STAT 626 – Methods in Time Series Analysis

TERM PROJECT

One Proc at A Time

11/27/18

Table of Contents

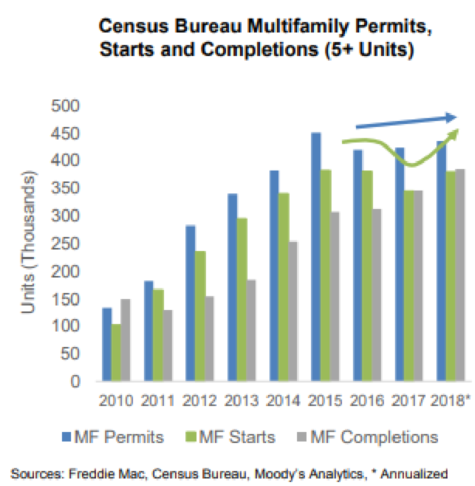
1. Introduction

Commercial construction projects often span years, from initial concept development to final sales and leasing. Changes in micro- and macro-economic variables over a multi-year period are difficult to predict, and investors are in constant demand for accurate and valid information to support investment decisions. The multi-family segment of the commercial construction industry is one of the most complex to predict, due to the vast number of interdependent variables including economic variables at many levels, demographic and lifestyle trends, government incentives, and population shifts.

In order to identify, better understand, and predict national-level trends in multi-family construction, Team One Proc at a Time performed a time-series analysis of 24 years of multi-family construction cost data made available by the U.S. Census Bureau.

Areas of application for this analysis include governmental policy, investment decisions, and social and economic research. The category 'Multi-family' ranges from two-unit duplexes, to large apartment complexes with hundreds of units, to mixed-use developments that combine residential units with storefronts and/or office spaces. In 1993, multi-family housing accounted for just 7% of new residential construction, but that number has increased to 18% in recent years[1]. Currently, 69% of all households in the US live in single family homes[2], but the share of multi-family properties is expected to grow steadily for many reasons, including[3]:

- 75 million Baby Boomers are headed into retirement
- Many of today's apartment complexes may be converted to retirement communities in the future
- Many millennials aren't buying homes
- It's getting more expensive to build new apartment units



1,2 Source: US Census Bureau website

3 List source: Business Insider article by Grant Cardone

2. Executive Summary

Using twenty-four years of monthly multi-family construction cost data from the U.S. Census Bureau, valid time-series models were developed that accurately: 1) fit the source data and 2) predict monthly costs for future periods. Using JMP for initial model selection, then SAS for further refinement of the 2 most accurate and valid models, the best model was the 'ARIMA(1,1,2)(1,1,1)12 No Intercept', which was able to predict the last 12-monthly values with an average 98% accuracy. However, additional modeling attempts conducted in SAS Forecast Studio resulted in an 'ARIMA(0,1,0)(1,0,1)12 No Intercept' model with even better accuracy at 98.95%.

To compare model accuracy, Schwarz Bayesian Criterion (SBC) scores were first used. To measure predictive accuracy, 12 periods predicted by the model were compared to actual values for those same periods. Those 12 accuracy measurements were then averaged for the final accuracy metric.

The modeling methodology began with an initial review of the source data's structure and correlations in JMP. Interpreting the ACF, PCF, and residual patterns suggested likely combinations of time-series parameters (p,d,q, seasonal periods) needed to produce a valid and accurate model. Still using JMP, 54 models were run in order to test each of the likely combinations, and after reviewing all the results, 2 were selected based on the models' ability to meet valid model assumptions.

Having selected the 2 best models in JMP, SAS was then used to recreate the same models. Additional outlier and level-shift refinements were performed with a maximum number = 20 at an alpha of 0.005 over multiple iterations until no more outliers were found. The resulting models were predicted with high accuracy, including the previously stated 98% prediction accuracy of the best model from JMP/SAS.

Following this initial selection of the best model using JMP and SAS, a second, separate modeling process was performed using SAS Forecast Studio. The intent of this second process in Forecast Studio was to independently verify the results of the first process in JMP and SAS. Forecast Studio did not identify the same best model, and in fact selected a model (ARIMA(0,1,0)(1,0,1)12 No Intercept) that did not make it onto the list for selection in JMP. The SAS Forecast Studio model produced the best results in terms of both SBC and prediction error rates.

The 'ARIMA(0,1,0)(1,1,1)12 No Intercept' model ultimately selected as the best model, is an ARIMA model with an autoregressive difference (p) of 0 periods, a moving average difference (q) of 0 periods, and differencing (d) of 1 period, combined with a Seasonal ARIMA (SARIMA) using P of 1, Q of 1 and D of 1 and a cycle of 12 periods.

As Section 8's discussion of Model 3 explains, the best model deviated slightly from the white noise probability test, however the deviations were not significant enough to reject the model. Altogether the chosen model provided a reliable and accurate predictive tool for new multi-family construction deliveries at a monthly level.

3. Business Context

3.1 Business Question

How accurately can the team model and predict 12 months of multi-family construction at a national level, based on time series data for the preceding 24 years?

3.2 Business Setting

Multi-family operators, investors, and general contractors use both macro- and micro-economic data to inform their predictions and business decisions. Dominant performance from assets requires only a slight edge in predictions over competitors, and this analysis aims to enhance decision making abilities. As a barometer for effects on inventory these predictions become useful for situational assessments. Once a valid, national-level predictive time-series model is developed, a logical next step is to apply these same steps towards regional and local-level analyses.

The performance of multi-family real estate is dependent on how supply and demand curves interact. Decisions made about supply will affect performance within the multi-family industry. This analysis looks at supply, which has experienced a full decade of strong growth nationally, jump-started by high renter demand after the housing crisis. Population growth in urban areas has led to a higher percentage of Americans living in urban areas as opposed to rural areas, for the first time in American history. And the demand for multi-family developments in urban areas continues to strengthen.

