United States Drug Prices

Analysis and Forecasting

Abstract

- We want to analyze the medical drug market in the United States between the dates 12/11/2013 to 12/30/2015
  - Create a metric for and analyze market volatility based on the drug price change log
  - Analyze how drug rate change characteristics impact prices and change rates
- We want to forecast market trends and individual drug prices
  - Correlation analysis of market volatility
  - Time series analysis of individual drugs

Data Collection

- We have a log of price changes submitted to the National Average Drug Acquisition Cost (NADAC) database.
  - Database updated on a weekly basis
  - Dates range from 12/11/2013 to 12/30/2015
- We will access the data using the medicaid API
  - https://data.medicaid.gov/resource/a4y5-998d.csv
  - API accepts SoQL queries within the URL

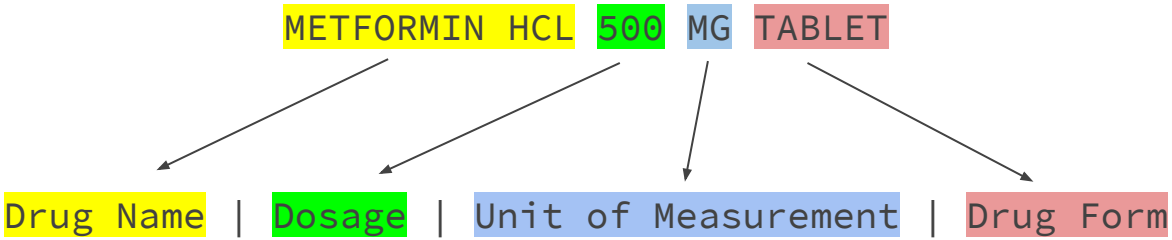
NADAC Database Metadata

Metadata

Column Name	Description
ndc_description	Identifies the drug name, strength, and dosage form of the drug product.
ndc	The National Drug Code (NDC) is a numerical code maintained by the FDA that includes the labeler code, product code, and package code. The NDC is an 11-digit code.
old_nadac_per_unit	The National Average Drug Acquisition Cost per pricing unit from the previous NADAC Reference File.
new_nadac_per_unit	Indicates whether the NDC was considered brand ('B') or generic ('G') for the NADAC rate calculation process. If the NDC was considered brand ('B') and approved under an Abbreviated New Drug Application (ANDA), the indicator is shown as ('B-ANDA').
classification_for_rate_setting	The difference between the New NADAC Per Unit and the Old NADAC Per Unit, divided by the Old NADAC Per Unit.
percent_change	The difference between the New NADAC Per Unit and the Old NADAC Per Unit, divided by the Old NADAC Per Unit.
primary_reason	Describes the primary reason for the NADAC Per Unit change, see explanation below for each reason: Survey Rate: The NADAC Per Unit has been updated using information from the most recently completed pharmacy survey. • WAC Adjustment: The NADAC Per Unit has been updated to reflect changes in published pricing. • Help Desk Adjustment: The NADAC Per Unit has been updated as a result of an inquiry to the help desk. • Brand Generic Change: The NADAC Per Unit has been updated as a result of a change in the Classification for Rate Setting. Rate Group Change: The NADAC Per Unit has been updated due to placement into a new NADAC drug grouping because of a change in the NDC attributes. NDC attributes that may result in a rate group change include package size updates, brand/generic designation redeterminations and revisions to the active ingredient(s), strength, dosage form or route of administration.
start_date	Date change was submitted to the NADAC database.

