Can ARIMA predict Currency Exchange Rates?

Euro to US Dollar

by Maren Beckman

Contents

Inspiration

“4 Ways to Forecast Exchange Rates” by Joseph Nguyen from investopedia.com

- No method has proven more successful
- Purchasing Power Parity - the economist’s choice, compares value of goods (McDonald’s Big Mac)
- Relative Economic Strength - supplements other methods with interest rate data
- Econometrics - incorporated various data estimated to impact currency value

Inspiration

continued

- Time series - uses only historical values to make predictions
- Most popular time series model is ARIMA
- This investigation focuses on ARIMA

What is ARIMA model?

The ARIMA model takes three parameters:

- "AR" for auto-regressive or how the data relates to its prior values
- "I" for integrated or the level of differencing required to make the data stationary
- "MA" for moving average or how the data trends

Data

Source

- Dataset collected from *US Federal Reserve Exchange Rates* at https://www.federalreserve.gov/data.htm
- Euros to US Dollars
- Euro introduced in January 2002
- Data through August 2018

Data

Exploration

	count	mean	std	min	25%	50%	75%	max
Rate	4189	1.25222	0.141234	0.8594	1.1534	1.2684	1.3494	1.601

Figure 1: Dollar to Euro Exchange Rate Over Time

Data

Exploration Continued

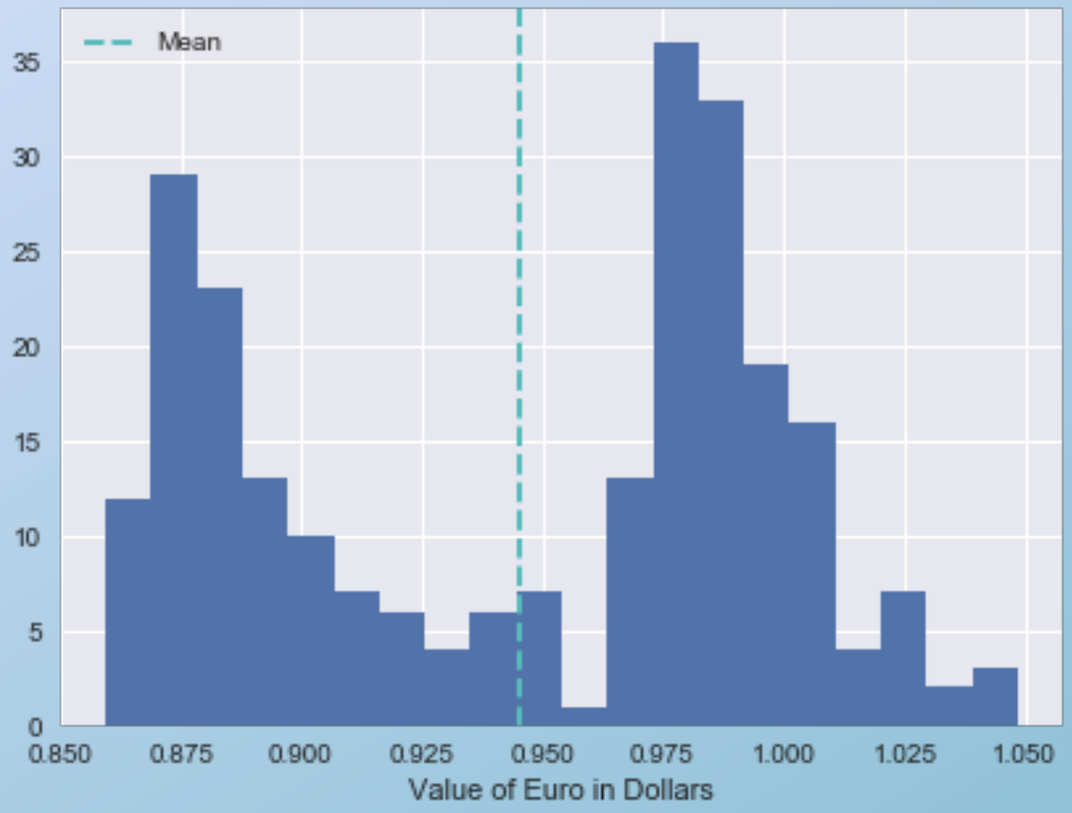
Euro value burn-in?

The lowest values are in the first year as the Euro gains confidence against US currency.

