Team 17 Project, Stat E-150 David Modjeska
Harvard University, Spring 2016 Dominic Murphy

Predicting foreign exchange rates between the US and Canada

1. Abstract

Anecdotally, fluctuations in the exchange rate between Canadian and US dollars are influenced by commodity prices and economic factors. This report presents a model combining those influences. A multiple regression model using oil futures open price, gold price, and Canadian GDP is found useful in explaining the variability in USD/CAD FX rate.

Keywords: Foreign exchange, oil futures, oil price, GDP, gold price, commodity prices

2. Introduction

Foreign exchange (FX) allows goods and services to move around the world. The FX market is global and decentralized. It is the world's largest market, with the main participants being large banks.

The FX market's three main participants' are traders, corporate CFO's, and governments. For traders, rates support hedging risk, pursuing profits, pricing FX derivatives, and algorithmic trading. For CFO's, rates support assessing risk and cross-border ventures. For governments, rates support budgeting and setting monetary policy.

Given the criticality of FX rates to the global economy, and the sizeable profit-making opportunities, both explanation and prediction are interesting and valuable. Explanation offers the promise of helping with key tasks above. Prediction offers more -- the allure of profit-making. This report aims to create an explanation model that will form a foundation for our aspirational goal of prediction. (Prediction is a bit harder.)

This report will focus on the exchange rate between US and Canadian dollars (USD and CAD). The US is a major trading partner for much of the world, and the USD is often seen as the benchmark currency in global markets. The Canadian dollar is a commodity currency with high volatility and trading volume. This is attractive to traders as there is liquidity and profit potential. (Also, the US is Canada's largest trading partner.)

3. Problem Definition and Obtaining Data

The project proceeded as follows. Having identified currencies of interest, we conducted informal research in public data, news articles, and white papers. Our data was then sourced from central banks, government statistics, and economic databases. We combined multiple data sources to unify grain, timing, and availability. Finally, we proceeded with formal analysis and verification, as described in the remainder of this report. (Data wrangling was done using Excel and R. Data sources are listed in the bibliography.)

David Modjeska
Dominic Murphy

On the basis of this research, we identified potentially useful variables: oil price, gold price, USD FX rates (relative to JPY, GBP, SEK, CHF, and EUR), Canadian interest rates, and Gross Domestic Product (GDP) by country and industry sector. Resource data includes both spot (current) and forward (futures) prices. The study population includes all days after the introduction of the Euro in 1999. From this population, we examined the last 10 years (i.e., 2006-2016), for which reliable and consistent data is available. As our study sample contains 3,596 observations, we used a 10% random sample for constructing the model. The aim of sampling was to avoid auto-correlation, improve the generalizability of results, and offset the statistical power of large sample sizes. With 360 cases in the random sample and approximately 35 variables, the ratio of cases to variables remains above the 5-to-1 threshold recommended by Anscombe. (Bootstrapping is discussed below.) Our observational unit was one day in this time series.

The study question is as follows: How can we best predict the average FX rate between the US and Canadian dollars? Our hypotheses are as follows:

- Oil price will explain a significant amount of the variability in USD/CAD FX rate, after accounting for other variables in the model. Oil accounts for 20% of Canada's exports, so an increase in the price of oil moves the balance of trade favorably and increases the value of the CAD. (Lafrance, Zhang, Issa R, & Helliwell, 2010)
- Gold price will explain a significant amount of the variability in USD/CAD FX rate, after accounting for other variables in the model. USD and Gold are inversely related: as the USD falls relative to its trading partners, the CAD increases against the USD. (Gilroy, 2014)
- Canadian GDP will explain a significant amount of the variability in USD/CAD FX rate, after accounting for other variables in the model. Poor GDP data releases effects trader sentiment. When GDP decreases, the value of the CAD decreases. (Cofnas, 2011)

4. Formal Analysis

In order to choose, fit, assess, verify, and contextualize a multiple regression model for average USD/CAD FX rate, we'll proceed in seven steps:

1. To start, we'll get a general overview of possible predictor variables via a correlations matrix.
2. We'll attempt stepwise regression to leverage automated tools.
3. If needed, we'll choose and fit a simple model upon which to base further exploration.
4. We'll run a series of nested F-tests to determine which candidate predictor subsets explain significant variability in the response.
5. Using the core model and nested F-tests, we'll explore which variables explain significant additional variability without introducing excessive collinearity.
6. We'll assess and check the final model.

