|  |
| --- |
| COMM350 |
| Forecasting Toronto Housing Prices |
| Peter Sephton |

|  |
| --- |
| Andrew Kosc & Michael Krakovsky  12-2-2016 |

Table of Contents

Introduction3

**The Theory4**

**Data Methodology6**

Stationarity Testing7

Original Variables7

*Statistical* *Analysis*8

First Difference10

*Statistical* *Analysis*11

**ARIMA Model12**

ARIMA Holdout Period Forecast13

**Cointegrating14**

Engle-Granger15

Johansen16

*Lag Selection*16

Vector Error Correction Model18

*VECM Output*19

*VECM Holdout Period Forecast*21

**Comparing ARIMA and VECM22**

Out of Sample Forecast24

**Conclusion26**

**Appendixes28**

**References39**

**Introduction**

The housing market is a core driver of the Canadian economy. In the past year, the growth in real-estate prices accounted for 30% of Canada’s annual GDP growth (Staff, 2016). As such, major fluctuations in the housing market not only affects individuals and their wealth, but also influences the strength of the entire economy. According to The Globe and Mail, Canadians spend more income on housing than “almost anyone around the world” (Luciw, 2014). Over the past few years, we have seen a steep rise in the prices of Canadian housing, driven primarily by the hot markets in Toronto and Vancouver. This has made these two cities a much more difficult place for young people to purchasing housing. Thus, as many of our colleagues prepare to begin their professional lives in Toronto, the decision to invest in a Toronto property is both a pertinent consideration and sizable life-decision. When buying a property, most individuals are buying their largest capital asset, and thus it is one of the most important investments a person can make (Boleat, 2012). It not only provides us with a sense of security but also serves as a source of capital wealth. However, what makes the purchase of a property either a wise or unwise investment, is primarily predicated on one thing, price.

Our paper will investigate the relationship between Toronto housing prices and multiple economic performance indicators. We intend to gather relevant statistics to model the trend that housing prices will follow in the next 2 years[[1]](#footnote-1). With news sources constantly discussing whether Toronto is in a housing bubble and which direction the market is heading, we’ve decided to build our own model to forecast the Toronto housing market.

Using both ARIMA and VECM modelling techniques, the results of our forecasts will provide insight that will be helpful in making an educated real estate decision. Therefore, the goal of this report is to determine where Toronto housing prices will be in 2 years, in turn, providing a helpful insight for individuals considering the purchase of a Toronto home as a worthwhile investment.

**The Theory**

In order to properly identify what data is relevant when examining fluctuations in housing prices, we decided to conduct industry research. After consolidating the insights gained from Investopedia (Nguyen, 2016), the International Monetary Fund (Zhu, 2014), and Royal Bank of Canada (RBC, 2016), we identified that demographics, interest rates, the economy & money supply, and government policies are key factors that affect housing markets.

With these factors in mind, we gathered data on thirteen variables (see Figure 1) that reflected the key drivers of housing prices. However, understanding that the presence of too many variables would greatly limit our degrees of freedom, we decided to select the four most appropriate variables from the data set. After taking into consideration the variables’ correlation coefficients (Exhibit 1), their ability to test cointegration with our dependant variable[[2]](#footnote-2), and their alignment with the factors found in our research, we concluded that we will use the following four independent variables for our project:

1. Canadian GDP 2. Personal Income Tax 3. Canadian Building Permits 4. Prime Lending Rate

|  |  |  |  |
| --- | --- | --- | --- |
| **Independent Variables** | **Correlation Coefficient**  **(Exhibit 1)** | **Stationarity**  **(Exhibit 2)** | **Alignment with Research** |
| Household Disposable Income | 0.9606 | I(2) | Variable that accounts for the health of Canadian economy & money supply |
| Canadian GDP (Income Based) | 0.9345 | I(1) | Variable that accounts for the health of the Canadian economy & money supply |
| Canadian Population | 0.9277 | I(2) | Variable that account for demographics |
| Personal Income Tax | 0.9261 | I(1) | Variable that account for the health of the Canadian economy & money supply |
| Canadian Building Permits | 0.9221 | I(1) | Variable that accounts for demographics and government policies |
| Ontario Population | 0.9045 | I(2) | Variable that account for demographics |
| Ontario Building Permits | 0.8902 | I(1) | Variable that accounts for demographics and government policies |
| Prime Lending Rate | (0.7576) | I(1) | Variable that accounts for the interest rate |
| Canada Unemployment | (0.5603) | I(1) | Variable that accounts for health of Canadian economy |
| Canada Total Units | 0.5599 | I(1) | Variable that account for demographics |
| Ontario Total Units | 0.3707 | I(1) | Variable that account for demographics |
| Ontario Unemployment | (0.2793) | I(1) | Variable that accounts for health of economy |
| Household Net Savings | 0.2224 | I(1) | Variable that accounts for money supply |

**Figure 1 – Assessment of Original Data**

We are using the variable Canadian GDP (Income Based) as it is a variable that reflects the health of the economy and the money supply in the hands of Canadians, both of which influence the housing market. Next, Personal Income Tax reflects the income levels of Canadians and their purchasing power; as more income tax suggests more income. Furthermore, Canadian Building Permits not only accounts for demographic changes and the additional level of supply, but also relate to government policies, as more building permits would be congruent with a government outlook supporting growth. Lastly, interest rates effect an individual’s ability to borrow, lend, and buy, which is a key driver of the housing market, and thus the Prime Lending Rate was also included. Given our thorough selection process, we are confident that the four independent variables chosen to help model our dependant variable are relevant in reflecting the big-picture of what factors impact housing prices.

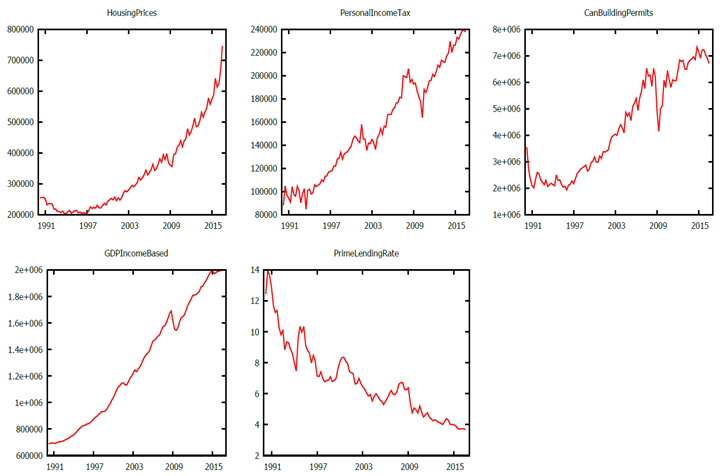
**Data Methodology**

The data for the entire project was retrieved from three main sources: the Toronto Real Estate Board, CANSIM Databases, and Stats Canada. All three of these databases are very reputable sources for statistics and thus we were confident that we will be working with and presenting accurate data. All data files were time-series data; however, the data sheets varied with respect to data frequency and start-points. Thus, all data was converted to quarterly intervals starting in 1990 yielding a total of 106 data points. We did this by using the “compact data” function in Gretl to convert the monthly series to quarterly[[3]](#footnote-3). Next, we felt approximately 25 years of data was enough of a time horizon to accurately predict housing prices, especially since it captured the 2008 housing crisis. We therefore chose 1990 as our start point, as all of the data series extended from at least Q1 1990 until Q3 2016. In addition, a holdout period of 5 data points (5 quarters) was chosen to effectively measure the strength of our forecasting model.