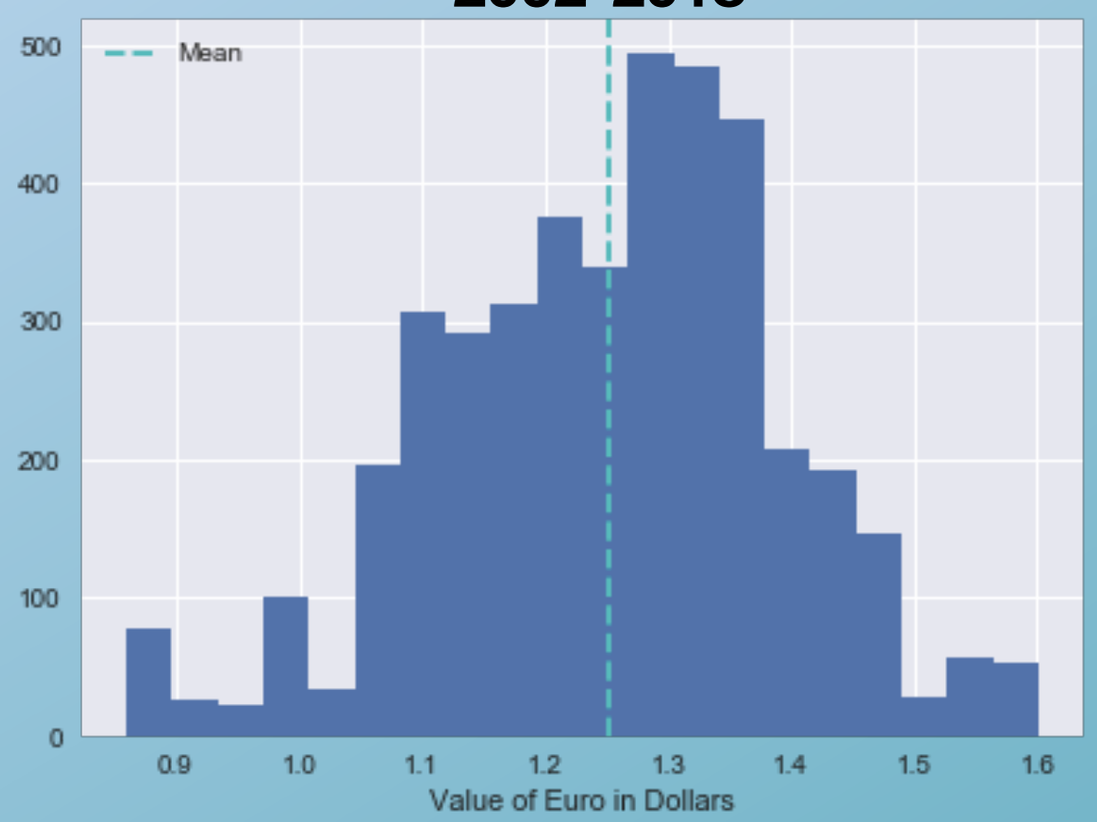
Data

Exploration Continued

2002



2002-2018

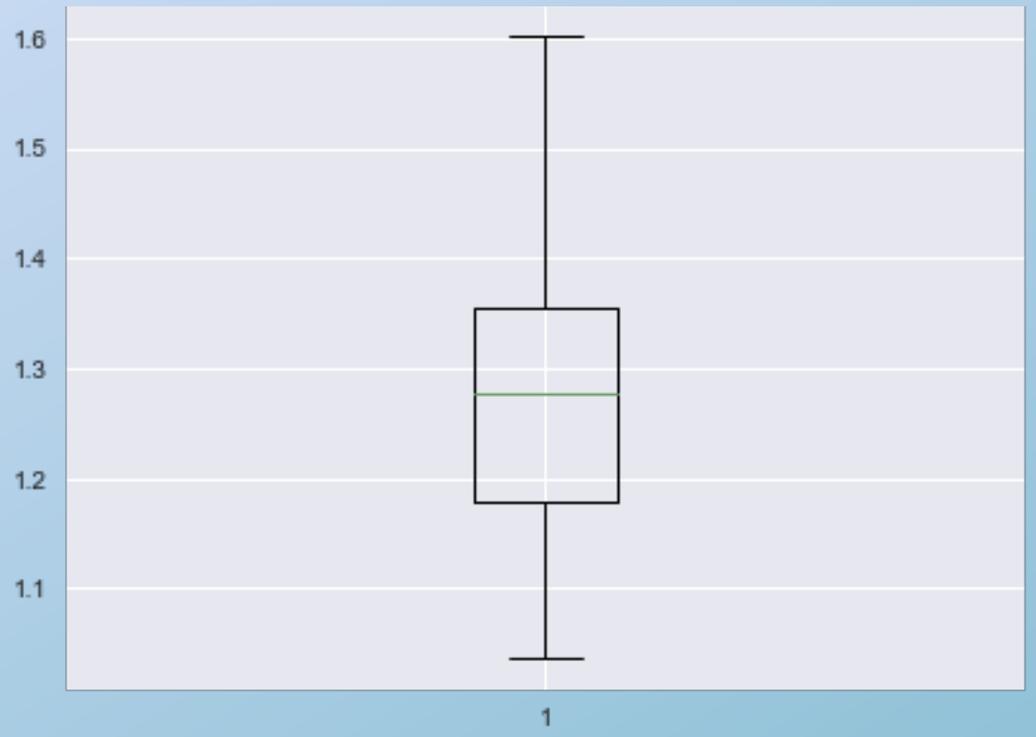


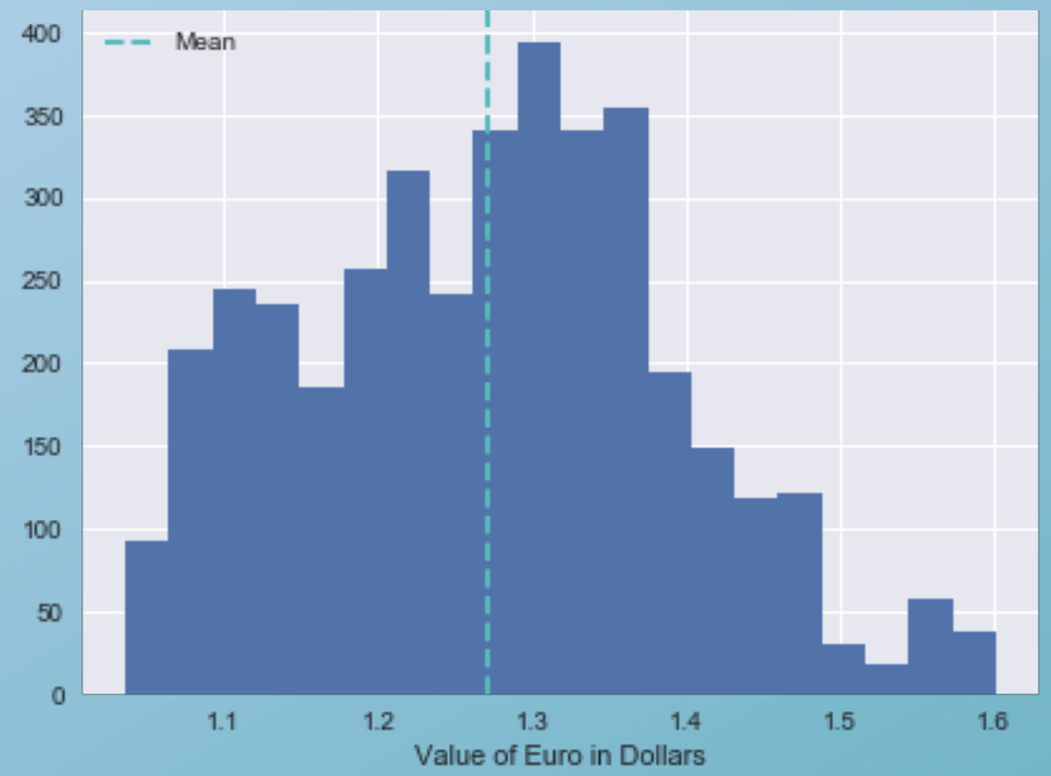
- 2002 values are inconsistent with the rest of the data
- We'll drop the first year of data to get a more established value of the exchange rate

Data

Exploration Continued

2003-2018





- Mean of 1.27
- Range from 1.04 to 1.60
- No outliers, we have normal distribution

Time Series Analysis

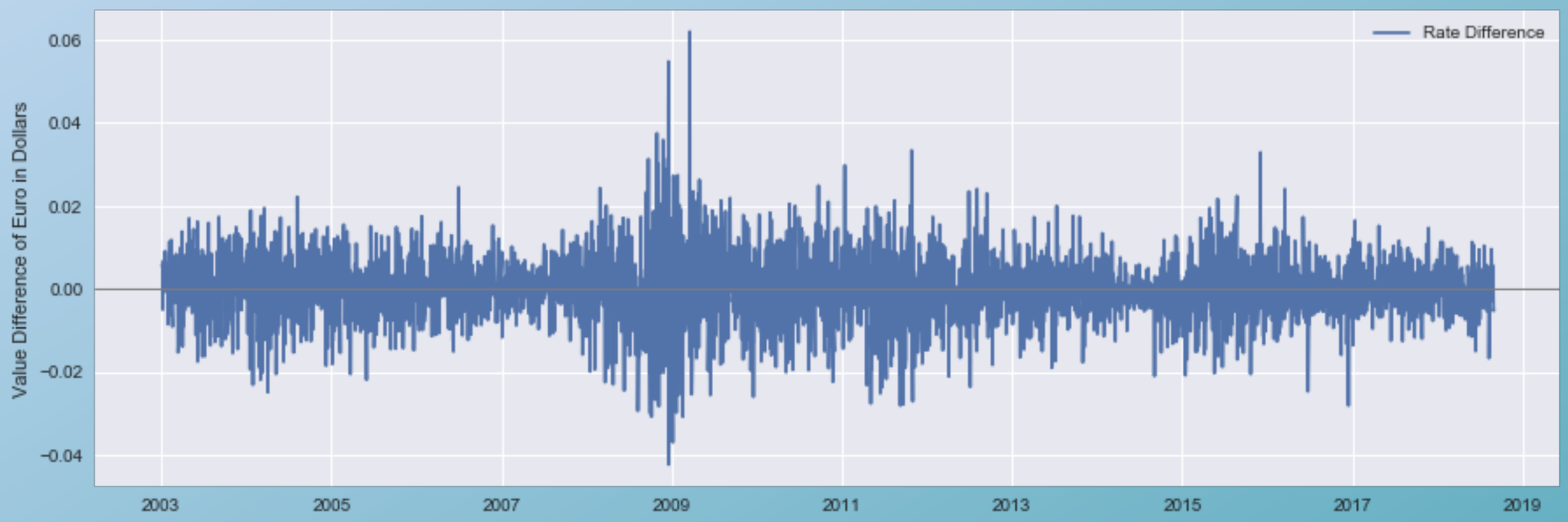
	count	mean	std	min	25%	50%	75%	max
Rate	3938.0	1.271777	0.121051	1.0361	1.17882	1.27795	1.35387	1.601

- Final data
- Euro is more valuable than Dollar
- Euro ranges from $1.04 to $1.60 (in 2008)

Analysis

Stationarity

- ARIMA needs stationary data
- Take the difference of each data point from the prior value
- Visualization of the results of **one degree** of differencing