Recently, multi-family supply has shown signs of softening. This restriction of supply may help keep rent growth from tapering off, but because of the multi-year time frame associated with developing new apartment buildings, it is important to distinguish supply shifts early. An early adaptation to a trend may mean deploying (or staving off) extensive resources upfront and gaining an edge on competition.

The context in which these models' predictions are useful is based on the vantage that each user adds or processes value within the industry. For instance:

- Individuals may look at these predictions for weighing rent vs own decisions. A surplus of supply places downward pressure on rental costs, which may result in rent becoming increasingly less expensive compared to the cost of home ownership. Individuals sensitive to near-term housing decisions will find the supply information, and its effects on the housing market, useful for housing decisions.
- Active and passive investors rely on assets' rent performance to realize returns. If there are unoccupied units or rent vacations due to surplus, then an asset's performance suffers. Producers may better optimize their limited resources if they had better predictions of supply-side curves.
- A developer in NYC will experience different local trends than a developer in Houston, San Francisco, etc. All developers will find variation in local markets compared to the national trends evaluated in this report.

We recognize the limited data in this project's analysis will not provide a comprehensive decision basis. However, identifying and accurately predicting national trends in multi-family supply is an important input for situational awareness. There are more data available from the Census Bureau to make this framework useful at a local level (e.g. region or city).

4. Data Description

The data is made available by the US Census Bureau. The estimated value of construction put in place is published every month by the Bureau, two months after the reference date. Surveys by privately-owned multi-family owners contribute most of the estimated data.

MULTI-FAMILY:

Includes new apartments and condominiums. The classification excludes residential units in buildings that are primarily nonresidential.

4.1 Variables

The original data set included the 74 variables shown in Appendix D, including the target variable 'New multi-family' and the date field 'Date' containing months from January 1993 through August 2018. Each of the variables that is not a date field contains numeric time series data.

4.2 Outcome Variable

The outcome variable "New multi-family" is a monthly measure of what the Census Bureau refers to as "Value of Construction Put In Place" for multi-family construction projects.

Value of Construction Put In Place is defined as: "The cost of labor and materials; cost of architectural and engineering work, overhead costs; interest and taxes paid during construction; and contractor's profits."

"Multi-family" construction projects are construction projects in which multiple separate housing units for residential inhabitants are contained within one building or several buildings within one complex.

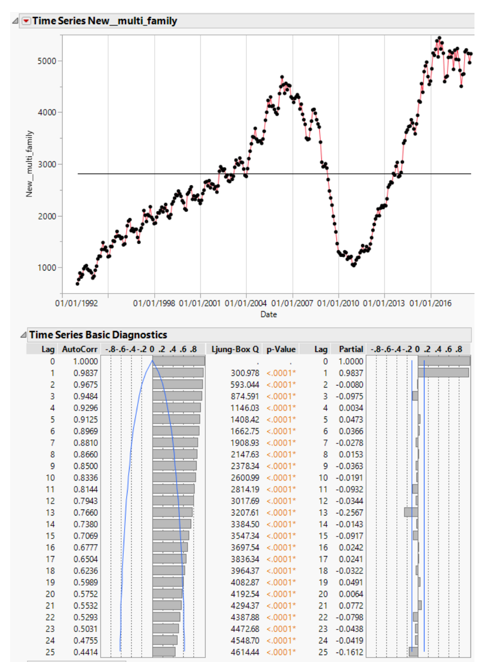
4.3 Timeframe

The data contained monthly amounts from January 1993 to August 2018.

5. Steps Taken – Project Level

The following process steps were followed to develop, identify, evaluate, and select the 2 best models:

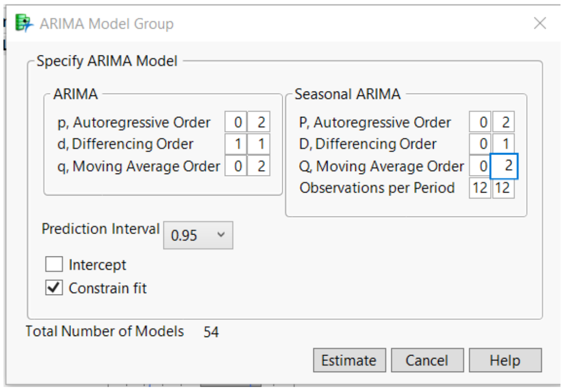
1) The data set was prepared to capture the desired 12 months of forecasts by:
   a) Removing the last 12 observations of the target variable
   b) Adding a new column for the model's forecast values.
2) In JMP, the data was graphed and then an initial time-series model was run using 'New_multi_family' as the target ("Y, Time Series") and 'Date' as the "X, Time ID". The data was found to be non-stationary due to the following characteristics:
   a) The mean of the multi-family homes was not constant over time.
   b) Variability in the residuals was not constant over time.
   c) The correlation structure of Yt changed over time.
3) Additional test models were run using variations of order 1 differencing and lag and square root transformations. The transformations did not help substantially, but differencing did help.
4) Seasonality was evident with with ACFs peaking every 12 periods (months). An ARIMA model group was run without intercepts*

