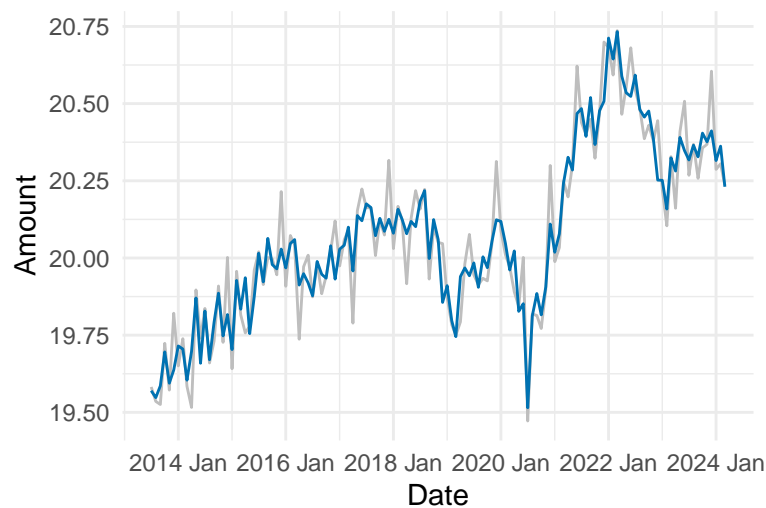


Figure 5: *Seasonally adjusted of log Land Transfer Duty in Victoria*

more consistent, although some heterogeneity still remains.

The grey line in Figure 5 represents the original log-transformed total land transfer duty, while the blue line represents the seasonally adjusted log-transformed total land transfer duty. Figure 5 suggests that by adjusting for the seasonal pattern, some local peaks and troughs have been smoothed out, resulting in a more consistent series.

Sales

Figure 6 illustrates a non-linear trend, showing fluctuations in sales units over the ten-year period from January 2014 to January 2024. The plot exhibits considerable variability with both high and low peaks scattered throughout the timeline. The values range from approximately 5000 to over

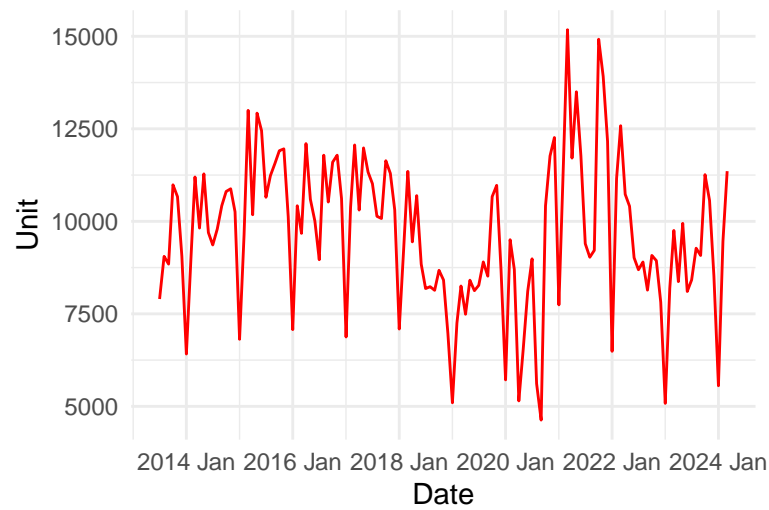
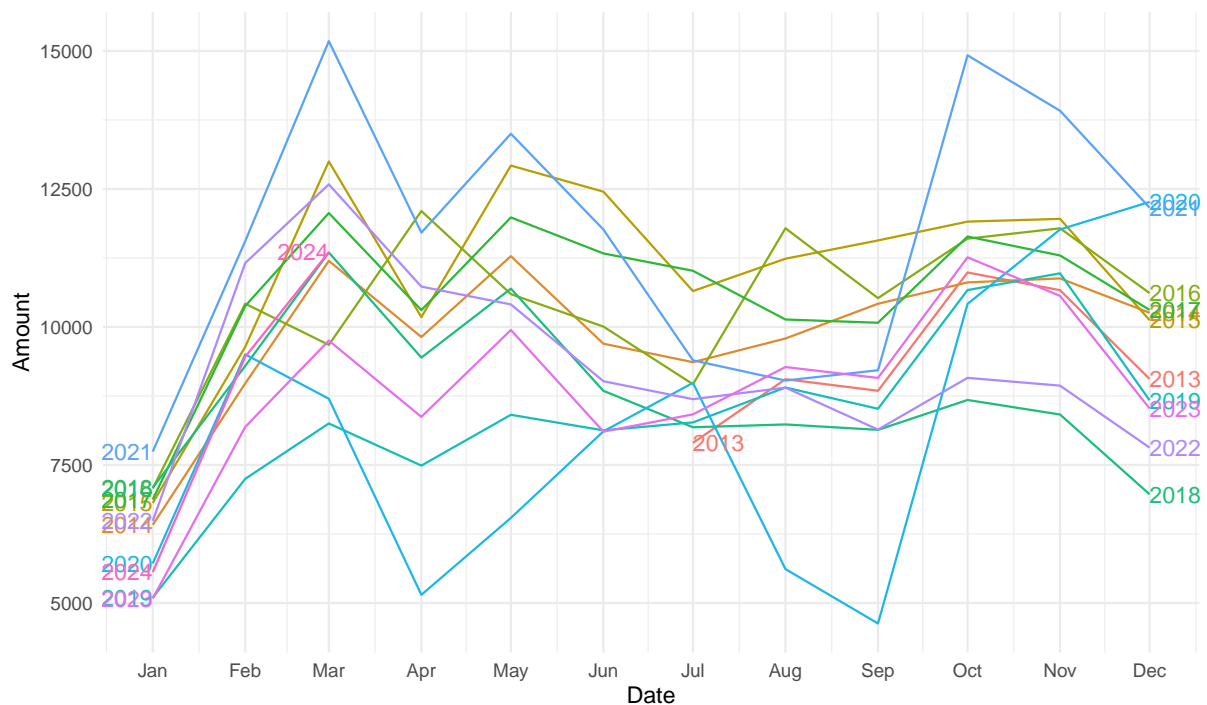


Figure 6: Time plot of Sales in Victoria over time

15000, indicating significant changes in sales units.