Breaking Apart ndc_description

- Drug name, dosage, units of measurement, and drug form are encapsulated in the ndc description.
- We needed to extract these features to scale drug prices by dosage.
  - Correlates amount of active ingredient

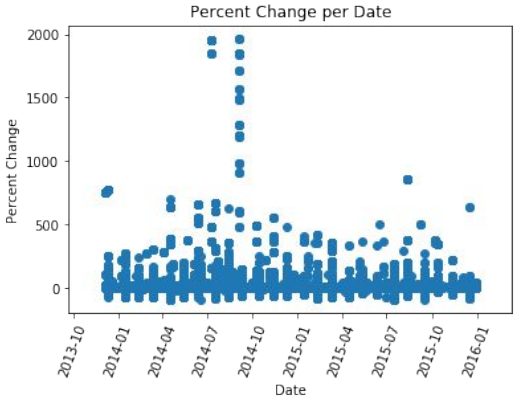


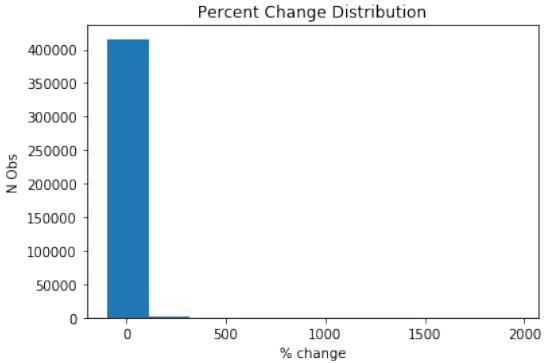
Exploratory Data Analysis: At a Glance

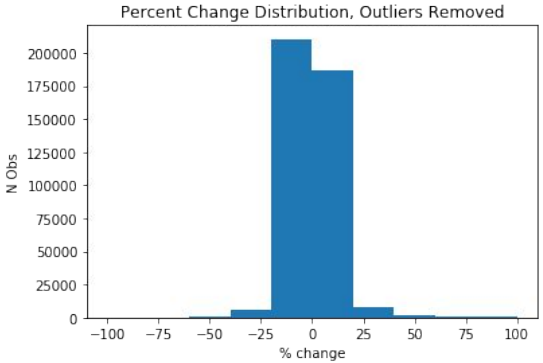
- Number observations per ndc number is tiny
  - Over the entire ~7 years of data, no ndc number has more than 64 observations
  - NDC number encapsulates labeler, product code and package code, information too specific to be useful
  - Better to use drug name extracted from ndc description
- ~13% data loss from dropping missing values
  - Entirely through ndc description fall through.

Exploratory Data Analysis: Percent Change

- Initial Assessment:
  - Several Outliers, wide range
- Outliers Removed:
  - Slight negative tilt, as expected

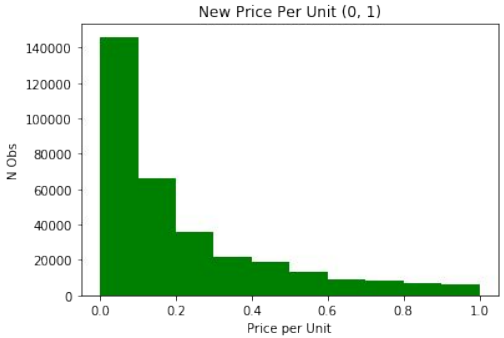






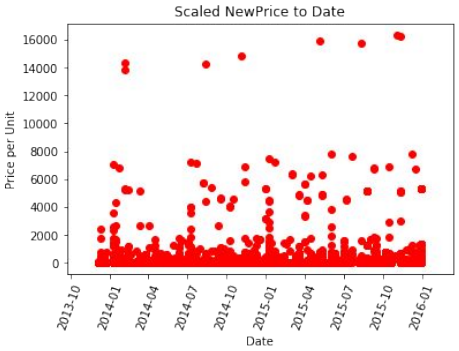
Exploratory Data Analysis: New Prices per Unit