**Stationarity Tests**

Original Variables

We began our forecasting process by plotting our variables to visually assess their time series trends and check for any signs of stationarity. We would recognize stationarity if any variable, independent of time, showed a constant mean and variance.



**Figure 2 - Time Series Plot: All Variables**

As seen in Figure 2, the variables appear to be non-stationary, as all the data plots show signs of a trend and lack a constant variance.

We also examined the correlograms of the variables, and got complimentary results (Exhibit 2). Seeing linear decay in each variable’s ACF, as opposed to exponential decay, was indicative of our data being non-stationary and needing to be differenced.

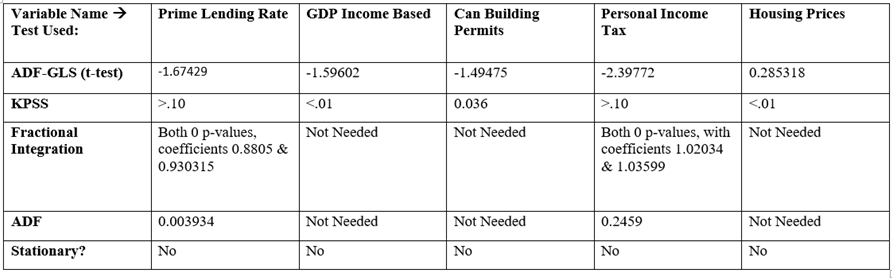
To further confirm our preliminary assessments, however, we needed to run statistical tests.

*Statistical Analysis*

The tests used for our statistical analysis were the ADF-GLS test and the KPSS test. The null hypothesis for the ADF-GLS test is that the data contains a unit root, which points to a non-stationary data set. The KPSS test, on the other hand, tests a null hypothesis of stationarity. Thus, we proceeded to run both tests for each variable, and compared the results. If both tests aligned on the stationarity of the variable, this was a conclusive result. Any conflicting conclusions provided by the tests were followed by an ADF test and, more importantly, a fractional integration test. The fractional integration test is used on variables where its series may be stationary, but it takes a long time to revert back to its mean, thus the accurate inference might not be recognized in the ADF/ADG-GLS or KPSS tests.

Given that the ADF-GLS test provides more reliable results than the ADF test, the ADF test was simply used as an additional data point to consider. The fractional integration test was the ultimate determinant of which prior result should be believed.

The ADF-GLS and ADF testing was done with a lag value of 5; therefore, we removed the first 5 values of the sample size to account for the lag variables[[4]](#footnote-4). KPSS testing was done utilizing a lag value of 4[[5]](#footnote-5). However, we ensured that we still conducted the tests with 5 data points removed as both tests needed to be conducted over the same sample size. As required, we allowed Gretl to choose the appropriate lag order for the fractional integration test. Every test was run against a 5% confidence level; therefore, p-values higher than 5% were not rejected while values under 5% were rejected against the null hypothesis. The trend option was selected for all variables, as the presence of a trend was identified in every time series plot.

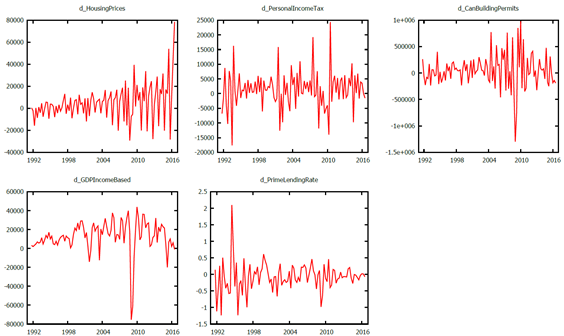
****

**Figure 3 – Unit Root & Stationarity Tests: All Variables**

Examining the results of our tests, presented in Figure 3, we see that GDP Income Based, Canada Building Permits, and Housing Prices are all non-stationary. We also see that Prime Lending Rate and Personal Income Tax generate conflicting ADF-GLS and KPSS results, but after fractional integration both variables are in fact non-stationary as well. Thus, because the stationarity tests conclude that all our series are non-stationary, it is necessary for us to difference the data.

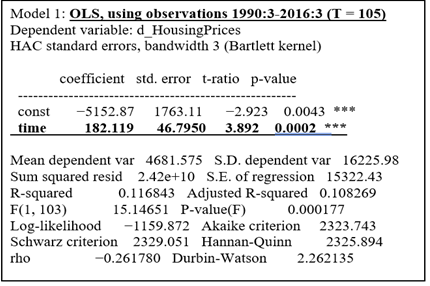
First Difference

To be used in our forecasting models, the variables have to be stationary I(1), or stationary I(0) in the first difference. As such, we must take the first difference of our variables, and test them for stationarity. As shown in Figure 4 below, we began by plotting the differenced series. All the variables appeared stationary, with the exception of the differenced Housing Prices. The differenced Housing Prices plot seemed to display a bit of a trend.



**Figure 4 – Time Series Plot: First Differences**

To determine whether to use a trend in our unit root test for the first difference of Housing Prices, we ran a simple OLS regression with a time trend variable and a constant.



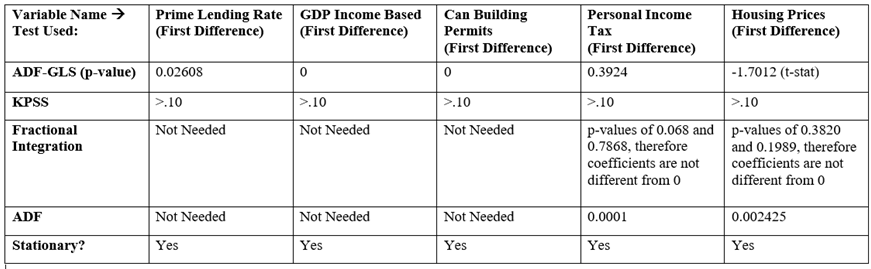
**Figure 5 – OLS Test: Housing Prices**

The OLS test in Figure 5 shows that a time trend variable with a p-value of 0.0002 is significant at the 5% level. Therefore, we added a trend for only the first difference of Housing Prices while ran the other tests with just a constant.

*Statistical Analysis*

We followed the same process on our first differences as we did with our initial variables. The only change to note would be that we now removed 6 data points from our sample to account for both the lags as well as the first difference of our variable.

Figure 6 below shows the results of our tests, which indicate that all the series are stationary in their first difference.



**Figure 6 – Unit Root & Stationarity Tests: First Differences**

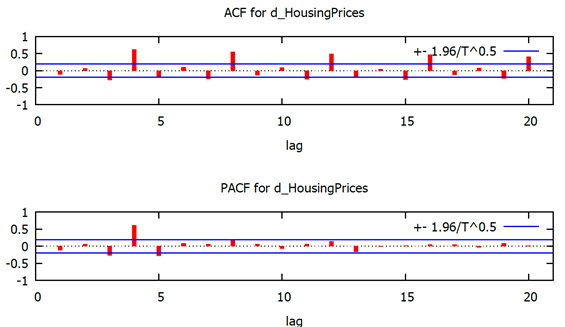
Ultimately, these tests provide us with the inferences that determine which models can be used to forecast our data. Given that all of our variables are integrated to the same order I(1), they can now be used to build both VECM and ARIMA models.

