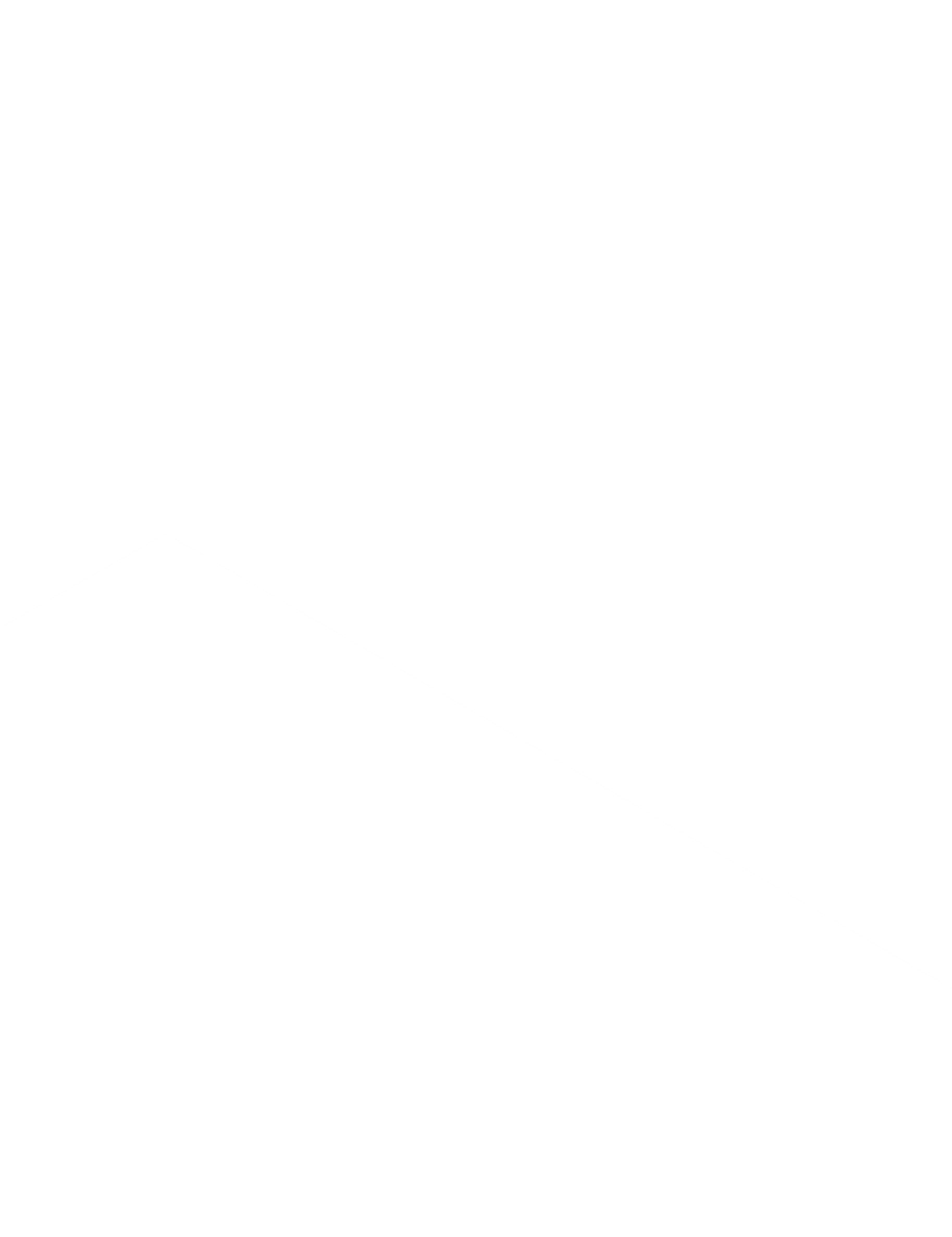
Model Risk Guideline

Effective: June 2020

Version 2.0

Protected – Operations / Proprietary



Multi-Unit Volume Projection Model

Technical Documentation

Multi-Unit & Assisted Housing Oversight and Risk Analytics

Sector of the Chief Risk Officer

Version Number: 1.1

Version Date: March 31, 2021

Version History

Changes made to the model documentation are listed below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Updated By | Approved By | Description of Change | Date |
| 1.0 | Moody’s Analytics |  | First version of the model documentation and model | Nov. 06, 2020 |
| 1.1 | Qianyi Feng |  | Second version of the model documentation and model | Mar.31, 2021 |

# TABLE OF CONTENTS

[1. TABLE OF CONTENTS 3](#_Toc55555021)

[2. TABLE OF TABLES 5](#_Toc55555022)

[3. TABLE OF FIGURES 5](#_Toc55555023)

[4. TERMINOLOGY 6](#_Toc55555024)

[4.1 Acronyms 6](#_Toc55555025)

[4.2 Variable naming convention 6](#_Toc55555026)

[5. MODEL EXECUTIVE SUMMARY 7](#_Toc55555027)

[6. MODEL THEORY AND APPROACH 8](#_Toc55555028)

[6.1 Product and Region Segmentation 8](#_Toc55555029)

[6.2 Temporal Segmentation 9](#_Toc55555030)

[6.3 Modeling Framework 11](#_Toc55555031)

[6.3.1 Unit root tests 11](#_Toc55555032)

[6.3.2 Units models 12](#_Toc55555033)

[6.3.3 Price models 12](#_Toc55555034)

[6.3.4 Econometric Model 13](#_Toc55555035)

[7. DATA 14](#_Toc55555036)

[7.1 Data Types and Data Sources 14](#_Toc55555037)

[7.1.1 Loans data 14](#_Toc55555038)

[7.1.2 Economic Data 14](#_Toc55555039)

[7.2 Data Issues, Cleaning and Exclusions 15](#_Toc55555040)

[7.2.1 Outlier loans 15](#_Toc55555041)

[7.2.2 RPI in Atlantic issue 16](#_Toc55555042)

[7.3 Economic Variables Unit Root Tests 17](#_Toc55555043)

[8. MODEL ESTIMATION 19](#_Toc55555044)

[8.1 Model Development Process 19](#_Toc55555045)

[8.1.1 Serial correlation testing 19](#_Toc55555046)

[8.1.2 Economic variables transformations and lags 19](#_Toc55555047)

[8.1.3 Candidate economic variables selection 20](#_Toc55555048)

[8.1.4 Dynamic predictions 20](#_Toc55555049)

[8.1.5 Model selection 20](#_Toc55555050)

[8.2 Units Models 21](#_Toc55555051)

[8.2.1 Serial correlation results – basic model 21](#_Toc55555052)

[8.2.2 Final models 22](#_Toc55555053)

[8.3 Price Models 24](#_Toc55555054)

[8.3.1 Serial correlation results – basic model 24](#_Toc55555055)

[8.3.2 Final Model 25](#_Toc55555056)

[9. INITIAL MODEL TESTING 28](#_Toc55555057)

[9.1 Units Models 28](#_Toc55555058)

[9.1.1 Final Units models fit stats 28](#_Toc55555059)

[9.1.2 Final Units models fit plots 28](#_Toc55555060)

[9.2 Price Models 31](#_Toc55555061)

[9.2.1 Final Price models fit stats 31](#_Toc55555062)

[9.2.2 Final Price models fit plots 31](#_Toc55555063)

[10. MODEL USER MANUAL 34](#_Toc55555064)

[10.1 Economic Data Preparation 34](#_Toc55555065)

[10.1.1 CMHC sources 34](#_Toc55555066)

[10.1.2 DataBuffet data 35](#_Toc55555067)

[10.2 Model implementation 36](#_Toc55555068)

[10.2.1 Codes 36](#_Toc55555069)

[11. REFERENCES 38](#_Toc55555070)

# TABLE OF TABLES

[Table 1: List of Acronyms 6](#_Toc55554999)

[Table 2: Projections product and region segments 8](#_Toc55555000)

[Table 3: Segments data coverage analysis- share of quarters with 3 loans or less, % 8](#_Toc55555001)

[Table 4: Unit root tests for Units and Price 12](#_Toc55555002)

[Table 5: Loan origination variables from mollea Master Data 14](#_Toc55555003)

[Table 6: Economic variables used in model development 14](#_Toc55555004)

[Table 7: Outlier loans with segment aggregate metrics and reason for exclusion 16](#_Toc55555005)

[Table 8: Correlation between Condominium RPI and Single-Family RPI by Region 16](#_Toc55555006)

[Table 9: Economic variables units root tests 18](#_Toc55555007)

[Table 10: Universe of economic variables considered for modeling 19](#_Toc55555008)

[Table 11: Units AR models by product 21](#_Toc55555009)

[Table 12: Final Units models 22](#_Toc55555010)

[Table 13: Price AR models by product 24](#_Toc55555011)

[Table 14: Final Price models 25](#_Toc55555012)

[Table 15: Final Units models fit stats 28](#_Toc55555013)

[Table 16: Final Price models fit stats 31](#_Toc55555014)

# TABLE OF FIGURES

[Figure 1: National share of loans with below 7 units by three aggregate products with 3-year moving average, % 8](#_Toc55554949)

[Figure 2: Quarterly number of units originated by region and product, # 9](#_Toc55554950)

[Figure 3: Quarterly average price of units originated by region and product, $000 10](#_Toc55554951)

[Figure 4: Condominium RPI and Single-Family RPI by Region over time 17](#_Toc55554952)

[Figure 5: Final Units models residuals 24](#_Toc55554953)

[Figure 6: Final Price models residuals 27](#_Toc55554954)

[Figure 7: Final Units - Rental New model fit plots 29](#_Toc55554955)

[Figure 8: Final Units - Rental Existing model fit plots 29](#_Toc55554956)

[Figure 9: Final Units - Refinance model fit plots 30](#_Toc55554957)

[Figure 10: Final Price - Rental New model fit plots 31](#_Toc55554958)

[Figure 11: Final Price - Rental Existing model fit plots 32](#_Toc55554959)

[Figure 12: Final Price - Refinance model fit plots 32](#_Toc55554960)

# TERMINOLOGY

## Acronyms

The following table collects important acronyms used in this document and their descriptions.

Table : List of Acronyms

|  |  |
| --- | --- |
| **Acronym** | **Description** |
| mollea | Multi-unit Lifecycle, Loss and Exposure Analyzer |
| OLS | Ordinary Least Squares |
| RPI | Repeated Prices Index |
| NSA | Not Seasonally Adjusted |
| SA | Seasonally Adjusted |
| SAAR | Seasonally Adjusted Annual Rate |
| AR | Autoregressive |

## Variable naming convention

The following variable naming convention is used in this document:

* “yd\_” before variable name denotes year-over-year growth.
* “yp\_” before variable name denotes year-over-year growth rate.
* “\_#” number (#) after variable name denotes lagged value in # of lagged years.

Examples:

* “yp\_gdp\_2” – year-over-year growth rate of gdp, lagged by 2 years
* “yd\_pop” – contemporaneous year-over-year growth in population.
* “stockp\_1” – stock prices in level, lagged by 1 year.