In recent years, there is a noticeable increase in variability, particularly around 2020 and onwards. This increasing volatility indicates that a transformation of the data may be necessary to better analyze and model these trends.



pronounced peaks and troughs.

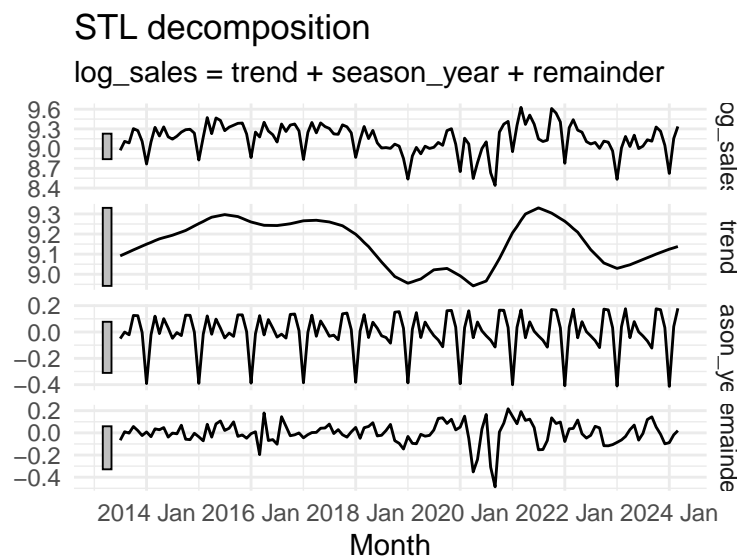


Figure 8: *Decomposition of log of sales in Victoria*

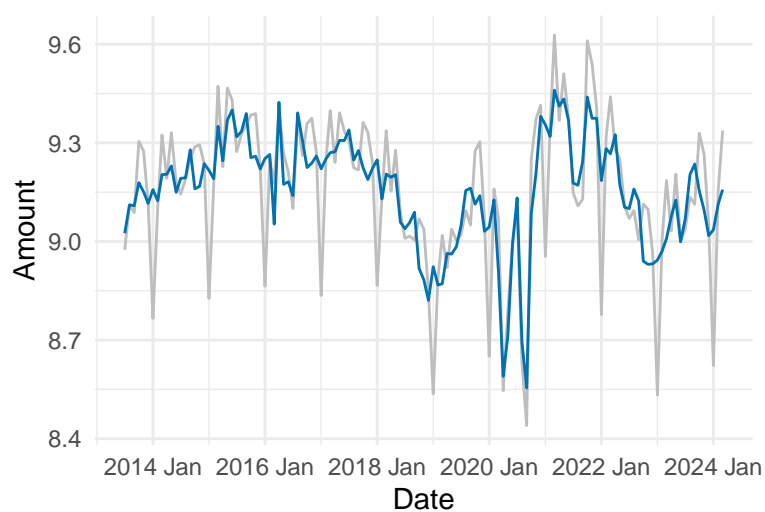


Figure 9: *Seasonally adjusted of log of sales in Victoria*

Figure 8 indicates a seasonal pattern for log-transformed sales. However, the large grey bar in the seasonal panel demonstrates that the variation in the seasonal component is the smallest compared to the overall variation in the data.

Moreover, Figure 8 demonstrates that after applying a log transformation, the variability becomes more consistent, although some heterogeneity still remains.

The grey line in Figure 9 represents the original log-transformed sales, while the blue line represents the seasonally adjusted log-transformed sales. Figure 9 suggests that there is a huge difference between seasonally adjusted sales and original sales, especially in periodic drops.

Home Value Index

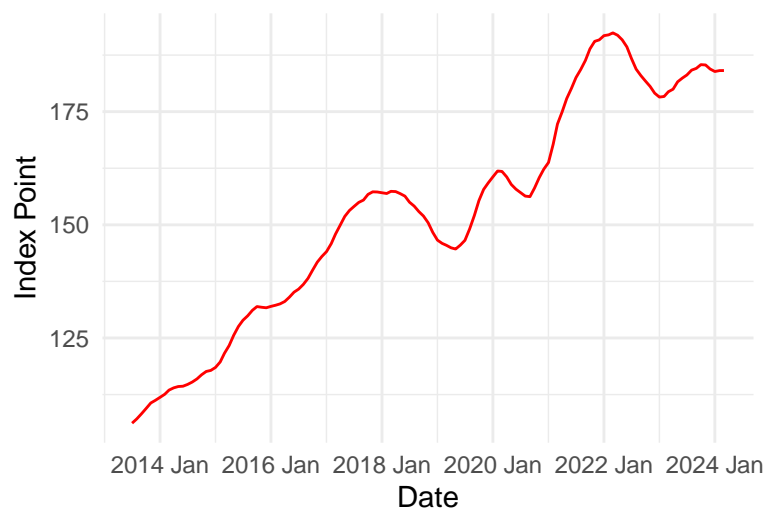


Figure 10: Time plot of Home Value Index in Victoria over time

Figure 10 illustrates an increasing trend, with some considerable variability throughout the timeline.