**ARIMA Model**

ARIMA Forecast

The ARIMA model consists of three parameters, the AR process, the integration, and the MA component. It utilizes past data points and past disturbances to attempt to explain the future of a series. Therefore, Housing Prices will be modelled based on its own historical pattern.

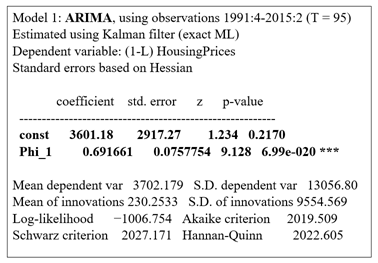
We have found in the stationarity tests that we need to use the first difference of the variable. The next step is to determine which AR and MA to utilize in the ARIMA model. In order to find the appropriate values, we graphed the correlogram of the first difference of the Housing Prices.



**Figure 7 – Correlogram: First Difference of Housing Prices**

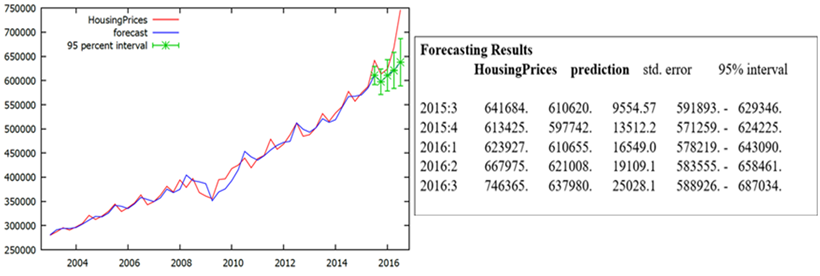
In Figure 7, the correlogram displays a pattern in the ACF. A spike emerges every 4th quarter above the confidence interval which leads us to believe that Housing Prices are seasonal. In between the large seasonal spikes, the lags oscillate consistently through the entire correlogram. The combination of spiking and oscillating lags allows us to confidently assume that there is the presence of a seasonal AR(1). Within the PACF, the correlogram displays a single spike at lag 4; however, no pattern exists, allowing us to assume an MA (0). We can therefore determine that the ARIMA (1, 1, 0) model will be the best fit. The I in AR**I**MA will be used in the non-seasonal section of the test while the AR section of **AR**IMA in the seasonal section of the ARIMA model[[6]](#footnote-6).

The ARIMA model output (Figure 8) provides a constant that is not significantly different from 0 with a p-value of 0.2170. In addition, it includes a Phi\_1 with a coefficient of 0.6917 that is significantly different from 0 with a p-value of ~0. The presence of AR(1) means that the errors in the equation might be related to each other over time. The constant should still hold its coefficient because the function should not be plotted through the point of origin. Therefore, we can now plot the ARIMA model over a holdout period and, in turn, judge its accuracy[[7]](#footnote-7).



**Figure 8 – ARIMA Model Output**

Once the predicted data was modelled against the actual data in the holdout period, the ARIMA model failed to accurately predict Housing Prices within a 95% confidence interval in three of five quarters. Housing Prices exceeded our confidence interval by $12,338 in Q1, $9,514 in Q4, and $59,331 in Q5.

  
**Figure 9 – ARIMA Holdout Period Forecast**

Given the ARIMA forecast output, we will now create a VECM model and can then see which approach has more accurate results.

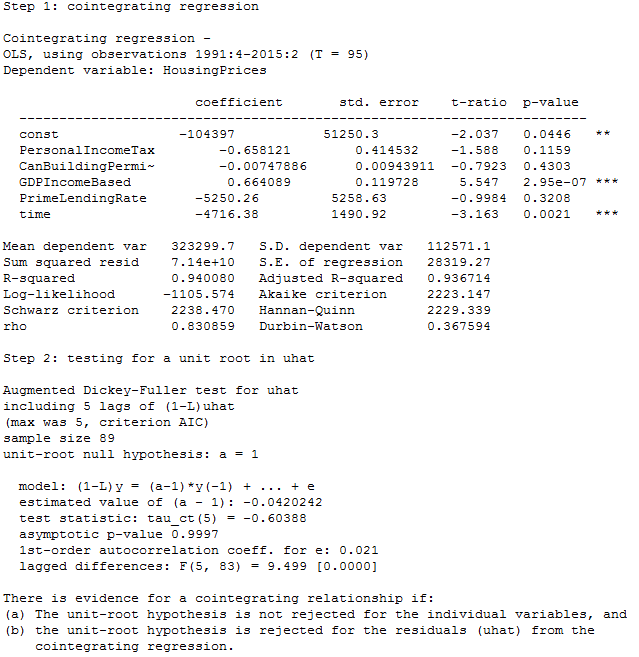
**Cointegration Model**

Since our ARIMA model failed to provide us with definitive results, we will now attempt to use the VECM model to forecast housing prices. In order to create the VECM model, we need to ensure the variable I(1), which was done in the beginning of the report. In addition, the variables need to be co-integrated with each other. Co-integration entails that the variables restore the model to equilibrium in the long-term. The engle-granger test as well as the more powerful Johansen test will display whether these variables respond to fluctuations in the housing market.

Engle-Granger Test

The Engle-Granger test examines the null hypothesis of non-cointegration by testing the residual from the cointegrating regression to see if it contains a unit root. The critical value of these unit root tests is influenced by the deterministic components in the equation (whether the data has a constant and/or a trend) as well as any lagged dependent variables.

We begin our co-integration testing with the Engle-Granger approach - ensuring a lag order of 5 given our sample size. Given our data plot, we determined that running the test with a constant and a trend was the most appropriate.

  
**Figure 10 - Engle-Granger Results**

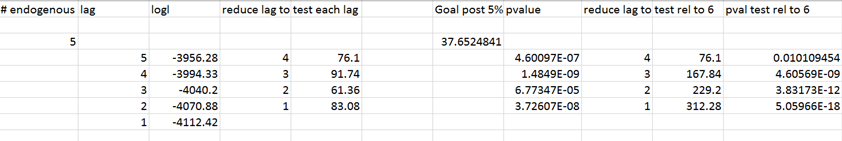
Examining Figure 10, we see that it resulted in an asymptotic p-value of 0.997, which is not lower than 0.05. Thus, we do not reject the null hypothesis that the residuals have a unit root, thereby indicating that the data is not co-integrated. However, the Engle-Granger Test is not as reliable as the Johansen Test when more than 2 variables are present, and thus it is important to also conduct the Johansen test, and prioritize those results as more credible as well.

Johansen Test

The Johansen test examines how many cointegrating vectors we have within our variables. The Johansen approach should lead us to the same inferences as the Engle-Granger approach when using only two variables. However, because we have more than two variables, we will rely on the approach of the Johansen test as it is more credible.