David Modjeska
Dominic Murphy

7. Bootstrapping will be used to estimate how much variability in the response the final model explains, accounting for sampling variability.

a. Correlations Matrix

In order to determine which variables in the data set might be most useful as predictors, we can first consider a bivariate matrix of Pearson correlations. The results show that the variables most highly correlated with *FXRate* are *OilPrice* (.816), the mining variables (.415 to .503), some of the Canadian government statistics (.346 to .708), some of the FX variables (.249 to .868), and some of the oil futures variables (approximately .835). The results also show that many of the predictors are highly correlated with each other. So the challenges in choosing and fitting a model are two: first, to include the most explanatory predictors; second, to exclude the excessively collinear predictors. (The SPSS output for correlations with *FXRate* is shown in Appendix 2 for space reasons.)

b. Stepwise Regression with Verification

Since our data set includes approximately 40 variables, choosing and fitting a multiple regression model could in principle proceed best by an automated method such as stepwise regression. The SPSS output is shown below, condensed for efficiency to show just the last model in the process, model 21. (Writing out the fitted model is cumbersome, because there are 13 predictors, but all of the coefficients are shown in the 'B' column of the table below.)

[Note: SPSS output below has been condensed - showing the last model only]

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
21	(Constant)	1.392	.097		14.283	.000		
	FX_USD_Per_AUD	.291	.033	.370	8.861	.000	.056	17.776
	CA_GDP_Growth	.004	.001	.066	4.324	.000	.420	2.380
	USDInterest3M	-.300	.035	-.240	-8.652	.000	.128	7.838
	OilAndGasPercentOfGDP	-.006	.003	-.036	-2.173	.031	.357	2.803
	USDInterestON	.151	.031	.116	4.821	.000	.168	5.954
	FX_USD_Per_GBP	-.150	.027	-.109	-5.618	.000	.262	3.818
	FX_USD_Per_SEK	-.035	.005	-.311	-6.466	.000	.042	23.613
	GovDebt_CA	-.006	.002	-.721	-3.643	.000	.003	399.503
	MetalOre	-3.908E-6	.000	-.107	-3.237	.001	.090	11.095
	FX_USD_Per_EUR	-.118	.034	-.153	-3.415	.001	.049	20.544
	Debt_GDP_Ratio_CA	.007	.003	.459	2.548	.012	.003	331.081
	Oil_Future_Last	-.005	.002	-.873	-2.490	.014	.001	1254.414
	Oil_Future_Open	.006	.002	1.104	3.151	.002	.001	1250.610

a. Dependent Variable: FXRate

[Note: SPSS output below has been condensed - showing the last model only]

Model Summary[v]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
21	.989	.979	.978	.012099972

v. Dependent Variable: FXRate

[Note: SPSS output below has been condensed - showing the last model only]

ANOVA[a]

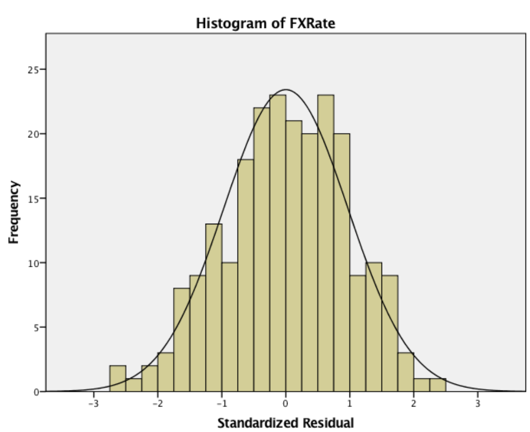
Model		Sum of Squares	df	Mean Square	F	Sig.
21	Regression	1.462	13	.112	767.881	.000[v]
	Residual	.031	214	.000		
	Total	1.493	227			

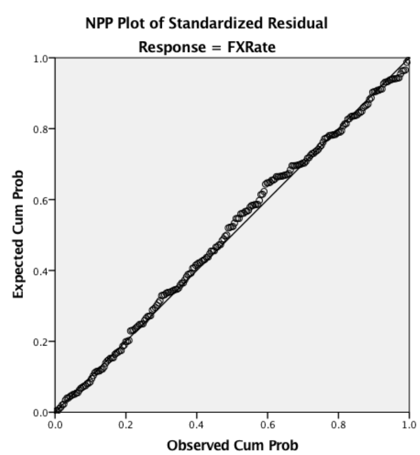
a. Dependent Variable: FXRate

At first glance, the stepwise model appears promising. The model summary shows that the adjusted R-squared value is .978, which indicates that the model explains a high percentage of the variability in USD/CAD FX rate, after accounting for the number of predictors and the sample size. The standard error of the estimate is small at approximately 0.01 USD/CAD, which indicates a high degree of accuracy in predicting the average FX rate.

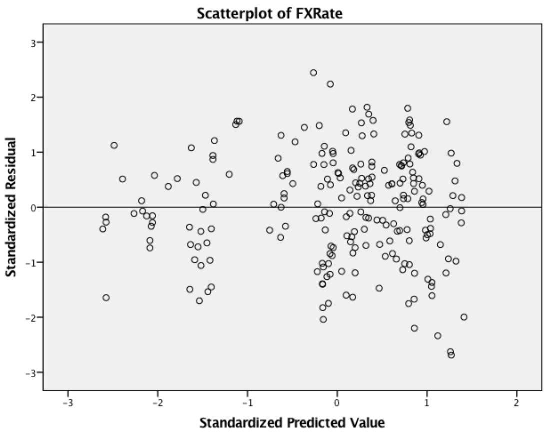
Is the stepwise model useful for predicting average USD/CAD FX rate?