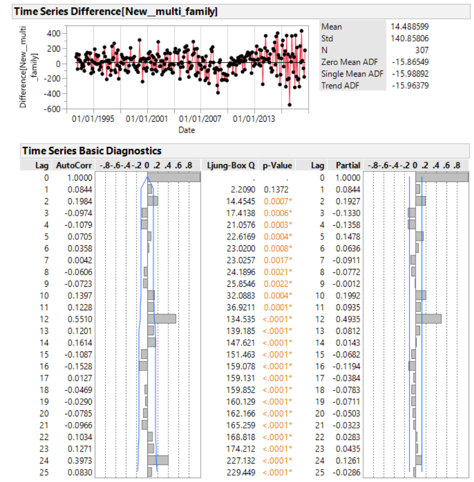


Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	0.9837	300.978	<.0001*	1	0.9837
2	0.9675	593.044	<.0001*	2	-0.0080
3	0.9484	874.591	<.0001*	3	-0.0975
4	0.9296	1146.03	<.0001*	4	0.0034
5	0.9125	1408.42	<.0001*	5	0.0473
6	0.8969	1662.75	<.0001*	6	0.0366
7	0.8810	1908.93	<.0001*	7	-0.0278
8	0.8660	2147.63	<.0001*	8	0.0153
9	0.8500	2378.34	<.0001*	9	-0.0363
10	0.8336	2600.99	<.0001*	10	-0.0191
11	0.8144	2814.19	<.0001*	11	-0.0932
12	0.7943	3017.69	<.0001*	12	-0.0344
13	0.7660	3207.61	<.0001*	13	-0.2567
14	0.7380	3384.50	<.0001*	14	-0.0143
15	0.7069	3547.34	<.0001*	15	-0.0917
16	0.6777	3697.54	<.0001*	16	0.0242
17	0.6504	3836.34	<.0001*	17	0.0241
18	0.6236	3964.37	<.0001*	18	-0.0322
19	0.5989	4082.87	<.0001*	19	0.0491
20	0.5752	4192.54	<.0001*	20	0.0064
21	0.5532	4294.37	<.0001*	21	0.0772
22	0.5293	4387.88	<.0001*	22	-0.0798
23	0.5031	4472.68	<.0001*	23	-0.0438
24	0.4755	4548.70	<.0001*	24	-0.0419
25	0.4414	4614.44	<.0001*	25	-0.1612

using the following parameter ranges including differencing and seasonality of 12:

*Intercepts were removed because the data was differenced both for the ARIMA and for the Seasonal ARIMA, and to align with the 'non-constant' model specification used in SAS.





5) After reviewing all 6 models for significant parameters, parsimony, and random residuals (see table below), only 3 met the criteria: 'ARIMA(1,1,2)(1,1,1)12 No Intercept', 'ARIMA(2,1,2)(1,1,1)12 No Intercept', and 'ARIMA(2,1,1)(1,1,1) No Intercept'. The first two had the best SBC scores of the three, so they were selected as the 2 best models for further refinement.

Model Comparison

Report	Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights	.2 .4 .6 .8	MAPE	MAE
☑	☐	Seasonal ARIMA(1, 1, 1)(1, 1, 1)12 No Intercept	279	10889.283	3448.7547	3463.3365	0.992	3440.7547	0.016057		3.271293	80.398682
☑	☐	Seasonal ARIMA(0, 1, 0)(1, 1, 1)12 No Intercept	281	11202.87	3456.9789	3464.2698	0.992	3452.9789	0.000263		3.357616	82.515312
☑	☐	Seasonal ARIMA(1, 1, 2)(1, 1, 1)12 No Intercept	278	10740.74	3446.3464	3464.5736	0.992	3436.3464	0.053534		3.228131	79.364446
☑	☐	Seasonal ARIMA(1, 1, 1)(0, 1, 2)12 No Intercept	279	10965.903	3450.1054	3464.6872	0.992	3442.1054	0.008173		3.291369	80.707556
☑	☐	Seasonal ARIMA(2, 1, 2)(1, 1, 1)12 No Intercept	277	10556.191	3443.0272	3464.8999	0.992	3431.0272	0.281436		3.197073	78.667001
☑	☐	Seasonal ARIMA(2, 1, 1)(1, 1, 1)12 No Intercept	278	10756.089	3446.7831	3465.0103	0.992	3436.7831	0.043033		3.238397	79.511812
☐	☐	Seasonal ARIMA(1, 1, 1)(0, 1, 1)12 No Intercept	280	11297.673	3454.6242	3465.5606	0.992	3448.6242	0.000853		3.334941	81.778410
☐	☐	Seasonal ARIMA(1, 1, 2)(0, 1, 2)12 No Intercept	278	10814.324	3447.6265	3465.8537	0.992	3437.6265	0.028226		3.247836	79.691853
☐	☐	Seasonal ARIMA(1, 1, 2)(0, 1, 1)12 No Intercept	279	11110.535	3451.4314	3466.0132	0.992	3443.4314	0.004211		3.294034	80.754392
☐	☐	Seasonal ARIMA(2, 1, 1)(0, 1, 2)12 No Intercept	278	10829.242	3448.0637	3466.2909	0.992	3438.0637	0.022684		3.257902	79.831821
☐	☐	Seasonal ARIMA(2, 1, 1)(0, 1, 1)12 No Intercept	279	11136.095	3452.1014	3466.6832	0.992	3444.1014	0.003013		3.302967	80.893903
☐	☐	Seasonal ARIMA(2, 1, 2)(0, 1, 2)12 No Intercept	277	10665.801	3444.9133	3466.7859	0.992	3432.9133	0.109603		3.224204	79.115510
☐	☐	Seasonal ARIMA(2, 1, 2)(0, 1, 1)12 No Intercept	278	10996.387	3449.3964	3467.6237	0.992	3439.3964	0.011650		3.286724	80.605952
☐	☐	Seasonal ARIMA(0, 1, 0)(0, 1, 2)12 No Intercept	281	11456.819	3461.3974	3468.6883	0.991	3457.3974	0.000029		3.400559	83.414202
☐	☐	Seasonal ARIMA(1, 1, 1)(2, 1, 1)12 No Intercept	278	10916.535	3450.4916	3468.7188	0.992	3440.4916	0.006738		3.265110	80.362100
☐	☐	Seasonal ARIMA(1, 1, 1)(1, 1, 2)12 No Intercept	278	10921.676	3450.6000	3468.8272	0.992	3440.6	0.006382		3.268095	80.375259
☐	☐	Seasonal ARIMA(0, 1, 0)(2, 1, 1)12 No Intercept	280	11211.48	3458.4332	3469.3695	0.992	3452.4332	0.000127		3.361533	82.558415
☐	☐	Seasonal ARIMA(0, 1, 0)(1, 1, 2)12 No Intercept	280	11223.758	3458.6397	3469.5760	0.992	3452.6397	0.000115		3.359602	82.529934
☐	☐	Seasonal ARIMA(1, 1, 2)(2, 1, 1)12 No Intercept	277	10759.274	3447.9085	3469.7812	0.992	3435.9085	0.024513		3.214433	79.183276
☐	☐	Seasonal ARIMA(1, 1, 0)(1, 1, 1)12 No Intercept	280	11248.479	3458.9237	3469.8600	0.992	3452.9237	0.000099		3.356688	82.517122
☐	☐	Seasonal ARIMA(0, 1, 1)(1, 1, 1)12 No Intercept	280	11247.546	3458.9332	3469.8695	0.992	3452.9332	0.000099		3.356784	82.515793