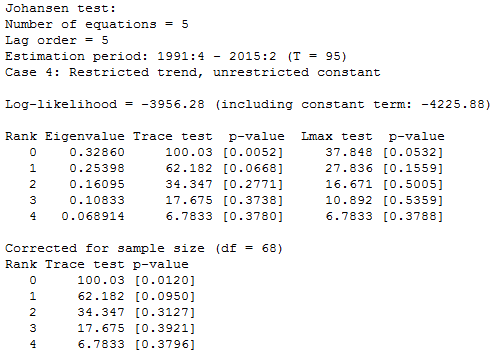
*Lag Selection*

In order to ensure we input the proper lag order into the Johansen test, we used the Likelihood Ratio Test (Professor Sephton’s Lag Selection excel model). We ran through each Johansen test and inputted the logl values into the model. Since we reject every value from lags and onwards, we can infer that the proper amount of lags to use is 5.



**Figure 11 - Likelihood Ratio Test**

Given the fact that the appropriate lag has been determined, we can now use it in the Johansen Test to determine the amount of cointegrated vectors in our data[[8]](#footnote-8). Given that the trends in our series appear to be different/not cancelling out in the cointegrating regression (Exhibit 3), we used a restricted trend and an unrestricted constant.

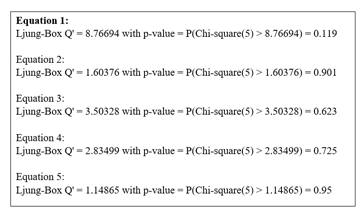
  
**Figure 12 - Johansen Test Results**

Examining Figure 12, we follow the Rank Trace Test which shows us that we do not reject the null of one cointegrating vector. This is because rank 1, with a p-value of 0.0950, is the lowest rank to be significant at the 5% level, which indicates that our data is co-integrated with one cointegrating vector. The Alternative Trace Test and the Lmax Test, also determine a rank of 1, as their p-values (0.0668 and 0.1559 respectively) are both also greater than an alpha of 5%.

Now that we have identified the appropriate number of lags as well as the number of cointegrating vectors in our data, we are able to move forward to the creation of a VECM.

Vector Error Correction Model

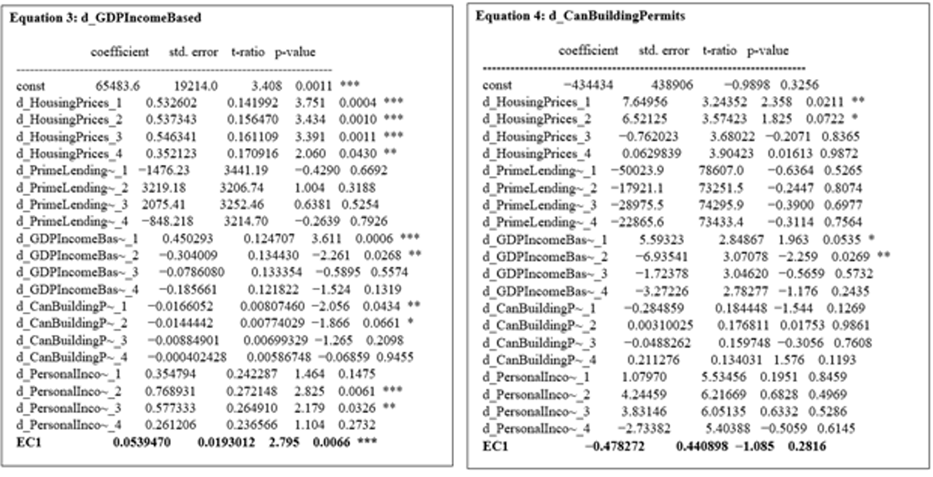
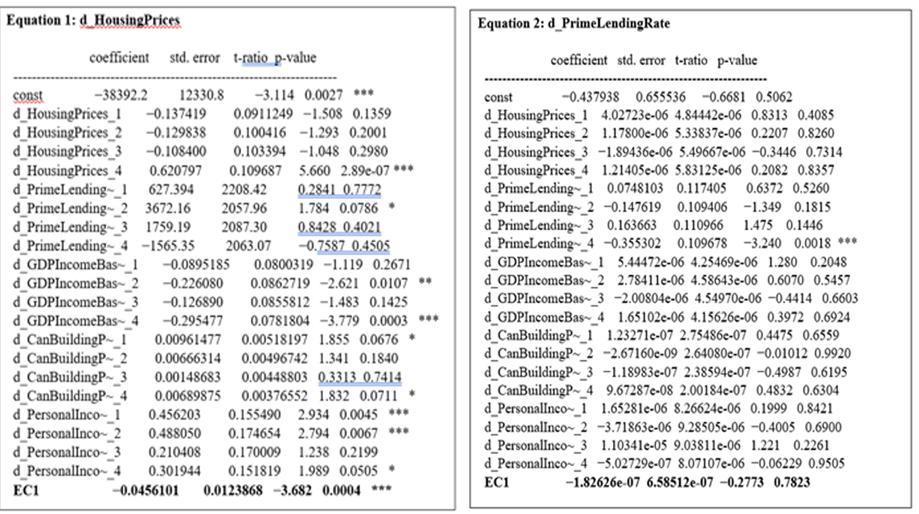
Since we posses the correct lag amount and the trace order, we can insert these values into our VECM model[[9]](#footnote-9). We used a model with a restricted trend due to the appearance of our series (Exhibit 3), which also maintains consistency with our choice for the Johansen test. After running the VECM model, we tested for autocorrelation with our lag order of 5.

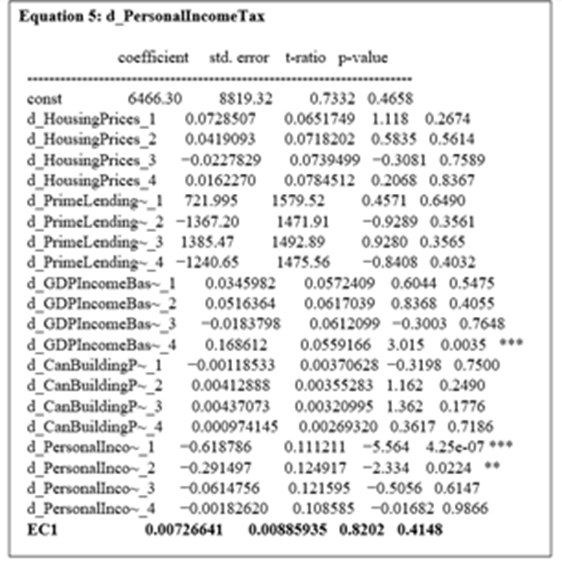


**Figure 13 – VECM Autocorrelation Test**

As seen in Figure 13, we are urged to not reject the null hypothesis of no autocorrelation, as all the Ljung Box Q’ p-values are greater than 5%.

Now that we have established there is no autocorrelation, we can examine the results of the VECM model to determine which variables respond to the error correction term and which are weakly exogenous. In a cointegrated system like ours, at least one of the variables has to respond to the error correction term, otherwise the series would not be drawn back to equilibrium (thus failing to be cointegrated).

  
See Appendix C for more details of VECM Model.



**Figure 14 – VECM Model Output**

The results in Figure 14 indicate that there are two variables that respond to the VECM model at the 5% confidence interval: Housing Prices with a p-value of 0.0004 and GDP with a p-value of 0.0066. These variables therefore contribute to the VECM model, since we reject the null hypothesis that the coefficients are not significantly different from 0. The other three variable are weakly exogenous, and thus do not help the model return to equilibrium in the long-run. Therefore, any shocks will be restored by the Housing Prices and GDP Income Based.

After being assured of the presence of cointegrating terms as well as the existence of no autocorrelation, we can now confidently forecast our VECM model[[10]](#footnote-10).

