

VUMC ELECTIVE SURGERY SCHEDULE ANALYSIS



IDS 506 WEEK 5

Vanderbilt University Medical Center (VUMC) Background and Business Problem

- VUMC has 55 adult operating rooms (ORs).
- Number of surgeries are scheduled weeks in advance
 - 6% of all surgeries scheduled the day of as add-ons
 - Actual surgeries not known until the day of
- Scheduling staff up until the day before is not reliable and results in over or understaffed, increasing costs to patients and VUMC
- Predicting the monthly number of surgeries is within 5% accuracy but predicting daily is “much to be desired”.

Desired Solution

- VUMC is wondering if it can use past data of number of surgeries to predict daily surgeries in the future
 - Predictions should give VUMC adequate time to schedule while being accurate



**VANDERBILT
UNIVERSITY
MEDICAL
CENTER**

Data and Overview of Time Series and Machine Learning Algorithms Used

The Data

- 241 days with number of actual surgeries
- Day of the Week
- Lag variables from $T - 28$ to $T - 1$

Algorithms Considered

- Autoregression, Moving Average (ARMA)
- Autoregression, Integrated, Moving Average (ARIMA)
- Vector Autoregression (VAR)
- Multiple Linear Regression

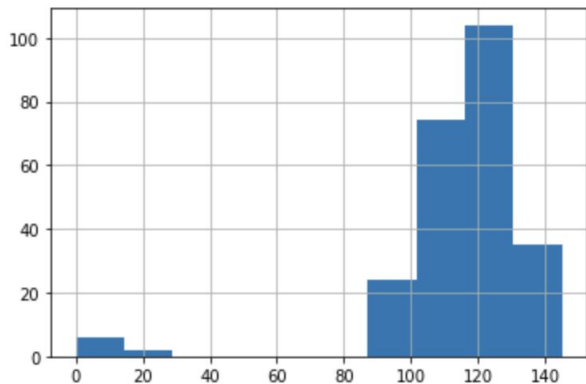
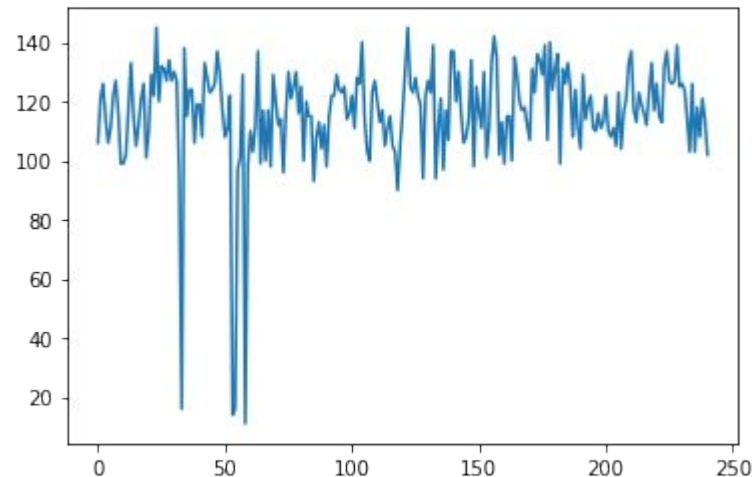
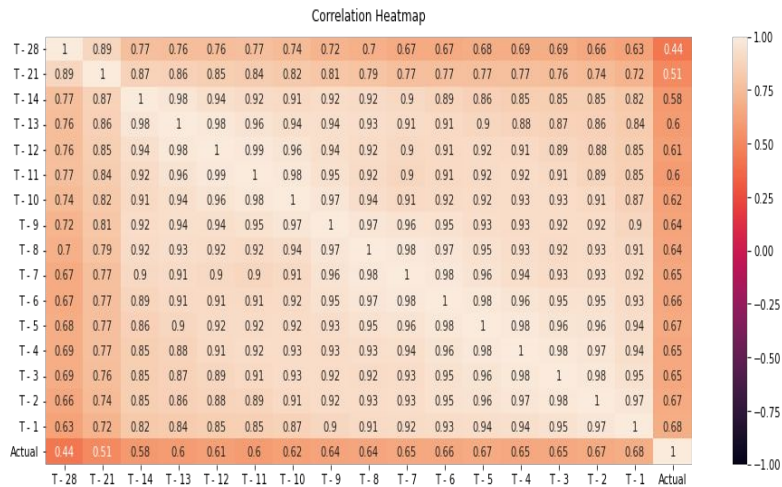
Algorithm Used

- Autoregression, Moving Average with Exogenous Regressors (ARMAX)