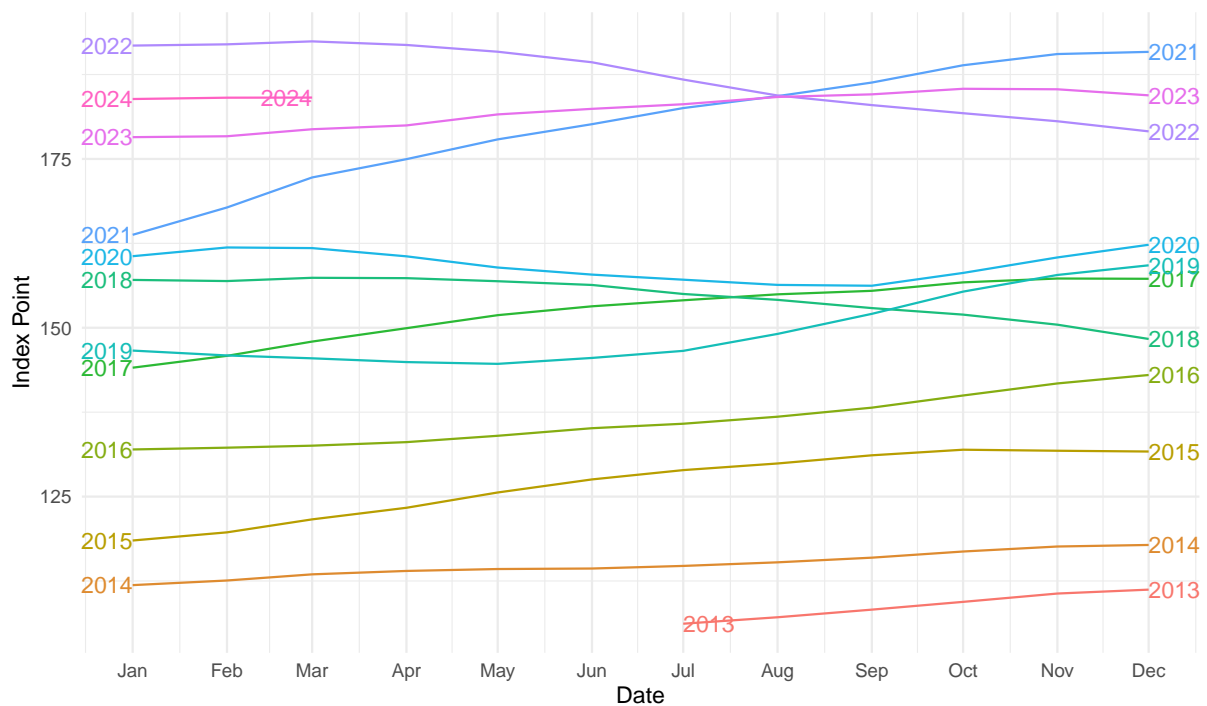


Figure 11: Seasonal trend of Home Value Index in Victoria

Figure 11 depicts no clear sign of seasonal pattern in home value index, suggesting a stable index point maintained throughout the period.

Figure 12 indicates a seasonal pattern, but the large grey bar in the seasonal panel shows that the

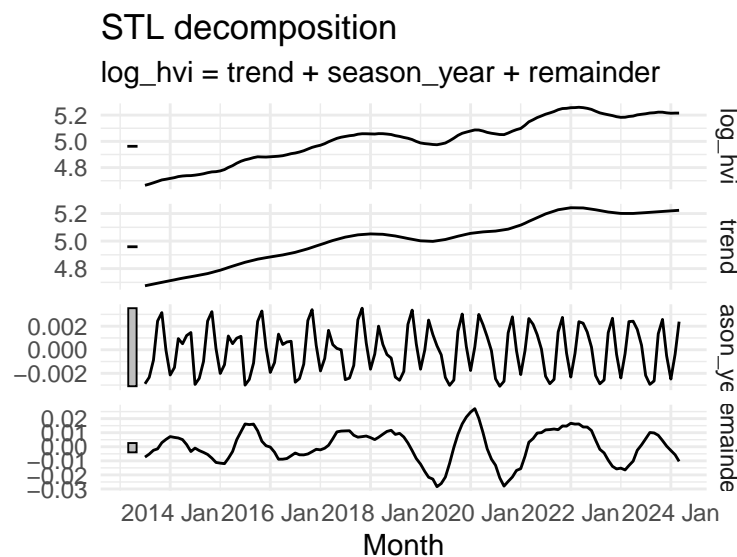


Figure 12: *Decomposition of log of home value index in Victoria*

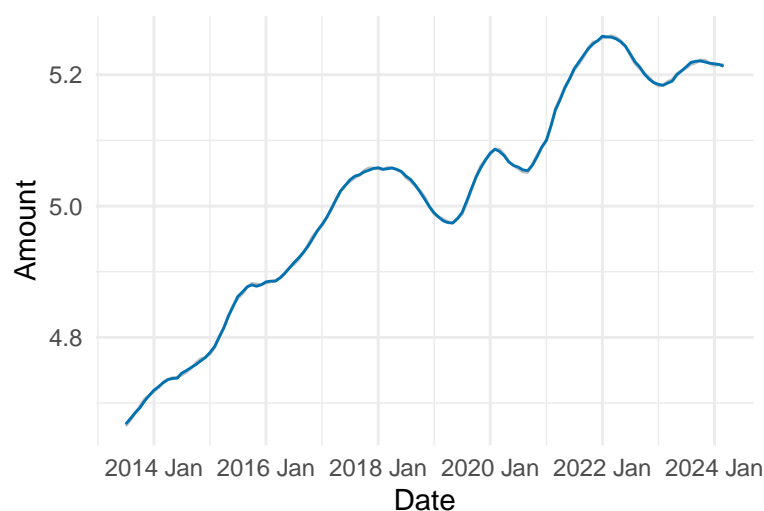


Figure 13: *Seasonally adjusted of log of home value index in Victoria*

variation in the seasonal component is minimal, similar to the pattern observed in sales and land transfer duty.

