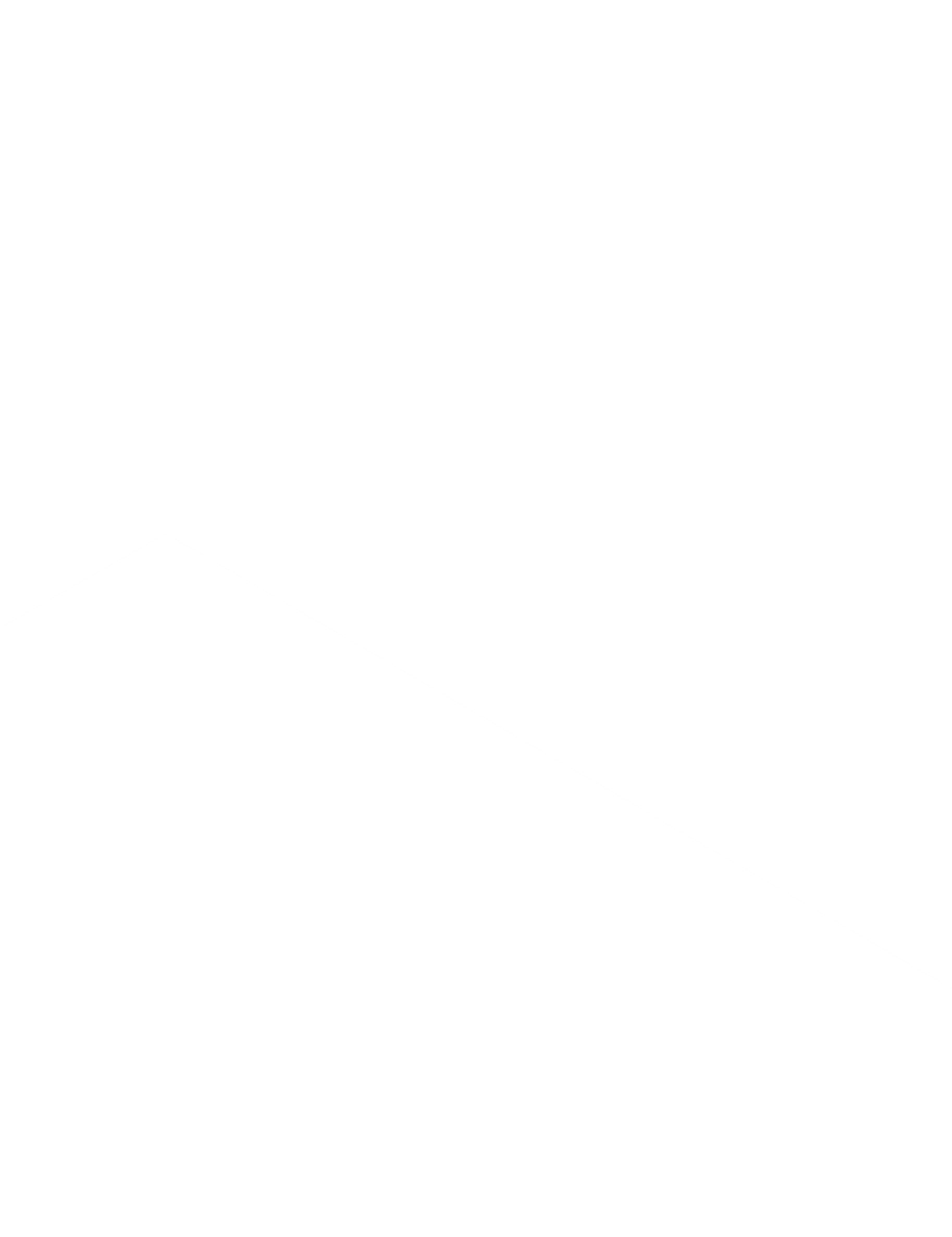
Model Risk Guideline

Effective: June 2020

Version 2.0

Protected – Operations / Proprietary



User Manual for Multi-Unit Volume Projection Model

Multi-Unit & Assisted Housing Oversight and Risk Analytics

Sector of the Chief Risk Officer

Version Number: 1.1

Version Date: July, 2021

# TABLE OF CONTENTS

[1. TABLE OF CONTENTS 2](#_Toc84974371)

[1.1 Summary 3](#_Toc84974372)