H_0: $\beta_1 = \beta_2 = \ldots = \beta_n = 0$. H_a: at least one of the betas does not equal zero. The ANOVA table shows that $F(13,214) = 767.881$, $p < .001$. Since $p < .05$, we reject the null hypothesis. It appears that this model explains a significant amount of the variability in FX rate.





David Modjeska
Dominic Murphy



The stepwise model appears to be useful. Does it meet the assumptions for multiple regression?

- **Randomness**: The data for the model were chosen as a random 10% sample of the population years included in our time-series data. So the randomness assumption is met.
- **Independence**: Because of the random sampling mentioned above, each data point in the time series is unlikely to have strong impact on other points. So the independence assumption is met.
- **Zero mean**: This assumption is automatically met by the least-squares regression method.
- **Constant variance**: The scatterplot of predicted values vs. residuals shows no obvious pattern. So the assumption of constant variance is met.
- **Normality**: The histogram of residual distribution is approximately normal. Also, the NPP plot shows that the data points are close to the line. So the assumption of normality for residuals is met.

A few of the points in the scatterplot have standardized residuals near ± 2.5. Still, these points fit with the pattern of the rest of the data and seem to be well described by the model. So we need not classify these points as outliers.

On the basis of the preceding tests, we can say that all of the assumptions for multiple regression are met. Unfortunately for the usefulness of the stepwise model, the collinearity of predictors is high, as expected. The coefficients table above shows approximately 80% of the predictors in the model have a variance inflation factor (VIF) above the heuristic threshold of 5. Given this high level of collinearity, we should seek a simpler model.

c. Simple Oil Model

Our first hypothesis concerns the centrality of oil in the Canadian economy, so a key factor in the USD/CAD FX rate. As a basis for modeling, let's choose and fit a model using oil spot price. The fitted model and SPSS output are below.

$$\widehat{FXRate} = 0.686 + 0.003*OilPrice$$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.816[a]	.666	.666	.046993030

a. Predictors: (Constant), OilPrice
b. Dependent Variable: FXRate

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15.823	1	15.823	7165.191	.000[b]
	Residual	7.937	3594	.002		
	Total	23.760	3595			

a. Dependent Variable: FXRate
b. Predictors: (Constant), OilPrice

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.686	.003		233.429	.000
	OilPrice	.003	.000	.816	84.647	.000

a. Dependent Variable: FXRate

This simple oil model appears to be a promising start for the exploration process. The model summary shows that the adjusted R-squared value is .666, so the model explains a fair percentage of the variability in USD/CAD FX rate, after accounting for the number of predictors and the sample size. The standard error of the estimate is approximately 0.05 USD/CAD, which indicates fair accuracy in predicting the average FX rate.

Does *OilPrice* predict a significant amount of the variability in USD/CAD FX rate?

> H_0: $\beta_1 = 0$. H_a: $\beta_1 \neq 0$. The Coefficients table shows that t = 84.647, p < .001. Since p < .05, we reject the null hypothesis. It appears that *OilPrice* explains a significant amount of the variability in FX rate. For each increase of 1 USD/barrel, the USD/CAD FX rate increases on average by $0.003.

This simple oil model is less useful than the more complex stepwise model. In particular, the adjusted R-squared value for the simple oil model is only .666, while the value for the stepwise model is .978. Similarly, the standard error of the estimate for the simple oil model is 0.047 USD/CAD, while the error is 0.012 USD/CAD for the stepwise model. Let's explore which additional predictors might be useful to add to the simple oil model, while keeping VIF low.

d. Nested F-tests

To explore adding predictors to the simple oil model, we can proceed with a series of nested F-tests using different variable subsets. The goal of using these tests is to reduce the odds of a Type I error in testing many potential predictors individually. For efficiency, the first F-test will be described in detail, while the later F-tests will be summarized.

As a group, the exchange rates between the USD and key non-CAD currencies look interesting. Fitting a new model that adds these predictors to the simple oil model produces the SPSS output below.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	22.033	8	2.754	5720.359	.000[b]
	Residual	1.727	3587	.000		
	Total	23.760	3595			

a. Dependent Variable: FXRate
b. Predictors: (Constant), FX_USD_Per_Basket, FX_USD_Per_AUD, FX_USD_Per_JPY, FX_USD_Per_CHF, OilPrice, FX_USD_Per_GBP, FX_USD_Per_SEK, FX_USD_Per_EUR

Does this new model explain significantly more variability in USD/CAD FX rate than the simple oil model?

H_0: $\beta_2 = \beta_3 = \ldots = \beta_8 = 0$. H_a: at least one of these betas does not equal zero.

$F = ((SSE_{Reduced} - SSE_{Full}) / \#\text{ predictors tested}) / (SSE_{Full} / n - k - 1)$

$= ((7937 - 1727) / 7) / (1727 / 3587) = 1842.607$

Using the upper tail of an F-distribution, the p-value for $F(7,3587)$ is less than .001 .