VANDERBILT
UNIVERSITY
MEDICAL
CENTER

EDA



DIVIDE THE DATA INTO TWO PARTS AND CHECK THE MEAN AND VARIANCE

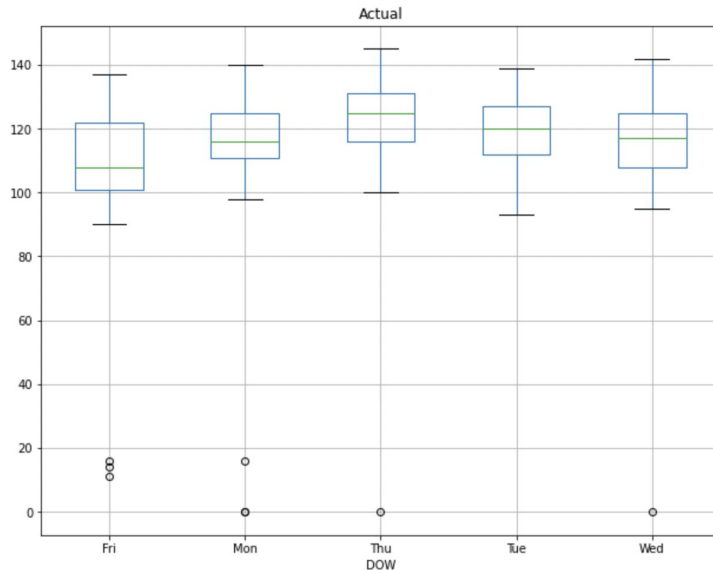
mean1=112.221311, mean2=116.723577

variance1=562.483808, variance2=471.712208

EDA

	SurgDate	DOW	T - 28	T - 21	T - 14	T - 13	T - 12	T - 11	T - 10	T - 9	T - 8	T - 7	T - 6	T - 5	T - 4	T - 3	T - 2	T - 1	Actual
0	2011-10-10	Mon	38	45	60	63	65	70	73	73	73	80	84	89	94	98	100	104	106
1	2011-10-11	Tue	35	47	65	68	78	82	82	82	86	89	92	95	99	99	99	114	121
2	2011-10-12	Wed	26	43	54	62	72	72	72	74	87	94	96	101	102	102	106	114	126
3	2011-10-13	Thu	28	48	65	70	72	72	72	82	87	91	94	94	94	97	98	103	114
4	2011-10-14	Fri	31	40	50	50	50	54	62	68	71	73	73	73	78	83	87	94	106

Boxplot grouped by DOW



IMPUTE OUTLIERS WITH VALUES OF 0 and KEEP OUTLIERS AS IT IS. FREQUENCY = 49 with PERIOD = 5

MISSING ROWS
 24-Nov-11 Thu
 28-May-12 Mon
 04-Jul-12 Wed
 03-Sep-12 Mon

Missing Rows are all holidays

OUTLIERS
 11/25/2011 Fri
 12/23/2011 Fri
 12/26/2011 Mon
 12/30/2011 Fri

Ignoring the Outliers because they will show the same pattern every year

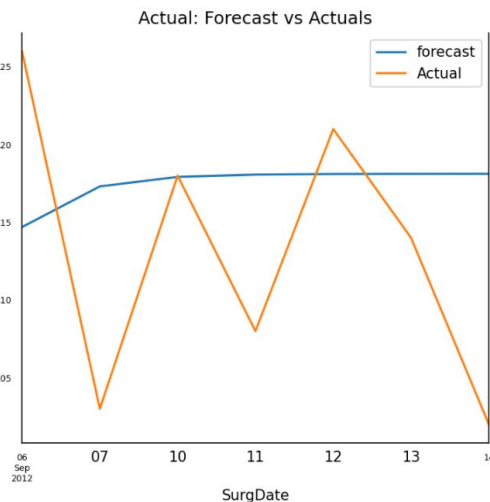
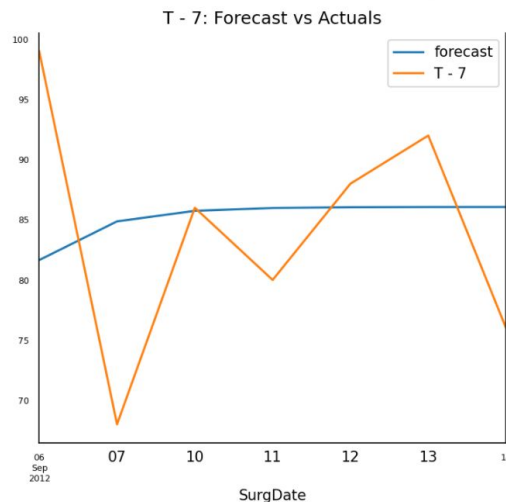
Vector Autoregression (VAR)

Reason not chosen as final model

- Algorithm used for multiple predictions where we only need to predict Actual surgeries
- Finds relationship between inputs where we are only interested in relationship of inputs to output which is Actual surgeries

Results:

VAR predicting actual surgeries from 9/06/12 to 9/14/12 (7 days)



Forecast Accuracy of: T - 7

mape	: 0.1029
me	: 1.0588
mae	: 8.3448
mpe	: 0.0272
rmse	: 10.4316
corr	: -0.4617
minmax	: 0.0929

Forecast Accuracy of: Actual

mape	: 0.0772
me	: 4.3328
mae	: 8.4132
mpe	: 0.0446
rmse	: 10.1356
corr	: -0.5279
minmax	: 0.0704