# MODEL EXECUTIVE SUMMARY

The multi-unit volume projection model developed by using the mollea (Multi-unit Lifecycle Loss and Exposure Analyzer) loans data from 2004 to 2019 and economic data coming from CMHC internal sources and Moody’s Analytics DataBuffet. The objective of this model is to forecast the multi-unit business volumes to:

* More accurately determine future MU volumes, and in doing so, allow CMHC to retain only the necessary economic capital given expected losses
* Support the credit assessment team in determining the application volumes they will face in the coming years, thus allowing them to better plan their underwriter needs
* Assess the supply of affordable housing coming on the market in the coming years, to determine whether it will meet the demand

In terms of the model development data, the source of the modelling data for loan originations is the mollea Master Dataset created as part of the mollea Data Preprocessing Analyzer (DPA). The mollea Master Dataset include data on 51,587 unique loans originated over the period 2000-2019. However, from a modelling data perspective and the data limitation of refinance segment, the final sample for modelling development is from 2004 to 2019 and the final modeled number of loans after exclusion and removal of pre-2004 is 44,784.

For the multi-unit volume projection model, Modeling, Oversight, and Risk Analytics (referred to as MORA hereafter) in conjunction with Moody’s Analytics developed two distinct model sets, one for Units and one for Price. Each of the modeling sets is divided into three product specific models, so in total this modeling effort had six separate models. And, this model set out to forecast volumes and values for six multi-unit products over five regions.

A panel time-series regression is chosen for the model framework, the data is an aggregated time series per segment (region and product) and should be modeled as such. This methodology conforms to industry standards and best practices and academic literature. In particular, considering data volatility and the short time-series available, regions were pooled into a panel and estimated together in six national models.

The model performance results are provided for the Units and Price model, respectively. MORA team calculated the divergence between the final model outputs with the actual data and compared the model divergence between the old business model and the multi-unit volume projection model by using the most current data from 2016 to 2020.

In order to assist model users implement this model, MORA team develops the user interface and manual to smoothly transit this model into production. The multi-unit volume projection model is developed by using the RStudio (Version 1.2.5033) software. Additionally, the MORA team migrates this model to the CMHC AIP Rstudio platform.

In summary, this document describes the development of the volume projection model, including the data used, the modeling techniques, the model performance results, the model development process, and model user guidelines.

# MODEL THEORY AND APPROACH

## Product and Region Segmentation

This model set out to forecast volumes and values for 6 multi-unit products over 5 regions:

Table : Projections product and region segments

|  |  |
| --- | --- |
| **Products** | **Regions** |
| Rental New - Above 7 units | Atlantic |
| Rental New - Below 7 units | Quebec |
| Rental Existing - Above 7 units | Ontario |
| Rental Existing - Below 7 units | Prairies and Territories |
| Refinance - Above 7 units | British Columbia |
| Refinance - Below 7 units |  |

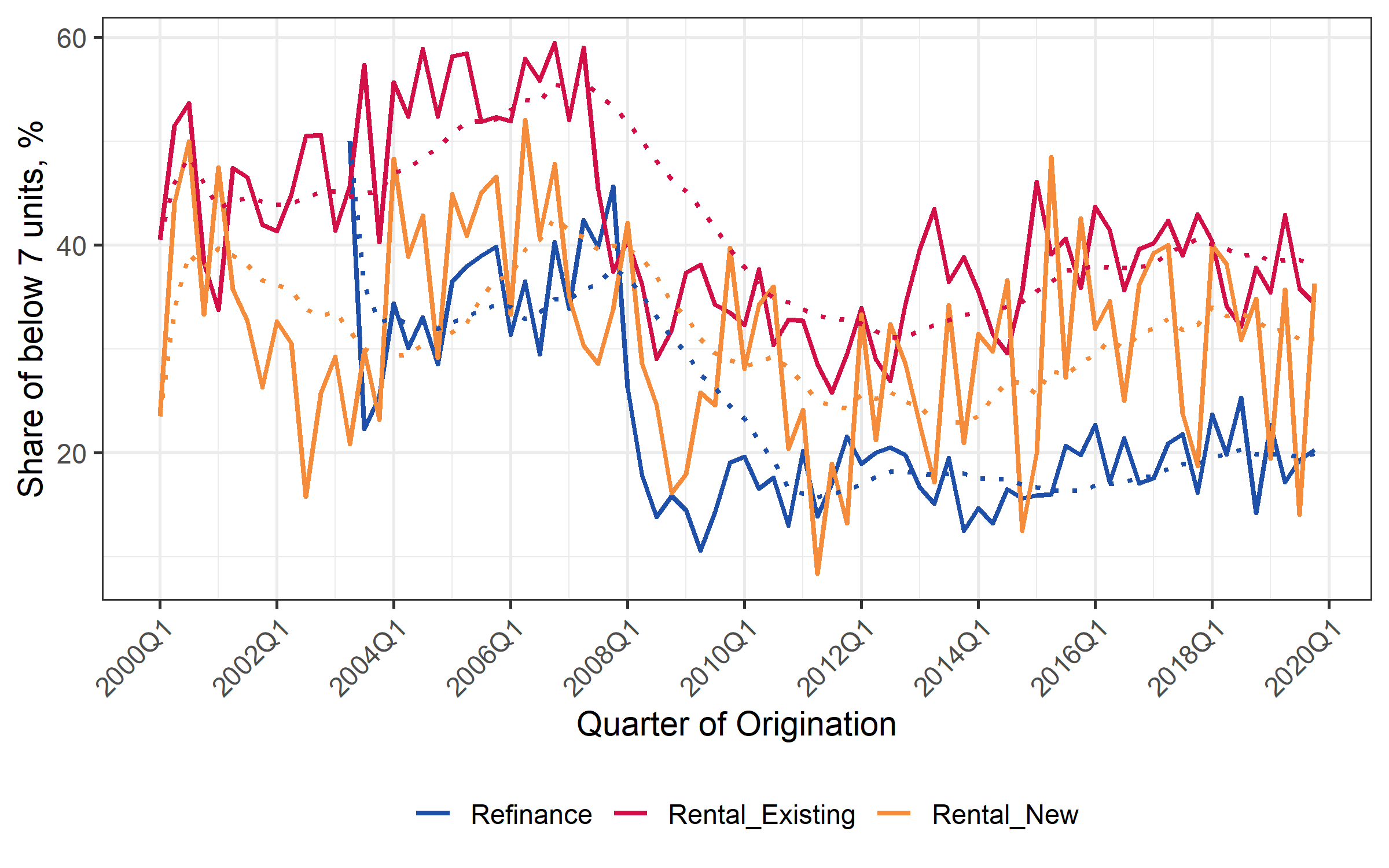
Development data included data from 2000 to 2019 (Refinance products starting at Q2 of 2003). Initial data analysis examined data coverage for the above 30 segments (5 regions times 6 products) when loan originations are aggregated on a quarterly basis. Analysis revealed that data is too thin and too volatile for some segments, with some having no loans at all for specific quarters. As presented in Table 3, many region-product segments have many observations (quarters) with 3 loans or less which is highly susceptible to high volatility.

Table : Segments data coverage analysis- share of quarters with 3 loans or less, %

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Refinance** | | **Rental Existing** | | **Rental New** | |
| **Region** | **Below 7 units** | **above 7 units** | **Below 7 units** | **above 7 units** | **Below 7 units** | **above 7 units** |
| Atlantic | 70.1% | 1.5% | 57.5% | 17.5% | 87.5% | 6.3% |
| British Columbia | 98.5% | 1.5% | 85.0% | 0.0% | 98.8% | 71.3% |
| Ontario | 16.4% | 1.5% | 25.0% | 0.0% | 98.8% | 13.8% |
| Prairies and Territories | 68.7% | 1.5% | 57.5% | 8.8% | 87.5% | 32.5% |
| Quebec | 1.5% | 1.5% | 0.0% | 0.0% | 2.5% | 1.3% |

Figure 1 presents the share of loans from below 7 units products over the total historical data available (2000-2019) including a 3-year moving average. The average share of above and below 7 units is stable over the past few years for most products and regions. Considering that, for forecasting purposes the data is aggregated to three products: Rental New, Rental Existing and Refinance. The forecasted numbers will be shared down to the original 6 products using the average share of below/above 7 units over the past 3 years.

Figure : National share of loans with below 7 units by three aggregate products with 3-year moving average, %



## Temporal Segmentation

As seen in the charts below, even after aggregating data into 3 product segments, quarterly data was very volatile and contained a few missing observations. An attempt was made to estimate models for both number of units and price using the quarterly data. Quarterly models required several dummies to control for extreme values and fit metrics were lower than annual models. In addition, most series did not have any seasonal pattern and prices have a trend, so modeling at the annual frequency is a sensible solution.