Since $p < .05$, we reject the null hypothesis. It appears that the full model explains significantly more variability in USD/CAD FX rate than does the simple model.

Using the testing process above, similar positive results were obtained for the mining variables, the Canadian government statistics, the oil futures variables, quadratic terms, and interaction terms.

e. Exploration of predictors

On the basis of the simple oil model and nested F-tests above, let's explore which predictors in the variable subsets add significant explanatory power, without introducing excessive collinearity or violating regression assumptions. The outcome of these explorations will be summarized for space purposes, and one interesting result will be analyzed in detail. (The final model will be described, tested, and verified in the following section of this report.)

Of the variable subsets, it was found that that currency variables introduced excessive collinearity, since their impact is already reflected elsewhere in the model. Among the government statistics, *GDP_CA* was identified as the most useful. Among the mining variables, *GoldPrice* was useful. However, *MetalOre* would introduce excessive collinearity. Finally, among the oil futures variables, *Oil_Futures_Open* turned out to be useful enough to justify replacing *OilPrice* in the simple oil model. Let's consider this replacement in detail. The fitted model and SPSS output are below.

$\widehat{FXRate} = 0.604 + 0.004 * Oil\_Future\_Open$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.834[a]	.695	.695	.044884015

a. Predictors: (Constant), Oil_Future_Open
b. Dependent Variable: FXRate

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.604	.004		166.816	.000
	Oil_Future_Open	.004	.000	.834	90.554	.000

a. Dependent Variable: FXRate

Does *Oil_Future_Open* predict a significant amount of the variability in USD/CAD FX rate?

> H_0: $\beta_1 = 0$. H_a: $\beta_1 \neq 0$. The coefficients table shows that t = 90.554, p < .001. Since p < .05, we reject the null hypothesis. It appears that *Oil_Future_Open* explains a significant amount of the variability in FX rate. For each increase of 1 USD/barrel, the USD/CAD FX rate increases on average by $0.004.

The new simple oil model is incrementally better than the old simple oil model. In particular, the adjusted R-squared value for the old simple oil model was approximately 67%, while the value for the new simple oil model is approximately 70%. Similarly, the standard error of the estimate for the new simple oil model is approximately 0.04 USD/CAD, while the error was approximately 0.05 USD/CAD for the old simple oil model.

To conclude this section, investigation of quadratic terms for predictors after our exploration showed that *Oil_Future_Open_Squared* is a useful addition to the model. Investigation of interaction terms did not find any useful additions to the model.

f. Final Model with Verification

On the basis of the new simple oil model, nested F-tests, and exploration of individual predictors, it seems we have built a viable model. Let's test significance and verify assumptions for linear regression. The fitted model and SPSS output are below.

$\widehat{FXRate} = 0.625 + 0.010*Oil\_Future\_Open - 3.831*10^{-5}*Oil\_Future\_Open^2 + 1.16*10^{-4}*GoldPrice - 0.025*GDP\_CA$

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.962[a]	.926	.925	.022724127

a. Predictors: (Constant), GDP_CA, Oil_Future_Open_Sq, GoldPrice, Oil_Future_Open
b. Dependent Variable: FXRate

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1.735	4	.434	840.185	.000[b]
	Residual	.138	267	.001		
	Total	1.873	271			

a. Dependent Variable: FXRate
b. Predictors: (Constant), GDP_CA, Oil_Future_Open_Sq, GoldPrice, Oil_Future_Open

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.625	.038		16.328	.000		
	Oil_Future_Open	.010	.001	2.012	18.241	.000	.023	44.138
	Oil_Future_Open_Sq	-3.831E-5	.000	-1.346	-12.571	.000	.024	41.574
	GoldPrice	1.16E-4	.000	.375	17.539	.000	.604	1.657
	GDP_CA	-.025	.002	-.251	-12.604	.000	.694	1.440

a. Dependent Variable: FXRate

The final model is a balance between the complexity of the stepwise model, and the simplicity of the old simple oil model. The final model summary shows an adjusted R-squared value of .925, which indicates that the model explains a high percentage of variability in USD/CAD FX rate, after accounting for the number of predictors and the sample size. The standard error of the estimate is approximately 0.02 USD/CAD, which indicates good accuracy in predicting the average FX rate.

Is the final model useful for predicting average USD/CAD FX rate?

H0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. Ha: at least one of the betas does not equal zero. The ANOVA table shows that $F(4,267) = 840.185$, $p < .001$. Since $p < .05$, we reject the null hypothesis. It appears that this model explains a significant amount of the variability in USD/CAD FX rate.