In Figure 13, the blue and grey lines completely overlap, indicating that there is no seasonal pattern in the home value index.

5.2 Cointegration Analysis

Based on the observations from Figure 2, which shows the trend and increasing variance pattern in land transfer duty, Figure 6, which exhibits an increasing pattern and a somewhat seasonal pattern in sales, and Figure 10, which illustrates the trend in the home value index, we can conclude that these

three series are not stationary.

Furthermore, to justify the use of the Vector Error Correction Model (VECM) by the *Department of Treasury and Finance*, it is essential to test for the presence of cointegration patterns (or long-term equilibrium relationships) among these three series.

The Johansen procedure will be employed to test for cointegration between log transformed of land transfer duty, sales, and the home value index series. This procedure tests the null hypothesis that no cointegration relationship exists among the series.

```
#####  
# Johansen-Procedure #  
#####
```

Test type: trace statistic , with linear trend

Eigenvalues (lambda):

```
[1] 0.40113747 0.12862617 0.01767753
```

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 2		2.25	6.50	8.18 11.65
r <= 1		19.60	15.66	17.95 23.52
r = 0		84.20	28.71	31.52 37.22

Eigenvectors, normalised to first column:

(These are the cointegration relations)

	ltd_total.l3	sales.l3	hvi.l3
ltd_total.l3	1.000000	1.000000	1.000000
sales.l3	-1.083331	0.7360273	0.6437896
hvi.l3	-1.374889	-1.1062996	-6.1359829

Weights W:

(This is the loading matrix)

	ltd_total.l3	sales.l3	hvi.l3
ltd_total.d	-0.643445282	-0.099265524	-0.0087039883
sales.d	0.499922944	-0.230174120	-0.0146337444
hvi.d	0.002144461	-0.004709971	0.0004088504

The test results, using 3 lags as chosen by the DTF, allow us to reject the null hypothesis that $r \leq 1$, yet we fail to reject the null hypothesis that $r \leq 2$ at 5% level of significance. This suggests that among the three variables, the rank of the matrix exceeds 2, indicating the presence of at least two cointegration relationships.

Furthermore, a rank of 2 implies that we need a combination of at least two time series to form a stationary series.

To create such a linear combination, we can utilize the components of the eigenvector associated with the largest eigenvalue. According to the Johansen test summary, the largest eigenvalue is approximately 0.4011375. This corresponds to the eigenvectors under the column `ltd_total.l3`, which is approximately equal to (1, -1.0833311, -1.3748894). By forming a linear combination of the series using these eigenvector components, we can achieve a stationary series.

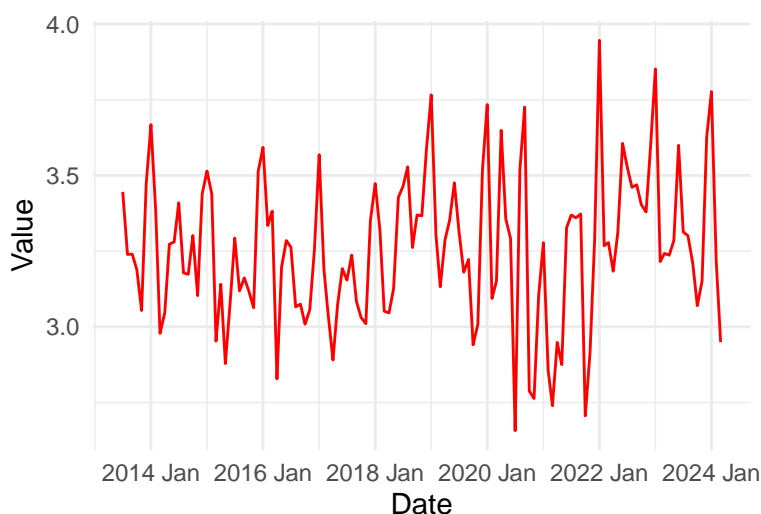


Figure 14: Stationary series formed via a linear combination of 3 time series

As shown in Figure 14, the linear combination appears to be more stationary, although there remains a slight indication of varying variances over time.

We can further evaluate this by applying the Augmented Dickey-Fuller (ADF) test:

Augmented Dickey-Fuller Test

```
data: linear_comb  
Dickey-Fuller = -3.2199, Lag order = 5, p-value = 0.08772  
alternative hypothesis: stationary
```

The ADF test statistic indicates that we can reject the null hypothesis of a unit root, providing evidence of a stationary series formed from the linear combination at the 10% level of significance. Therefore, we can conclude that it is appropriate to fit a Vector Error Correction Model (VECM) to these three time series.

6 Methodology

6.1 Cross-sectional, temporal and cross-temporal hierarchies

Upon analyzing the data characteristics and the disaggregation of LTD, it becomes evident that a three-level hierarchical structure can be established. There is utility in generating forecasts at various levels of aggregation, driven by diverse reasons and objectives. For example, each level of aggregation may exhibit distinct characteristics; for instance, transactions involving residential properties might vary from those involving non-residential properties due to differences in market dynamics or market size for each property type.

In this context, total land transfer duty is divided into two categories: residential and non-residential properties. Non-residential properties are further disaggregated into three sub-categories: commercial, industrial, and other, which predominantly includes agricultural properties. In addition to the total land transfer duty for all property types, the government and the *Department of Treasury and Finance* may also be interested in forecasts for each category and sub-category. Figure 15 illustrates the cross-sectional hierarchical structure of land transfer duty as discussed.