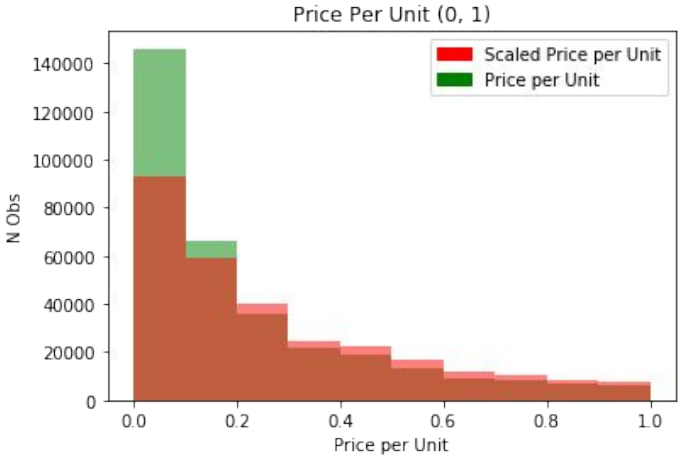
- Exponential Distribution
  - Significance testing should be done non-parametrically
- Majority of Entries Within the Range 0 - 0.2
- Scatterplot shows notable outliers
  - Stelara: treats Crohn's disease
  - Neulasta: post-chemotherapy immune system stimulant

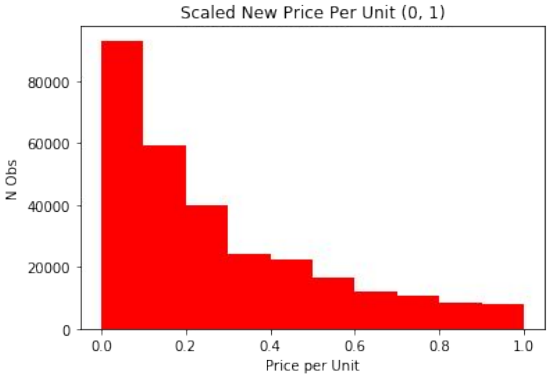


Exploratory Data Analysis: Scaled Price

- We scaled price to dosage
  - Same issue with outliers
  - Normal slope less aggressive in the 0 to 1 range.



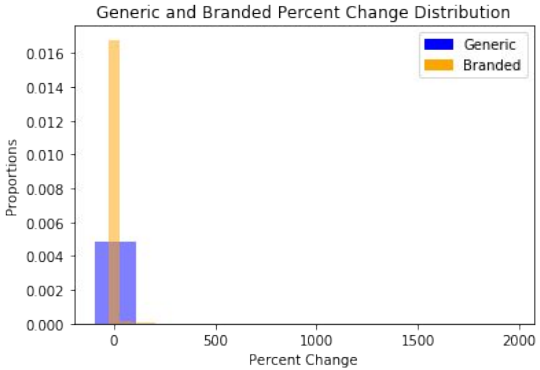


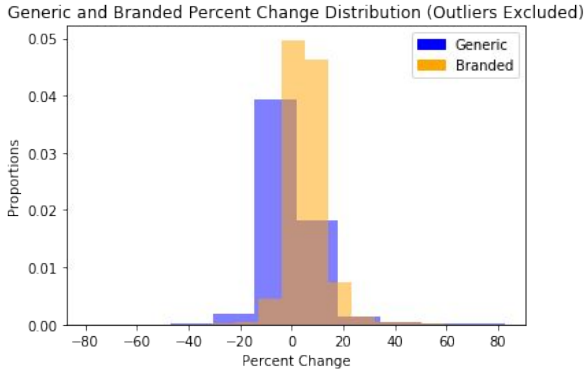


Generic vs Branded Drugs: Percent Change

Percent Change:

We have run a T-test comparing the percent change between generic and branded drugs and found a significant difference