Top 6 JMP Models	Parameter Estimates Valid	Parsimonious	Model Valid	SBC
Seasonal ARIMA(1,1,1)(1,1,1)12 No Intercept	Yes	No	No	3463.3
Seasonal ARIMA(0,1,0)(1,1,1)12 No Intercept	Yes	No	No	3464.3
Seasonal ARIMA(1,1,2)(1,1,1)12 No Intercept	Yes	Yes	Yes*	3464.6
Seasonal ARIMA(1,1,1)(0,1,2)12 No Intercept	Yes	No	No	3464.7
Seasonal ARIMA(2,1,2)(1,1,1)12 No Intercept	Yes	Yes	Yes*	3464.9
Seasonal ARIMA(2,1,1)(1,1,1)12 No Intercept	Yes	Yes	Yes*	3465.0

* See Appendixes A and B for JMP screenshots showing residuals of reviewed models.

6) SAS was used to identify outliers and level shifts for the two best models, and to improve the forecast accuracy. The process and results of the SAS evaluations are explained in Sections 6 and 7 below.
7) Independent of the JMP and SAS best model selection process, SAS Forecast Studio was used to identify the best valid model. The results of this evaluation are explained in Section 8 below.

Ancillary Steps:
In addition to the steps above, several attempts were made to find better models in other ways including:

a. Forecast Studio was used in a separate attempt to identify the best valid model. Unfortunately, the resulting models were no better than JMP, and in most cases worse. Notable is the residuals for the best models in JMP looked much worse in Studio.
b. Two transformations of the target were attempted to smooth the data: log and square root. Neither technique improved the model results and were not needed in the end.

6. Model 1 ARIMA(2,1,2)(1,1,1)12 No Intercept (SBC = 3410.37)

6.1 Steps Taken After Selection

Once ARIMA (2,1,2)(1,1,1)12 was selected as one of the two valid models through the JMP process (see Appendix A), the following steps were performed in SAS:

1) The team identified 8 level shifts at and 9 outliers over 2 iterations as part of model validation. A third iteration resulted in no added outliers. Summaries of the additive outliers, the level shifts, and the final iteration are shown to the right.

2) Model Validation: Team decided to change the method from ML to CLS because a convergence error was discovered. Additionally, when we checked the results using ML, we had to reject the white noise null hypothesis due to probability violation.

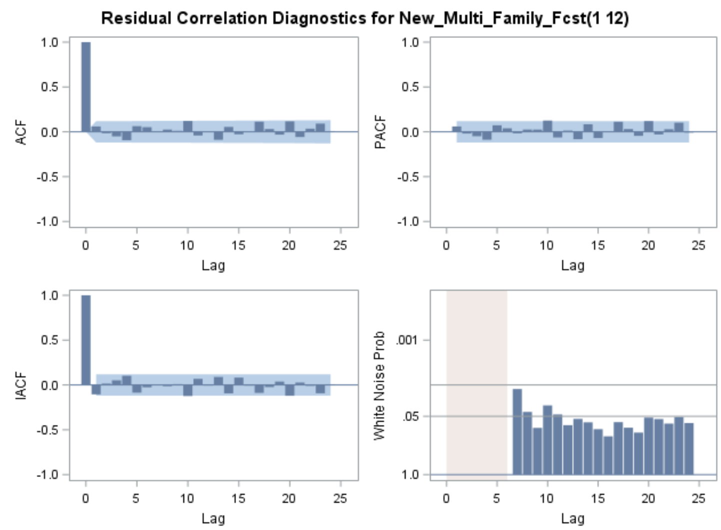
Using CLS, we see that parameter estimates are all statistically significant. See Appendix A.