[2. DESKTOP RStudio MODEL USER MANUAL 4](#_Toc84974373)

[2.1 Folder Creation 4](#_Toc84974374)

[2.2 Economic and Loan Data Preparation 4](#_Toc84974375)

[2.2.1 CMHC sources 4](#_Toc84974376)

[2.2.2 DataBuffet data 5](#_Toc84974377)

[2.2.3 Raw loans data (DPA) preparation 6](#_Toc84974378)

[2.3 Model Run 6](#_Toc84974379)

[2.3.1 Codes introduction 6](#_Toc84974380)

[2.3.2 Run “E\_model\_implementation.R” ONLY 7](#_Toc84974381)

[3. AIP RStudio MODEL USER MANUAL 10](#_Toc84974382)

[3.1 Folder Creation 10](#_Toc84974383)

[3.2 Economic and Loan Data Preparation 11](#_Toc84974384)

[3.2.1 CMHC sources 11](#_Toc84974385)

[3.2.2 DataBuffet data 11](#_Toc84974386)

[3.2.3 Raw loans data (DPA) preparation 11](#_Toc84974387)

[3.3 Model Run 11](#_Toc84974388)

[3.3.1 Codes introduction 11](#_Toc84974389)

[3.3.2 Run “E\_model\_implementation.R” ONLY 12](#_Toc84974390)

## Summary

In order to run this model in the AIP platform and desktop machine, the MORA team develop two versions for the multi-unit volume projection model: the AIP version and the Desktop version of RStudio.

For a model user, there are four parts of the multi-unit volume projection model:

* “A\_data\_creation.R” – code that creates the data for all analysis and model implementation. This code is triggered by running one of the other codes and all inputs are automatically derived from the triggering code.
* “B\_data\_analysis\_documentation.R” – generates all the analysis presented above except model estimation and testing.
* “C\_model\_development\_units.R” – executes model development procedures for the Units models.
* “D\_model\_development\_price.R” - executes model development procedures for the Price models.
* “E\_model\_implementation.R” - executes models implementation, generates models stats and forecasts.
* “ ecca\_template.R”- generates certain formats of graphs and tables

Please also note:

* Since Mollea’s DPA (Data Preprocessing Analyzer) dataset is not fully migrated to the AIP Sandbox, we alternatively extract and stage the Mollea DPA production dataset into a temporary dataset: prod\_multi\_full\_19q4\_newLBA.RData in AIP, saving in the “data” folder. Accordingly, this dataset will be updated directly when the DPA production dataset has been updated
* Make sure to have the AIP access via [ServiceNow](https://cmhc.service-now.com/cmhc?id=sc_home) before running this model in AIP platform

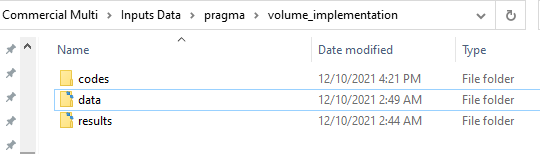
# DESKTOP RStudio MODEL USER MANUAL

This section contains user guide model implementation. This model can be implemented in the desktop version of RStudio, codes are commented throughout and all model implementation inputs are to be entered at the beginning of the code.

## Folder Creation

To run this model, three folders need to be created in a certain production folder as follows;

1. “codes” folder saves save all the codes of this model
2. “data” folder contains all the economic data such as “vacancy\_rates\_pr.xlsx”, “condo\_rpi\_pr.xlsx”, “singles\_rpi\_pr.xlsx” and aggregated loan data such as “data\_aggr.RData”
3. “results” folder saves model parameters and the final output



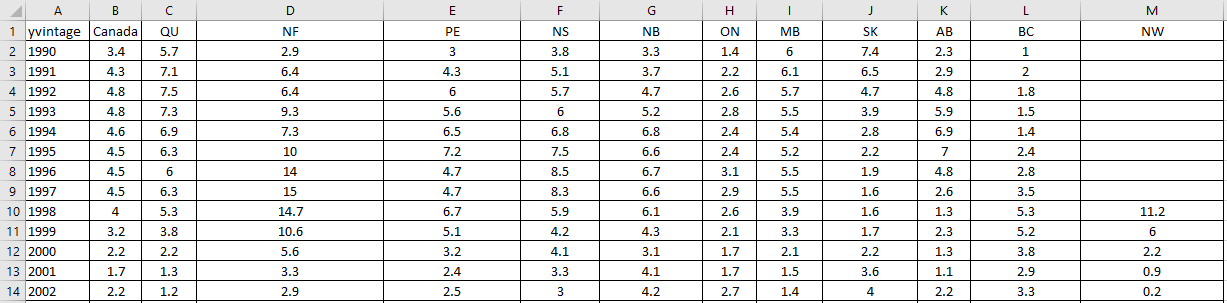
Note: It’s not necessary to create any subfolder under these three folders above since the model itself will automatically create them when running it.