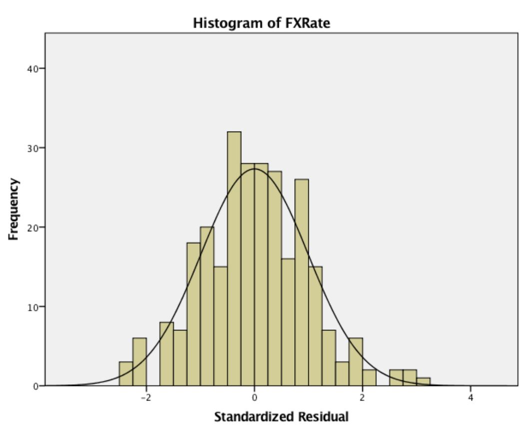
Does each predictor explain a significant amount of the variability in USD/CAD FX rate?

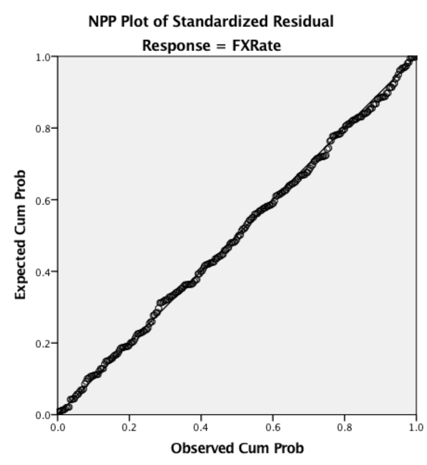
- **Oil_Future_Open:** H_0: $\beta_1 = 0$. H_a: $\beta_1 \neq 0$. The coefficients table shows that $t = 18.241$, $p < .001$. Since $p < .05$, we reject the null hypothesis that $\beta_1 = 0$.

- **Oil_Future_Open_Squared:** H_0: $\beta_2 = 0$. H_a: $\beta_2 \neq 0$. The coefficients table shows that $t = -12.571$, $p < .001$. Since $p < .05$, we reject the null hypothesis $\beta_2 = 0$.
- **GoldPrice:** H_0: $\beta_3 = 0$. H_a: $\beta_3 \neq 0$. The coefficients table shows that $t = 17.539$, $p < .001$. Since $p < .05$, we reject the null hypothesis $\beta_3 = 0$.
- **GDP_CA:** H_0: $\beta_4 = 0$. H_a: $\beta_4 \neq 0$. The coefficients table shows that $t = -12.604$, $p < .001$. Since $p < .05$, we reject the null hypothesis $\beta_4 = 0$.

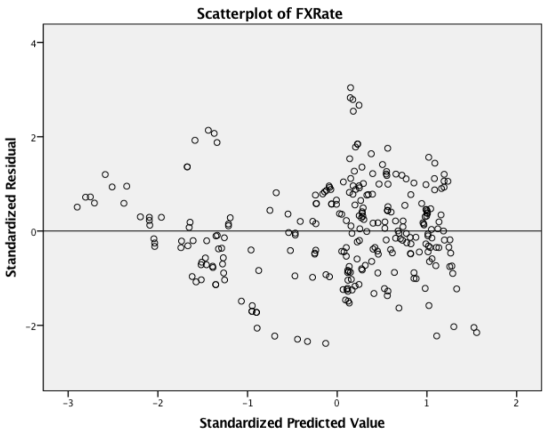
It appears that each of the predictors explains a significant amount of the variability in USD/CAD FX rate.

- **Oil_Future_Open** and **Oil_Future_Open_Squared:**
  - For each increase of 1 USD/barrel in the oil futures open price, the average USD/CAD FX rate increases by $0.010 - 3.831 \times 10^{-5} \times$ *Oil_Future_Open*, holding non-oil predictors in the model constant.
  - As a baseline, when *Oil_Future_Open* equals zero (hypothetically): for each increase of 1 USD/barrel in the oil futures open price, the average USD/CAD FX rate increases by 0.010, holding non-oil predictors in the model constant.
  - Since the linear term's beta is positive and the quadratic term's beta is negative, as *Oil_Future_Open* increases, the average USD/CAD FX rate generally increases at a decreasing rate (within the range of our oil futures data), holding non-oil predictors in the model constant.
- **GoldPrice:** For each increase of 1 USD/ounce in gold price, the average USD/CAD FX rate increases by 1.16×10^{-4}, holding other predictors in the model constant.
- **GDP_CA:** For each increase of one billion dollars in Canadian GDP, the average USD/CAD FX rate decreases by -.025, holding other predictors in the model constant.





David Modjeska
Dominic Murphy



The final model seems useful, and each predictor is significant. Does the model meet assumptions for multiple regression?

- **Randomness**: The data for the model were chosen as a random 10% sample of the population years included in this time-series data. So the randomness assumption is met.
- **Independence**: Because of the random sampling mentioned above, each data point in the time series is unlikely to have strong impact on other points. So the independence assumption is met.
- **Zero mean**: This assumption is automatically met by the least-squares regression method.
- **Constant variance**: The scatterplot of predicted values vs. residuals shows no obvious pattern. So the assumption of constant variance is met.
- **Normality**: The histogram of residual distribution is approximately normal. Also, the NPP plot shows that the data points are close to the line. So the assumption of normality for residuals is met.