*VECM Holdout Period Forecast*

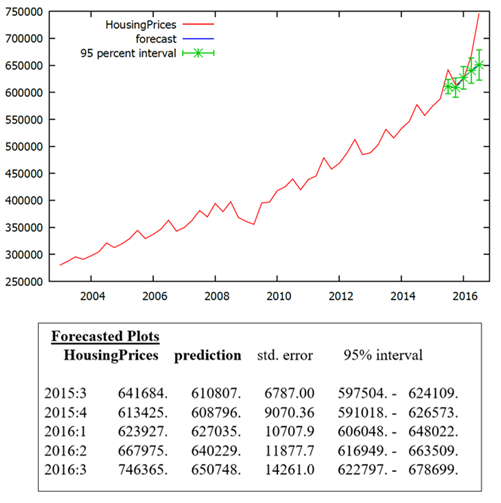
  
**Figure 15 – VECM Holdout Period Forecast**

Figure 15 shows the graphical forecast as well as the numerical output for our VECM model. The housing prices fall out of the 95% confidence interval range three times (in Q1, Q4, and Q5 ). This happens when the VECM model under-projects the housing prices by $17575, $4466, and $67666 respectively.

We can now compare the performance of both our VECM and ARIMA models to determine which modelling approach did the best job in capturing our series on the hold-out period.[[11]](#footnote-11)

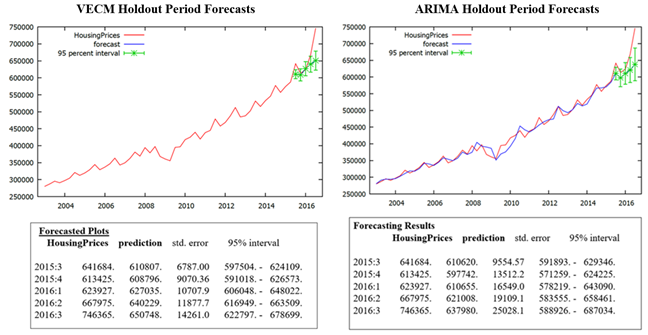
**Comparison of ARIMA Model and VECM**

Comparing the forecasts of the ARIMA and the VECM model within our holdout period will allow us to select the better modelling approach, which can then be used to project Housing Prices into the future. The evaluation statistics used to determine the superior model will be the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Theil’s U.

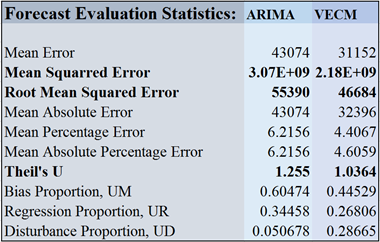
The MSE and the RMSE quantify the difference between the predicted values and the true values of the variable being estimated. As such, the lower the statistic, the more accurate the model is. Theil’s U is an accuracy measure that compares the forecasted results with the results of forecasting with minimal historic data. Similarly, the lower this statistic, the better the model.

Figure 16 shows the VECM and ARIMA forecasts and Figure 17 directly compares the evaluative statistics of the modelling approaches.

The results indicate that the VECM model is the superior approach when forecasting Housing Prices. Firstly, its Mean Squared Error of 2.18E+09 is lower than the Mean Square Error of 3.07E+09 for the ARIMA model. Next, its Root Mean Squared Error of 46684 is lower than the ARIMA model’s 55390 Root Mean Squared Error. Lastly, its Theil's U score is 0.2186 points lower than that of the ARIMA model. Consequently, because the VECM model performs better across all three evaluation statistics, it considered to be the better model. The out-of-sample VECM prediction will, therefore, be viewed as more reliable than that of the ARIMA model.



**Figure 16 – Comparison of ARIMA and VECM: Forecasts**



**Figure 17 – Comparison of ARIMA and VECM: Forecast Evaluation Statistics**

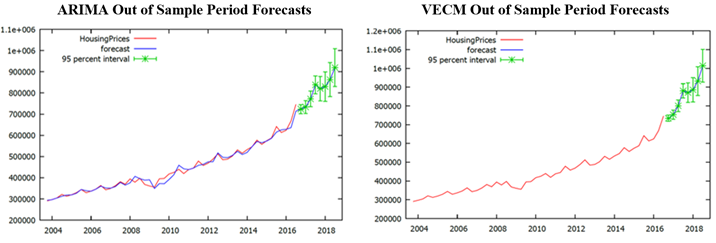
It is important to note that both models underestimate the prices of the housing market – whereby neither model captures the rapid price acceleration that has occurred in recent quarters. What this means is that the pace at which housing prices are increasing is not consistent with its historical behaviour, nor is it effectively being restored to equilibrium by the other cointegrating variables. We see this as a potential warning sign that there might be a housing bubble, as the sellers-market is driving unreasonably high valuations for properties.

Assuming the prices return to our confidence interval, we will create two out of sample forecasts to attempt to predict the future housing prices.

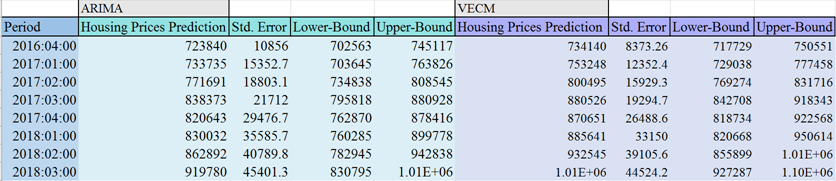
**Out of Sample Forecasts**

In order to predict housing prices, we started by choosing an out of sample forecast length of 8 quarters (2 years). This forecast length was chosen as it provides us with an appropriate time horizon to consider making an investment decision into real estate post-graduation, as well as captures two full cycles of seasonality.

The ARIMA model and VECM model were used to predict housing prices in the future with the same parameters of the models created for the holdout period forecasts. Prior to creating the model, the data was set to the original sample size, and the first 6 data points remained removed to account for lags and the first difference. Figure 18 shows the forecasted results and Figure 19 shows the statistical comparison.

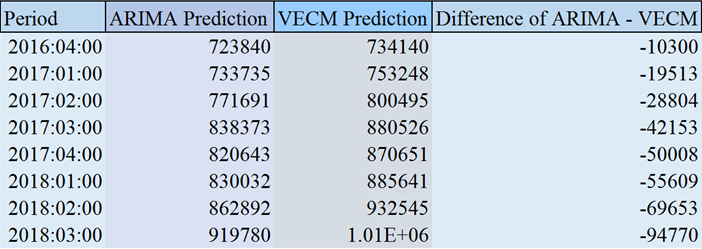


**Figure 18 – Out of Sample Forecasts: ARIMA and VECM**



**Figure 19 – Out of Sample Forecast Comparison: ARIMA and VECM**

Both forecasts expect Housing Prices to increase over the next two years, but ARIMA does predict a lower result than the VECM in every quarter. Figure 20 highlights how far the predictions of the VECM and ARIMA model were from each other. We see that the VECM model has both a higher slope as well as begins at a higher forecasted price. Ultimately, the VECM models predicts a value of $1,010,000 and the ARIMA model predicts a value of $919,780 for housing prices by Q3 2018.