Figure : Quarterly number of units originated by region and product, #

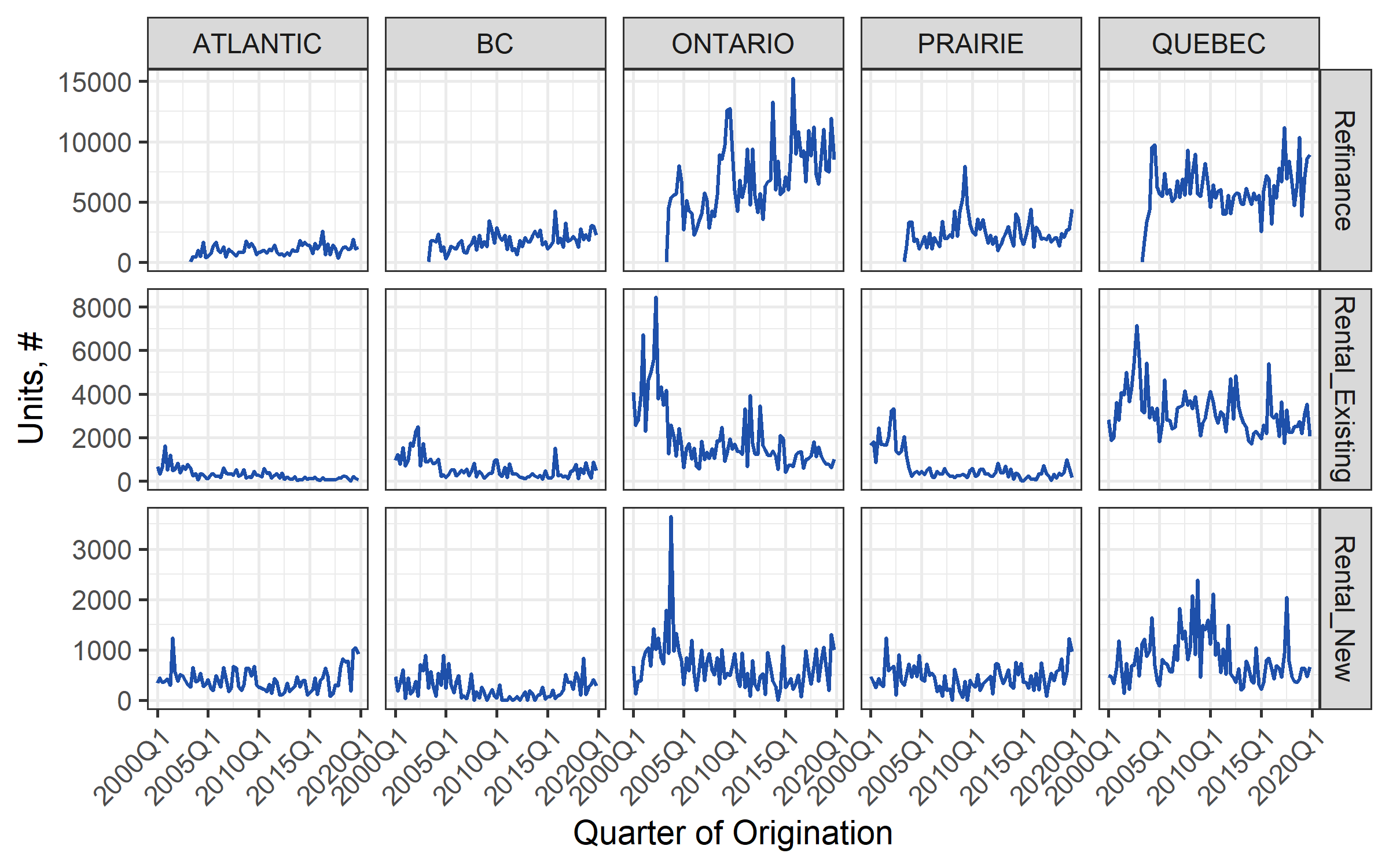
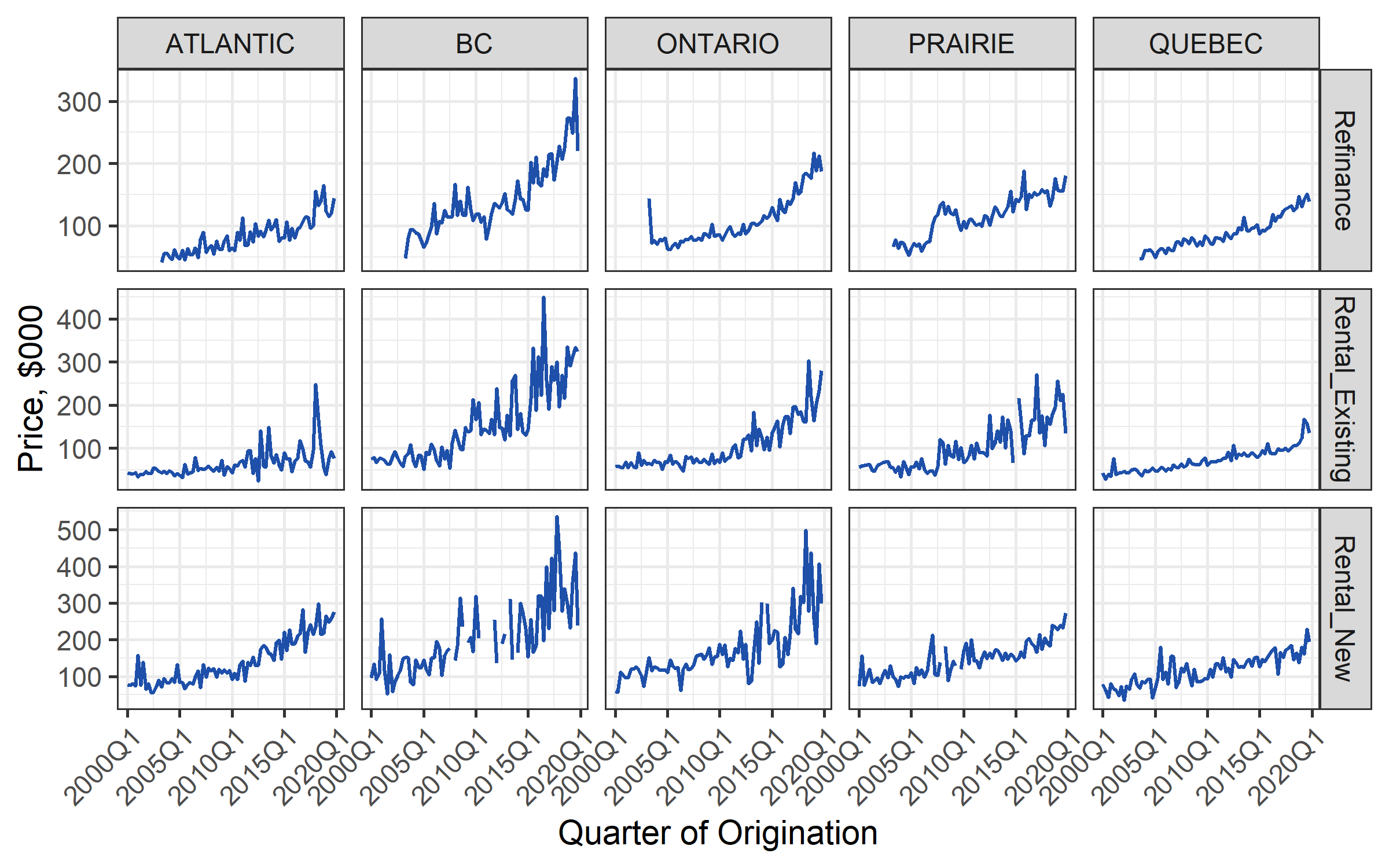


Figure : Quarterly average price of units originated by region and product, $000



As a result of the above analysis, Modeling, Oversight and Risk Analytics in conjunction with Moody’s Analytics have chosen a regional panel model for each of the three products segmentation using annual aggregation for modeling.

## Modeling Framework

In order to forecast loan originations by number of units and loan value, Modeling, Oversight and Risk Analytics in conjunction with Moody’s Analytics have chosen to model the number of units and their corresponding decomposed price per unit. Appropriately, two distinct model sets were developed, one for Units and one for Price. Each of the modeling sets are divided into three product specific models, so in total this modeling effort had six separate models. The general framework chosen for the analysis is a panel time-series regression, the data is an aggregated time series per segment (region and product) and should be modeled as such. This methodology conforms to industry standards and best practices and to academic literature. Considering data volatility and the short time-series available, regions were pooled into a panel and estimated together in six national models. As detailed below, each of the two model sets (Units and Price) had a separate modeling framework appropriate to its distinct features. All models were tested using appropriate panel time-series analysis tests (i.e. stationarity, serial-correlation, spatial dependence).

### Unit root tests

The first step in choosing the econometric framework was to decide on the appropriate modeled variable. Which transformation should be used for Units and Prices? For that, unit root tests were performed on the modeled variables in levels and transformed. Table 4 below shows panel unit root tests after product data was aggregated into the 5 regions. The reason for aggregation is missing data in some segment observations causing an unbalanced panel. Units is stationary in levels as well as transformed. Price in levels is as expected non-stationary but is stationary in growth and growth rate.

Table : Unit root tests for Units and Price

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Maddala and Wu** (1999) | **PM – Choi** (2001) | **invnormal - Choi** (2001) | **logit - Choi** (2001) |
| **Hypothesis** | | **H1: stationarity** | **H1: stationarity** | **H1: stationarity** | **H1: stationarity** |
| **Units** | Level | 0.046 | 0.028 | 0.054 | 0.042 |
| Annual growth | 0.000 | 0.000 | 0.000 | 0.000 |
| Annual growth rate | 0.000 | 0.000 | 0.010 | 0.000 |
| **Price** | Level | 0.987 | 0.948 | 1.000 | 1.000 |
| Annual growth | 0.000 | 0.000 | 0.000 | 0.000 |
| Annual growth rate | 0.000 | 0.000 | 0.000 | 0.000 |

For Units all three options were considered as possible candidates, as levels were considered as stationary. For Price only the transformed versions were considered as a candidate for the modeled variable. In the next sections, the selection of the final dependent variable is discussed.

### Units models

In order to decide on an appropriate econometric framework, units in levels, annual growth, and growth rate were all estimated using a basic panel model with fixed effects and a few candidates economic variables. RMD estimated a basic model for each of these options, with either units in levels, annual growth or growth rate on the Left hand side and checked which had a better fit out-of-sample using the fit metrics. Therefore, the Units in levels gave the best modeling and forecasting results and was selected. The results showed that modeling Units in levels gave the best fit. In addition, it had a much simpler and more efficient code as it did not require dynamic un-differencing the predicted values back to Units in levels.

Units models econometric framework:

* Models – 3 models, one for each product: Rental New, Rental Existing, and Refinance.
* Data structure – national panel time-series for the 5 regions: Atlantic, Quebec, Ontario, Prairies and Territories, and British Columbia.
* Modeled variable – Units in levels.
* Estimation method – Ordinary Least Squares (OLS).
* Fixed effects – yes.
* Robust standard errors – yes.

### Price models

As all economic variables are used in real terms, real prices were used by deflating nominal prices with regional CPI. Forecasted prices are multiplied by CPI in order to transform them back to their nominal values. Based on commonly used techniques, annual growth rates were picked as the modeled variable.

Price models econometric framework:

* Models – 3 models, one for each product: Rental New, Rental Existing, and Refinance.
* Data structure – national panel time-series for the 5 regions: Atlantic, Quebec, Ontario, Prairies and Territories, and British Columbia.
* Modeled variable – annual growth rate of real Price.
* Estimation method – Ordinary Least Squares (OLS).
* Fixed effects – no, were insignificant.
* Robust standard errors – yes.

## Econometric Model

The intuitive approach described above can be formalized using a mathematical specification and estimated using statistical techniques to construct high-quality forecasts. As our general approach we consider a model of the form:

where indicates Canada’s 5 regions, is the year of origination time series indicator and is the dependent variable of interest, Units in levels or Price growth rate. The vector is an unknown random error term. We can further decompose into its components:

The vectors, and and contain independent variables thought to explain the behavior of . Regional fixed-effects are represented by , these effects do not change over time and contain regional time-invariant characteristics not captured by other variables. The vector contain factors that impact loan originations and change both over time and over space, such as regional unemployment rate or RPI. Other type of factors are national economic variables which do not change between regions, such as interest rates and stock prices, those are represented by the vector.

# DATA

## Data Types and Data Sources

Modeling data has two general types and sources: loans data, which comes from mollea Master Dataset and economic data, which comes from both CMHC and outside sources.

### Loans data

The source of the data for loan originations is the mollea Master Dataset created as part of the mollea - Data Preprocessing Analyzer (please see corresponding documentation for further information). The mollea Master Dataset used is the *prod\_multi\_full\_19q4\_newLBA* file, located in the following sub-folder:

[\\Prdermfs01\erm\Insurance\Models\Inputs Data\](file:///\\Prdermfs01\erm\Insurance\Models\Inputs%20Data\)pragma\20191231\Production\

The mollea Master Dataset contain data on 51,587 unique loans originated over the period 2000-2019. As detailed below, 7 loans were excluded because of extreme Units or Prices numbers. The final modeled number of loans after exclusion and removal of pre-2004 data is 44,784.