- **Two degree** of differencing was also done

Stationarity

Dickey Fuller Test

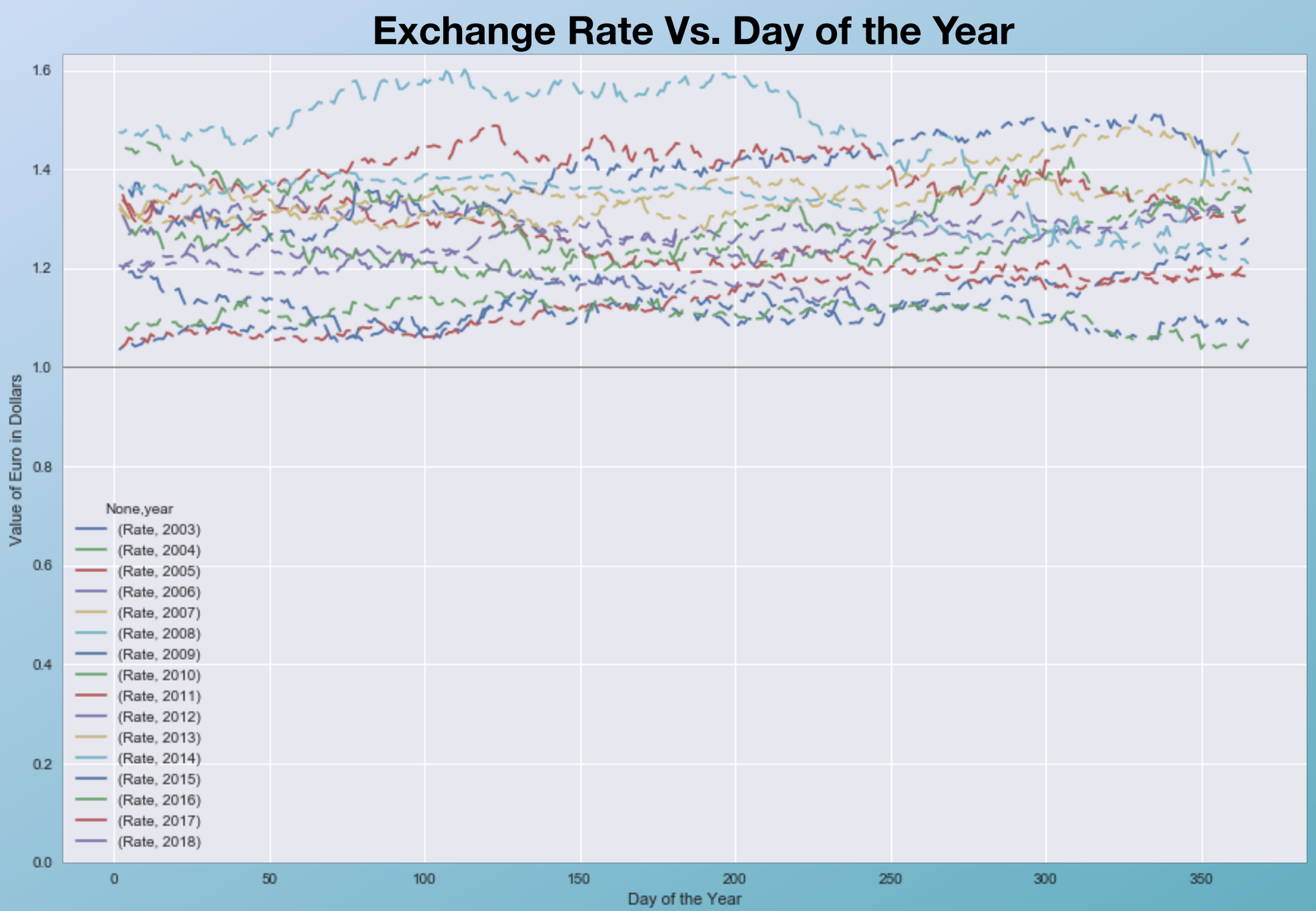
	Statistic	p-Value
Original Data	-2.403107	0.140861
First Difference	-62.348214	0.000000
Second Difference	-18.877696	0.000000

- Negative statistic indicates strength of stationarity
- Small p-value indicates significance of statistic
- First difference - lowest statistic and smallest p-value

This confirms that the first difference best establishes stationarity.

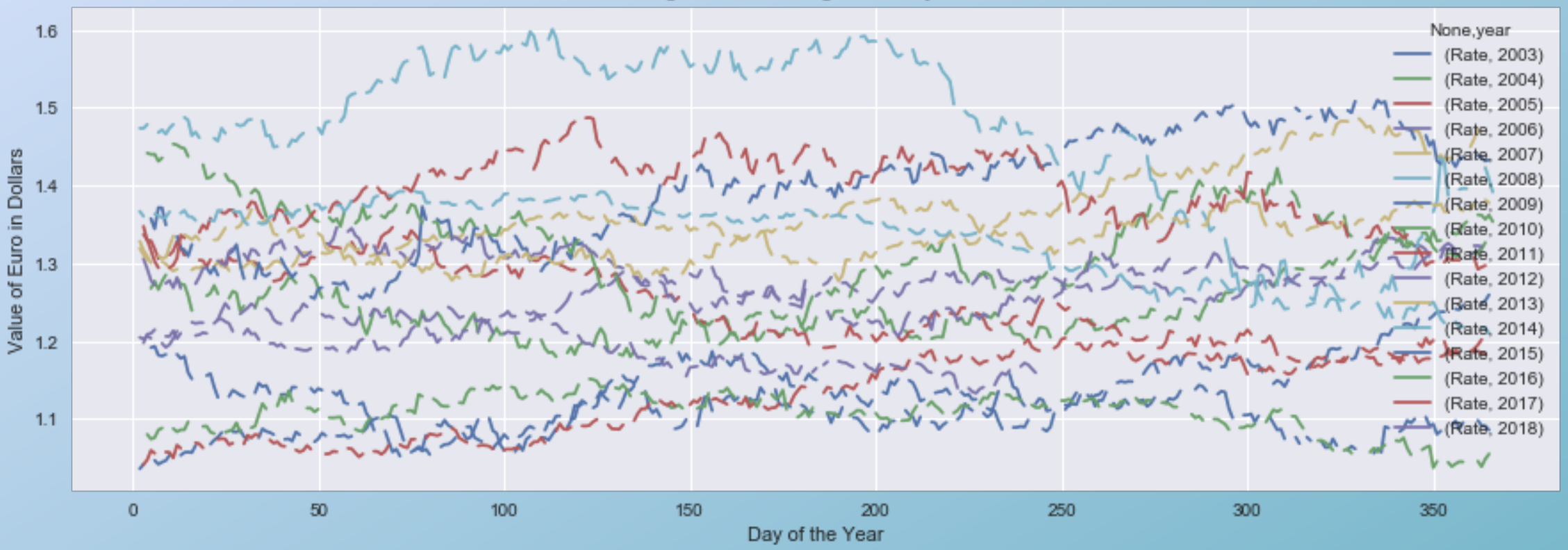
Analysis

Seasonality



Analysis

Seasonality continued

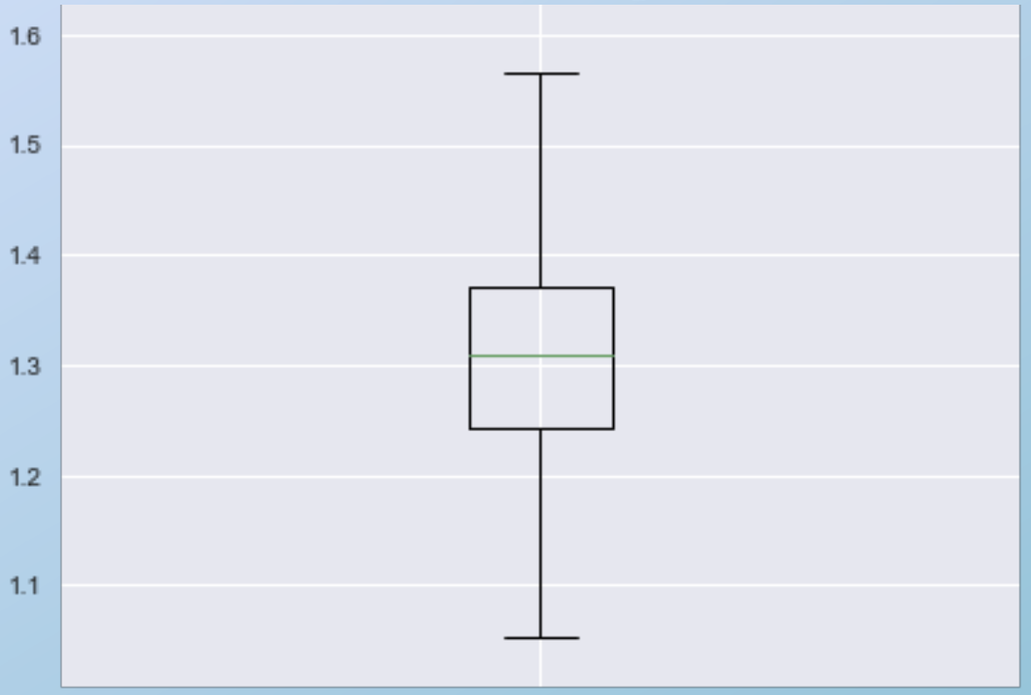


- No visible patterns - the records criss cross each other with no repetition - no seasonality
- Non-seasonal ARIMA model will best suit the data.