Given that this is a straightforward cross-sectional hierarchical structure, we will also consider temporal hierarchies to further enhance forecast accuracy. Land transfer duty is collected monthly, and forecasts can be generated at bi-monthly, quarterly, four-monthly, semi-annual, and annual frequencies. Various temporal hierarchies can be constructed with monthly land transfer duty treated as the bottom level. Figure 16 presents an example of temporal hierarchies with monthly land transfer duty as the bottom level.

Although forecasts using cross-sectional and temporal hierarchies have demonstrated substantial improvements (Kourentzes & Athanasopoulos, 2019), these approaches have typically been used

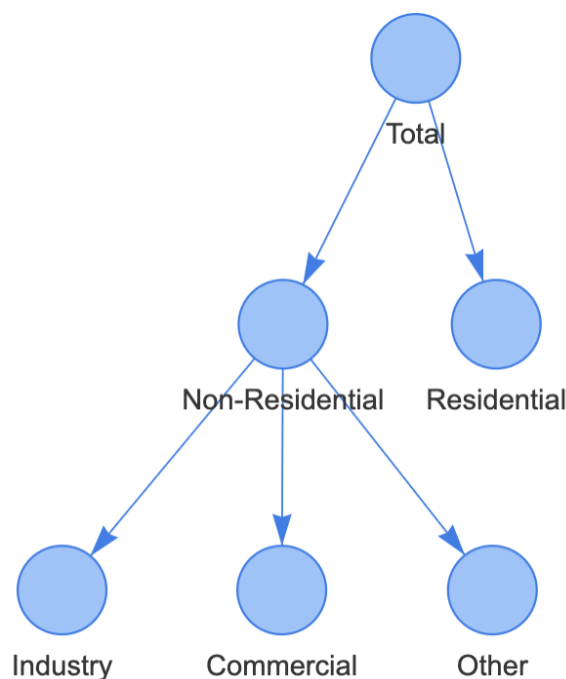


Figure 15: *Cross-sectional hierarchy*

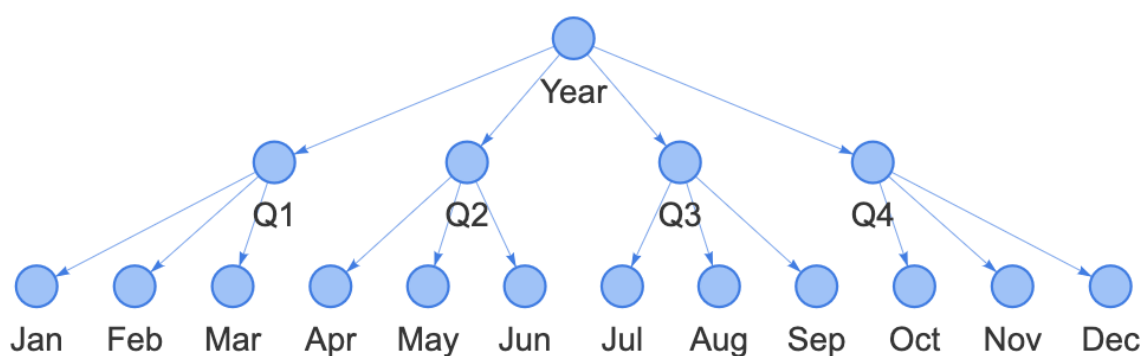


Figure 16: *Temporal hierarchy*

separately. By combining cross-sectional hierarchies and temporal hierarchies, referred to as cross-temporal hierarchies, forecast accuracy can be further improved. Figure 17 from **Kourentzes and Athanasopoulos, 2019** shows an example of a cross-temporal hierarchical structure.

6.2 Forecast reconciliation

From the cross-temporal hierarchies, forecasts can be produced at all levels for every nodes. And in an ideal scenario, forecasts from different levels of aggregation could seamlessly sum up to the top level. However, practical implementation often reveals incoherent among independently produced forecasts. Consequently, it becomes vital for forecasts to align and aggregate according to the hierarchical