A few of the points in the scatterplot have standardized residuals near 3. Still, these points fit with the pattern of the rest of the data and seem to be well described by the model. So we need not classify these points as outliers.

On the basis of the preceding tests, we can say that all of the assumptions for multiple regression are met.
In passing, we should consider collinearity in the final model. The coefficients table shows that each predictor has VIF below the heuristic threshold of 5, apart from the usual collinearity between the first- and second-order terms for the oil futures open price. So the VIF level is acceptable in this model.

g. Investigating sampling variability with bootstrapping

To better understand the utility of the final model, let's investigate sampling variability via bootstrapping. This process was done with 5000 samples, BCa confidence intervals at the 95% level, and simple sampling. The SPSS output is below (condensed to show only the predicted value, for efficiency).

Residuals Statistics[a]

			Bootstrap[b]			
					BCa 95% Confidence Interval	
		Statistic	Bias	Std. Error	Lower	Upper
Predicted Value	Minimum	.69404846				
	Maximum	1.05041897				
	Mean	.92603570	.00004747	.00501683	.91597458	.93609421
	Std. Deviation	.080023979	-.000230597	.003217184	.073746352	.085615544
	N	272	0	16	243	303

a. Dependent Variable: FXRate

b. Unless otherwise noted, bootstrap results are based on 5000 bootstrap samples

Bootstrapping results agree with the final model, with a mean R-squared value of .926. The bias of $4.7*10^{-4}$ is small. We have 95% confidence that this model explains between 91.6% and 93.6% of the variability in USD/CAD FX rate.

5. Conclusion

This study has validated our three hypotheses: that each of oil price, gold price, and Canadian GDP will explain a significant amount of the variability in exchange rate between the US and Canadian dollars, after accounting for other variables in the model. The result is a multiple regression model with three predictors: oil futures open price (with linear and quadratic terms), gold spot price, and Canadian GDP. In addition to each of these predictors being useful individually, the overall model explains approximately 93% of the variability in USD/CAD FX rate.

As expected, this study validated the public view of Canada's as a resource-based economy, particularly focused on oil. In our analysis, a simple linear regression model with a single predictor – oil price – accounted for approximately 67% of the variability in USD/CAD FX rate. Despite the importance of oil as a predictor, however, other factors strongly affect the USD/CAD FX rate, as shown by our final model.

Several surprises arose during this analysis:

1. The oil futures open price turned out to be a more powerful predictor than the oil spot price. It appears that traders embed additional knowledge about market conditions in futures pricing, relative to spot pricing.

2. The collinearity of possible predictors was very high. This finding gives an impression of a financial market with substantial, overlapping, and mutually reinforcing information. The anecdotal volatility and complexity of these markets seems to be reflected in the data available for analysis.
3. Relative to the large number of predictors available, a small model explained much of the variability in USD/CAD FX rate. The final model contained only three predictors (one with linear and quadratic terms).

In an analysis of this size, problems naturally arose. For the most part, we were able to address these problems effectively. Opportunities remain for detailed follow-up and future analysis.

1. The source data contained mixed frequencies, both daily and quarterly. FX rates and commodity prices are available daily, while government statistics such as GDP are released monthly or quarterly. A simple fill of GDP data over each period was found useful in our modeling. (Interpolation matches the situational knowledge of traders less well, so that approach was avoided here.) In general, we would expect the FX market's reaction to fresh GDP data to be greatest early in a release cycle, while shrinking in the lag days after a release. The interaction of GDP and lag days expressed as a beta polynomial weighting function could be used to model the decaying influence of GDP between releases (Armesto, Engemann, & Owyang, 2010).

2. The number of potential predictors available for analyzing foreign exchange rates is large. Sources include industry, markets, and governments. Techniques exist for winnowing down a potentially large predictor set, some of which were used in our modeling. Such techniques include correlation matrices, scatterplot matrices, stepwise regression, and nested F-tests. None of these techniques proved sufficient in isolation, but the combination was able to direct our predictor exploration toward productive avenues.

If approaching this analysis again, we would probably make a few modifications to our approach. First, we would consider a broader range of commodity indices, both energy and otherwise. Second, we would explore techniques for reconciling data with different frequencies, possibly embedding a specialized technique inside a more general regression approach. Finally, we would survey the range of available predictors more systematically, in order to assemble as much of the data set as possible early in the analysis process, to avoid redundant "data wrangling".

There are promising avenues for future research. First, opportunities exist to consider new time relationships among predictors and foreign exchange rates. Opportunities include validating prediction on daily, monthly, and quarterly frequencies with rolling windows. In a similar vein, seasonality may have an impact on FX rates. Second, other national economies have characteristics broadly similar to Canada's. Two countries of interest are Australia and Sweden. Each has an economy with a substantial resources component, as well as advanced industrial and service sectors. This results in reliable, plentiful, and public data for research.

David Modjeska
Dominic Murphy

6. Bibliography

a. References

1. Cofnas, A. (2011). Trading Binary Options: Strategies and Tactics. In A. Cofnas, *Trading Binary Options: Strategies and Tactics* (p. 60). New York: Bloomberg Press.