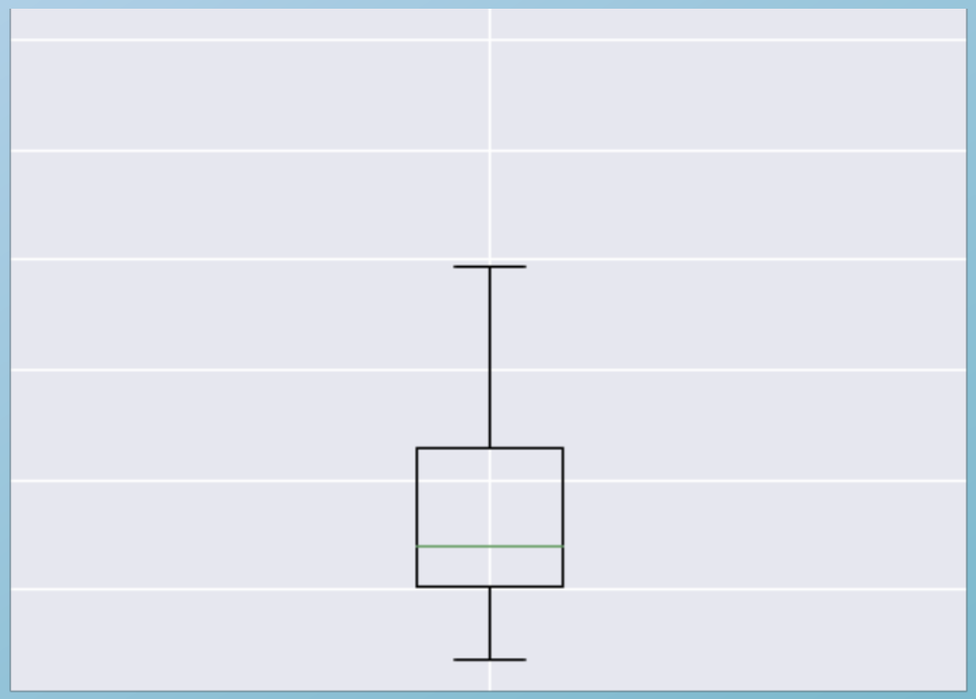
Model Preparation

70/30 Training/Test

Training



Test

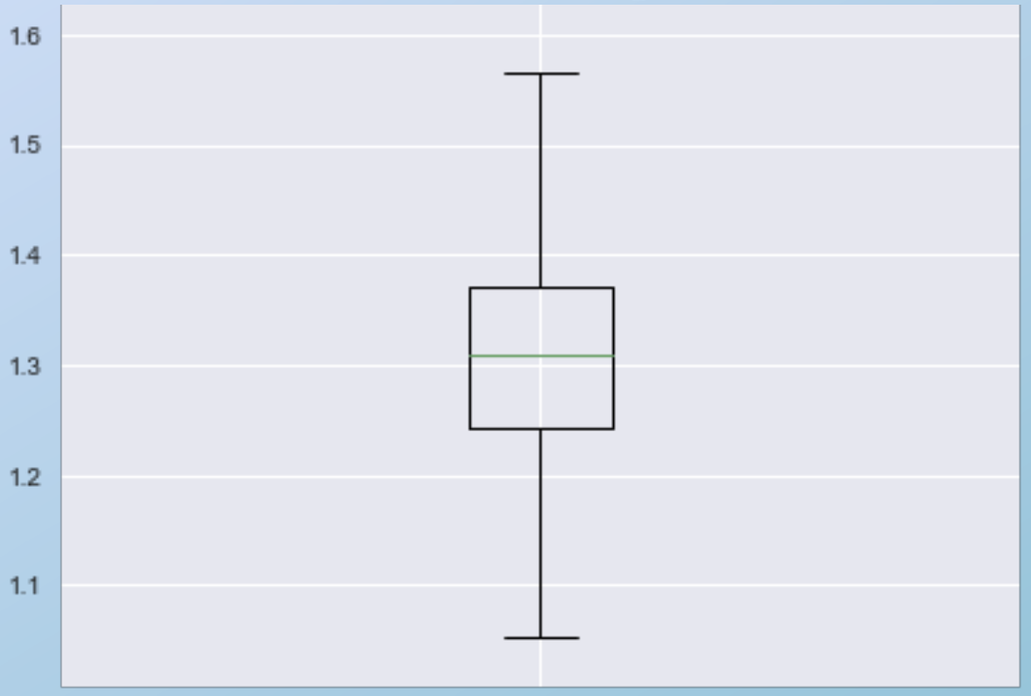


	Count	Start	End	Minimum	Maximum	Median
Training	2756	Jan 2003	Dec 2013	$1.04	$1.60	$1.31
Test	1182	Dec 2013	Aug 2018	$1.04	$1.39	$1.18

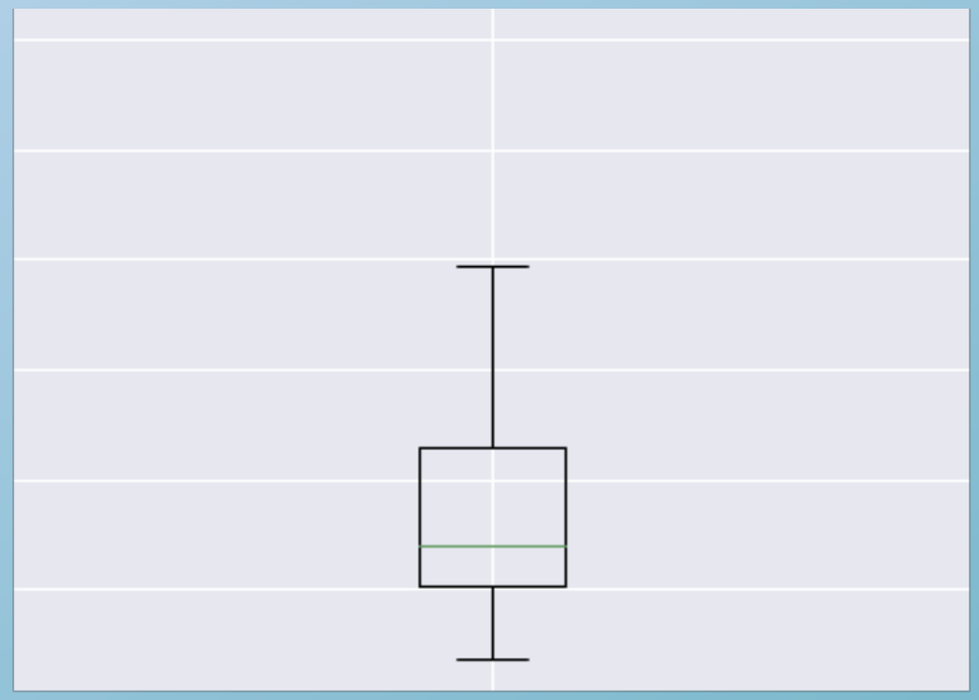
Model Preparation

70/30 Training/Test

Training



Test



	Count	Start	End	Minimum	Maximum	Median
Training	2756	Jan 2003	Dec 2013	$1.04	**$1.60**	**$1.31**
Test	1182	Dec 2013	Aug 2018	$1.04	$1.39	$1.18