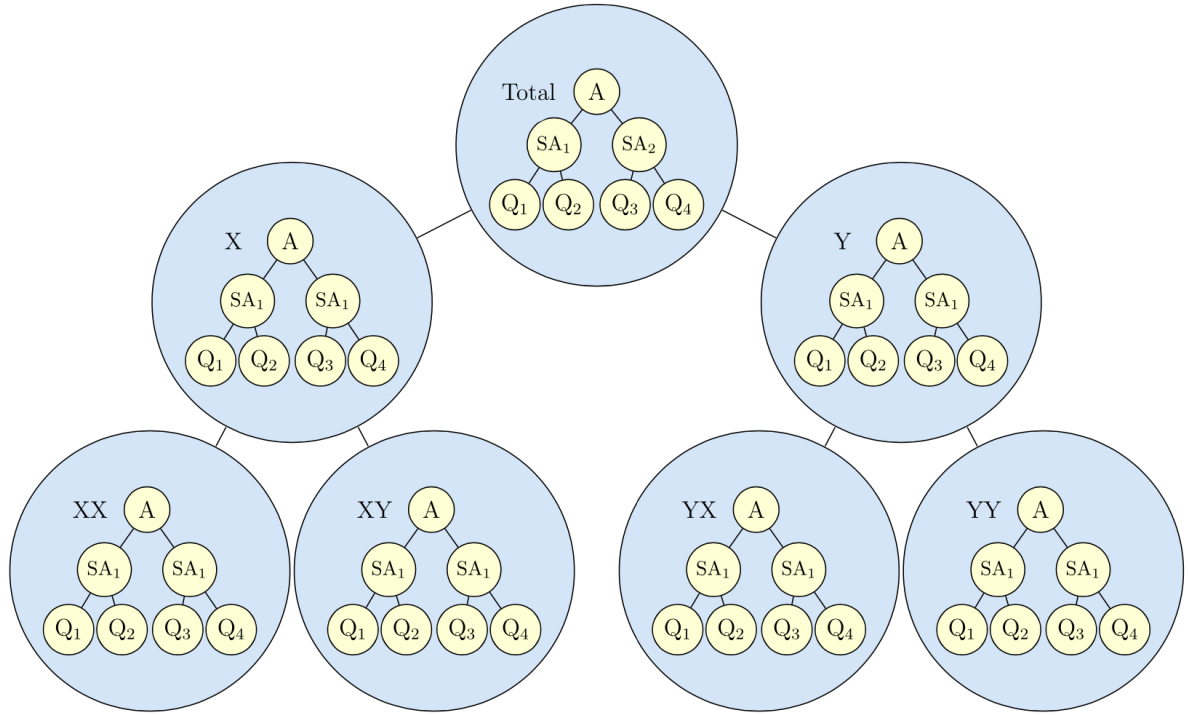


Figure 17: Cross-temporal hierarchy

structure organizing the array of time series. As a solution to this challenge, forecast reconciliation emerges as one of the most effective methodologies employed today.

We first consider cross-sectional forecast reconciliation. Recall from Figure 15, we can construct this hierarchical structure for LTD in a matrix form like this:

$$\begin{bmatrix} \text{Total}_t \\ \text{Non-residential}_t \\ \text{Residential}_t \\ \text{Commercial}_t \\ \text{Industrial}_t \\ \text{Other}_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \text{Residential}_t \\ \text{Commercial}_t \\ \text{Industrial}_t \\ \text{Other}_t \end{bmatrix}$$

or in a more compact notation:

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t,$$

where \mathbf{S} represents the summing matrix defining how bottom-level series are aggregated.

Then for any set of base forecast, denoted as $\hat{\mathbf{y}}_h$, where h is the forecast horizon, all coherent forecasting approaches, denoted as $bf \hat{\mathbf{y}}_h$, can be represented as:

$$\tilde{\mathbf{y}}_h = \mathbf{SG}\hat{\mathbf{y}}_h,$$

where \mathbf{G} is a matrix that maps the base forecasts into the bottom level (Hyndman & Athanasopoulos, 2021).

The equation shows that pre-multiplying any set of base forecasts with \mathbf{SG} will return a set of coherent forecasts. So the idea is to find that matrix \mathbf{G} in such a way that it minimises the variance between your reconciled forecast and your original forecast

Within the domain of hierarchical time series forecasting, there are three traditional single level approaches for generating forecasts for hierarchical time series. The first, known as the *bottom-up* approach, initiates by producing forecasts for each series at the lowest level and subsequently aggregates these to generate forecasts for the upper levels of the hierarchy. For this approach, \mathbf{G} can be defined as:

where the first 2 columns zero out the base forecast of the series above the bottom level.

Conversely, the *top-down* approach starts with a forecast at the highest level, which is then disaggregated to lower levels using predetermined proportions—typically based on historical data distributions (Gross & Sohl, 1990). For this approach, \mathbf{G} can be defined as:

where the first column includes the set of proportions that distribute the base forecasts of the top level to the bottom level

Lastly, the *middle-out* approach is a combination of both the bottom-up and top-down methods. However, the traditional single level approaches may have their limitations since only base forecast at one level is used. In response to this, Wickramasuriya et al., 2019 introduced the *MinT* (Minimum Trace) optimal reconciliation methodology, which devises a \mathbf{G} matrix aimed at minimizing the total forecast variance within the coherent forecast set.

Wickramasuriya et al., 2019 show that the variance-covariance of the h -step-ahead coherent forecast errors is given by:

$$\mathbf{V}_h = \text{Var}[\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h] = \mathbf{SGW}_h\mathbf{G}'\mathbf{S}',$$

where $\mathbf{W}_h = \text{Var}[\mathbf{y}_{T+h} - \hat{\mathbf{y}}_h]$ is the variance-covariance matrix of the corresponding base forecast errors.

Wickramasuriya et al., 2019 also show that:

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1},$$

minimises the trace of \mathbf{V}_h subject to $\mathbf{S}\mathbf{G}\mathbf{S} = \mathbf{S}$

Therefore, the optimally reconciled forecasts are given by:

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_h$$

which refers as the **MinT**

Now, a challenge with \mathbf{G} is that it requires an estimation of \mathbf{W}_h , the forecast error variance of h -step-ahead base forecasts. There are four simplifying approximations in place that have been shown to work well:

1. **OLS (Hyndman et al., 2011)**: $\mathbf{W}_h = k_h\mathbf{I}$ for all h , where $k_h > 0$
2. **WLS_S (Athanasopoulos et al. 2017)**: $\mathbf{W}_h = k_h\mathbf{\Lambda}$ for all h , where $k_h > 0$, $\mathbf{\Lambda} = \text{diag}(\mathbf{S}\mathbf{1})$, and $\mathbf{1}$ is unit vector of dimension m (the number of bottom-level series)
3. **WLS_V (Hyndman et al. 2016)**: $\mathbf{W}_h = k_h\text{diag}(\hat{\mathbf{W}}_1)$ for all h , where $k_h > 0$,

$$\hat{\mathbf{W}}_1 = \frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t',$$

and \mathbf{e}_t is an n -dimensional vector of residuals of the models that generated the base forecasts stacked in the same order as the data.

4. **MinT_S (Wickramasuriya et al., 2019)**: $\mathbf{W}_h = k_h\hat{\mathbf{W}}_{1,D}^*$ for all h , where $k_h > 0$, and $\hat{\mathbf{W}}_{1,D}^* = \lambda\hat{\mathbf{W}}_{1,D}^* + (1-\lambda)\hat{\mathbf{W}}_1$ is a shrinkage estimator with diagonal target $\hat{\mathbf{W}}_{1,D}^*$, a diagonal matrix comprising the diagonal entries of $\hat{\mathbf{W}}_1$, and λ the shrinkage intensity parameter.