For each unique loan in the Master Dataset, data on its origination was extracted from the below variables:

Table : Loan origination variables from mollea Master Data

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Definition** | **Variable type** | **Use** |
| start\_date | Date of loan origination | Date | Aggregation into origination year |
| product\_n | Product code number | Factor | Aggregation into product segment |
| province\_n | Province code number | Factor | Aggregation into region segment |
| multi\_value | Loan value at origination | Numeric | Average price per unit for modeling |
| units\_n | Number of housing units at origination | Numeric | Number of units for modeling |

Unique loans were aggregated into annual product-region segments as discussed above. For each observation, number of housing units in loans and average price per unit were calculated.

The Refinance product was introduced in 2003 and appears in the data staring the second quarter of the year. The introduction of the new product had impact on the two modeled Rental products. As a result, in first iteration of the modeling effort we had to control for Refinance using dummies in the two Rental models. This approach did not add to model out of sample fit and added unnecessary complication, and so it was decided to remove those years from the data for the two Rental products. As Refinance was only picking up originations during 2003 numbers and volatility was high in that year. Because of all the above, final sample for all three products is 2004-2019.

### Economic Data

Economic data comes from two sources: CMHC internal sources and Moody’s Analytics DataBuffet. The below economic variables were used in the model development process. Transformations and lags of these variables will be discussed in section 8.1.2.

Table : Economic variables used in model development

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Description** | **Units** | **Geo** | **Source** | **Original Source** |
| vac\_rate | Vacancy rate | %, NSA | Regional | CMHC | CMHC |
| condo\_rpi\* | Condominium Repeated Sales Index | Index Jan-2001=100, NSA | Regional | CMHC | CMHC |
| singles\_rpi | Single family homes Repeated Sales Index | Index Jan-2001=100, NSA | Regional | CMHC | CMHC |
| ump\_rate | Labor Force: Unemployment Rate | %, SA | Regional | DataBuffet | Statistics Canada |
| emp | Labor Force: Total employed | Thousand #, SA | Regional | DataBuffet | Statistics Canada |
| gdp | Real Gross Domestic Product | Billion Chained 2012 CAD, SAAR | Regional | DataBuffet | Statistics Canada |
| hstarts | Housing starts | Thousand #, SAAR | Regional | DataBuffet | CMHC |
| cpi | Consumer Price Index: Total | Index 2002=100, SA | Regional | DataBuffet | Statistics Canada |
| cpi\_rent | Consumer Price Index: Shelter - Rent | Index 2002=100, NSA | Regional | DataBuffet | Statistics Canada |
| pop | Population: Total | Thousand #, SA | Regional | DataBuffet | Statistics Canada |
| rmort5y | Average Residential Mortgage Lending Rate - 5 Year | %, NSA | National | DataBuffet | CMHC |
| rprime | Interest Rate: Prime Business Rate | %, NSA | National | DataBuffet | Statistics Canada; Bank of Canada |
| rmort\_spread | Residential Mortgage Prime rate (rmort5y - rprime) | %, NSA | National | Calculated | Statistics Canada; Bank of Canada |
| rgt5y | Interest Rate: 5-year Bond Yield | %, NSA | National | DataBuffet | Statistics Canada; Bank of Canada |
| rgt10y | Interest Rate: 10-year Bond Yield | %, NSA | National | DataBuffet | Statistics Canada; Bank of Canada |
| stockp | Stock Market: S&P/TSX Composite Index | Index, NSA | National | DataBuffet | SIX Financial Information |
| cpwti | Futures Price: NYMEX Light Sweet Crude Oil - Contract 1 | USD per barrel, NSA | National | DataBuffet | U.S. Energy Information Administration (EIA) |

notes: NSA – Not Seasonally Adjusted; SA – Seasonally Adjusted; SAAR - Seasonally Adjusted Annual Rate

\* Missing for Atlantic, see section 7.2.2

## Data Issues, Cleaning and Exclusions

### Outlier loans

As stated above, eight loans in the mollea Master Dataset were identified as extreme outliers because of extreme price values, which do not align with average numbers for the specific region and time. Table 7 below depicts those loans, their number of units and average price and the same aggregate metrics for all other loans in that segment and year of origination. Seven of the outlier loans were excluded as they presented either very high of very low price per unit and the segment had enough data from other loans. The eighth loan (loan id = 90495565) came from a thin segment and thus further analysis was performed to avoid loss of data in that observation. The analysis revealed that the loan total value included also a non-residential part, as a fix the non-residential share of the loan was removed, so that average price per unit will include only residential units value. This should be made because the models are for residential properties only, so the non-residential value had to be taken out. The per housing unit value of this loan is extremely high so we adjusted the value to residential only.

Table : Outlier loans with segment aggregate metrics and reason for exclusion

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **Region** | **Product** | **Outlier loans** | | | **Corresponding segment (all other loans)** | | | **Decision and reason** |
| **Loan ID** | **Units, #** | **Price, $000** | **Sum of loans, #** | **Sum of units, #** | **Average price, $000** |
| 2002 | Quebec | Rental Existing | 90157157 | 60 | 520,733.3 | 652 | 20,530 | 45.9 | Exclusion: very high price |
| 2006 | Quebec | Rental New | 90256124 | 231 | 3.0 | 146 | 3,003 | 132.6 | Exclusion: very low price |
| 2006 | Ontario | Rental New | 90261033 | 128 | 8.0 | 32 | 2,344 | 122.1 | Exclusion: very low price |
| 2009 | Quebec | Refinance | 90403684 | 402 | 472,321.2 | 789 | 26,875 | 73.6 | Exclusion: very high price |
| 2009 | Prairie and Territories | Rental New | 90433434 | 172 | 345.4 | 11 | 827 | 125.6 | Exclusion: very high price |
| 2013 | Ontario | Rental New | 90565540 | 256 | 30.4 | 17 | 2,085 | 181.9 | Exclusion: very low price |
| 2013 | Ontario | Rental New | 90566043 | 96 | 55.7 | 17 | 2,085 | 181.9 | Exclusion: very low price |
| 2011 | BC | Rental New | 90495565 | 72 | 408.1 | 3 | 21 | 254.5 | Apply fix: high price, thin data |

We summarized the treatments for exclusions and outliers as follows:

* Excluded loans with extreme price values that do not align with average numbers for the specific region and time
* Removed non-residential part of a specific loan since the models are for the residential properties
* Removed data before 2004 for all products

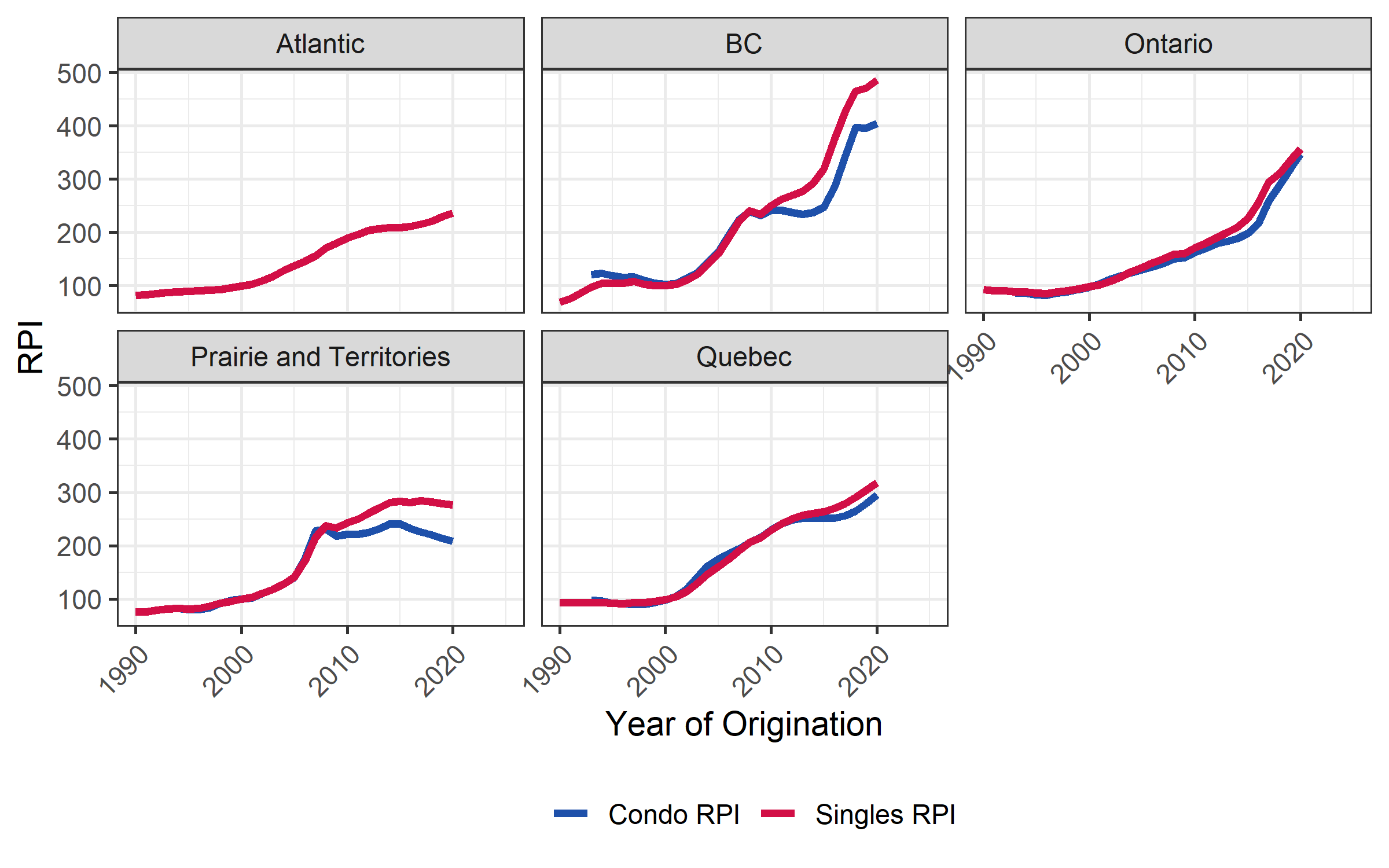
### RPI in Atlantic issue

Because condominium prices are expected to impact multi-family rental units’ originations and prices, the option of using CMHC’s condominium Repeated Prices Index (RPI) was explored. CMHC’s condominium RPI does not include any data on the Atlantic region so substituting it with single-family RPI was tested. The below table and charts show the high correlation between the two series for all regions over the modeled period.

Table : Correlation between Condominium RPI and Single-Family RPI by Region

|  |  |
| --- | --- |
| **Region** | **Correlation** |
| BC | 0.989 |
| Ontario | 0.995 |
| Prairie and Territories | 0.976 |
| Quebec | 0.994 |

Figure : Condominium RPI and Single-Family RPI by Region over time



Because condominium RPI is expected to have stronger impact on loan originations the single-family RPI it was decided to use the former. Considering the high correlation between condominium RPI and Single-Family RPI over all regions, for the Atlantic region the missing condominium RPI was supplemented by single-family RPI.