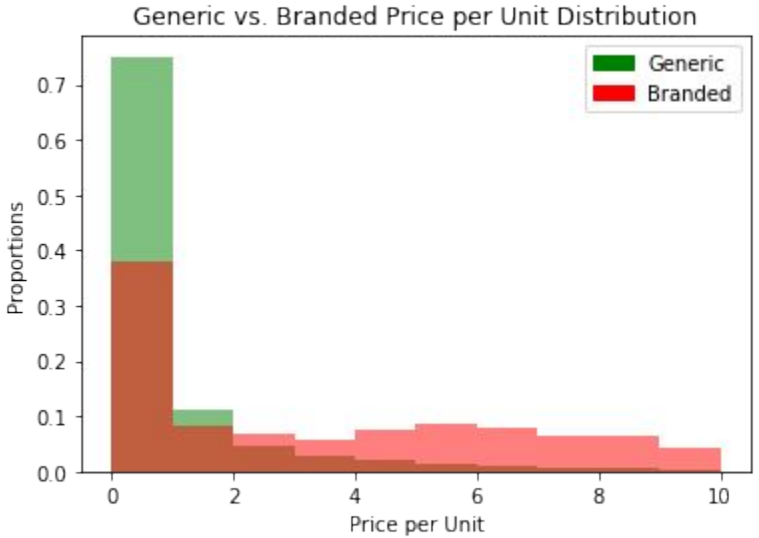



```
Mean Percent Change for Generic Drugs: 1.23 percentage points
Mean Percent Change for Branded Drugs: 6.29 percentage points
Standad Deviation for Generic Drugs: 27.12 percentage points
Standad Deviation for Branded Drugs: 17.64 percentage points
T-Statistic: -21.0
P-Value: 0.0
```