## Economic and Loan Data Preparation

### CMHC sources

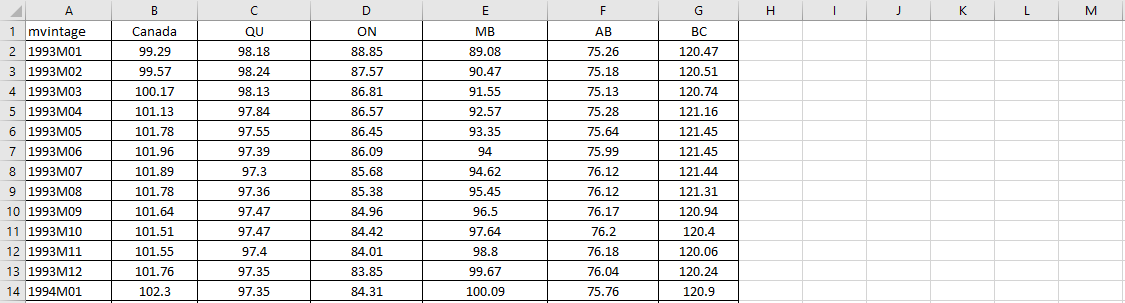
Please make sure to update the following economic raw data to the end of your forecast period. For example, if your forecast window is between 2020 to 2025 (The following section will take 2020-2025 as an example. You can set up your forecast window up to 2030, but you have to update the following economic inputs to 2030), then you have to update all the monthly and annual data to the year 2025, accordingly. For filling those forecast economic inputs, a simple way is to take the same annual/monthly growth rate compared to the previous year/month. CMHC sourced economic data should be formatted as below and saved in a designated “data” folder:

1. Vacancy rates (downloaded manually from CMHC’s Housing Market Information Portal):
   1. File name: “vacancy\_rates\_pr.xlsx”
   2. File structure:

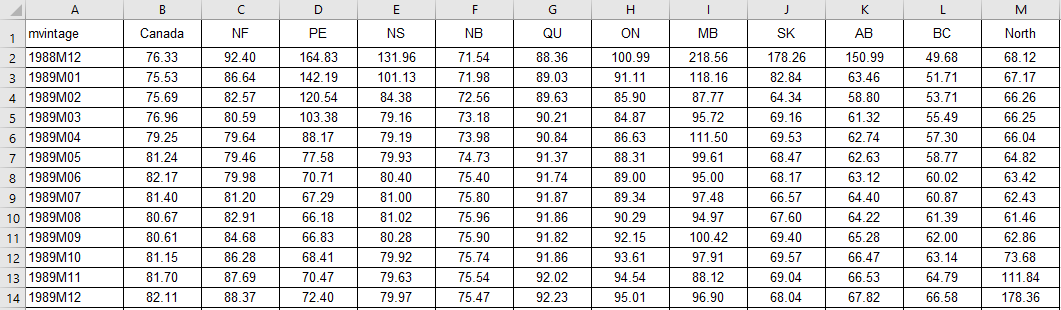


* 1. Update annual data to the end of the forecast window (e.g, update to 2025)

1. Condominium RPI data:
2. File names: “condo\_rpi\_pr.xlsx”
3. File structure:



1. Update monthly data to the end of the forecast window (e.g, update to 2025M12)
2. Single-family RPI data:
3. File names: “singles\_rpi\_pr.xlsx”
4. File structure:



1. Update monthly data to the end of the forecast window (e.g, update to 2025M12)

### DataBuffet data

Data from DataBuffet should be downloaded following the below steps and saved in the designated “data” folder (DataBuffet Data was already updated to the year 2050 and it seems no need to update for now!):