Stronger Dollar value in test set

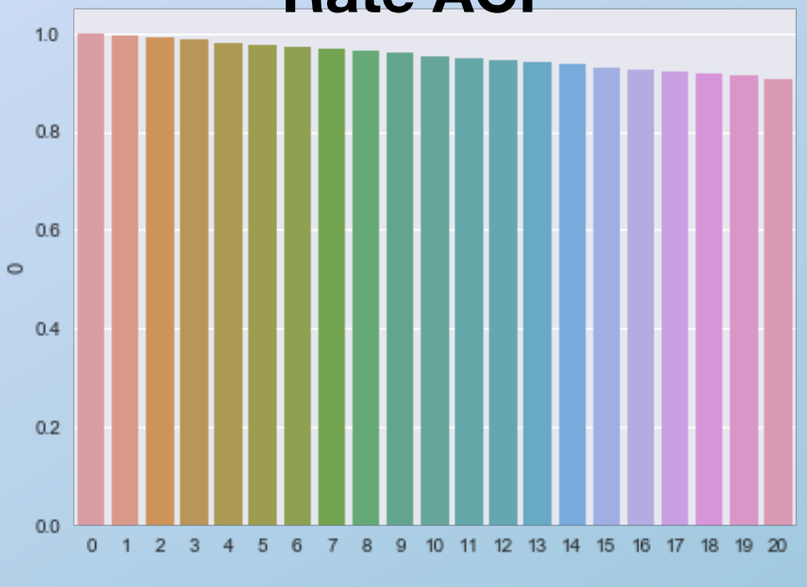
ARIMA Model

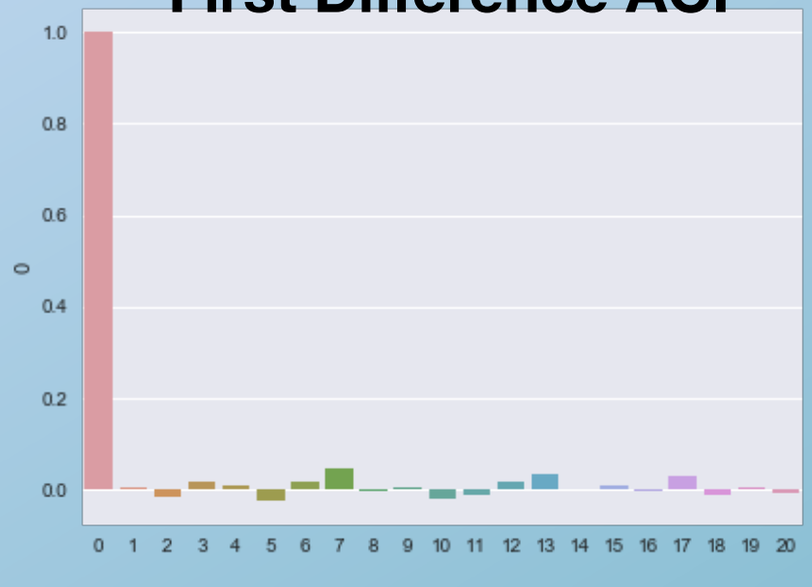
Estimating Parameters

ARIMA parameters to identify:

- Auto-Regressive (‘AR’) - AutoCorrelation Function (ACF) - exploring correlations with prior values
- Integrated (‘I’) - Stationarity
- Moving Average (‘MA’) - Partial AutoCorrelation Function (PACF) - exploring isolated correlations between individual values

ACF and PACF

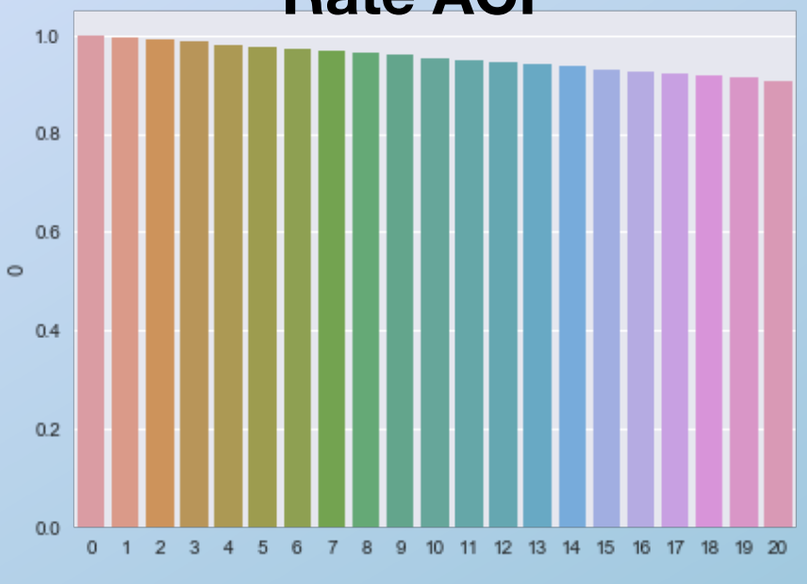




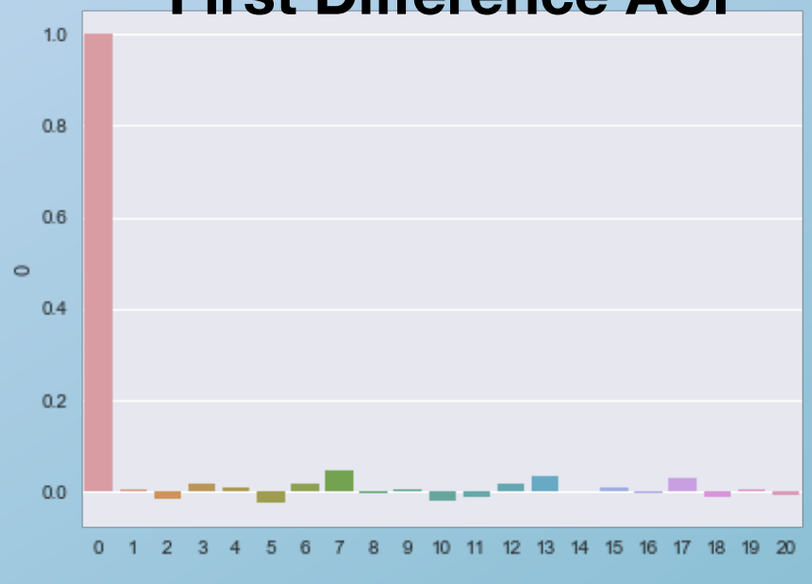
- Differenced data shows no correlation
- No clear AR parameter value