Generic vs Branded Drugs: Price per Unit

This is not a normal distribution

We used the Mann-Whitney U test to determine significance

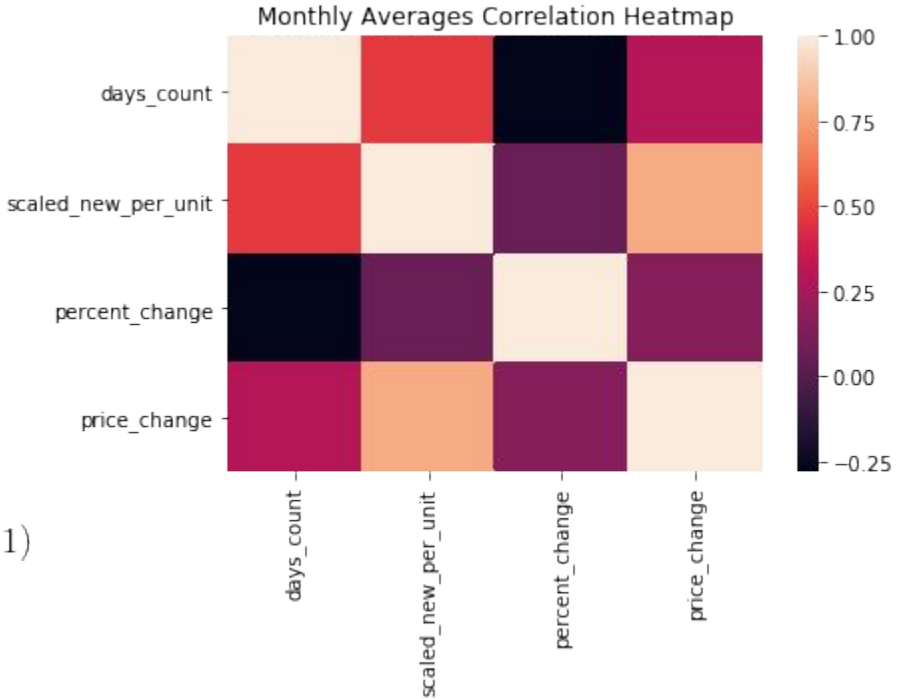


Correlation Testing: Inferring Relationships

- Price per unit increasing every month
- Percent change decreasing every month

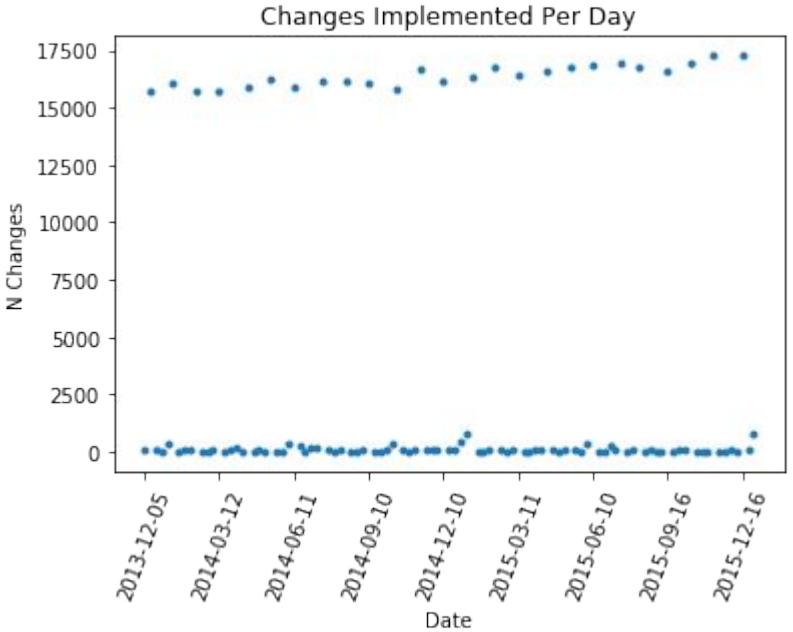
PPU as an approximate function of day count:

$$PCNT(day) \cong -3.49 * 10^{-03} * day + 2.715$$
$$PPU(day) \cong (-3.49 * 10^{-05} * day + 1.027) * PPU(day - 1)$$



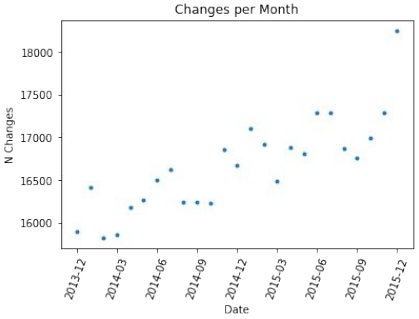
Metrics For Market Volatility: Changes per Day

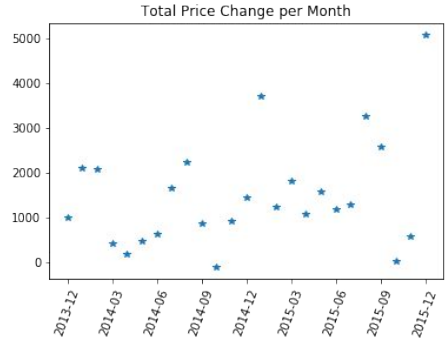
- Lows and Highs:
  - Most medicines appear to be updated on a monthly basis
- Grouping by Month:
  - More directly related to our business goals

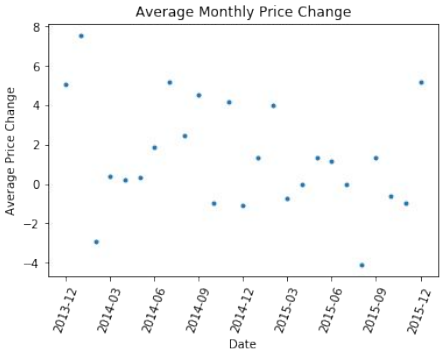


Metrics For Market Volatility: Monthly Analysis

- At a Glance
  - Changes per Month: strongest correlation with date
  - Total Price Change per Month: loose correlation
  - Average Monthly Price Change: slight negative correlation, as expected with negative percent change to days count