2. Gilroy, A. (2014, 09 22). *Understanding Gold Price Drivers*. Retrieved from Market Realist: http://marketrealist.com/2014/09/must-know-guide-investing-in-gold/

3. Lafrance, R., Zhang, Q., Issa R, R., & Helliwell, J. F. (2010). *NEMO: An Equation for the Canadian Dollar.* Ottawa, Ontario: Bank of Canada.

b. Data Sources

FX Rates	US Federal Reserve	https://www.federalreserve.gov/releases/h10/hist/
CAN FX Futures	Quandl	https://www.quandl.com/data/CHRIS/CME_CD2
US Federal Debt	US Treasury	http://www.treasurydirect.gov/NP/debt/search?startMonth=01&startDay=01&startYear=2006&endMonth=01&endDay=01&endYear=2016
US GDP	US Bureau of Economic Analysis	http://www.bea.gov/national/pdf/nipaguid.pdf
US Interest Rates	Federal Reserve Bank of St. Louis	https://research.stlouisfed.org/fred2/series/USDONTD156N
CAN Federal Debt	Statistics Canada	http://www.statcan.gc.ca/
CAD GDP	Statistics Canada	http://www.statcan.gc.ca/
US Interest Rates	Federal Reserve Bank of St. Louis	https://research.stlouisfed.org
Oil Spot Price	US Energy Information Administration	https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm
Oil Futures Price	US Energy Information Administration	https://www.eia.gov/dnav/pet/pet_pri_fut_s1_d.htm
Gold Spot Price	Quandl	https://www.quandl.com/collections/markets/gold
Gold Futures Price	Quandl	www.quandl.com

Appendix 1: Data Snapshot (first 5 rows)

Columns 1-9

Date	FXRate	OilPrice	GoldPrice	CADInterestON	USDInterestON	USDInterest3M	CADInterest3M	GDPOilAndGas
14-Apr-06	0.870322	69.53	593	3.7218	2.41063	4.91	3.95167	86331
15-Apr-06	0.870322	69.53	593	3.7218	2.41063	4.91	3.95167	86331
16-Apr-06	0.870322	69.53	593	3.7218	2.41063	4.91	3.95167	86331
17-Apr-06	0.870322	70.3	593	3.7218	2.40094	4.91	3.95167	86331
18-Apr-06	0.873134	71.28	614.75	3.7283	4.80188	4.9225	3.96167	86331

Columns 10-15

OilAndGasPercentOfGDP	CA_GDP_Growth	US_GDP_Growth	FX_USD_Per_AUD	FX_USD_Per_EUR	FX_USD_Per_GBP
NA	NA	6.36422	0.7289	1.2106	1.751
NA	NA	6.36422	0.7289	1.2106	1.751
NA	NA	6.36422	0.7289	1.2106	1.751
NA	NA	6.36422	0.7375	1.2267	1.7716
NA	NA	6.36422	0.7385	1.2274	1.7762

Columns 16-22

FX_USD_Per_JPY	FX_USD_Per_SEK	FX_USD_Per_CHF	FX_USD_Per_Basket	Oil_Future_Open	Oil_Future_High	Oil_Future_Low
118.65	7.6934	1.299	2.75567151	71.15	72.29	71.15
118.65	7.6934	1.299	2.75567151	71.15	72.29	71.15
118.65	7.6934	1.299	2.75567151	71.15	72.29	71.15
117.68	7.579	1.2776	2.775929688	73.15	73.15	73.15
117.75	7.5842	1.2759	2.778064904	73.77	73.77	73.77

Columns 23-29

Oil_Future_Last	Oil_Future_Settle	Oil_Future_Volume	Oil_Future_Open_Int	PriceDeflator_CA	PriceDeflator_US	PriceDeflator_Di
72.29	72.29	40	6048	95.918	94.587	1.3
72.29	72.29	40	6048	95.918	94.587	1.3
72.29	72.29	40	6048	95.918	94.587	1.3
73.15	73.15	40	6048	95.918	94.587	1.3
73.77	73.77	40	6048	95.918	94.587	1.3

Columns 30-36

GDP_US	GovDebt_US	Debt_GDP_Ratio_US	GDP_CA	GovDebt_CA	Debt_GDP_Ratio_CA	MetalOre
13800	8407	61	NA	NA	NA	17958
13800	8407	61	NA	NA	NA	17958
13800	8407	61	NA	NA	NA	17958
13800	8367	61	NA	NA	NA	17958
13800	8374	61	NA	NA	NA	17958

Appendix 2: Bivariate Pearson Correlations with *FX Rate*

		FXRate	OilPrice	GoldPrice	CADInterestON	USDInterestON
FXRate	Pearson Correlation	1	.816**	.415**	.088**	-.005
	Sig. (2-tailed)		.000	.000	.000	.782
	N	3596	3596	3596	3596	3596