Cross-sectional reconciliation

Temporal reconciliation

Cross-temporal reconciliation

6.3 Model Selection

As depicted in figure ??, land transfer duty (LTD) exhibits mutual correlations with variables such as sales and the home value index (HVI). It is important to note that these relationships are not

unidirectional; factors like sales and HVI may influence LTD, but changes in LTD can reciprocally impact these variables. This dynamic interaction is also observed with economic indicators like inflation and interest rates.

The Vector Autoregression (VAR) model, which predicts future values of multiple time series based on their historical data, is well-suited for capturing these complex interactions. Furthermore, as illustrated in , the detection of cointegration patterns indicating a stable long-term equilibrium relationship among variables such as LTD, sales, and HVI, which necessitates the application of the Vector Error Correction Model (VECM). This highlights the pertinence of employing both VAR and VECM to adequately model these relationships.

One limitation of these models is their reliance on ample data to produce reliable parameters. Monthly data spanning over a decade for LTD has been found to be sufficient. Nevertheless, the subsequent section on cross-validation will elaborate on the precise sizing of the training dataset required to ensure compatibility with the VAR and VECM models.

In addition to VAR and VECM, the Autoregressive Integrated Moving Average (ARIMA) model has also been employed, primarily for purposes of comparison and validation of improvements. The efficacy of this comparison will be assessed through time series cross-validation, utilizing various accuracy metrics such as the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Mean Absolute Scaled Error (MASE).

7 Time Series Cross-Validation

7.1 Definition and rationale

Time series cross-validation represents a sophisticated adaptation of the conventional training/test set approach for model selection, as stated by Hyndman and Athanasopoulos (2021). This methodology is particularly well-suited to time series data because it exclusively includes observations from periods prior to those being forecasted, distinguishing it from traditional cross-validation techniques.

In the scope of this project, time series cross-validation is employed to rigorously evaluate whether the VAR/VECM or ARIMA models yield more accurate forecasts for this dataset across forecasting horizons ranging from 1 to 12 steps ahead. Additionally, this approach is used to gauge the extent of improvement introduced by the reconciliation process in comparison to the b base forecasts. Another critical application of time series cross-validation in this context is to generate rolling forecasts over a 12-month period for various time intervals. This is essential for the Department of Treasury and Finance (DTF) to analyze the effects of market dynamics or policy changes on land transfer duty (LTD).

7.2 Forecasting

To create folds for the time series cross-validation procedure, we will use `stretch_tsibble()` function with user defined value for two arguments, `.init` for size of initial training set, and `.step` to define how many steps training sets will roll forward each time.

For example, `stretch_tsibble(.init = 10, .step = 1)` produces series of training sets, where first training set has size of 10 observations and will roll 1 step forward each time, i.e. the size of second set and third set will be 11 and 12, respectively. Each folds will be denoted by a unique `.id` value.

Generating forecast

To enhance forecast accuracy, this methodology extends beyond a cross-sectional hierarchical structure to incorporate a temporal dimension, culminating in cross-temporal reconciliation forecasting. This approach integrates forecasts across different time structures, and its efficacy is evaluated by comparing it with cross-sectional and temporal reconciliation forecasts individually to determine any improvements.

Given the monthly frequency of the data, additional aggregation levels such as bi-monthly, quarterly, four-monthly, semi-annually, and annually are established, with the annual aggregation representing the top level of the temporal hierarchy.

Models are fitted at each cross-sectional level across these varying temporal frequencies. A `for` loop is utilized to generate forecasts and residuals, which are subsequently organized into a matrix for each temporal frequency: monthly (denoted as `k1`), bi-monthly (`k2`), quarterly (`k3`), four-monthly (`k4`), semi-annually (`k6`), and annually (`k12`). These matrices are then collectively stored within a list.

This structured approach facilitates the fitting of different models tailored to each temporal frequency, allowing for adjustments in argument values as necessary. Moreover, if a distinct model is required for various cross-sectional levels, an `if` condition is employed within each temporal frequency loop to ensure this customization. The forecasts and residuals are then allocated to `base` and `res` data structures, respectively.

ARIMA The ARIMA model will be implemented across all levels of the temporal hierarchical structure and at every cross-sectional level using the `auto.arima()` function from the `forecast` package. This function efficiently determines the optimal parameters for the autoregressive lag, differencing, and moving average components of the model by automatically adjusting them for the error terms.

Forecasts will be generated up to one year, or 12 months into the future, utilizing the `forecast()` function. This function is designed to produce both the mean point forecasts and the associated residuals, thereby providing a comprehensive output that includes predictions and their accuracy measures.

VAR/VECM Fitting VAR and VECM models to complex hierarchical structured data necessitates a more intricate approach, as there is no standard method readily available for this specific application. Additionally, these models require the incorporation of external variables, such as sales and the home value index (HVI). To address these challenges, the chosen strategy involves the creation of two separate user-defined functions for VAR and VECM. These functions are designed to process the input data and output both the mean point forecasts and residuals, mirroring the functionality provided by the `auto.arima()` function. Given the significant differences in the scales of sales, HVI, and LTD data, it is advisable to employ logarithmic transformations prior to model fitting, followed by re-transformation to their original scales. This method will be embedded within both functions to ensure accurate scale representation.

Vector Autoregression (VAR) Model The VAR fitting function, designated as `var_forecast_fun()`, accepts six parameters:

- `train`: A series representing LTD data across various hierarchical levels.
- `sales`: The time series data for sales.
- `hvi`: The time series data for the Home Value Index (HVI).
- `period`: The temporal frequency for generating nested Date variables, such as month, 2 months, quarter, etc.
- `length`: The total number of rows in the input data.
- `fc_range`: The forecast range, which may include 12 months for monthly data, or 6 two-month periods for bimonthly data, etc.