Understanding Periodicity: Fourier Transform

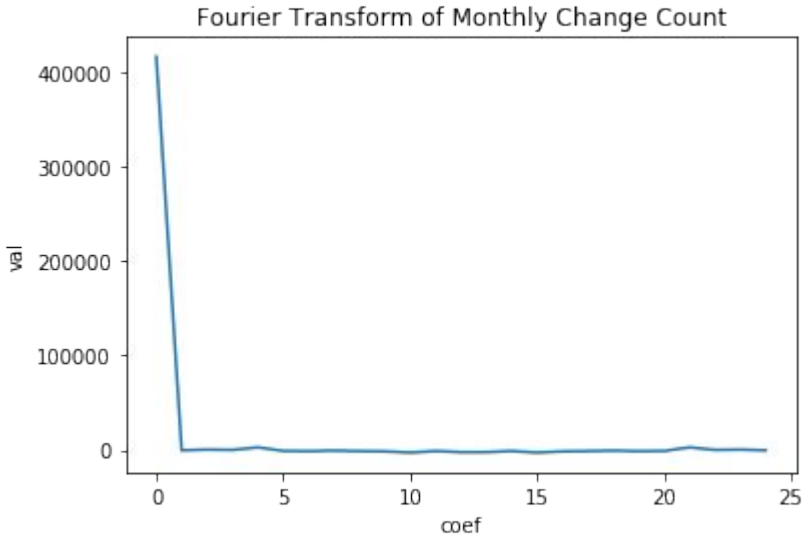
Hard to pick up periodicity at such low resolutions

We can attempt to formalize periodicity using the fourier transform

I don’t know what is going on here!

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \, e^{2\pi i \left(\frac{n}{T}\right) x}$$

$$e^{ix} = \cos x + i \sin x$$



Forecasting Volatility

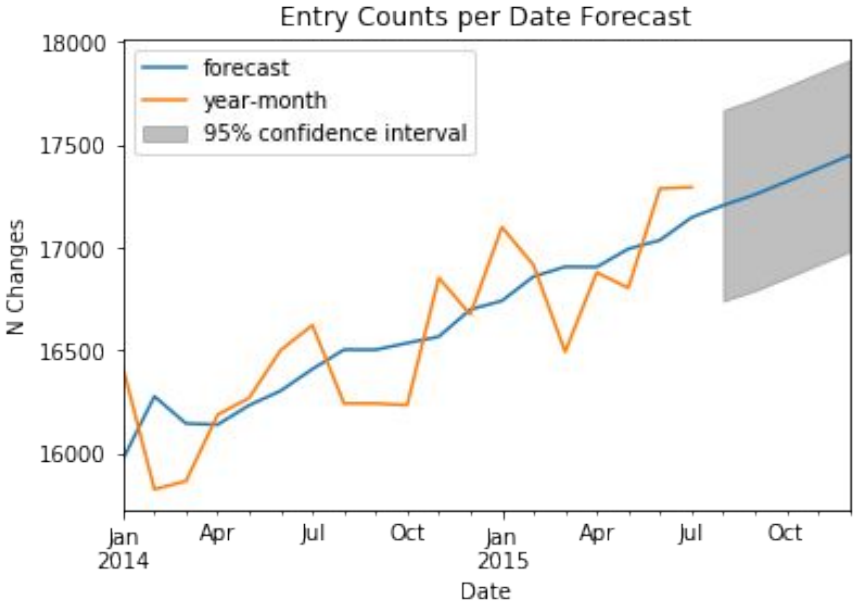
ARIMA Model Results

Dep. Variable:	D.year-month	No. Observations:	19
Model:	ARIMA(1, 1, 1)	Log Likelihood	-132.259
Method:	css-mle	S.D. of innovations	236.817
Date:	Tue, 10 Sep 2019	AIC	272.518
Time:	01:01:30	BIC	276.296
Sample:	01-01-2014 - 07-01-2015	HQIC	273.157

	coef	std err	z	P>|z|	[0.025	0.975]
const	63.7735	9.889	6.449	0.000	44.391	83.156
ar.L1.D.year-month	0.0808	0.234	0.345	0.735	-0.378	0.539
ma.L1.D.year-month	-0.9999	0.139	-7.175	0.000	-1.273	-0.727