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MA1,1	1.84234	0.02372	77.66	<.0001	1	New_Multi_Family_Fcst	0
MA1,2	-0.97247	0.02340	-41.55	<.0001	2	New_Multi_Family_Fcst	0
MA2,1	0.84199	0.07072	11.91	<.0001	12	New_Multi_Family_Fcst	0
AR1,1	1.77225	0.04416	40.14	<.0001	1	New_Multi_Family_Fcst	0
AR1,2	-0.84831	0.04419	-19.19	<.0001	2	New_Multi_Family_Fcst	0
AR2,1	0.44303	0.10473	4.23	<.0001	12	New_Multi_Family_Fcst	0
NUM1	-385.44384	82.12017	-4.69	<.0001	0	LS288	0
NUM2	261.63337	62.09242	4.21	<.0001	0	LS185	0
NUM3	-191.19174	52.18294	-3.66	0.0003	0	AO280	0
NUM4	249.57224	63.50715	3.93	0.0001	0	LS150	0
NUM5	187.42631	51.51731	3.64	0.0003	0	AO21	0
NUM6	211.36455	65.96993	3.20	0.0015	0	LS270	0
NUM7	-158.91841	51.55114	-3.08	0.0023	0	AO212	0
NUM8	177.82728	51.80116	3.43	0.0007	0	AO259	0
NUM9	180.50126	51.81791	3.48	0.0006	0	AO53	0
NUM10	-163.66213	51.81053	-3.16	0.0018	0	AO68	0
NUM11	-287.10241	66.90318	-4.29	<.0001	0	LS43	0
NUM12	236.08208	67.92649	3.48	0.0006	0	LS273	0
NUM13	142.99035	51.64947	2.77	0.0060	0	AO251	0
NUM14	154.85760	51.75799	2.99	0.0030	0	AO58	0
NUM15	-167.72046	59.88719	-2.80	0.0055	0	LS191	0
NUM16	186.00157	67.20904	2.77	0.0061	0	LS40	0
NUM17	126.41507	51.78604	2.44	0.0153	0	AO50	0

Variance Estimate	6896.925
Std Error Estimate	83.04773
AIC	3326.52
SBC	3410.365
Number of Residuals	283

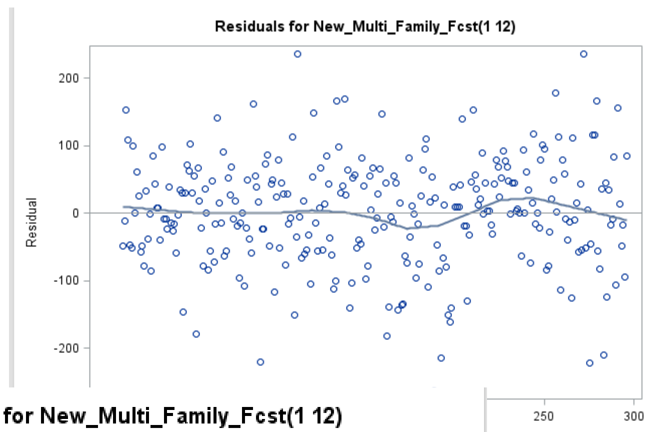
Outlier Detection Summary	
Maximum number searched	20
Number found	0
Significance used	0.005

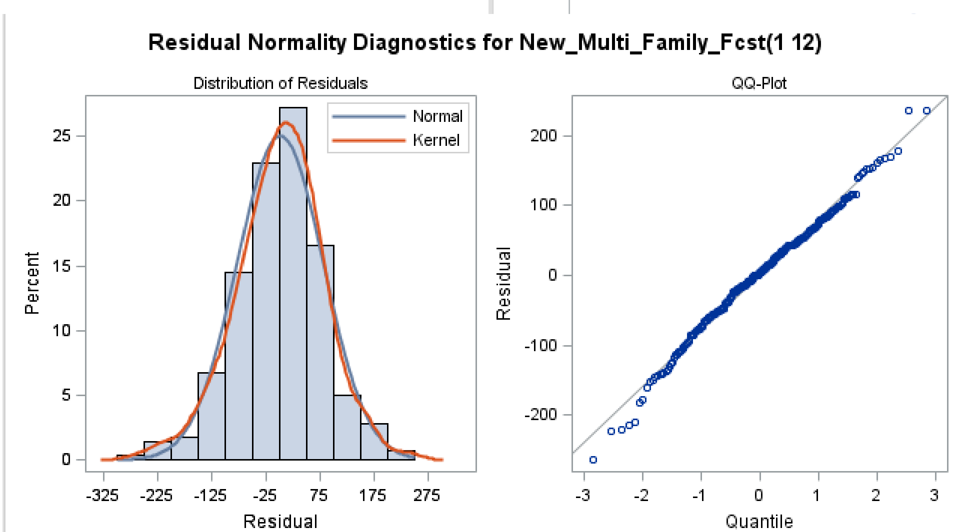
The white noise probabilities at a few of the lags are less than .05 indicating that they are statistically significant. The assumption of white noise has been violated and the model is not (strictly) valid. We decided not to reject the model but to continue and assess the model's forecasting ability.



In terms of other validations, we see that residuals do appear to be constant over time.

Additionally, a review of distribution does show that the residuals are normally distributed and the QQ plot appears to contain constants between -2 and 2.





6.2 Weaknesses & Responses

As noted above, the model does not perfectly meet stationarity and white noise assumptions. However, the model's results are close enough to validity that the slight issues were accepted .

6.3 Predictions & Business Conclusions

The team removed the last 12 months of actual data to determine if the model was able to predict Multi Family homes. Comparing actuals vs predicted values over the twelve-month period, the average percent error was 3.84%. The model showed a tendency to over predict rather than under predict.