1. Login using username and password at: <https://www.economy.com/databuffet>
2. Create a new data basket from the top menu.
3. Paste the below list of Mnemonics into the basket interface:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| FLBRQ.ICAN\_AB |  | FPOPQ.ICAN\_NL |  | FCPIRENTQ.ICAN\_ON |
| FLBRQ.ICAN\_BC |  | FPOPQ.ICAN\_NS |  | FCPIRENTQ.ICAN\_PE |
| FLBRQ.ICAN\_MB |  | FPOPQ.ICAN\_ON |  | FCPIRENTQ.ICAN\_QC |
| FLBRQ.ICAN\_NB |  | FPOPQ.ICAN\_PE |  | FCPIRENTQ.ICAN\_SK |
| FLBRQ.ICAN\_NL |  | FPOPQ.ICAN\_QC |  | FCPIQ.ICAN\_AB |
| FLBRQ.ICAN\_NS |  | FPOPQ.ICAN\_SK |  | FCPIQ.ICAN\_BC |
| FLBRQ.ICAN\_ON |  | FRPRIMEQ.ICAN |  | FCPIQ.ICAN\_MB |
| FLBRQ.ICAN\_PE |  | FCPWTI.IUSA |  | FCPIQ.ICAN\_NB |
| FLBRQ.ICAN\_QC |  | FRGT5YQ.ICAN |  | FCPIQ.ICAN\_NL |
| FLBRQ.ICAN\_SK |  | FRGT10YQ.ICAN |  | FCPIQ.ICAN\_NS |
| FGDP$Q.ICAN\_AB |  | FLBEQ.ICAN\_AB |  | FCPIQ.ICAN\_ON |
| FGDP$Q.ICAN\_BC |  | FLBEQ.ICAN\_BC |  | FCPIQ.ICAN\_PE |
| FGDP$Q.ICAN\_MB |  | FLBEQ.ICAN\_MB |  | FCPIQ.ICAN\_QC |
| FGDP$Q.ICAN\_NB |  | FLBEQ.ICAN\_NB |  | FCPIQ.ICAN\_SK |
| FGDP$Q.ICAN\_NL |  | FLBEQ.ICAN\_NL |  | FHSTQ.ICAN\_AB |
| FGDP$Q.ICAN\_NS |  | FLBEQ.ICAN\_NS |  | FHSTQ.ICAN\_BC |
| FGDP$Q.ICAN\_ON |  | FLBEQ.ICAN\_ON |  | FHSTQ.ICAN\_MB |
| FGDP$Q.ICAN\_PE |  | FLBEQ.ICAN\_PE |  | FHSTQ.ICAN\_NB |
| FGDP$Q.ICAN\_QC |  | FLBEQ.ICAN\_QC |  | FHSTQ.ICAN\_NL |
| FGDP$Q.ICAN\_SK |  | FLBEQ.ICAN\_SK |  | FHSTQ.ICAN\_NS |
| FRMORT5YQ.ICAN |  | FCPIRENTQ.ICAN\_AB |  | FHSTQ.ICAN\_ON |
| FSTOCKPQ.ICAN |  | FCPIRENTQ.ICAN\_BC |  | FHSTQ.ICAN\_PE |
| FPOPQ.ICAN\_AB |  | FCPIRENTQ.ICAN\_MB |  | FHSTQ.ICAN\_QC |
| FPOPQ.ICAN\_BC |  | FCPIRENTQ.ICAN\_NB |  | FHSTQ.ICAN\_SK |
| FPOPQ.ICAN\_MB |  | FCPIRENTQ.ICAN\_NL |  |  |
| FPOPQ.ICAN\_NB |  | FCPIRENTQ.ICAN\_NS |  |  |

1. Set the “Frequency” to “Annual”.
2. Set “Date Range” to begin with 1/1/1990.
3. Save the basket.
4. Download the basket.
5. Save the downloaded file in the designated “data” folder.

For more detailed instructions on data retrieval from DataBuffet and using baskets follow this link: <https://www.moodysanalytics.com/landing/2019/data-buffet/module-3>

### Raw loans data (outputs of Data Preprocessing Analyzer)

The MORA team will provide the raw loans data at: //Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/20191231/Production/prod\_multi\_full\_19q4\_newLBA.csv. This dataset is updated regularly (for now, it’s at an annual basis update) and please ask the MORA team for this dataset.