## Economic Variables Unit Root Tests

All economic variables were tested for unit roots in order to determine which transformation is needed for a stationary time-series estimation. Because panel unit root tests tend to have unreliable results, especially in short panels, regional-level variables were tested both in their regional form and in nationally aggregated form. The below table depicts the results of units root tests for all economic variables in both levels and transformed. For each variable either annual growth or annual growth rate transformation was selected according to commonly used econometric principles (i.e. growth rate for gdp, growth for rates). The conclusion column states the conclusion taken according to tests results for each variable, the results could be either I(0) – stationary in levels or I(1) – difference stationary.

Table : Economic variables units root tests

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Transformation** | **National data tests** | | **Regional data panel tests** | | | **Conclusion** |
| **ADF Test** (1984) | **Phillips-Perron Test** (1988) | **Maddala and Wu** (1999) | **PM - Choi** (2001) | **invnormal - Choi** (2001) |
| rpi | level | 0.670 | 0.938 | 0.000 | 0.000 | 0.000 | I(1) |
| growth rate | 0.601 | 0.370 | 0.000 | 0.000 | 0.022 |
| vac\_rate | level | 0.470 | 0.364 | 0.009 | 0.001 | 0.017 | I(1) |
| growth | 0.053 | 0.041 | 0.000 | 0.000 | 0.000 |
| ump\_rate | level | 0.428 | 0.242 | 0.000 | 0.000 | 0.095 | I(1) |
| growth | 0.011 | 0.052 | 0.001 | 0.000 | 0.577 |
| emp | level | 0.643 | 0.384 | 0.009 | 0.001 | 0.515 | I(1) |
| growth rate | 0.036 | 0.214 | 0.000 | 0.000 | 0.000 |
| pop | level | 0.990 | 0.990 | 0.939 | 0.904 | 0.998 | I(1) |
| growth rate | 0.642 | 0.525 | 0.000 | 0.000 | 0.000 |
| gdp | level | 0.438 | 0.675 | 0.616 | 0.662 | 0.937 | I(1) |
| growth rate | 0.337 | 0.190 | 0.000 | 0.000 | 0.095 |
| hstarts | level | 0.452 | 0.348 | 0.698 | 0.728 | 0.937 | I(1) |
| growth rate | 0.089 | 0.010 | 0.000 | 0.000 | 0.000 |
| cpi\_rent | level | 0.973 | 0.295 | 0.023 | 0.008 | 0.771 | I(1) |
| growth rate | 0.298 | 0.345 | 0.000 | 0.000 | 1.000 |
| rmort5y | level | 0.220 | 0.197 | - | | | I(1) |
| growth | 0.010 | 0.055 | - | | |
| rprime | level | 0.040 | 0.064 | - | | | Possibly I(0) |
| growth | 0.010 | 0.038 | - | | |
| rmort\_spread | level | 0.044 | 0.022 | - | | | Possibly I(0) |
| growth | 0.018 | 0.010 | - | | |
| rgt5y | level | 0.424 | 0.165 | - | | | I(1) |
| growth | 0.010 | 0.049 | - | | |
| rgt10y | level | 0.428 | 0.206 | - | | | I(1) |
| growth | 0.010 | 0.037 | - | | |
| stockp | level | 0.155 | 0.051 | - | | | I(1) |
| growth rate | 0.059 | 0.010 | - | | |
| cpwti | level | 0.781 | 0.633 | - | | | I(1) |
| growth rate | 0.215 | 0.010 | - | | |

Notes: Phillips-Perron Test has an H0: unit root hypothesis, while all other variables has a H1: stationarity hypothesis.

As seen above, most variables could be regarded as difference stationary variables (I(1)) with the exception of prime rates and mortgage rate spread. In order to avoid any possible non-stationary process in one of the estimated regions, all economic variables were used transformed to either growth or growth rate.

# MODEL ESTIMATION

## Model Development Process

Model development process for both Units and Price contained four major steps for each product. First, serial correlation of the panel time-series was analyzed and Autoregressive (AR) terms were selected as a basic model. Secondly, candidate transformations and lags for each economic variable were determined. Then, a process was run to select the best candidate economic variables. Lastly, candidate models were specified of which a winner model was selected. The below sub-sections elaborate the methods used in each of the above steps. In-sample analysis was done by estimating and predicting results for the entire development period: 2004-2019. Out-of-sample analysis was performed by estimating a model on a subset of the development data (training data): 2004-2014 and then predicting on the out-of-sample subset: 2015-2019.

### Serial correlation testing

In order to assess the required AR terms for each product model, the modeled variable was estimated with increasing AR terms up to four terms. Data was trimmed so all four models will have the same sample size. For each of the four AR models the following stats were assessed: AR terms significance, AIC, BIC and residuals panel Breusch-Godfrey AR tests (Breusch T. , 1978; Godfrey, 1978) with alternating order of correlations. Despite assessing all of the above stats, the final decision criteria for best AR model was minimal BIC.

### Economic variables transformations and lags

The economic variables presented in section 0 were converted into their stationary transformation(s). Transformation options were either annual growth or annual growth rate, for each variable transformations were selected according to industry standards and economic theory. Each transformed variable was considered as an explanatory variable with up to two annual lags. As seen in Table 10 below, 60 different distinct variables were considered in the analysis for each model.

Table : Universe of economic variables considered for modeling

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Transformation(s) | Number of considered variables (contemporaneous + 2 lags) | Expected sign |
| rpi | yp | 3 | + |
| rpi\_national | yp | 3 | + |
| vac\_rate | yd | 3 | - |
| vac\_rate\_national | yd | 3 | - |
| ump\_rate | yd | 3 | - |
| emp | yd & yp | 6 | + |
| pop | yd & yp | 6 | + |
| gdp | yp | 3 | + |
| hstarts | yd & yp | 6 | + |
| cpi\_rent | yp | 3 | + |
| rmort5y | yd | 3 | - |
| rprime | yd | 3 | - |
| rmort\_spread | yd | 3 | - |
| rgt5y | yd | 3 | - |
| rgt10y | yd | 3 | - |
| stockp | yp | 3 | +/- |
| cpwti | yp | 3 | +/- |
| Total |  | 60 |  |

When needed, additional transformation was explored for specific models, such as rolling averages. In addition, interactions between region dummies and economic variables were used to capture cases were a variable has a different effect between regions.

### Candidate economic variables selection

In order to choose the best candidate economic variables efficiently and accurately the following method was used for every model:

1. Looped through all 60 considered variables and estimate them as an explanatory variable in a base model with AR terms and fixed effects (for Units).
2. Coefficient and p-value were obtained for each of the 60 variables.
3. Only variables with p-value < 0.1 were kept.
4. Only variables with coefficient sign conforming to economic intuition were kept.
5. Variables were ranked by significance level.
6. For each variable only the lag with the lowest p-value was kept.
7. For some variable with relatively large p-value were kept (some regional variables in Units model). In terms of some regional variables, they are the “Fixed-effects” part of the panel model econometric framework. In this kind of estimations each panel member (region) has a dummy variable (or equivalently a factor variable for all members together) capturing the members effect not captured in other variables. In the cases in which regions don’t have significant difference those variables will not be significant. However, because this is a basic part of the framework, those variables are always kept.

The above method enabled the selection of the variables most likely to have an impact on the modeled variable while controlling for serial correlation. The method also made it easy to sort out variables with impacts that do not make economic sense. By the end of this stage, each model has a shortlist of candidate economic variables.

### Dynamic predictions

Predictions for each model for either in-sample analysis, out-of-sample analysis or model implementation were done dynamically. The reason for the dynamic process is that most models required AR terms, which need to be calculated from the lagged prediction of the modeled variable.

The dynamic prediction method followed the below steps for each year in either in-sample or out-of-sample period, the code looped through all the years from first to last:

1. Estimate model on training period.
2. Start loop over predicted years:
   1. Select year for prediction.
   2. Generate predicted values (Units in levels Price in growth rates) just for this year.
   3. Un-transform result for Price.
   4. Generate AR terms using lagged values from “b”.
   5. Loop back to “a” until all years have been generated.
3. Un-deflate predicted results for Price

### Model selection

The model selection stage used the candidate shortlist obtained in the previous stage. A number of candidate models were specified for each segment by iteratively mixing variables from the candidate list, and adding regional dummy interactions if needed. Each candidate model was tested using standard panel time-series residuals tests to make sure OLS assumptions are being kept. Selection also used fit metrics of both in-sample and dynamically created out-of-sample predictions. Model selection process followed the bellow steps:

1. Estimated a number of candidate specifications all with significant variables with economically sound signs.
2. Test models’ residuals for:
   1. Unit roots.
   2. Serial correlation.
   3. Cross-sectional dependence.
   4. Homoscedasticity
3. Calculated in-sample predicted values for the entire development data (2004-2019)
4. Calculated out-of-sample predicted values for the out-of-sample period (2015-2019) by estimating the model on a training period (2004-2014)
5. Selected provisional “winner” model using fit metrics for both in-sample and out-of-sample.
6. Plotted provisional “winner” model residuals against candidate economic variables to assess which variables are good candidates to add to the model.
7. Checked residuals for break points to find policy shifts or extreme outliers.
8. Iterated back to “1” using results from “6” until a final “winner” model was selected based on fit metrics and parsimony.

## Units Models

### Serial correlation results – basic model

The table below presents the results of the serial correlation analysis on which AR terms were included in the Units’ three products models.

Table : Units AR models by product

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rental New** | | | | | | | | | | | |
| **Model** | **AR Terms** | | | | **Fixed-effect** | **N** | **BIC** | **Breusch-Godfrey Test (order = )** | | | |
| **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **1** | **2** | **3** | **4** |
| AR(4) | 0.581\*\*\* | 0.203 | -0.189 | -0.163 | Yes | 60 | 992.926 | 0.003 | 0.001 | 0.003 | 0.006 |
| AR(3) | 0.611\*\*\* | 0.219 | -0.321\*\*\* |  | Yes | 60 | 990.012 | 0.001 | 0.001 | 0.002 | 0.006 |
| AR(2) | 0.608\*\*\* | 0.030 |  |  | Yes | 60 | 991.033 | 0.011 | 0.002 | 0.006 | 0.005 |
| AR(1) | 0.621\*\*\* |  |  |  | Yes | 60 | 986.986 | 0.101 | 0.193 | 0.135 | 0.038 |
| **Rental Existing** | | | | | | | | | | | |
| **Model** | **AR Terms** | | | | **Fixed-effect** | **N** | **BIC** | **Breusch-Godfrey Test (order = )** | | | |
| **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **1** | **2** | **3** | **4** |
| AR(4) | 0.249 | -0.063 | 0.144 | 0.016 | Yes | 60 | 1061.165 | 0.953 | 0.834 | 0.889 | 0.929 |
| AR(3) | 0.251 | -0.065 | 0.146 |  | Yes | 60 | 1057.087 | 0.940 | 0.882 | 0.922 | 0.974 |
| AR(2) | 0.243 | -0.041 |  |  | Yes | 60 | 1054.291 | 0.900 | 0.990 | 0.847 | 0.936 |
| AR(1) | 0.236 |  |  |  | Yes | 60 | 1050.300 | 0.932 | 0.953 | 0.898 | 0.964 |
| **Refinance** | | | | | | | | | | | |
| **Model** | **AR Terms** | | | | **Fixed-effect** | **N** | **BIC** | **Breusch-Godfrey Test (order = )** | | | |
| **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **1** | **2** | **3** | **4** |
| AR(4) | 0.221 | 0.197\*\* | -0.445\*\* | 0.184 | Yes | 60 | 1190.705 | 0.060 | 0.002 | 0.006 | 0.011 |
| AR(3) | 0.130 | 0.232\* | -0.390\*\* |  | Yes | 60 | 1189.061 | 0.841 | 0.091 | 0.081 | 0.148 |
| AR(2) | 0.063 | 0.123 |  |  | Yes | 60 | 1197.677 | 0.609 | 0.003 | 0.000 | 0.000 |
| AR(1) | 0.097 |  |  |  | Yes | 60 | 1194.822 | 0.243 | 0.012 | 0.001 | 0.003 |

Notes: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1

Above results indicate that Rental New model need only the first AR(1) term, Rental Existing needs none (AR(1) term is insignificant) and Refinance requires 3 terms.