Dates		Model 1	ARIMA(2,1,2)(1,1,1)12NoIntercept		
Date	Obs	Actual	Forecast	Variance	Variance %
9/1/17	297	5,024	5,028	4	0.08%
10/1/17	298	5,232	5,122	-110	-2.10%
11/1/17	299	5,015	4,978	-37	-0.73%
12/1/17	300	4,810	4,836	26	0.55%
1/1/18	301	4,502	4,893	391	8.69%
2/1/18	302	4,730	4,951	221	4.68%
3/1/18	303	4,741	5,245	504	10.62%
4/1/18	304	5,170	5,393	223	4.32%
5/1/18	305	5,205	5,338	133	2.56%
6/18/18	306	5,138	5,425	287	5.59%
7/18/18	307	4,959	5,194	235	4.74%
8/18/18	308	5,131	5,493	362	7.06%

	Avg Variance	Rank
Model 1	3.84%	3

6.4 Recommended Business Actions

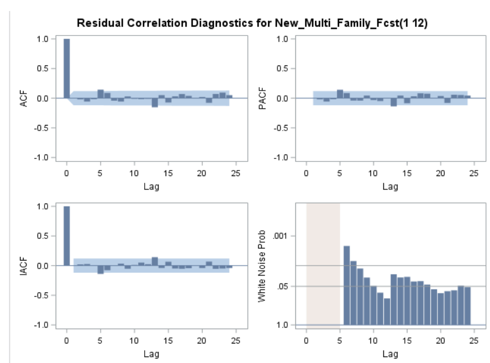
The team recommended evaluating other models to try to reduce the average percent forecast error.

7. Model 2 ARIMA(1,1,2)(1,1,1)12 No Intercept (SBC = 3426.27)

7.1 Steps Taken After Selection

ARIMA (1,1,2)(1,1,1)12 was selected as the best model through the JMP process due to its SBC score or rank, 3rd best, white noise probabilities not being significant, residuals plot showed mean of zero and non-constant variance (see Appendix B). The following steps were performed in SAS:

1. 5 level shifts and 9 outliers over 3 iterations were identified using an alpha of 0.005. See Appendix B.
2. The estimation method was changed from Maximum likelihood (ML) to Conditional Least Squares (CLS) due to a convergence error.
3. 12 Months were forecasted after identifying all level shifts and outliers.



7.2 Weaknesses & Responses

In the SAS-produced model the residual plots show a mean of zero and evidence constant variance, but the white noise probability plot shows that the white noise null hypothesis is violated. However, the model's prediction results are close enough that the slight issues were accepted. The forecasted values are between -2.5% and 7% error range that provides assurance that the model is predicting within an acceptable range.

Variance Estimate	7629.229
Std Error Estimate	87.34546
AIC	3360.656
SBC	3426.274
Number of Residuals	283

7.3 Predictions & Business Conclusions

Using SAS the team forecasted the last 12 months that were removed from the original dataset to determine if the model was able to predict Multi Family homes. Based on the comparison between actuals and forecasts the model was off by 2% on average or between -2.5% and 7.9%. This can be considered a high degree of accuracy and reinforced the decision to accept slight deviations from valid model assumptions, such as white noise probabilities.

Dates		Model 2	ARIMA(1,1,2)(1,1,1)12NoIntercept		
Date	Obs	Actual	Forecast	Variance	Variance %
9/1/17	297	5,024	5,031	7	0.13%
10/1/17	298	5,232	5,100	-132	-2.51%
11/1/17	299	5,015	4,961	-54	-1.08%
12/1/17	300	4,810	4,808	-2	-0.05%
1/1/18	301	4,502	4,830	328	7.30%
2/1/18	302	4,730	4,865	135	2.85%
3/1/18	303	4,741	5,117	376	7.92%
4/1/18	304	5,170	5,243	73	1.41%
5/1/18	305	5,205	5,189	-16	-0.30%
6/18/18	306	5,138	5,277	139	2.71%
7/18/18	307	4,959	5,060	101	2.03%
8/18/18	308	5,131	5,320	189	3.68%

	Avg Variance	Rank
Model 2	2.01%	2

7.4 Recommended Business Actions

As noted above, the model was able to predict 12 months with a high degree of accuracy, 2% error rate on average. Therefore, the model's target audience should consider referencing the predictions to help guide business decisions.

8. Model 3 ARIMA(0,1,0)(1,0,1)12 SAS No Intercept (SBC = 2754.33)

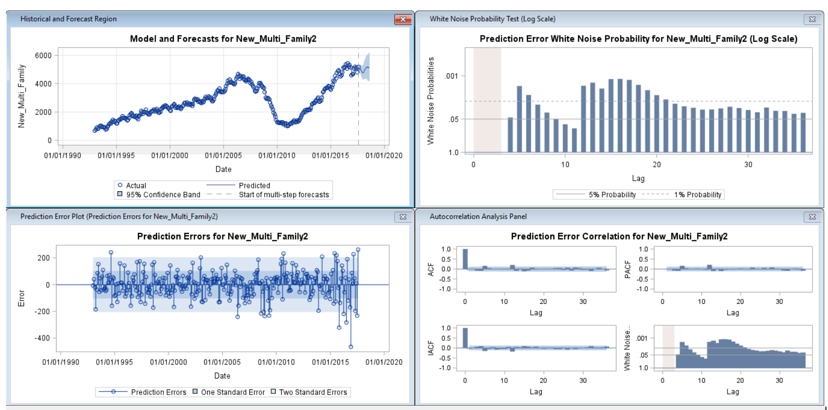
Model 3 was developed in a separate modeling effort in SAS Forecast Studio.

8.1 Steps Taken

To validate the model selections, the team decided to run the data in SAS Forecast studio. See Appendix C which graphically presents the methodology used.

The team decided to configure the model using a 12-month cycle and set outlier detection filters that aligned to Model 1 and Model 2's configurations.

Forecast Studio identified 'ARIMA(0,1,0)(1,0,1)12 No Intercept' as the best model based upon SBC. This did not agree with the selected Model 1 and model 2 from JMP/SAS, and was taken as a data point for consideration, but not a reason to stop using the previous two models.



	Selection Statistics of Fit							
Model	AADJRSQ	ADJRSQ	AIC	AICC	SBC	SMAPE	SSE	UMSE
Generated ARIMA Model (LEAF_0)	0.99225	0.99238	2753.537707	2753.675638	2768.285609	3.63	3249081	11165
Generated Smoothing Model (LEAF_1)	0.99279	0.99289	2743.263462	2743.345654	2754.334541	3.41	3072098	10485

Like Model 1, the white noise probabilities at a few of the lags are less than .05 indicating that they were statistically significant. The violation in the white noise test meant that the model was not (strictly) valid, however, the results were close enough that the model's predictions were accepted.