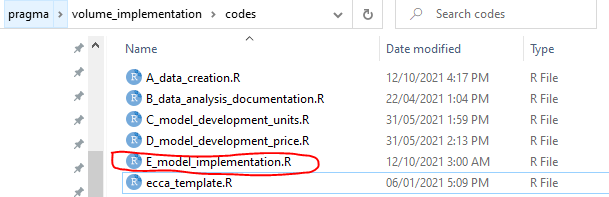
## Model Run

The following steps will implement the model using the provided codes. The below instructions also match data analysis and model development codes and can be used to generate all the results and charts in this document. However, for model users/owners, please only change or modify the inputs in step “E\_model\_implementation.R” ONLY.

### Codes introduction

This model has the below processes, all in R:

* “A\_data\_creation.R” – code that creates the data for all analysis and model implementation. This code is triggered by running one of the other codes and all inputs are automatically derived from the triggering code. **(Please don’t modify!)**
* “B\_data\_analysis\_documentation.R” – generates all the analysis presented above except model estimation and testing. **(Please don’t modify!)**
* “C\_model\_development\_units.R” – executes model development procedures for the Units models. **(Please don’t modify!)**
* “D\_model\_development\_price.R” - executes model development procedures for the Price models. **(Please don’t modify!)**
* “E\_model\_implementation.R” - executes models implementation, generates models stats and forecasts.

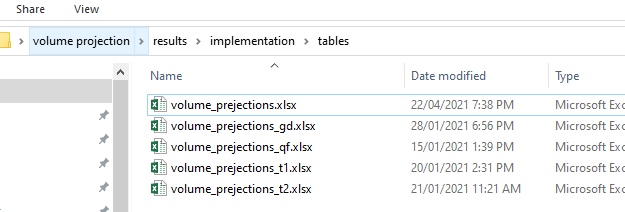


### Run “E\_model\_implementation.R” ONLY

1. Specify the following inputs under the “Script setup / Options” section of the “E\_model\_implementation.R” (Change the input file location in line with your group drive’s name, here below is referred to Mollea’s group drive):

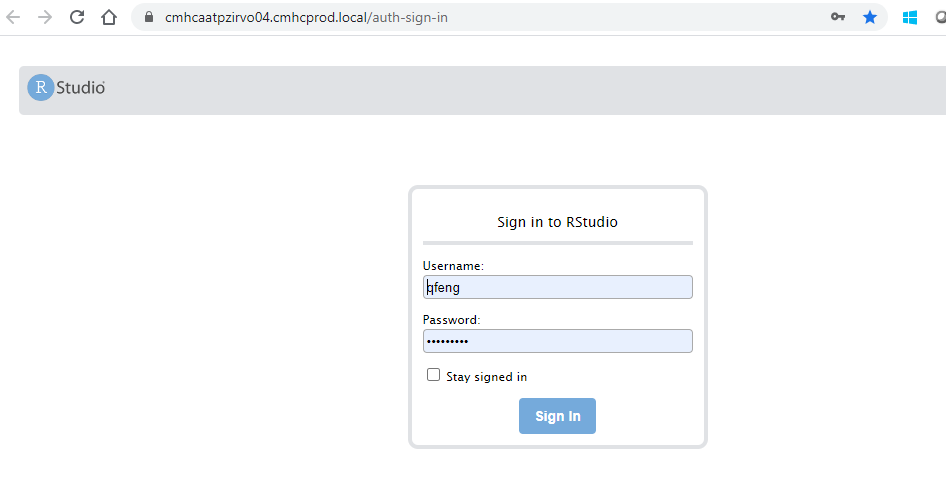
|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Possible input** | **Description** | **Example** |
| dir\_main | Folder location in quotation marks | location of main model folder (all results will be placed in sub-folders here) | dir\_main <- "//Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/volume\_implementation/" |
| dir\_codes | Folder location in quotation marks | Location of model code folder | dir\_codes <- "//Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/volume\_implementation/codes/" |
| dir\_data | Folder location in quotation marks | Location of model data folder | dir\_data <- "//Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/volume\_implementation/data/" |
| dir\_winner | Folder location in quotation marks | Location of winner model for Units and Price | dir\_winner <- "//Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/volume\_implementation/results/model\_dev/" |
| dir\_raw\_data | Folder location in quotation marks | Location of mollea master data (raw loans data, DPA) | dir\_raw\_data <- "//Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/20191231/Production/ |
| raw\_data\_name | File name without extension in quotation marks | Raw csv data file name | raw\_data\_name <- "prod\_multi\_full\_19q4\_newLBA" |
| db\_data\_name | File name without extension in quotation marks | DataBuffet downloaded basket file name | db\_data\_name <- "Basket\_2020-07-16\_14h\_13m" |
| winner\_units\_name | File name without extension in | Units model development results (winner models) | winner\_units\_name <- "winner\_models\_units" (Please don’t change) |
| winner\_price\_name | File name without extension in | Price model development results (winner models) | winner\_price\_name <- "winner\_models\_price" (Please don’t change) |
| new\_data | "yes"/"no" | Create new aggregated data - "yes" will run "A\_data\_creation.R" to create a fresh data file; "no" will load an old version of "data\_aggr.Rdata" from data folder | new\_data <- "yes" |
| new\_raw\_data | "yes"/"no" | Load new raw MOLLEA data - "yes" will load new raw MOLLEA data in "A\_data\_creation.R"; "no" will use saved raw data from data folder | new\_raw\_data <- "no" |
| f\_hist | YYYY | Define first year of historical data | f\_hist <- 2000 |
| l\_hist | YYYY | Define last year of historical data | l\_hist <- 2019 |
| f\_dev\_Rental\_New | YYYY | Define first year of Rental New model development (No need to change) | f\_dev\_Rental\_New <- 2004 |
| f\_dev\_Rental\_Existing | YYYY | Define first year of Rental Existing model development (No need to change) | f\_dev\_Rental\_Existing <- 2004 |
| f\_dev\_Refinance | YYYY | Define first year of Refinance model development (No need to change) | f\_dev\_Refinance <- 2004 |
| l\_dev | YYYY | Define last year of model development | l\_dev <- 2019 |
| oos\_start | YYYY | Define first year of out-of-sample analysis (any year between f\_hist and l\_hist) and it won’t affect the final outputs (No need to change) | oos\_start <- 2015 |
| l\_frcst | YYYY | Define end of forecast | l\_frcst <- 2025 |