ACF and PACF

Rate ACF



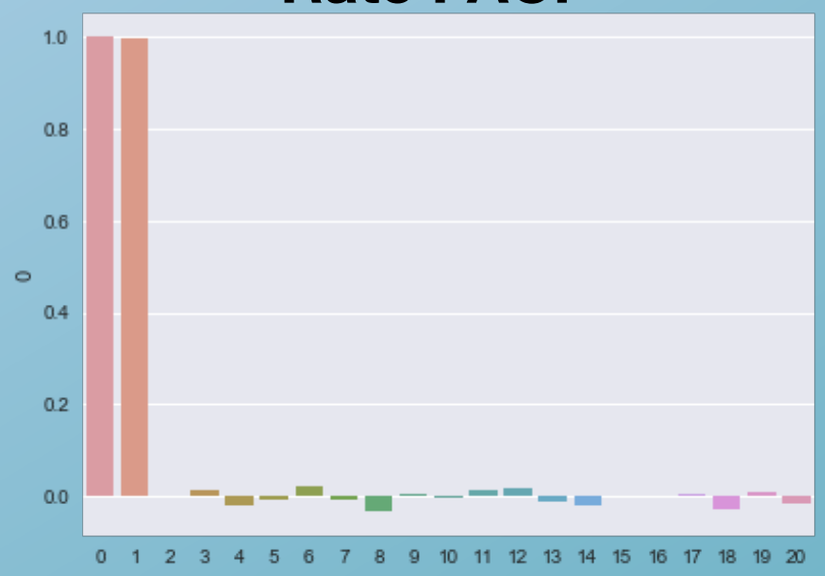
First Difference ACF



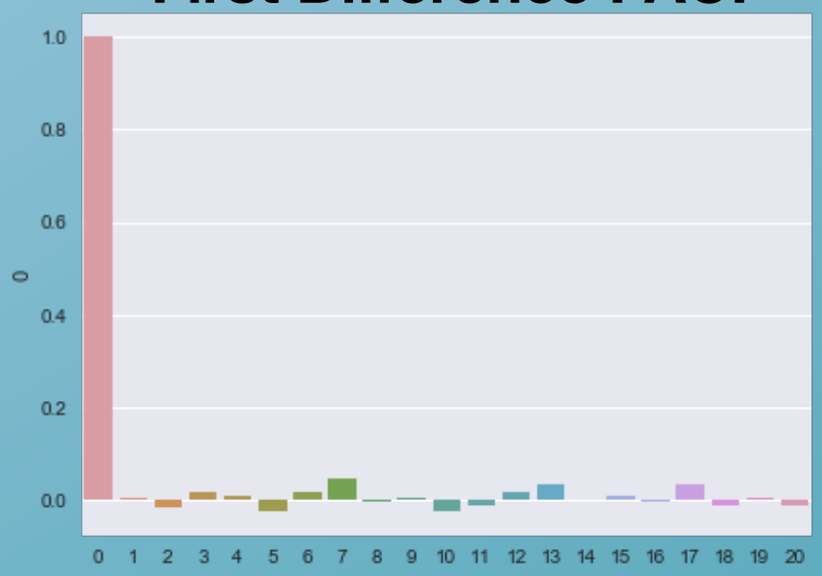
- Differenced data shows no correlation
- No clear AR parameter value

- Differenced data PACF shows no correlation
- No clear MA parameter value

Rate PACF



First Difference PACF



ARIMA Model

Testing Parameters - Auto Arima Grid Search

order			AIC	BIC
(0,	1,	0)	-18441.104	-18429.261
(1,	1,	0)	-18439.164	-18421.401
(0,	1,	1)	-18439.166	-18421.402
3,	1,	2)	-18437.215	-18395.767
(4,	1,	2)	-18435.327	-18387.957
(4,	1,	3)	-18433.663	-18380.372
(1,	2,	1)	-18409.738	-18386.055
(0,	2,	1)	-18410.152	-18392.389
(1,	2,	2)	-18423.269	-18393.664
(2,	2,	2)	-18414.372	-18378.848
(1,	2,	3)	-18404.859	-18369.334
(2,	2,	3)	-18399.697	-18358.251

AIC and BIC metrics - balancing complexity with data explanation

Each model tested captures the data with similar accuracy and parsimony.

Model Training & Prediction

Testing model efficacy

Train four models using different parameters:

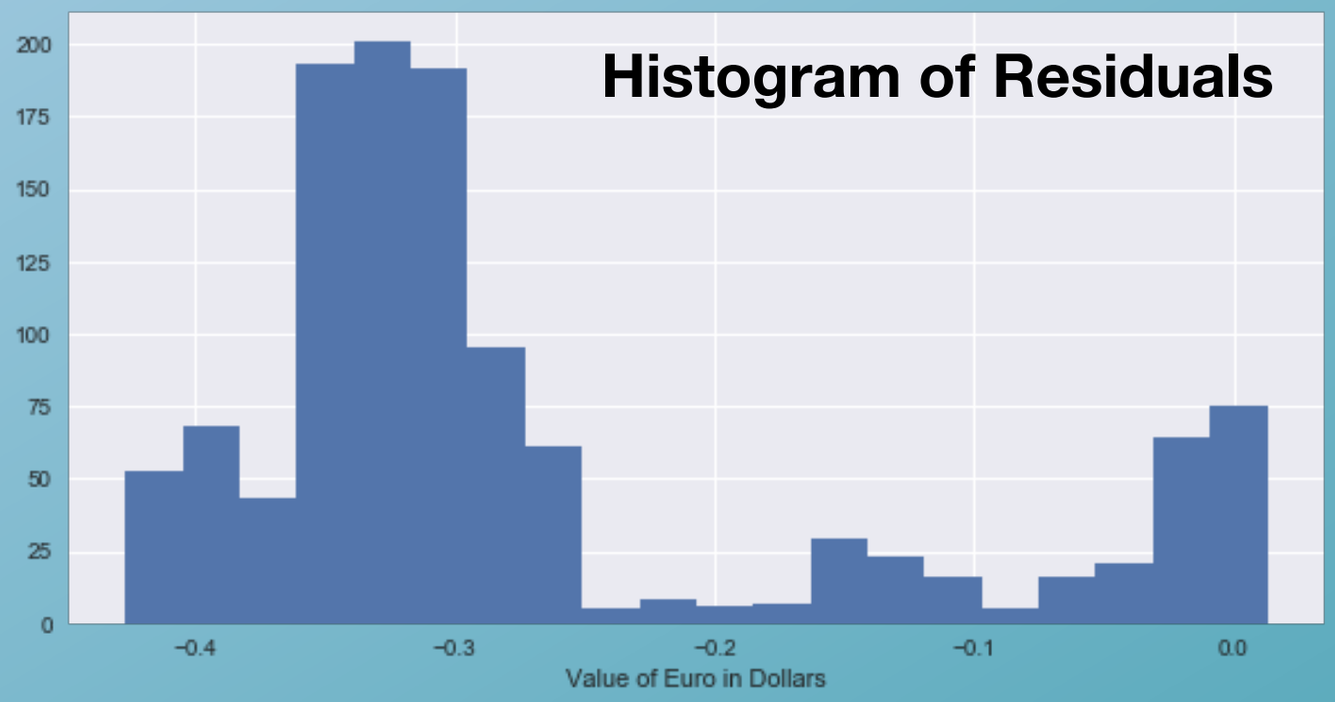
- (0, 1, 0)
- (0, 1, 1)
- (3, 1, 2)
- (1, 2, 2)

Out of Sample Forecasting

Predicting values for the test range and comparing those results to the actual values.

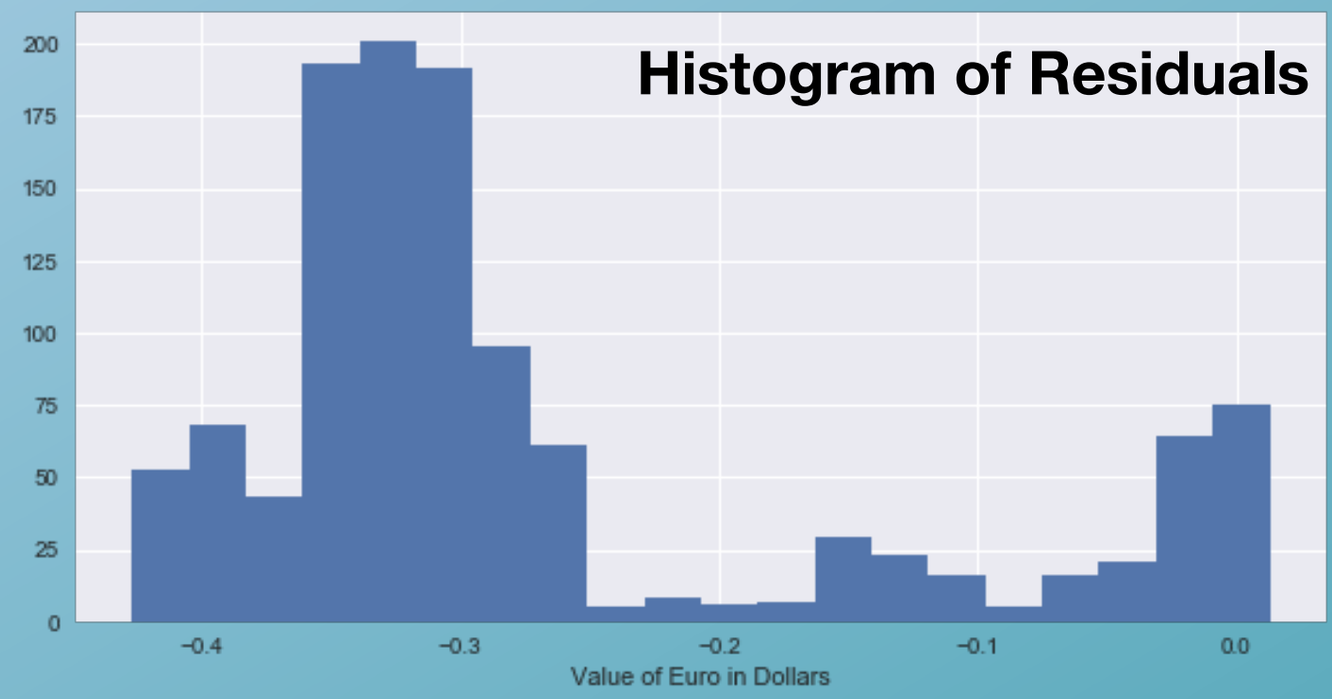
ARIMA (0,1,0)