8.2 Predictions & Business Conclusions

The team removed the last 12 months of actual data and used the model to predict these amounts. The prediction accuracy average for the 12 periods was an impressive 98.95%.

This further emphasized that we could live with some evidence of white noise in our model.

See Appendix C for SAS output.

Dates		Model 3 SAS Forecast Studio			
Date	Obs	Actual	Forecast	Variance	Variance %
9/1/17	297	5,024	5,036	12	0.24%
10/1/17	298	5,232	5,055	-177	-3.39%
11/1/17	299	5,015	4,984	-31	-0.62%
12/1/17	300	4,810	4,841	31	0.65%
1/1/18	301	4,502	4,783	281	6.24%
2/1/18	302	4,730	4,825	95	2.01%
3/1/18	303	4,741	4,973	232	4.90%
4/1/18	304	5,170	5,077	-93	-1.80%
5/1/18	305	5,205	5,147	-58	-1.11%
6/18/18	306	5,138	5,192	54	1.04%
7/18/18	307	4,959	5,108	149	3.01%
8/18/18	308	5,131	5,205	74	1.44%

	Avg Variance	Rank
Model 1	3.84%	3
Model 2	2.01%	2
Model 3	1.05%	1

9. Conclusion

9.1 Chosen Best Model

All three of the models had impressive forecasting accuracies. That said, based on SBC and the slightly better accuracy, Model 3 (ARIMA(0,1,0)(1,0,1)12 No Intercept) was chosen as the best model.

MODEL	SBC	ACCURACY
Model 1: ARIMA(2,1,2)(1,1,1,)12 No Intercept	3410.37	96.16%
Model 2: ARIMA(1,1,2)(1,1,1,)12 No Intercept	3426.27	97.99%
Model 3: ARIMA(0,1,0)(1,0,1,)12 No Intercept	2754.33	98.95%

9.2 Prediction Accuracy

As noted in the Model discussions sections above, Model 3 had an average 12-month forecast error of 1.05%.

9.3 Final Thoughts

The chosen data set from the Census Bureau website proved to be a good time series for analysis. Given that real-world data typically does not result in textbook models, the results from this project were impressive, especially the accuracy of the chosen models and the only slight deviations from model assumptions.

As mentioned previously, similar data exists at regional and local levels, and similar analyses at lower levels would provide more operationally-useful insight for businesses, investors, and individuals making decisions related to local multi-family trends.

There were also many additional targets available in the data set, which we would consider using these methods to derive similar analyses (e.g. time series for single family construction, shopping malls, transportation, and so on).

Appendices

Appendix A - Model 1 Exhibits

Model: Seasonal ARIMA(2, 1, 2)(1, 1, 1)12 No Intercept

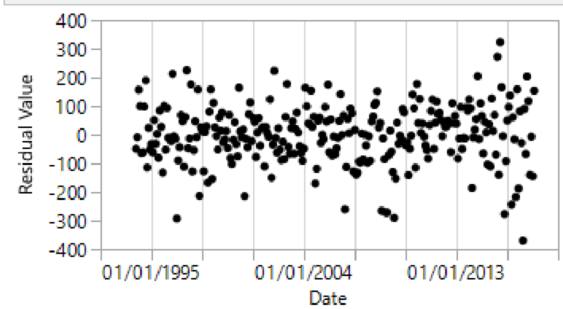
Model Summary

DF	277	Stable	Yes
Sum of Squared Errors	2924064.89	Invertible	Yes
Variance Estimate	10556.1909		
Standard Deviation	102.743325		
Akaike's 'A' Information Criterion	3443.02718		
Schwarz's Bayesian Criterion	3464.89986		
RSquare	0.99223751		
RSquare Adj	0.99209739		
MAPE	3.19707255		
MAE	78.6670007		
-2LogLikelihood	3431.02718		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	1.645388	0.1061370	15.50	<.0001*
AR1,2	1	2	-0.729708	0.1044800	-6.98	<.0001*
AR2,12	2	12	0.291932	0.0871186	3.35	0.0009*
MA1,1	1	1	1.704279	0.0821283	20.75	<.0001*
MA1,2	1	2	-0.843456	0.0765820	-11.01	<.0001*
MA2,12	2	12	0.876766	0.0600306	14.61	<.0001*

Residuals



Lag	AutoCorr	Ljung-Box Q	p-Value	Lag	Partial
0	1.0000	.	.	0	1.0000
1	-0.0004	0.0000	0.9951	1	-0.0004
2	0.0248	0.1767	0.9154	2	0.0248
3	-0.0048	0.1833	0.9802	3	-0.0048
4	-0.1028	3.2384	0.5187	4	-0.1035
5	0.1169	7.2033	0.2060	5	0.1184
6	0.0168	7.2852	0.2953	6	0.0214
7	-0.0715	8.7776	0.2690	7	-0.0815
8	0.0257	8.9720	0.3447	8	0.0176
9	0.0115	9.0106	0.4363	9	0.0422
10	0.0317	9.3067	0.5033	10	0.0183
11	0.0263	9.5120	0.5747	11	0.0040
12	-0.0246	9.6926	0.6429	12	-0.0040
13	-0.0280	9.9261	0.7000	13	-0.0255
14	0.0573	10.9100	0.6931	14	0.0536
15	-0.0304	11.1875	0.7392	15	-0.0315
16	-0.0299	11.4580	0.7804	16	-0.0406
17	0.0974	14.3322	0.6435	17	0.1068
18	-0.0080	14.3519	0.7059	18	0.0112
19	0.0512	15.1530	0.7128	19	0.0187
20	0.0681	16.5732	0.6805	20	0.0672
21	-0.0245	16.7576	0.7257	21	0.0113
22	-0.0154	16.8310	0.7726	22	-0.0498
23	0.0401	17.3291	0.7928	23	0.0467
24	0.0183	17.4339	0.8296	24	0.0415
25	-0.0159	17.5134	0.8623	25	-0.0479

JMP outputs that showed a valid and parsimonious model – Model 1.