1. Run step “E\_model\_implementation.R” in its entirety, the final outputs will be named as volume\_projections.xlsx and it’s saved at [dir\_main results/implementation/tables/](dir_main%20results/implementation/tables/). (In this example above, the final output is saved at: [\\Prdermfs01\erm\Insurance\Models\Commercial Multi\Inputs Data\pragma\volume\_implementation\results\implementation\tables](file:///\\Prdermfs01\erm\Insurance\Models\Commercial%20Multi\Inputs%20Data\pragma\volume_implementation\results\implementation\tables)\volume\_projections.xlsx)



# AIP RStudio MODEL USER MANUAL

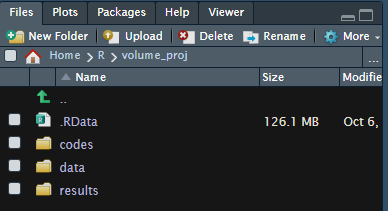
This section contains user guide model implementation in AIP RStudio. Please make sure to have access to AIP platform. Login to the AIP RStudio platform at [https://cmhcaatpzirvo04.cmhcprod.local/auth-sign-in](https://can01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fcmhcaatpzirvo04.cmhcprod.local%2Fauth-sign-in&data=04%7C01%7Cqfeng%40cmhc-schl.gc.ca%7C8a46ff94f3474333429c08d90a7311d0%7C38b7fc89dbe84ed1a78b39dfb6a217a8%7C1%7C0%7C637552311759119218%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=gOhIC43YyaC4gVUx45%2FCbkB5wAfHkym1YYNAYElIeJI%3D&reserved=0) to check by using your AIP credential. (Please refer to the user manual of AIP platform for more details.)

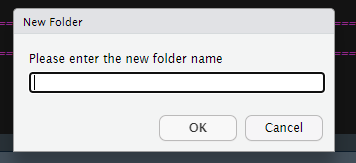


## Folder Creation

In the AIP RStudio interface, please create three folders by clicking “New Folder” button. Then name the new created folder. Three folders need to be created in a certain directory as follows;

1. “codes” folder saves save all the codes of this model
2. “data” folder contains all the economic data such as “vacancy\_rates\_pr.xlsx”, “condo\_rpi\_pr.xlsx”, “singles\_rpi\_pr.xlsx” and aggregated loan data such as “data\_aggr.RData”
3. “results” folder saves model parameters and the final output





## Economic and Loan Data Preparation

### CMHC sources

Same as section 2.2.1.