Multiple Linear Regression

Recursive Feature Elimination for Feature Selection

- Features labeled False are considered weak
- Kept the following features based on RFE
- Also did not want to choose a lag time too close to day of actual
 - Wanted to give VUMC most time to create schedule based on data
- Chosen features for Linear Regression
 - 'T - 13', 'T - 12', 'T - 11', 'T - 10', 'T - 7', 'T - 6'

RFE Results:

```
[('T - 28', False, 7),  
 ('T - 21', False, 6),  
 ('T - 14', False, 5),  
 ('T - 13', True, 1),  
 ('T - 12', True, 1),  
 ('T - 11', True, 1),  
 ('T - 10', True, 1),  
 ('T - 9', False, 4),  
 ('T - 8', False, 8),  
 ('T - 7', True, 1),  
 ('T - 6', True, 1),  
 ('T - 5', False, 2),  
 ('T - 4', False, 3),  
 ('T - 3', True, 1),  
 ('T - 2', False, 9),  
 ('T - 1', True, 1)]
```

Multiple Linear Regression

Multiple Linear Regression Results

	Actual value	Predicted value	Difference	Percent_Difference	lower_bound	upper_bound
SurgDate						
2012-09-06	126	126.18	0.18	0.00	111.997911	140.362613
2012-09-07	103	108.81	5.81	0.05	94.628723	122.993424
2012-09-10	118	121.92	3.92	0.03	107.734007	136.098709
2012-09-11	108	116.60	8.60	0.07	102.416205	130.780906
2012-09-12	121	120.79	-0.21	-0.00	106.607132	134.971833
2012-09-13	114	119.92	5.92	0.05	105.735000	134.099702
2012-09-14	102	112.18	10.18	0.09	97.999345	126.364047

R squared: 0.4814367215157894

Mean Absolute Error: 4.973687591086066

Mean Absolute Percentage Error: 0.04630307379639109

Mean Square Error: 37.40005359516734

Root Mean Square Error: 6.1155583224401795

THE ARMAX PROCESS

Univariate Analysis

Stationarity

ARMAX(WHY?)

Hyper
Parameters
Tuning

Exogenous
Variables

COMPARE(AIC
and MAE) after
Forecast

1. Stationarity using ADF test
2. ARMA(p,q) - tells us about the order of AR & MA part
3. Find the optimum p and q terms with Hyperparameter Tuning
4. Fit the ARMAX model with Exogenous Variables (T-X)s
5. Evaluate the models on the basis of AIC & MAE
6. Forecast the test sample based on the best Model

Test Statistic	-5.561718
p-value	0.000002
#lags used	5.000000
number of observations used	239.000000
critical value (1%)	-3.458011
critical value (5%)	-2.873710
critical value (10%)	-2.573256

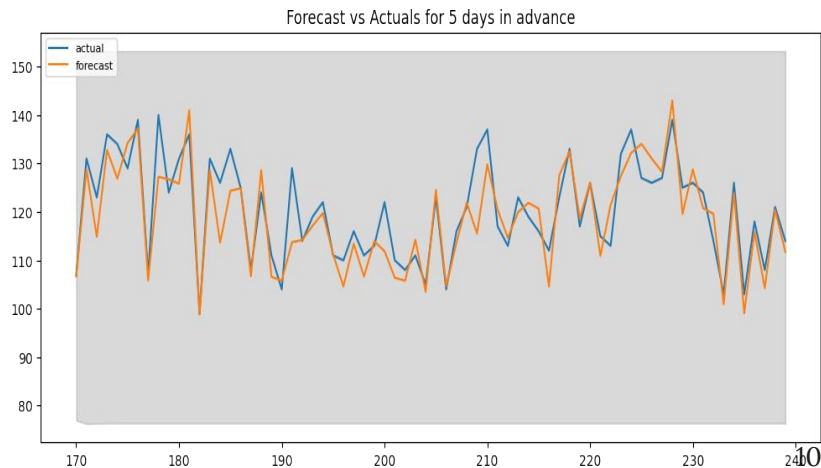
**EXOGENOUS VARIABLES ARE
VARIABLES THAT APPEAR IN THE
MODEL BUT NOT EXPLAINED BY
THE MODEL**

Model vs Metric Comparison

All auto_arima models from pmdarima

	Baseline	2 days ahead	3 days ahead	5 days ahead	7 days ahead	10 days ahead	From T-7 to T-10
AIC	1489	998	996	999.5	997	989	1247
MAE	9.69	3.93	3.93	3.96	4.25	4.68	8.4
		(T - 2 thru T - 1)	(T - 3 thru T - 1)	(T - 5 thru T - 1)	(T - 7 thru T - 1)	(T - 10 thru T - 1)	(T - 10 thru T - 7)

```
sx2model = pm.auto_arima(df_train, exogenous=df[['T - 1', 'T - 2', 'T - 3', 'T - 4', 'T - 5']][:170],
                          start_p=1, start_q=1,
                          max_p=3, max_q=3, # maximum p and q
                          m=52,             # frequency of series
                          d=None,
                          seasonal = False,
                          trace=True,
                          error_action='ignore',
                          suppress_warnings=True,
                          stepwise=True)
```





Thanks!

Any questions ?