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	12.3826	+0.0000j	12.3826	0.0000
MA.1	1.0001	+0.0000j	1.0001	0.0000

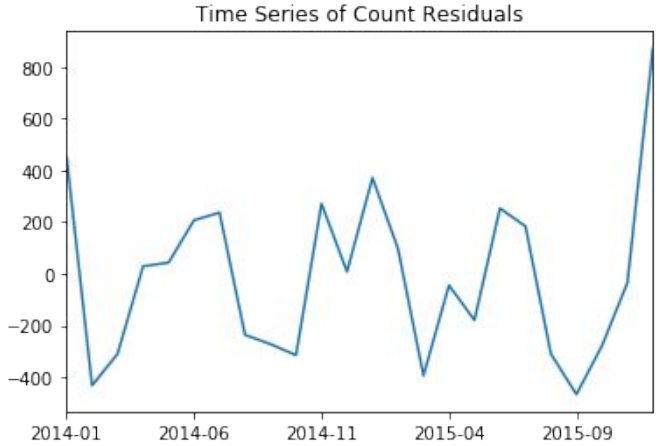


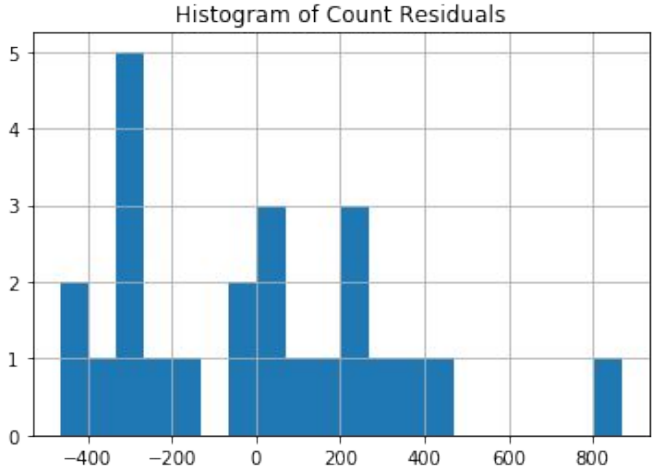
Model Performance

Low resolution

It may be difficult to come to any real conclusion given the low resolution

Count of residuals has seasonality, histogram of residuals has negative tilt.