**Figure 20 – Out of Sample Forecasts: ARIMA and VECM**

**Conclusion**

To conclude, we have created an ARIMA and VECM model that forecasts Housing Prices in Toronto over the next two years. By comparing these two modelling approaches over our holdout period from the beginning of Q3 2015 until the end of Q3 2016, we determined that the VECM model was more accurate than the ARIMA (1,1,0) based on the forecast evaluation statistics.

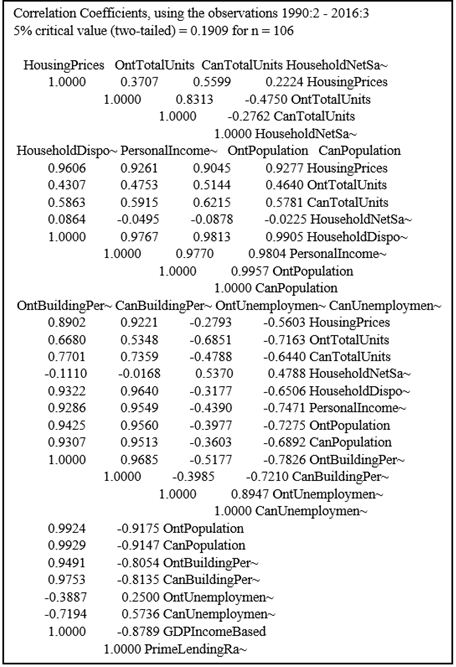
This lead us to predict that the housing prices in Toronto will reach an average of $1,010,000 by the third quarter of 2018 using our VECM model. When considering the ARIMA model, it forecasts a lower average price of $919,780 by the third quarter of 2018. Given that the current average price is $746,365, both out of sample forecasts predict a sizable price increases over the next two years.

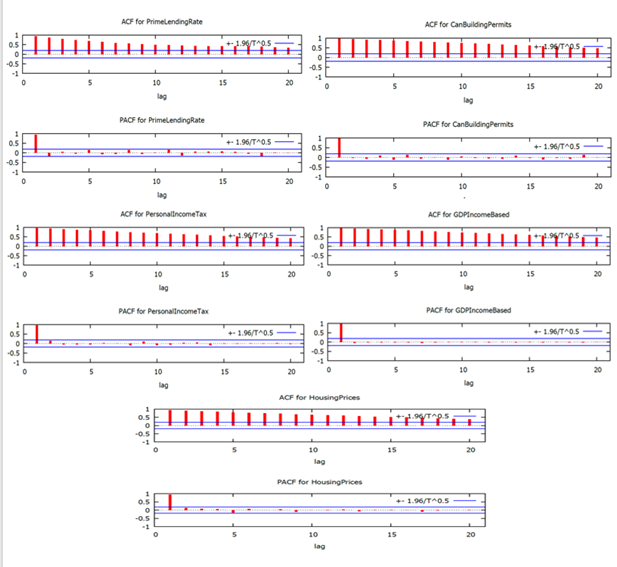
Although the results are encouraging, we are hesitant to make any investment decisions off these forecasts alone, as there are outstanding factors that still require consideration. Firstly, if we increased the frequency of our data, resulting in more data points, we would be more confident in our models. We believe the range of 25 years, covering the market crash in 2008, is more than sufficient, however using quarterly data limited us to 106 observations. Utilizing monthly data over the course of the 25 years would have made our data series a lot more robust. Next, if there is in fact a housing bubble (that has either been forming for years or is forming now), neither the ARIMA nor the VECM model could predict a possible burst or deflation. As such, the models do not properly account for these risks nor the slow down or down-turn in prices associated with them. Moreover, we did not include any variables that were specifically tracking Toronto data in our analysis, beyond the dependant variable itself. Given the difference in economic and demographic factors nation-wide vs. in Toronto exclusively, our model could have benefitted from having variables that represented Toronto data, and thus were more closely related to Toronto housing prices.

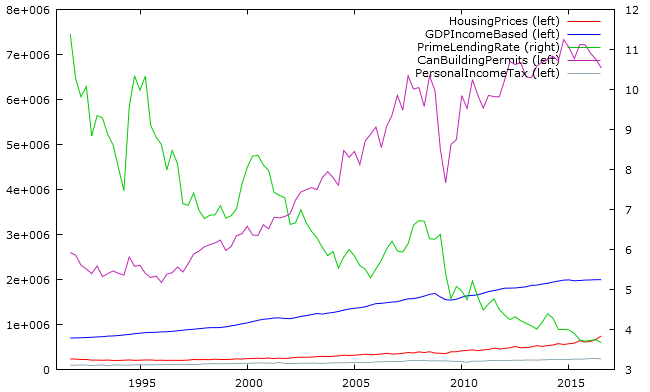
Overall, we are very satisfied with the forecasting analysis we completed. Our report provides insights into the potential future of housing prices in Toronto as well as the relationship housing prices have with other variables. We believe our model produces reasonably accurate results in terms of predicting housing prices over the next two years, yet the presence of too many external factors requires further analysis for an educated investment decision to be made.

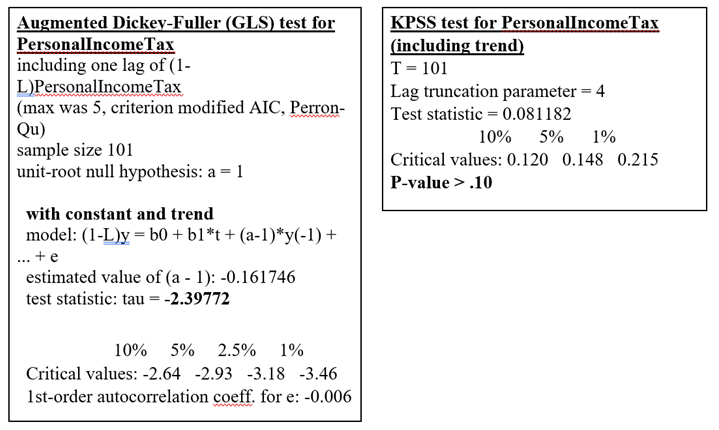
**Appendices:**

**Exhibit 1: Correlation Matrix All Variables**



**Exhibit 2: Correlograms for All Variables**

**Exhibit 3: Time Series Plot All Data**

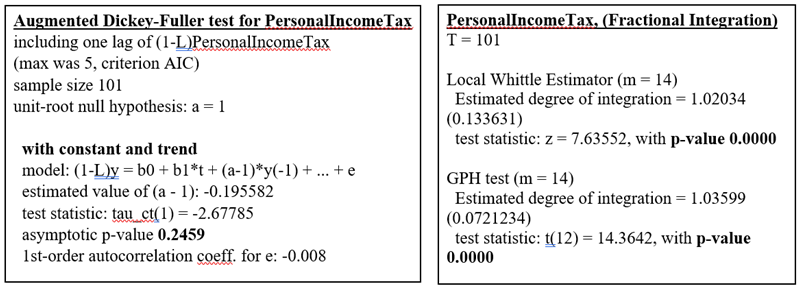
**Exhibit 4: Unit Root Tests of Original Variables**