The input data is consolidated into a single dataframe, or `tibble` object, with an appropriately nested Date variable. This `tibble` is subsequently transformed into a `tsibble` object utilizing the `as_tsibble()` function from the `tsibble` package. A VAR model is then applied to this `tsibble` using the `VAR(vars(train, sales, hvi))` syntax, where parameters such as lag are automatically determined by the function. However, temporal aggregation results in insufficient observations for semi-annual and annual frequencies if the lag exceeds two. To address this limitation, an `if` condition restricts the chosen lag for these data frequencies to one.

Post model fitting, forecasts and residuals are computed as standard. Nevertheless, due to the lag configuration, there might be instances of missing residual values. Given that the FoReco package does not support missing values in the residual input matrices, these missing values are substituted with the calculated mean of the residuals. Since the number of these substituted values ranges from one to three, their impact on the reconciled results is anticipated to be minimal.

Ultimately, the function is structured to output both the mean point forecast and residuals, aggregated into a list format for subsequent extraction.

Vector Error Correction Model (VECM) The VECM fitting function, denoted as `vecm_forecast_fun`, accepts the same parameters as `var_forecast_fun` along with an additional `lag` parameter. This parameter facilitates the selection of varying lag lengths. The procedure for transforming the input data into a `tsibble` object remains consistent with that of the VAR model.

A distinct feature that differentiates VECM from VAR is its accommodation for non-stationary variables or those exhibiting cointegration patterns. The Johansen test, executed via the `ca.jo()` function from the `urca` package, assesses the presence of cointegration. The `K` parameter of `ca.jo()`, which determines the lag order in the VAR model employed within the Johansen procedure, is specified by the `lag` argument. This parameter can also be chosen using cross-validation to ascertain the optimal value for each temporal frequency or cross-sectional level, employing information criteria such as Akaike (AIC) or Bayesian (BIC).

Subsequent to the cointegration test, a significance level of 5% helps determine the value of `r`, representing the number of cointegrating relationships to be included in the VECM model. Should the Johansen test indicate an absence of a cointegration pattern, particularly in annual frequency data due to insufficient observations, a VAR model is utilized instead. If cointegration is confirmed, the VECM fitting function used is `vec2var()` from the `vars` package.

Similar to the VAR model, the VECM procedure also involves the generation of forecasts and residuals. Additionally, it addresses missing residual values caused by lag length, in a manner analogous to the `var_forecast_fun` procedure. The output of this function comprises a list containing the mean point forecast and residuals.

7.3 Reconciliation Process

The reconciliation procedure is consistent across ARIMA and VAR/VECM cross-validation models. Initially, forecasts and residuals from all cross-sectional levels and temporal frequencies are amalgamated into `base` and `res` datasets. The FoReco package facilitates the reconciliation process, offering integrated solutions for cross-sectional, temporal, and cross-temporal reconciliation.

- **htsrec()**: Implements cross-sectional reconciliation utilizing `shr`, which denotes the use of a shrunk covariance matrix—specifically MinT-`shr`.
- **thfrec()**: Facilitates temporal reconciliation employing `struc` for the computation of structural variances.
- **octrec()**: Conducts optimal cross-temporal reconciliation, using `struc` to calculate cross-temporal structural variances.
- **tcsrec()**: Executes heuristic first-temporal-then-cross-sectional reconciliation, combining structural variances for temporal aggregation with a bottom-up approach for cross-sectional aggregation.

The output of this reconciliation process is a matrix where rows represent different cross-sectional levels, and columns span from a 1-step to a 12-step-ahead forecast, reflecting the original scale of land transfer duty (LTD).

Subsequently, this matrix is integrated into a nested structure within an array corresponding to its designated `.id` value. This entire forecasting and reconciliation process is then repeated for subsequent folds or `.id` values.

8 Results

8.1 Model Performance:

Presentation of model accuracy and comparison.

8.2 Forecast Results:

Detailed discussion of the forecasted values and their confidence intervals.

9 Discussion

9.1 Interpretation of Results:

Analysis of what this forecasts mean for DTF

9.2 Limitations and Assumptions:

Any limitations encountered during the forecasting process.

10 Conclusion and Recommendations

10.1 Summary:

Recap the findings and their implications.

10.2 Future Work:

Suggestions for improving future forecasts.

11 Appendix

Table 1: *Summary Statistics of variables*

	Min	Quartile_1	Median	Mean	Quartile_3	Max
ltd_total	286,162,171	434,646,740	501,902,870	542,225,112	634,212,641	1,013,171,645
ltd_nonres	42,909,678	76,682,419	97,505,851	109,079,081	135,555,977	357,222,264
ltd_comm	22,106,375	42,135,408	54,959,990	59,467,363	70,539,203	132,405,000
ltd_ind	5,229,382	17,094,782	24,748,954	31,256,622	37,482,400	214,372,653
ltd_other	3,325,394	9,464,705	14,776,865	18,252,446	24,363,583	74,469,171
ltd_res	238,100,812	346,798,966	406,948,825	433,146,031	489,072,115	761,393,438
sales	4,629	8,414	9,697	9,680	11,158	15,177
hvi	106.2	132.5	155	152.9	178.4	192.4

12 References

12.1 Bibliography:

Cite all data sources, literature, and software used in the report the insights generated by your work.

Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on .

https://treasury.gov.au/sites/default/files/2022-03/258735_real_estate_institute_of_australia.pdf

https://www.treasury.nsw.gov.au/sites/default/files/2021-06/the_economic_costs_of_transfer_duty_a_literature_rev