### Final models

Using a base model containing regional fixed-effects and the above AR terms candidate economic variables are chosen as depicted in section 8.1.3. Then the process presented in sections 8.1.4 and 8.1.5 is followed to create several candidate models of which the model with the best fit is chosen for each product.

Some details of the model selection process with statistical measures are added in the following bullet points.

* Rental Existing:
  + Model 1 and 2 has relatively the value higher values compared to the other three models
  + Model 2 has the smallest AIC values
  + Model 1, 3, 4, and 5 have non-significant variables that’s why they were not chosen. They had a better fit with some correlation which was insignificant so those variables were removed and the specification changed, eventually ending at Model 2.
* Rental New:
  + Model 2 and 3 has relatively the value higher values compared to the other three models
  + Model 2 and 3 have relatively small AIC
  + Model 1 and 2 have non-significant variables that’s why they were not chosen.
  + The final model is Model 3.
* Refinance:
  + Model 1, 2, and 3 have non-significant variables. The final model is Model 4.

Table : Final Units models

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Rental New** | **Rental Existing** | **Refinance** |
| Atlantic | 486.519 | -8864.495\*\*\* | -22427.470\*\*\* |
| (0.339) | (0.000) | (0.000) |
| BC | -227.651 | -8112.433\*\*\* | -22313.152\*\*\* |
| (0.657) | (0.000) | (0.000) |
| Ontario | 217.278 | -4529.799\*\*\* | 353.985 |
| (0.679) | (0.000) | (0.818) |
| Prairie | 188.185 | -8368.245\*\*\* | -20376.024\*\*\* |
| (0.716) | (0.000) | (0.000) |
| AR(1) | 0.416\*\*\* |  | 0.330\*\*\* |
| (0.000) |  | (0.001) |
| AR(2) |  |  | 0.126\* |
|  |  | (0.085) |
| AR(3) |  |  | -0.776\*\*\* |
|  |  | (0.000) |
| yd\_pop | 4.247\*\*\* |  | 68.389\*\*\* |
| (0.010) |  | (0.000) |
| yd\_emp \* Quebec | 26.242\*\*\* |  |  |
| (0.001) |  |  |
| yp\_rpi\_national\_1 | 3779.209\*\* |  |  |
| (0.017) |  |  |
| yd\_rprime \* Quebec | -928.745\*\* |  |  |
| (0.014) |  |  |
| yd\_rgt5y \* Quebec |  | -2053.382\*\*\* |  |
|  | (0.000) |  |
| yd\_rprime\_4avg \* Quebec |  | -2018.119\*\*\* |  |
|  | (0.005) |  |
| yd\_rprime\_4avg \* Ontario |  | -2736.674\*\*\* |  |
|  | (0.005) |  |
| yp\_cpwti\_2 \* Quebec |  | 3838.520\*\*\* |  |
|  | (0.002) |  |
| yd\_vac\_rate\_national |  | -488.536\*\* |  |
|  | (0.046) |  |
| yp\_rpi\_national \* Quebec |  | 20610.657\*\*\* |  |
|  | (0.005) |  |
| yd\_rmort5y |  |  | -1810.147\*\* |
|  |  | (0.046) |
| yd\_ump\_rate\_1 |  |  | -802.505\* |
|  |  | (0.097) |
| yd\_ump\_rate\_1 \* Ontario |  |  | -6280.500\*\*\* |
|  |  | (0.000) |
| (Intercept) | 181.927 | 9704.523\*\*\* | 27193.336\*\*\* |
| (0.717) | (0.000) | (0.000) |
| N | 80 | 80 | 65 |
|  | 0.796 | 0.961 | 0.953 |
|  | 0.770 | 0.955 | 0.943 |
| AIC | 1256.784 | 1330.540 | 1223.367 |
| Residuals Unit-Root IPS (2003) test | 0.000 | 0.000 | 0.001 |
| Residuals Unit-Root Hadri (2000) test | 0.276 | 0.655 | 0.422 |
| Residuals AR(1) Breusch-Godfrey Test | 0.548 | 0.659 | 0.387 |
| Residuals cross-sectional dependence Breusch-Pagan (1980) LM test | 0.118 | 0.457 | 0.523 |
| Residuals cross-sectional dependence Pesaran (2004) CD | 0.182 | 0.471 | 0.082 |
| Breusch-Pagan (1979) homoscedasticity test | 0.001 | 0.013 | 0.349 |

Notes: p-values in parentheses; \*\*\* p-value < 0.01; \*\* p-value < 0.05; \* p-value < 0.1

IPS (2003) Test has an H1: stationarity hypothesis, Hadri (2000) test has an H1: at least one series has a unit root hypothesis

All of the above winner models have stationary residuals, no serial correlation and no cross-sectional dependence. Rental New and Rental Existing models are not homoscedastic but reported results have robust standard errors.

Figure : Final Units models residuals

|  |  |
| --- | --- |
| **Rental New** | **Rental Existing** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_New\model_resid_is_regional_Rental_New.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_Existing\model_resid_is_regional_Rental_Existing.png |
| **Refinance** | |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Refinance\model_resid_is_regional_Refinance.png | |

## Price Models

### Serial correlation results – basic model

The table below presents the results of the serial correlation analysis on which AR terms were included in the Price’s three products models.

Table : Price AR models by product

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rental New** | | | | | | | | | | | |
| **Model** | **AR Terms** | | | | **Fixed-effect** | **N** | **BIC** | **Breusch-Godfrey Test (order = )** | | | |
| **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **1** | **2** | **3** | **4** |
| AR(4) | -0.474\*\* | -0.447\*\*\* | -0.215 | -0.296 | No | 55 | -20.500 | 0.475 | 0.237 | 0.363 | 0.497 |
| AR(3) | -0.488\*\* | -0.364\*\*\* | -0.050 |  | No | 55 | -21.428 | 0.945 | 0.121 | 0.167 | 0.212 |
| AR(2) | -0.474\*\* | -0.339\*\*\* |  |  | No | 55 | -25.297 | 0.982 | 0.112 | 0.170 | 0.246 |
| AR(1) | -0.346\* |  |  |  | No | 55 | -22.297 | 0.163 | 0.012 | 0.015 | 0.018 |
| **Rental Existing** | | | | | | | | | | | |
| **Model** | **AR Terms** | | | | **Fixed-effect** | **N** | **BIC** | **Breusch-Godfrey Test (order = )** | | | |
| **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **1** | **2** | **3** | **4** |
| AR(4) | -0.556\*\*\* | -0.443\*\*\* | -0.290\*\* | -0.072 | No | 55 | -6.343 | 0.004 | 0.017 | 0.037 | 0.050 |
| AR(3) | -0.547\*\*\* | -0.408\*\*\* | -0.268\* |  | No | 55 | -10.118 | 0.004 | 0.014 | 0.034 | 0.047 |
| AR(2) | -0.453\*\*\* | -0.332\*\*\* |  |  | No | 55 | -10.573 | 0.001 | 0.004 | 0.009 | 0.005 |
| AR(1) | -0.381\*\*\* |  |  |  | No | 55 | -8.953 | 0.001 | 0.002 | 0.004 | 0.006 |
| **Refinance** | | | | | | | | | | | |
| **Model** | **AR Terms** | | | | **Fixed-effect** | **N** | **BIC** | **Breusch-Godfrey Test (order = )** | | | |
| **AR(1)** | **AR(2)** | **AR(3)** | **AR(4)** | **1** | **2** | **3** | **4** |
| AR(4) | -0.219 | -0.253\*\* | -0.196\* | -0.274\*\*\* | No | 55 | -76.337 | 0.516 | 0.732 | 0.824 | 0.118 |
| AR(3) | -0.190 | -0.181\* | -0.123 |  | No | 55 | -74.492 | 0.486 | 0.784 | 0.912 | 0.050 |
| AR(2) | -0.166 | -0.155 |  |  | No | 55 | -77.349 | 0.808 | 0.971 | 0.811 | 0.046 |
| AR(1) | -0.145 |  |  |  | No | 55 | -79.504 | 0.702 | 0.838 | 0.781 | 0.054 |

Notes: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1

Above results indicate that Rental New model needs 2 terms, Rental Existing needs 2 terms and Refinance requires 1 term.

### Final Model

Using a base model containing regional fixed-effects and the above AR terms candidate economic variables are chosen as depicted in section 8.1.3. Then the process presented in sections 8.1.4 and 8.1.5 is followed to create a number of candidate models of which the model with best fit is chosen for each product.

Some details of the model selection process with statistical measures are added in the following bullet points.

* Rental Existing:
  + Model 2, 3, 4, and 5 have non-significant variables that’s why they were not chosen. They had a better fit with some correlation which was insignificant so those variables were removed and the specification changed, eventually ending at Model 1.
* Rental New:
  + Model 3 and 4 has relatively the higher value and smaller AIC values compared to the other three models
  + But model 4 has non-significant variables and then the final model is Model 3.
* Refinance:
  + Model 1, 2, and 3 have non-significant variables. The final model is Model 4.