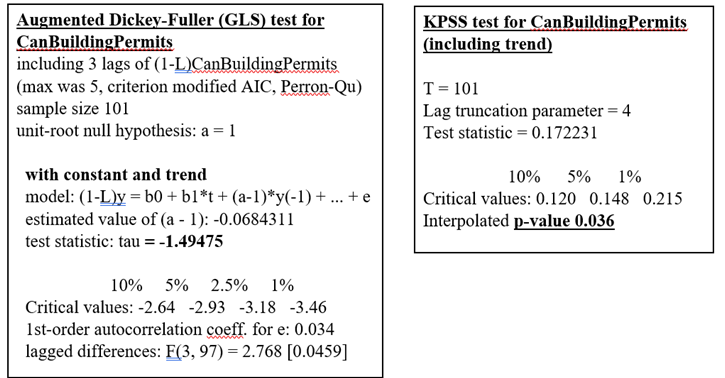
Fractional Integration for Personal Income Tax

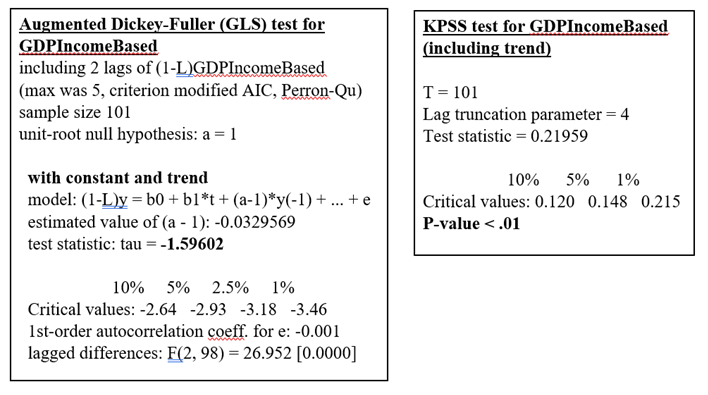
ADF Test for Personal Income Tax

KPSS Test for Personal Income Tax

ADF-GLS Test for Personal Income Tax





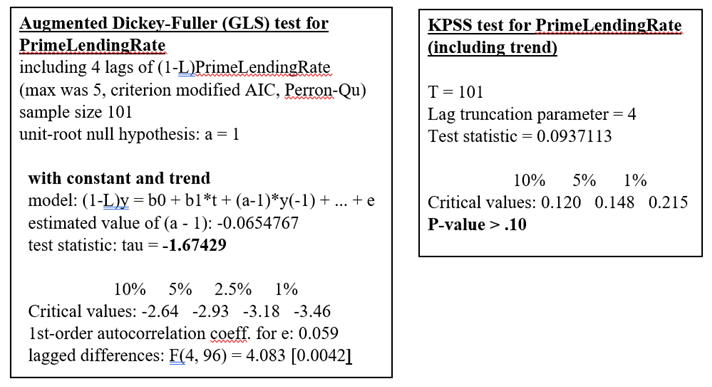
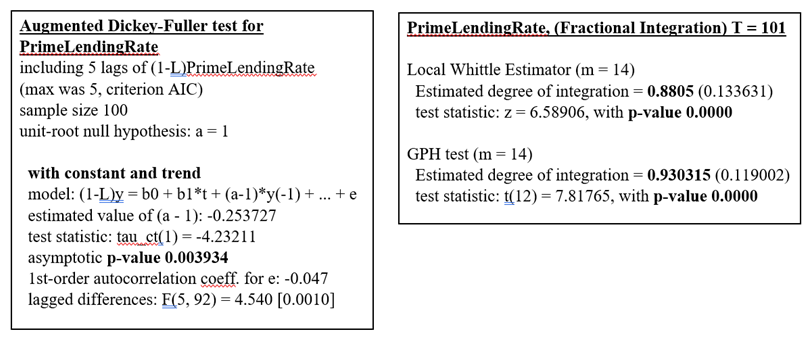


KPSS for GDP Income Based

ADF-GLS for GDP Income Based

KPSS for Can Building Permits

ADF-GLS for Can Building Permits

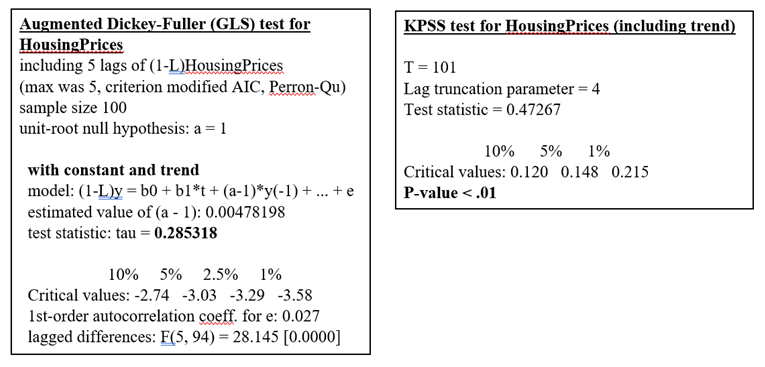


ADF-GLS for Prime Lending Rate

KPSS for Prime Lending Rate

Fractional Integration of Prime Lending Rate

ADF for Prime Lending Rate



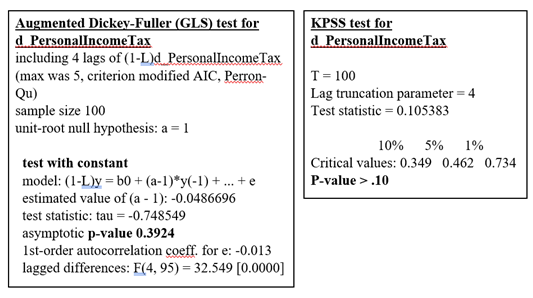
KPSS for Housing Prices

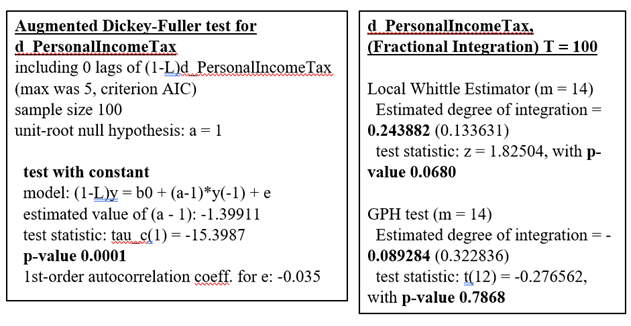
ADF-GLS for Housing Prices

**Exhibit 5: Unit Root Test on First Difference Variables**

KPSS for d\_Personal Income Tax

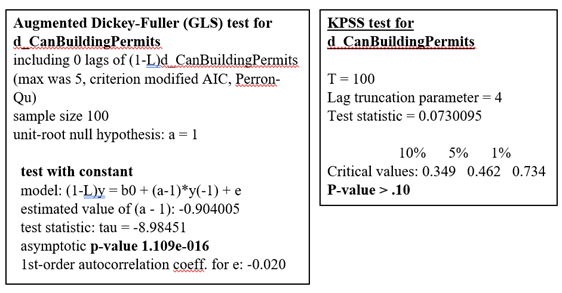
ADF-GLS for d\_Personal Income Tax

****

****

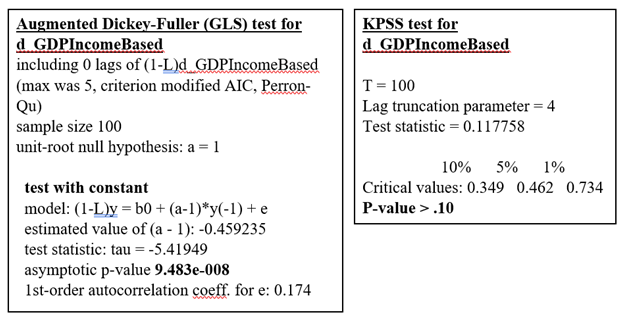
Fractional Integration for d\_Personal Income Tax