USDInterest3M	CADInterest3M	GDPOilAndGas	OilAndGasPercentOfGDP	CA_GDP_Growth	US_GDP_Growth
-.024	.022	-.422**	-.516**	.341**	.124**
.150	.187	.000	.000	.000	.000
3596	3596	3596	2969	2969	3596

FX_USD_Per_AUD	FX_USD_Per_EUR	FX_USD_Per_GBP	FX_USD_Per_JPY	FX_USD_Per_SEK	FX_USD_Per_CHF
.868**	.675**	.249**	-.639**	-.860**	-.268**
.000	.000	.000	.000	.000	.000
3596	3596	3596	3596	3596	3596

FX_USD_Per_Basket	Oil_Future_Open	Oil_Future_High	Oil_Future_Low	Oil_Future_Last	Oil_Future_Settle
.355**	.834**	.832**	.836**	.835**	.835**
.000	.000	.000	.000	.000	.000
3596	3596	3596	3596	3596	3596

Oil_Future_Volume	Oil_Future_Open_Int	PriceDeflator_CA	PriceDeflator_US	PriceDeflator_Diff	GDP_US
.035*	.085**	-.039*	-.238**	.530**	-.320**
.036	.000	.019	.000	.000	.000
3596	3596	3596	3596	3596	3596

GovDebt_US	Debt_GDP_Ratio_US	GDP_CA	GovDebt_CA	Debt_GDP_Ratio_CA	MetalOre
-.176**	-.063**	-.346**	-.708**	-.714**	-.503**
.000	.000	.000	.000	.000	.000
3596	3596	2969	2513	2513	3596

Appendix 3: Description of Variables

Variable	**Short Description**	**Example**	**In Final Model**
Date	Date (Daily)	March 3, 2016	
FXRate	USD/CAD exchange rate	0.747887	Y
OilPrice	Crude Oil Spot Price: West Texas Intermediate (WTI) - Dollars per Barrel.	37.9	Y
Oil Future Open	Daily Opening Oil Futures Forward Price in USD	44.15	Y
Oil Future Open Square	Daily Opening Oil Futures Forward Price in USD squared	1,949.22	Y
GoldPrice	Gold Spot Price USD per Ounce, seasonal adjusted	1,232	Y
GDP CA	Canadian Gross domestic product, quarterly (dollars x 1,000,000)	1,649	Y
CADInterestON	Canadian Overnight Interbank Interest Rate	0.5013	
USDInterestON	US Overnight Interbank Interest Rate	0.372	
USDInterest3M	Canadian 3 Month Interbank Interest Rate	0.438	
CADInterest3M	US 3 Month Interbank Interest Rate	1.052	
GDPOilAndGas	Canadian Gross Domestic Product - Oil and Gas Sector	101,216	
OilAndGasPercentOfGDP	Canadian Oil and Gas Sector as a % of Total GDP	6.138	
CA GDP Growth	Canadian GDP Annualized Growth rate	0.088	
US GDP Growth	US GDP Annualised Growth rate	3.02284	
FX USD Per AUD	Spot FX Rate - US Dollar / Australian Dollar	0.7572	
FX USD Per EUR	Spot FX Rate - US Dollar / Euro	1.118	
FX USD Per GBP	Spot FX Rate - US Dollar / Great British Pound	1.4426	
FX USD Per JPY	Spot FX Rate - US Dollar / Japanese Yen	113.66	
FX USD Per SEK	Spot FX Rate - US Dollar / Swedish Krona	8.3339	
FX USD Per CHF	Spot FX Rate - US Dollar / Swiss Franc	0.9813	
FX USD Per Basket	USD FX Rate , Trade Weighted with (AUD,EUR,GBP,JPY,SEK,CHF)	2.519	
Oil Future High	Daily Maximum Oil Futures Forward Price in USD	44.74	
Oil Future Low	Daily Minimum Oil Futures Forward Price in USD	44.01	
Oil Future Last	Daily Oil Futures Last bid price USD	44.74	
Oil Future Settle	Settlement price calculated at the end of the trading day	45.14	
Oil Future Volume	Daily Volume Oil Futures Forward Price in USD	901	
Oil Future Open Int	Oil Futures Opening Interest - Total number of contracts long or short in the	5,002	
PriceDeflator CA	Canadian Gross Domestic Product: Implicit Price Deflator Index	109.817	
PriceDeflator US	US Gross Domestic Product: Implicit Price Deflator Index	110.29	
PriceDeflator Diff	US Price deflator Less Canadian Price Deflator	0.473	
GDP US	US Gross Domestic Product	18.148	
GovDebt US	US Federal debt - Total Outstanding Treasury Bonds (billions of dollars)	18,922	
Debt GDP Ratio US	US Federal debt / Gross Domestic Product, %	104	
GovDebt CA	Canada General Government Net Debt, quarterly (dollars x 1,000,000)	463	
Debt GDP Ratio CA	Canada General Government Net Debt, % of GDP	28	
MetalOre	Canadian Gross Domestic Product - Metal Ore Sector	21099	