Table : Final Price models

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Rental New** | **Rental Existing** | **Refinance** |
| AR(1) | -0.518\*\*\* | -0.466\*\*\* | -0.516\*\*\* |
| (0.000) | (0.000) | (0.000) |
| AR(2) | -0.385\*\*\* | -0.240\*\*\* |  |
| (0.000) | (0.006) |  |
| yd\_vac\_rate\_national | -0.126\*\*\* |  |  |
| (0.002) |  |  |
| yp\_rpi | 0.595\* |  |  |
| (0.065) |  |  |
| yd\_vac\_rate\_2 |  | -0.070\*\* | -0.062\*\*\* |
|  | (0.049) | (0.000) |
| yp\_stockp |  | -0.616\*\*\* |  |
|  | (0.002) |  |
| yd\_rmort\_spread\_2 |  |  | -0.059\*\*\* |
|  |  | (0.001) |
| yp\_rpi\_national\_1 |  |  | 0.842\*\*\* |
|  |  | (0.005) |
| (Intercept) | 0.078\*\*\* | 0.164\*\*\* | 0.037\*\* |
| (0.000) | (0.000) | (0.043) |
| N | 80 | 80 | 70 |
|  | 0.378 | 0.368 | 0.383 |
|  | 0.345 | 0.335 | 0.345 |
| AIC | -73.837 | -50.033 | -116.971 |
| Residuals Unit-Root IPS (2003) test | 0.000 | 0.000 | 0.000 |
| Residuals Unit-Root Hadri (2000) test | 0.295 | 0.150 | 0.327 |
| Residuals AR(1) Breusch-Godfrey Test | 0.296 | 0.318 | 0.416 |
| Residuals cross-sectional dependence Breusch-Pagan (1980) LM test | 0.865 | 0.233 | 0.280 |
| Residuals cross-sectional dependence Pesaran (2004) CD | 0.293 | 0.439 | 0.671 |
| Breusch-Pagan (1979) homoscedasticity test | 0.279 | 0.477 | 0.125 |

Notes: p-values in parentheses; \*\*\* p-value < 0.01; \*\* p-value < 0.05; \* p-value < 0.1

IPS (2003) Test has an H1: stationarity hypothesis, Hadri (2000) test has an H1: at least one series has a unit root hypothesis

All of the above winner models have stationary residuals, no serial correlation, no cross-sectional dependence and are homoscedastic (reported results have robust standard errors). Below is a chart of model residuals by region.

Figure : Final Price models residuals

|  |  |
| --- | --- |
| **Rental New** | **Rental Existing** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_New\model_resid_is_regional_Rental_New.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_Existing\model_resid_is_regional_Rental_Existing.png |
| **Refinance** | |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Refinance\model_resid_is_regional_Refinance.png | |

# INITIAL MODEL TESTING

This section contain models fit results for both in-sample and out-of-sample. For each of the six models both regional and nationally aggregated results are presented. Definitions of below fit metrics are:

* – pseudo of actual values vs. dynamically generated predicted values
* RMSE – RMSE of actual values vs. dynamically generated predicted values
* Sum of residuals – Sum of actual values minus dynamically generated predicted values

In addition to the fit metrics, fir charts are presented both regionally and nationally.

## Units Models

### Final Units models fit stats

Table : Final Units models fit stats

|  |  |  |  |
| --- | --- | --- | --- |
| **Stat** | **Rental New** | **Rental Existing** | **Refinance** |
| In-sample | 0.805 | 0.966 | 0.936 |
| Out-of-sample 5yr | 0.440 | 0.901 | 0.968 |
| Out-of-sample 5yr Ontario | 0.943 | 0.543 | 0.136 |
| Out-of-sample 5yr Quebec | 0.318 | 0.366 | 0.731 |
| In-sample RMSE | 538.487 | 803.937 | 2787.652 |
| Out-of-sample RMSE 5yr | 852.728 | 1278.019 | 2379.885 |
| Out-of-sample RMSE 5yr Ontario | 337.875 | 1927.639 | 3108.049 |
| Out-of-sample RMSE 5yr Quebec | 1369.622 | 1856.817 | 2750.244 |
| Out-of-sample Sum of residuals 5yr | 148.792 | -9289.368 | 18882.286 |

### Final Units models fit plots

Figure : Final Units - Rental New model fit plots

|  |  |
| --- | --- |
| **In-sample National** | **In-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_New\is_national_Rental_New.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_New\is_regional_Rental_New.png |
| **Out-of-sample National** | **Out-of-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_New\oos_national_Rental_New.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_New\oos_regional_Rental_New.png |

Figure : Final Units - Rental Existing model fit plots

|  |  |
| --- | --- |
| **In-sample National** | **In-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_Existing\is_national_Rental_Existing.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_Existing\is_regional_Rental_Existing.png |
| **Out-of-sample National** | **Out-of-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_Existing\oos_national_Rental_Existing.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Rental_Existing\oos_regional_Rental_Existing.png |

Figure : Final Units - Refinance model fit plots

|  |  |
| --- | --- |
| **In-sample National** | **In-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Refinance\is_national_Refinance.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Refinance\is_regional_Refinance.png |
| **Out-of-sample National** | **Out-of-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\units\Refinance\oos_national_Refinance.png | Z:\zrubin\volume_projection\results\charts\model_dev\units\Refinance\oos_regional_Refinance.png |

## Price Models

### Final Price models fit stats

Table : Final Price models fit stats

|  |  |  |  |
| --- | --- | --- | --- |
| **Stat** | **Rental New** | **Rental Existing** | **Refinance** |
| In-sample | 0.870 | 0.737 | 0.953 |
| Out-of-sample 5yr | 0.851 | 0.907 | 0.912 |
| Out-of-sample 5yr Ontario | 0.893 | 0.873 | 0.974 |
| Out-of-sample 5yr Quebec | 0.667 | 0.874 | 0.762 |
| In-sample RMSE | 25.450 | 33.118 | 11.071 |
| Out-of-sample RMSE 5yr | 29.005 | 26.742 | 17.142 |
| Out-of-sample RMSE 5yr Ontario | 29.764 | 16.973 | 17.071 |
| Out-of-sample RMSE 5yr Quebec | 15.049 | 12.542 | 8.749 |
| Out-of-sample Sum of residuals 5yr | 114.388 | 192.014 | 215.882 |

### Final Price models fit plots

Figure : Final Price - Rental New model fit plots

|  |  |
| --- | --- |
| **In-sample National** | **In-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_New\is_national_Rental_New.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_New\is_regional_Rental_New.png |
| **Out-of-sample National** | **Out-of-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_New\oos_national_Rental_New.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_New\oos_regional_Rental_New.png |

Figure : Final Price - Rental Existing model fit plots

|  |  |
| --- | --- |
| **In-sample National** | **In-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_Existing\is_national_Rental_Existing.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_Existing\is_regional_Rental_Existing.png |
| **Out-of-sample National** | **Out-of-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_Existing\oos_national_Rental_Existing.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Rental_Existing\oos_regional_Rental_Existing.png |

Figure : Final Price - Refinance model fit plots

|  |  |
| --- | --- |
| **In-sample National** | **In-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Refinance\is_national_Refinance.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Refinance\is_regional_Refinance.png |
| **Out-of-sample National** | **Out-of-sample Regional** |
| Z:\zrubin\volume_projection\results\charts\model_dev\price\Refinance\oos_national_Refinance.png | Z:\zrubin\volume_projection\results\charts\model_dev\price\Refinance\oos_regional_Refinance.png |

# MODEL PERFORMANCE TESTING

This section provides the model performance testing results for the Units and Price model, separately. The MORA team compares the model forecast coming from the old business model that has been used for the multi-unit analytics team with the forecast from the multi-unit volume projection model over a one- to five-year projection by using the out-of-sample. We calculated the divergence between these two models and the actual data and compared the model lifting of the multi-unit volume projection model and the old business model. As evidenced by the model accuracy and stability shown below, the multi-unit volume projection model outperforms the old business model.

## Units Model Performance

For the Units model, it tends to show significant model lifting from 1 to 5 years projection, evidenced by the model accuracy improvement shown in Table 19. Most results are positive values, inferring obvious model lift in comparison to the old business model. In Table 19, few negative values are mainly driven by the unstable forecast of the old business model. The forecast of the old business model drives its divergence from actual data randomly distributed. For example, the forecast divergence for the year 2019 is ranging from 7.9% to 27.8% in green in Table 17. However, in terms of the model stability, the Units model performs more stably than the old business model. For example, the forecast of the Units model drives its divergence from actual data stably ranging from 11.0% to 12.0% in red in Table 18.

Table : The divergence of the old business model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Old Business Model Divergence | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 2014 | 18.1% | 15.9% | 19.0% | 11.4% | 26.4% | N/A |
| 2015 |  | 16.7% | 20.6% | 12.7% | 27.8% | 46.6% |
| 2016 |  |  | 13.5% | 3.8% | 21.5% | 42.1% |
| 2017 |  |  |  | 11.4% | 9.7% | 36.9% |
| 2018 |  |  |  |  | 7.9% | 36.1% |
| 2019 |  |  |  |  |  | 29.8% |

Table : The divergence of the multi-unit volume projection model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Volume Projection Model Divergence | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 2014 | 7.0% | 5.6% | 5.8% | 2.1% | 12.0% | N/A |
| 2015 |  | 6.4% | 6.3% | 1.2% | 11.5% | 39.7% |
| 2016 |  |  | 5.8% | 0.9% | 11.9% | 40.0% |
| 2017 |  |  |  | 0.6% | 11.0% | 42.0% |
| 2018 |  |  |  |  | 11.0% | 41.8% |
| 2019 |  |  |  |  |  | 40.9% |

Table : The units model lifting of the multi-unit volume projection model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model Performance Lifting | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 2014 | 11.1% | 10.3% | 13.2% | 9.3% | 14.4% | N/A |
| 2015 |  | 10.2% | 14.3% | 11.5% | 16.3% | 6.9% |
| 2016 |  |  | 7.8% | 2.9% | 9.6% | 2.1% |
| 2017 |  |  |  | 10.8% | -1.4% | -5.1% |
| 2018 |  |  |  |  | -3.0% | -5.8% |
| 2019 |  |  |  |  |  | -11.2% |

## Price Model Performance

Similarly, the Price model also tends to demonstrate significant model lifting from 1 to 5 years projection, evidenced by the model accuracy improvement shown in Table 22. Most results are positive values, inferring obvious model lift in comparison to the old business model. In Table 22, one negative value is mainly due to the unstable forecast of the old business model. The forecast of the old business model drives its divergence from actual data randomly distributed. For example, the forecast divergence for the year 2018 is ranging from 13.3% to 56.3% in blue in Table 20. However, in terms of the model stability, the Units model forecasts more stable results than the old business model. For example, the forecast of the Units model drives its divergence from actual data stably ranging from 22.7% to 30.5% in purple in Table 21.