ADF for d\_Personal Income Tax

****

KPSS for d\_Can Building Permits

ADF-GLS for d\_Can Building Permits

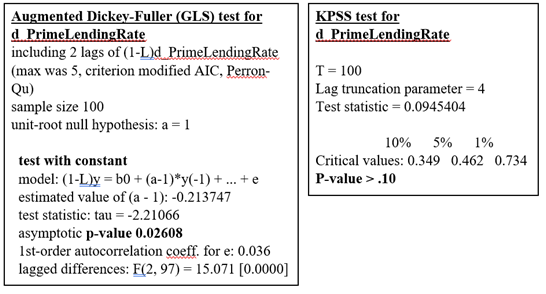
****

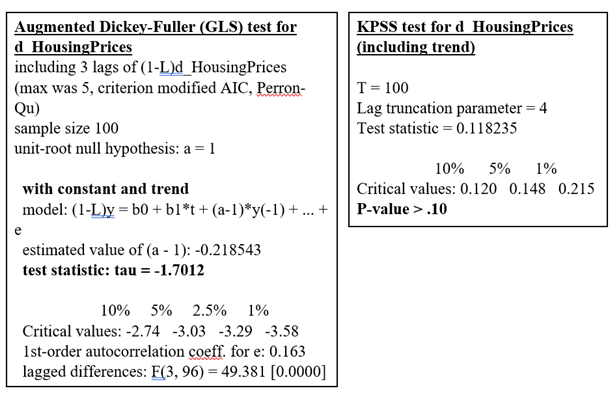
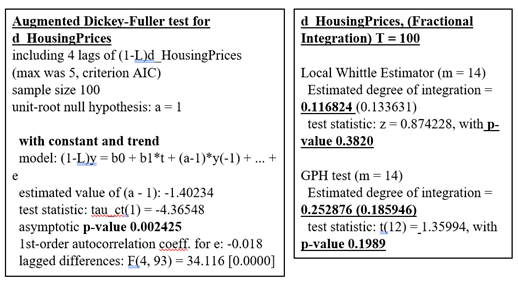
KPSS for d\_GDP Income Based

ADF-GLS for d\_GDP Income Based

KPSS for d\_Prime Lending Rate

ADF-GLS for d\_Prime Lending Rate

****

****

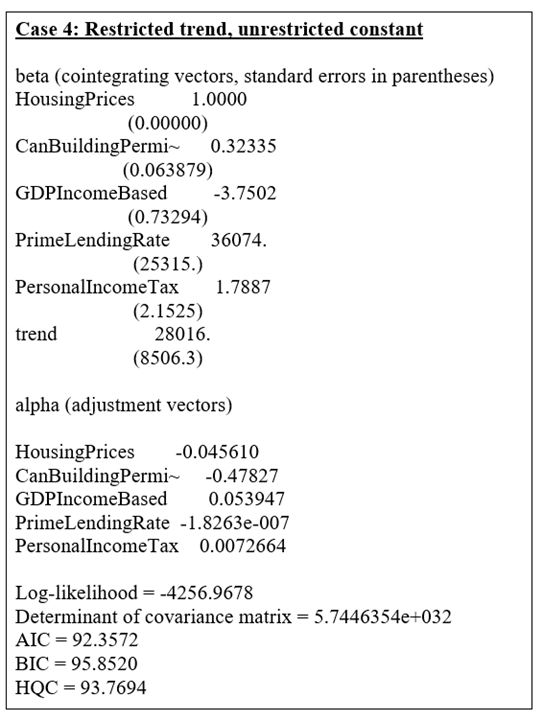
Fractional Integration for d\_Housing Prices

GLS for d\_Housing Prices

KPSS for d\_Housing Prices

ADF-GLS for d\_Housing Prices

**Exhibit 6: VECM Betas and Alphas**



**References**

Boleat, M. (2012). The Mortgage Market. [Novel] Retrieved December 01, 2016, from <https://books.google.ca/books?id=-U_wEoivFEcC&pg=PA42&lpg=PA42&dq=home+largest+capital+asset+people+own&source=bl&ots=iBlZeIyqmn&sig=8IP9P3KfERlrg4oB9UnLU3yTo5w&hl=en&sa=X&ved=0ahUKEwi2k_awgtTQAhXk54MKHa-UA644ChDoAQgaMAA#v=onepage&q&f=false>

Luciw, R. (2014, Oct 30) Canadians spend more income on housing than almost anyone in the world [Article] The Globe and Mail. Retrieved December 01, 2016, from

<http://www.theglobeandmail.com/globe-investor/personal-finance/household-finances/canadians-spend-more-of-their-income-on-housing-than-almost-anyone-in-the-world/article21369414/>

Nguyen, J. (2016). *4 Key Factors That Drive the Real Estate Market*. [Article]. Retrieved from<http://www.investopedia.com/articles/mortages-real-estate/11/factors-affecting-real-estate-market.asp>

Royal Bank of Canada (RBC). (2016). *Canadian Housing*. [Report]. Retrieved from:<http://www.rbc.com/economics/economic-reports/canadian-housing-forecast.html>

SF Gate. (2016). *How Does the Economy Affect the Housing Market?*. [Article]. Retrieved from:<http://homeguides.sfgate.com/economy-affect-housing-market-50583.html>

Staff, B. (2016, June 30). How Canada’s hot housing market is propping up GDP growth - Article - BNN. Retrieved December 01, 2016, from <http://www.bnn.ca/how-canada-s-hot-housing-market-is-propping-up-gdp-growth-1.518657>

Zhu, M. (2014). *Housing Markets, Financial Stability and the Economy*. [IMF Communications Department]. Retrieved from:<http://www.imf.org/external/np/speeches/2014/060514.htm>

1. The two-year forecast will run from the third quarter of 2016 until the third quarter of 2018. [↑](#footnote-ref-1)
2. We ran stationarity tests for all thirteen of the independent variables to determine which were eligible for modelling (testing for I(1) and I(2)). This was a non-core process, so the tests were omitted for the report. [↑](#footnote-ref-2)
3. Note: Although one can convert monthly data to quarterly, and quarterly to yearly, it is not advised to go the other way (ex: quarterly to monthly). [↑](#footnote-ref-3)
4. The cube root of our sample size, 106. [↑](#footnote-ref-4)
5. The fourth root of our sample size, 106. [↑](#footnote-ref-5)
6. The seasonal AR element of the modelling process was advised by Professor Sephton for our project. [↑](#footnote-ref-6)
7. ARIMA Model will be using 5 quarter holdout period [↑](#footnote-ref-7)
8. It’s good practice to set the maximum lag that you will consider, and only test down to reduce the lag length. We did not try and test up. [↑](#footnote-ref-8)
9. Correct lag amount is 5 and Trace Order is 1. [↑](#footnote-ref-9)
10. VECM model will be using a 5 quarter hold out period [↑](#footnote-ref-10)
11. It’s important to note that there is no guarantee the modelling approach will continue to be the best out of sample. [↑](#footnote-ref-11)