- Increasing straight line
- Predictions are not very accurate
- Residuals are not normally distributed



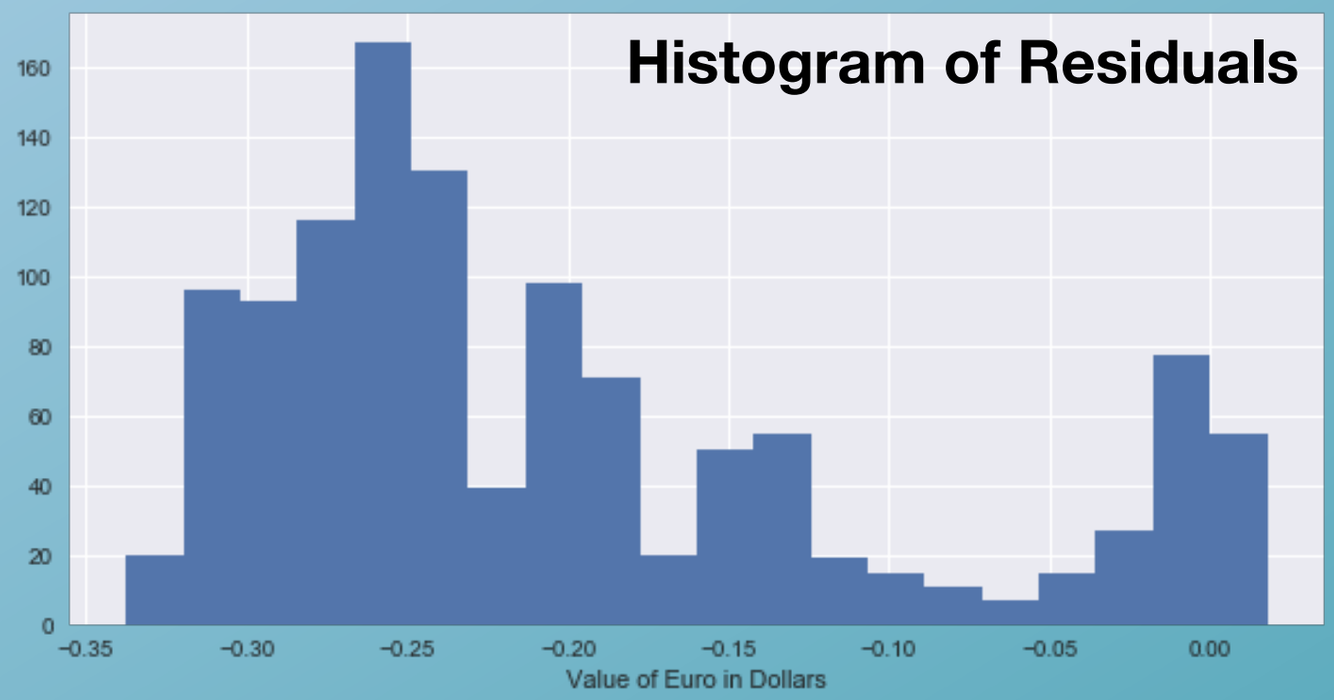
ARIMA (0,1,1)

- Results nearly identical to (0, 1, 1) model



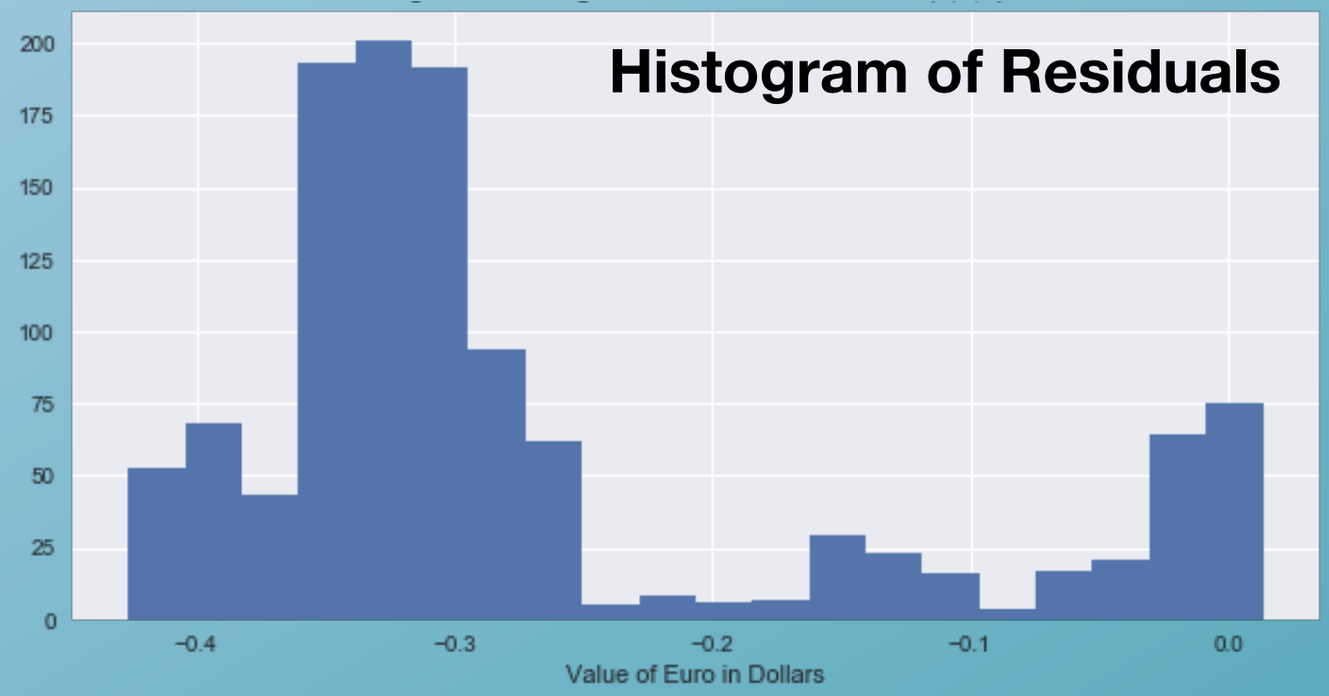
ARIMA (1,2,2)

- Flat straight line
- Reduced residual range, but still not normally distributed



ARIMA (3,1,2)

- Increasing straight line
- Residuals reflect inconsistency with actual values



Prediction Success

Mean Squared Errors

Model	MSE
(0, 1, 0)	0.087535
(0, 1, 1)	0.087527
(1, 2, 2)	0.049368
(3, 1, 2)	0.087295

- Similar accuracy with models (0,1,0), (0,1,1) and (3,1,2)
- Best accuracy with ARIMA (1,2,2)

Prediction Success

Visualizing Residuals over time

- Residuals are not randomly distributed
- Similar accuracy in the first half-year
- (1, 2, 2) has comparatively decreasing residuals