Table : The divergence of the old business model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Old Business Model Divergence | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 2014 | 41.6% | 44.2% | 52.3% | 53.8% | 64.9% | N/A |
| 2015 |  | 46.8% | 55.1% | 56.3% | 66.8% | 77.2% |
| 2016 |  |  | 49.5% | 49.4% | 62.7% | 74.5% |
| 2017 |  |  |  | 13.3% | 35.8% | 58.5% |
| 2018 |  |  |  |  | 27.4% | 53.2% |
| 2019 |  |  |  |  |  | 41.5% |

Table 1: The divergence of the multi-unit volume projection model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Volume Projection Model Divergence | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 2014 | 26.5% | 25.8% | 18.8% | 22.7% | 10.0% | N/A |
| 2015 |  | 26.6% | 21.3% | 26.2% | 13.5% | 24.7% |
| 2016 |  |  | 22.0% | 26.5% | 12.1% | 27.9% |
| 2017 |  |  |  | 30.5% | 16.6% | 26.0% |
| 2018 |  |  |  |  | 19.6% | 23.5% |
| 2019 |  |  |  |  |  | 23.9% |

Table 2: The price model lifting of the multi-unit volume projection model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model Performance Lifting | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
| 2014 | 15.1% | 18.3% | 33.6% | 31.2% | 54.9% | N/A |
| 2015 |  | 20.2% | 33.8% | 30.2% | 53.3% | 52.5% |
| 2016 |  |  | 27.5% | 22.9% | 50.7% | 46.6% |
| 2017 |  |  |  | -17.3% | 19.2% | 32.5% |
| 2018 |  |  |  |  | 7.9% | 29.7% |
| 2019 |  |  |  |  |  | 17.6% |

## Model performance and policy change in the year 2020

MORA team identified the underperformance in the pandemic year 2020 in comparison to the actual data and performs a further analysis from a perspective of data and multi-unit policy change in the following two sections.

### Model Performance in the year 2020

Table 23 shows the aggregated actual multi-unit loan data from 2016 to 2020 in non-refinance and refinance segments. Due to the multi-unit policy change in the pandemic year 2020, the units and the insured loan amounts of refinancing products in the year 2020 (refinance units and insured loan amount are 141.7k and $21.3 B in red in Table 23, respectively) dramatically increased 40% and 54% in comparison to the year 2019, respectively. Overall speaking, years before 2020 tend to perform a stable annual increase or decrease within 20%; however, the total units and total insured loan amounts in the year 2020 increased 35% and 48% in comparison to 2019, separately, mainly due to the impact from refinancing products. As the multi-unit projection model can’t capture the unexpected policy change in refinance products, a relatively large divergence from the multi-unit volume projection model is observed for the forecast in 2020.

Table 3: The aggregated actual multi-unit loan data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Multi-Unit Loans** | **2016** | **2017** | **2018** | **2019** | **2020** |
| **Non-Refinance** |  |  |  |  |  |
| Loans | 1,207 | 1,262 | 1,076 | 1,126 | 1,420 |
| Units | 26,525 | 33,539 | 30,838 | 34,867 | 41,533 |
| Insured Loan Amount | $3,089,551,099 | $4,440,195,637 | $4,655,624,390 | $5,796,142,458 | $7,767,258,674 |
| **Refinance** |  |  |  |  |  |
| Loans | 1,488 | 1,689 | 1,723 | 2,062 | 3,153 |
| Units | 79,015 | 88,602 | 81,409 | 100,973 | 141,742 |
| Insured Loan Amount | $7,322,334,467 | $9,331,223,497 | $9,933,079,713 | $13,835,589,898 | $21,302,129,706 |
| **Summary** |  |  |  |  |  |
| Loans | 2,695 | 2,951 | 2,799 | 3,188 | 4,573 |
| Units | 105,540 | 122,141 | 112,247 | 135,840 | 183,275 |
| Total Insured Loan Amount | $10,411,885,566 | $13,771,419,134 | $14,588,704,104 | $19,631,732,357 | $29,069,388,380 |

For example, when comparing to the actual data in 2020, the 1-year forecast results in Table 24 using 2004 to 2019 development window shows a dramatically large divergence in refinance segment, evidenced by the divergence ratio at 45.8% and 29.6% for units and insured loan amounts, respectively. Therefore, the large divergence in refinance segment in the year 2020 has a relatively large impact on the model performance in 2020.

Table 4: The aggregated predicted multi-unit loan data

|  |  |  |
| --- | --- | --- |
| **2004-2019 Development Window** | | |
| **1-Yr forecast** | **Units** | **Price** |
| Non-refinance | 31,342 | $7,115,944,751 |
| Refinance | 76,892 | $15,002,955,441 |
| Summary | 108,235 | $22,118,900,192 |
| **Divergence** | **Units** | **Price** |
| Non-refinance | 24.5% | 8.4% |
| Refinance | 45.8% | 29.6% |
| Summary | 40.9% | 23.9% |

### Multi-Unit policy change in the year 2020

Besides a quantitative analysis for the model performance in the year 2020, MORA also takes efforts to investigate the Multi-Unit policy changes in the pandemic year, aiming to provide more insights in this section.

In the pandemic year 2020, the multi-unit underwriting team implemented some changes to underwriting operations, which were aimed at providing for more efficient operational decisions and included a focus on risk-based underwriting. And, it mainly applied to the refinance products of the Multi-Unit MLI (Mortgage Loan Insurance) business.

In particular, the policy-related changes to restriction on the use of funds related to the Multi-Unit MLI refinance product were implemented in Q2 2020. Although it remains unclear if they were intended to be temporary or not, they do remain in effect until now. They appear to have had little to no impact on volumes as most borrowers seem able to demonstrate that funds will be re-invested in Multi-Unit rental housing (which was the new requirement imposed). In fact, despite the policy change in Q2 of 2020, volumes for 2020 were the highest ever recorded. So it's very unclear if the policy drove any outcomes in terms of volumes and we suspect economic factors also played a big role in the increased volumes and we have no real way to know what business was never initiated.

In summary, MORA finds the underestimated forecast in the year 2020 is mainly coming from the dramatically increasing volumes in refinancing loans and pandemic economic environment. Therefore, MORA will keep evaluating the impact of the policy change and monitoring the model performance 2020 onwards.

# SENSITIVITY ANALYSIS

## Sensitivity Methodologies

Sensitivity analysis is an important component in model developing exercise. By underlining risk factors that have the greatest impact on the model outcome, this analysis provides practical information for model builders and users. McCarthy et al. (1994) states that sensitivity analysis could highlight model parameter that ought to be the most accurately measured in order to maximise the precision of the model, give a general indication of the reliability of the model predictions, and highlight parameters and interaction that have the largest influence on the population to help determine effective management strategies. One of the commonly logical approaches of sensitivity analysis is to take consideration of the parameters, which are the most significant factor on the model results. Mathematical models are widely applied in various industries, such as Engineering, Physics, Environmental Science, Social Science, Economics and Finance. Hamby (1995) reviews the techniques of sensitivity analysis applied in environmental models; Cross and Beissinger (2001) discuss and compared sensitivity analysis methods for social science models. Hosman et al. (2010) introduce the sensitivity analysis that used on quantifying omitted variable bias. The technique of sensitivity analysis is applied to mathematical models of Finance, is to shock the risk factors or the ratio of dependent and independent (parameter estimates) variables one at a time based on supervisory experience, Roy et al. (2018) discuss the approach in great detail.

## Sensitivity Analysis Results

Sensitivity analysis for the volume projection models that measure the influences on model performance by shifting the parameter estimates by a certain fixed amount.

### The sensitivity for Units model by segments:

Figure 3: Population Sensitivity

Figure 4: RPI Sensitivity

Figure 5: Employment Rate Sensitivity

Figure 16: Unemployment Rate Sensitivity

After shocking a certain fixed percentage, the sensitivity results show that the direction of variables change impact is consistent with the model’s coefficient and intuitive.

### The sensitivity for Price model by segments:

Figure 7: RPI Sensitivity

Figure 8: Stock Price Sensitivity

Figure 8: Vacancy Rate Sensitivity

Similarly, the sensitivity results for Price model, the shocked macro-economic variables are in line with the Price dollar amounts.

# MODEL ASSUMPTIONS AND LIMITATIONS

## Key Assumptions

The key methodology assumptions in developing the model:

1. Units and Price data of the multi-unit business is an aggregated time series per segment (region and product) and can be modelled.
2. Changes in product features, policies, or account management strategies do not materially impact the model performance
3. Model segments, independent variables, and coefficient estimates are stable over time.
4. Data assumptions and treatments applied do not materially impact model performance.
5. Statistical assumptions of panel time-series regression models are satisfied.
6. Observations can be treated as independent records in the modelling dataset.

## Key Limitations

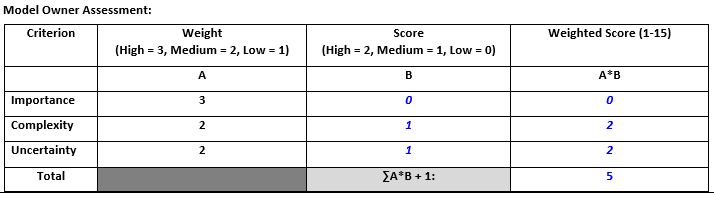
The main limitations of the models are summarized below:

1. **Data Support:** The model will be implemented to automate the volume projection for the Units and the Price of the Mollea DPA production dataset. The continued operation of the model is dependent on the continued availability of the data sources used to build the model.
2. **Model Timeframe:** The estimation sample timeframe is, due to limitations in the data quality, limited to loans information after 2004.Furthermore, from a modelling data perspective and the data limitation of refinance segment, the final sample for modelling development is from 2004 to 2019.
3. **Missing Explanatory Variables**: One of the key limitations of the model selection procedure is the list of potential explanatory variables. While developed under a strong framework (expert survey, literature review, etc.) additional variables need to be considered. Furthermore, key explanatory economic variables identified by the expert survey such as the provincial vacancy rate or average rent price are not included in the model due to limitations in the availability of simulated projections of the variables.
4. **Model Development:** The model development lacks rigour; cross-validation is done after the model development, only a single model selection procedure is used, search for potential non-linearity is limited to a few key variables and are not based on the optimal procedure, etc.

# MODEL RISK RATING

According to the Model Risk Policy Guideline for guidance, in this section, we propose the following comments to determine the risk rating score and corresponding rationale.

* **Low-Importance:** The model output of the multi-unit volume projection model will only affect the multi-unit line of business for new business volumes.
* **Medium-Complexity:** The basic time series AR methodology and unit roots were applied to develop the model.
* **Medium-Uncertainty:** The model output shows an intermediate degree of sensitivity due to model assumption changes.



In summary, the proposed Risk Rating is Class-C Material and the weighted score is five.

# REFERENCES

Breusch, T. (1978). Testing for Autocorrelation in Dynamic Linear Models. *Australian Economic Papers, 17*, 334-355.

Breusch, T., & Pagan, A. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica, 47*, 1287–1294.

Breusch, T., & Pagan, A. (1980). The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies, 47*, 239-253.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance, 20*(2), 249-272.

Godfrey, L. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables'. *Econometrica, 46*, 1293-1301.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *The Econometrics Journal, 3*(2), 148-161.

Im, K., Pesaran, M., & Shin, Y. (2003). Testing for unit roots in heterogenous panels. *Journal of Econometrics, 115*(1), 53-74.

Maddala, G., & Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics, 61*, 631-652.

Perron, P. (1988). Trends and Random Walks in Macroeconomic Time Series. *Journal of Economic Dynamics and Control, 12*, 297-332.

Pesaran, M. (2004). *General Diagnostic Tests for Cross Section Dependence in Panels.* CESifo Working Paper Series, 1229.

Said, S., & Dickey, D. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika, 71*, 599-607.