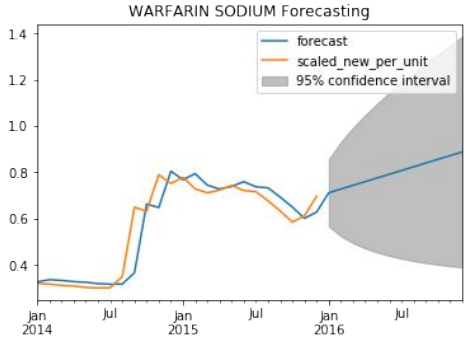


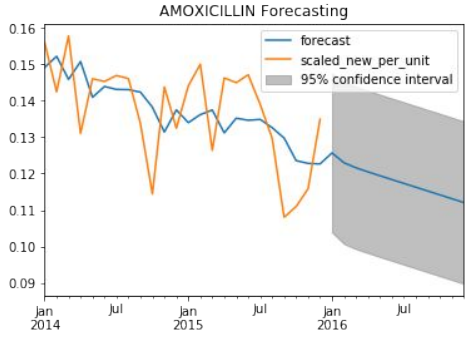


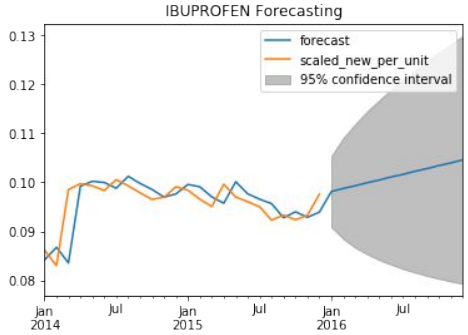
Individual Drugs: ARIMA Modelling

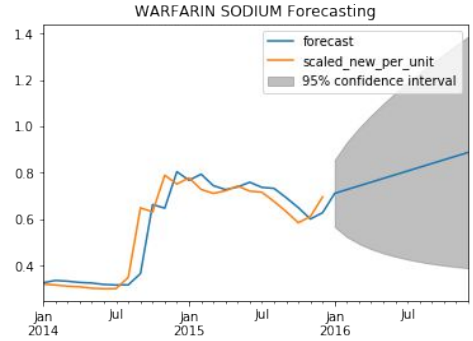
We used the top 25 most mentioned drugs in the database.

We see varying degrees of success in the models.







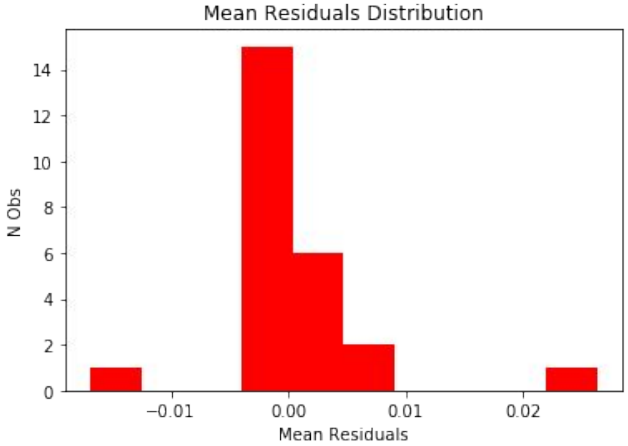


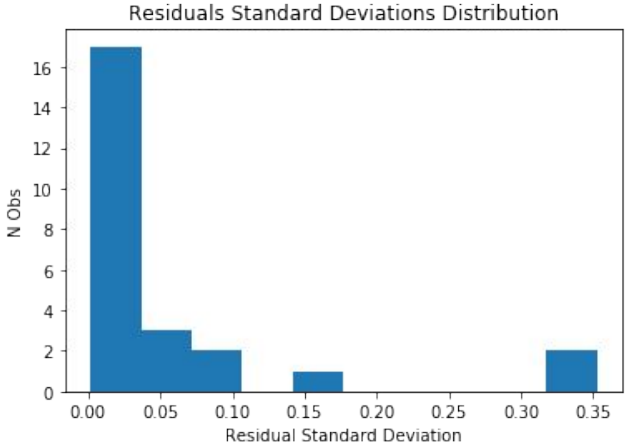
Individual Drugs: Validation

Mean: Majority lay between 0.00 and 0.01.

Standard Deviation: Majority between 0 and 0.1

Overall, decent.





Conclusion

- We Created a Pipeline:
  - Transformed price changelog into drug information.
  - Data can be expanded to current week.
- We Explored Trends in the Market:
  - Explored relationship between time, change count, percent change, price.
- We Created a Model for Volatility:
  - Time series analysis of predictors of change
- We Created a Model to Predict Individual Drugs:
  - Dictionary of models that can easily be expanded to include other drugs of interest.

Further Work

- Roll out model to include up to date data
  - Database has ~1.4 million entries and is growing
  - Distributed computing solutions such as spark are imperative
- Analyse production and consumption data to complete supply and demand curve
  - Useful for producing more tangible results regarding the current state of the drug market
- Integrate A/B testing of public policies regarding drugs and drug marketing
  - We will fit our analysis to a ‘public health’ metric