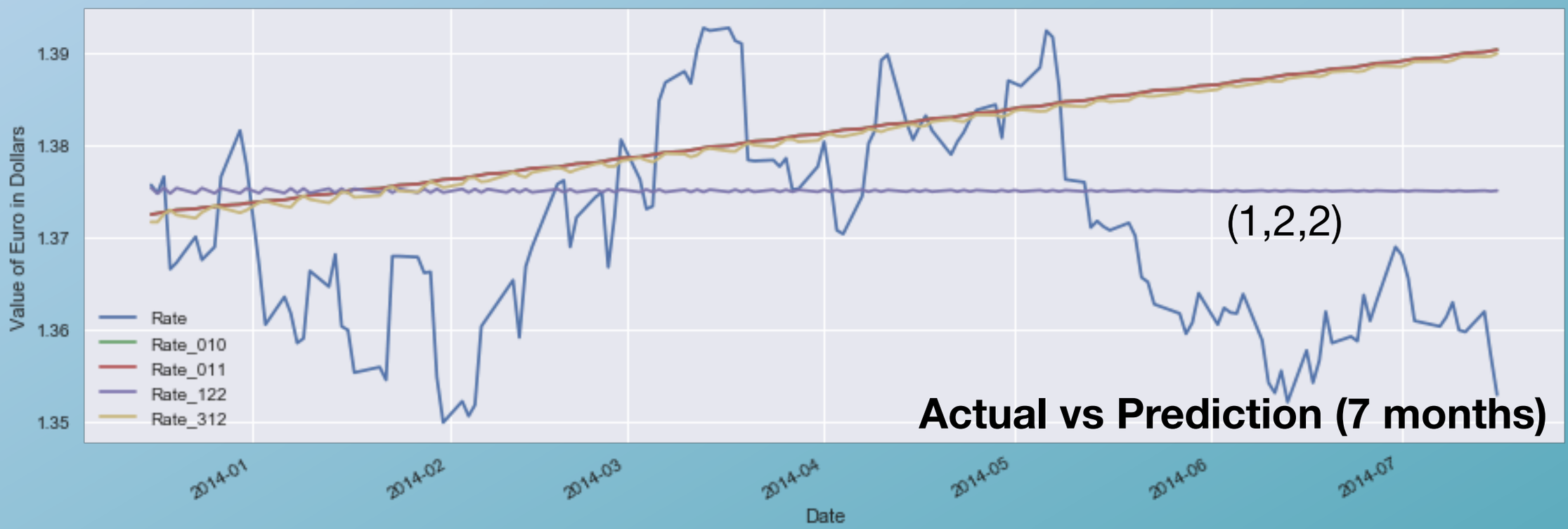
Prediction Success

Visual Comparison of Actual vs Prediction

- Three of four models predicted increasing rates
- Actual Euro value had a dramatic decrease in the first year
- Model (1,2,2) appears to be most accurate

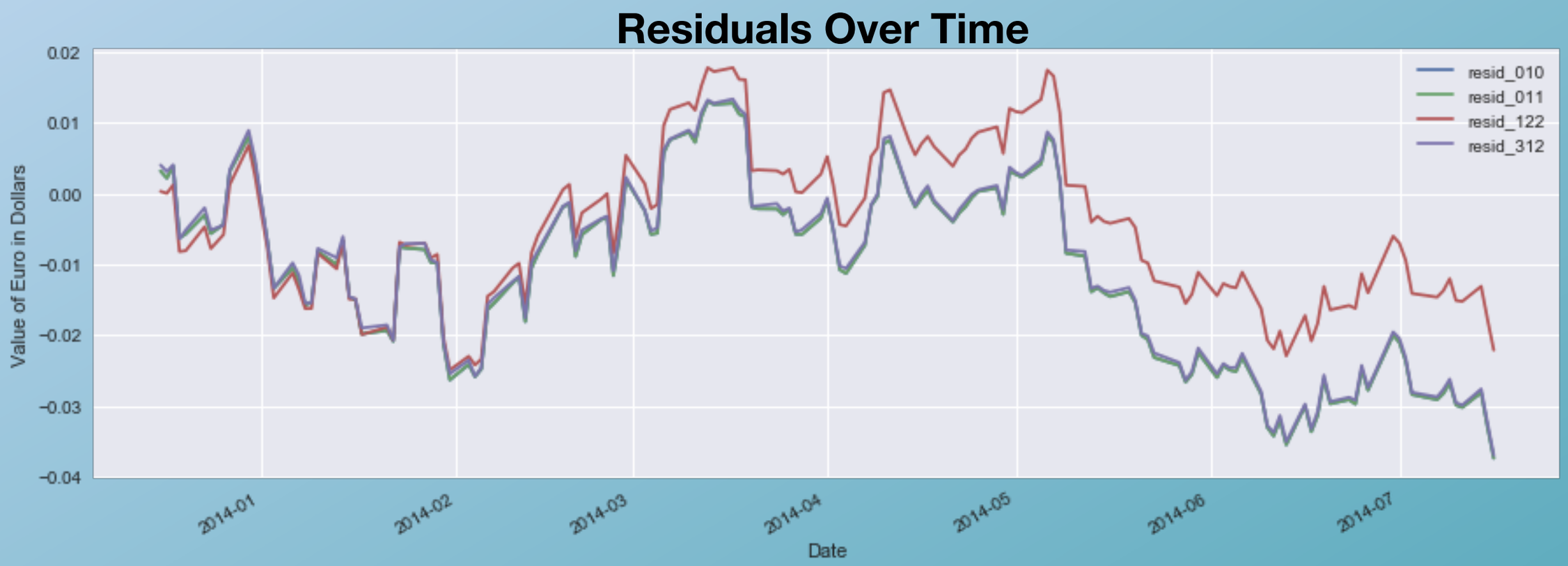
Early Predictions

- Euro value drops about seven months into the test data
- Trends in model predictions are already apparent



Early Prediction Success

Model	MSE
010	0.000282
011	0.000282
122	**0.000143**
312	0.000273



Conclusion

- ARIMA model applied to exchange rate of Euro to US Dollar
- Dickey Fuller Test, ACF and PACF used to target model parameters
- Four Models selected using Auto ARIMA grid search: (1, 2, 2), (3, 1, 2), (0, 1, 0), (0, 1, 1)
- Best model (1, 2, 2) - lowest Mean Squred Error (MSE)
- Observed weak fit - residuals were not randomly distributed, histogram of residuals was not normally distributed
- Poor performance over five year test period
- Acceptable accuracy with predictions on first seven months

Further Exploration

Facebook's Prophet

- Additive model gaining in popularity.

Bayesian Structural Time Series Models

- Greater transparency without differencing, lags and moving averages (suggested by Kim Larsen from Stitch Fix)

Linear Regression (Boosted Least Squared Regression)

- Benefit of normal distribution of the data
- Create features and predict future values without additional data

Rolling Forecast

- Add new data as available and update predictions
- ARIMA likely to show better accuracy in short-term predictions

References

4 Ways to Forecast Exchange Rates by Joseph Nguyen
https://www.investopedia.com/articles/forex/11/4-ways-to-forecast-exchange-rates.asp

Big Mac PP
https://www.investopedia.com/terms/b/bigmacppp.asp

What is an Autoregressive Integrated Moving Average - ARIMA
https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

US Federal Reserve Exchange Rates
https://www.federalreserve.gov/data.htm

Plot each year of a time series on the same x-axis using Pandas by Andrew Tedstone
http://atedstone.github.io/pandas-plot-seasons-time-series/

A Gentle Introduction to Autocorrelation and Partial Autocorrelation
https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/

A comprehensive Beginner's Guide to Time Series Forecasting by Aarshay Jain
https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/

Sorry ARIMA, but I'm Going Bayesian by Kim Larsen
https://www.predictiveanalyticsworld.com/patimes/sorry-arima-but-im-going-bayesian/7565/

Prediction of Foreign Exchange Rate using Regression Techniques
http://mtmi.us/rbtr/sept2017/06-Sharma-Hota-Handa-pp29-33.pdf

Time Series Analysis in Python by William Koehrsen
https://towardsdatascience.com/time-series-analysis-in-python-an-introduction-70d5a5b1d52a