### DataBuffet data

Same as section 2.2.2.

### Raw loans data (DPA) preparation

In AIP environment, the raw loans data (DPA) is not ready to be used. Then alternatively, the MORA team provides a temporary aggregated dataset- data\_aggr.RData to model user. Please make sure you have this data in your “data” folder before running the step “E\_model\_implementation.R”. Otherwise, this code may crash!!

Also, this dataset is updated regularly (for now, it’s at an annual basis update) and please request the MORA team for this dataset.

## Model Run

### Codes introduction

Same as section 2.3.1.

### Run “E\_model\_implementation.R” ONLY

1. Specify the following inputs under the “Script setup / Options” section of the “E\_model\_implementation.R”

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Possible input** | **Description** | **Example** |
| dir\_main | Folder location in quotation marks | location of main model folder (all results will be placed in sub-folders here) | dir\_main <- "/home/qfeng/R/volume\_proj/" |
| dir\_codes | Folder location in quotation marks | Location of model code folder | dir\_codes <- "/home/qfeng/R/volume\_proj/codes/" |
| dir\_data | Folder location in quotation marks | Location of model data folder | dir\_data <- "/home/qfeng/R/volume\_proj /data/" |
| dir\_winner | Folder location in quotation marks | Location of winner model for Units and Price | dir\_winner <- "/home/qfeng/R/volume\_proj /results/model\_dev/" |
| dir\_raw\_data | Folder location in quotation marks | Location of mollea master data (raw loans data, DPA) | dir\_raw\_data <- "//Prdermfs01/erm/Insurance/Models/Commercial Multi/Inputs Data/pragma/20191231/Production/ |
| raw\_data\_name | File name without extension in quotation marks | Raw csv data file name | raw\_data\_name <- "prod\_multi\_full\_19q4\_newLBA" |
| db\_data\_name | File name without extension in quotation marks | DataBuffet downloaded basket file name | db\_data\_name <- "Basket\_2020-07-16\_14h\_13m" |
| winner\_units\_name | File name without extension in | Units model development results (winner models) | winner\_units\_name <- "winner\_models\_units" (Please don’t change) |
| winner\_price\_name | File name without extension in | Price model development results (winner models) | winner\_price\_name <- "winner\_models\_price" (Please don’t change) |
| new\_data | "yes"/"no" | Create new aggregated data - "yes" will run "A\_data\_creation.R" to create a fresh data file; "no" will load an old version of "data\_aggr.Rdata" from data folder | new\_data <- "yes" |
| new\_raw\_data | "no" | In AIP, new raw MOLLEA data CANNOT be loaded and please keep “no” all the time | new\_raw\_data <- "no" (Please don’t change) |
| f\_hist | YYYY | Define first year of historical data | f\_hist <- 2000 |
| l\_hist | YYYY | Define last year of historical data | l\_hist <- 2019 |
| f\_dev\_Rental\_New | YYYY | Define first year of Rental New model development (No need to change) | f\_dev\_Rental\_New <- 2004 |
| f\_dev\_Rental\_Existing | YYYY | Define first year of Rental Existing model development (No need to change) | f\_dev\_Rental\_Existing <- 2004 |
| f\_dev\_Refinance | YYYY | Define first year of Refinance model development (No need to change) | f\_dev\_Refinance <- 2004 |
| l\_dev | YYYY | Define last year of model development | l\_dev <- 2019 |
| oos\_start | YYYY | Define first year of out-of-sample analysis (any year between f\_hist and l\_hist) and it won’t affect the final outputs (No need to change) | oos\_start <- 2015 |
| l\_frcst | YYYY | Define end of forecast | l\_frcst <- 2025 |

1. Run step “E\_model\_implementation.R” in its entirety, the final outputs will be named as volume\_projections.xlsx and it’s saved at [dir\_main results/implementation/tables/](dir_main%20results/implementation/tables/). (In this sample above, the final output is saved at: /home/qfeng/R/volume\_proj/results/implementation/tables/volume\_projections.xlsx)