288	4602.0604	83.0477	4439.2898	4764.8309	4593.0000	-9.0604
289	4595.4997	83.0477	4432.7291	4758.2702	4679.0000	83.5003
290	4806.2137	83.0477	4643.4432	4968.9843	4699.0000	-107.2137
291	4904.7395	83.0477	4741.9690	5067.5101	5060.0000	155.2605
292	5169.3046	83.0477	5006.5341	5332.0752	5183.0000	13.6954
293	5115.3124	83.0477	4952.5418	5278.0829	5066.0000	-49.3124
294	5167.1376	83.0477	5004.3670	5329.9082	5149.0000	-18.1376
295	4924.6642	83.0477	4761.8936	5087.4347	4830.0000	-94.6642
296	5115.1158	83.0477	4952.3452	5277.8863	5200.0000	84.8842
297	5027.8281	83.0477	4865.0575	5190.5986	.	.
298	5121.9599	113.4061	4899.6880	5344.2317	.	.
299	4978.1631	137.2013	4709.2534	5247.0727	.	.
300	4836.2841	159.9156	4522.8552	5149.7130	.	.
301	4893.1432	183.9406	4532.6263	5253.6601	.	.
302	4951.3806	210.4248	4538.9556	5363.8057	.	.
303	5244.5838	239.7147	4774.7517	5714.4159	.	.
304	5393.4644	271.5635	4861.2097	5925.7192	.	.
305	5338.0930	305.3222	4739.6726	5936.5135	.	.
306	5425.0664	340.1249	4758.4339	6091.6989	.	.
307	5194.2849	375.0502	4459.2001	5929.3697	.	.
308	5493.2828	409.2415	4691.1841	6295.3814	.	.

SAS output of predictions after additive outliers and level shifts are controlled for in Model 1.

Appendix B - Model 2 Exhibits

Model: Seasonal ARIMA(1, 1, 2)(1, 1, 1)12 No Intercept

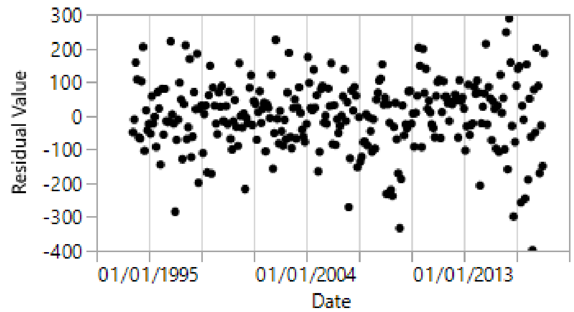
Model Summary

DF	278	Stable	Yes
Sum of Squared Errors	2985925.68	Invertible	Yes
Variance Estimate	10740.7398		
Standard Deviation	103.637541		
Akaike's 'A' Information Criterion	3446.34636		
Schwarz's Bayesian Criterion	3464.5736		
RSquare	0.99208851		
RSquare Adj	0.99197467		
MAPE	3.22813074		
MAE	79.3644463		
-2LogLikelihood	3436.34636		

Parameter Estimates

Term	Factor	Lag	Estimate	Std Error	t Ratio	Prob>|t|
AR1,1	1	1	0.9381210	0.0392190	23.92	<.0001*
AR2,12	2	12	0.2325957	0.0815047	2.85	0.0046*
MA1,1	1	1	0.9815078	0.0723200	13.57	<.0001*
MA1,2	1	2	-0.1345654	0.0636637	-2.11	0.0354*
MA2,12	2	12	0.8662310	0.0589682	14.69	<.0001*

Residuals



Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.0024		0.0017	0.9673	1	0.0024	
2	-0.0346		0.3455	0.8414	2	-0.0346	
3	-0.0348		0.6941	0.8746	3	-0.0347	
4	-0.0805		2.5671	0.6327	4	-0.0818	
5	0.1624		10.2212	0.0692	5	0.1615	
6	0.0643		11.4263	0.0761	6	0.0577	
7	-0.0409		11.9153	0.1034	7	-0.0376	
8	0.0296		12.1717	0.1437	8	0.0386	
9	0.0109		12.2070	0.2019	9	0.0400	
10	0.0163		12.2851	0.2664	10	-0.0012	
11	-0.0028		12.2875	0.3424	11	-0.0266	
12	-0.0231		12.4460	0.4106	12	-0.0067	
13	-0.0709		13.9489	0.3775	13	-0.0762	
14	0.0159		14.0245	0.4479	14	0.0021	
15	-0.0794		15.9201	0.3874	15	-0.0937	
16	-0.0709		17.4389	0.3578	16	-0.0774	
17	0.0700		18.9256	0.3328	17	0.0642	
18	-0.0262		19.1347	0.3836	18	-0.0123	
19	0.0275		19.3651	0.4336	19	0.0181	
20	0.0531		20.2300	0.4436	20	0.0743	
21	-0.0313		20.5317	0.4878	21	0.0235	
22	-0.0115		20.5724	0.5473	22	-0.0263	
23	0.0540		21.4767	0.5520	23	0.0682	
24	0.0306		21.7677	0.5931	24	0.0480	
25	-0.0090		21.7928	0.6477	25	-0.0394	

JMP outputs that showed a valid and parsimonious model – Model 2.

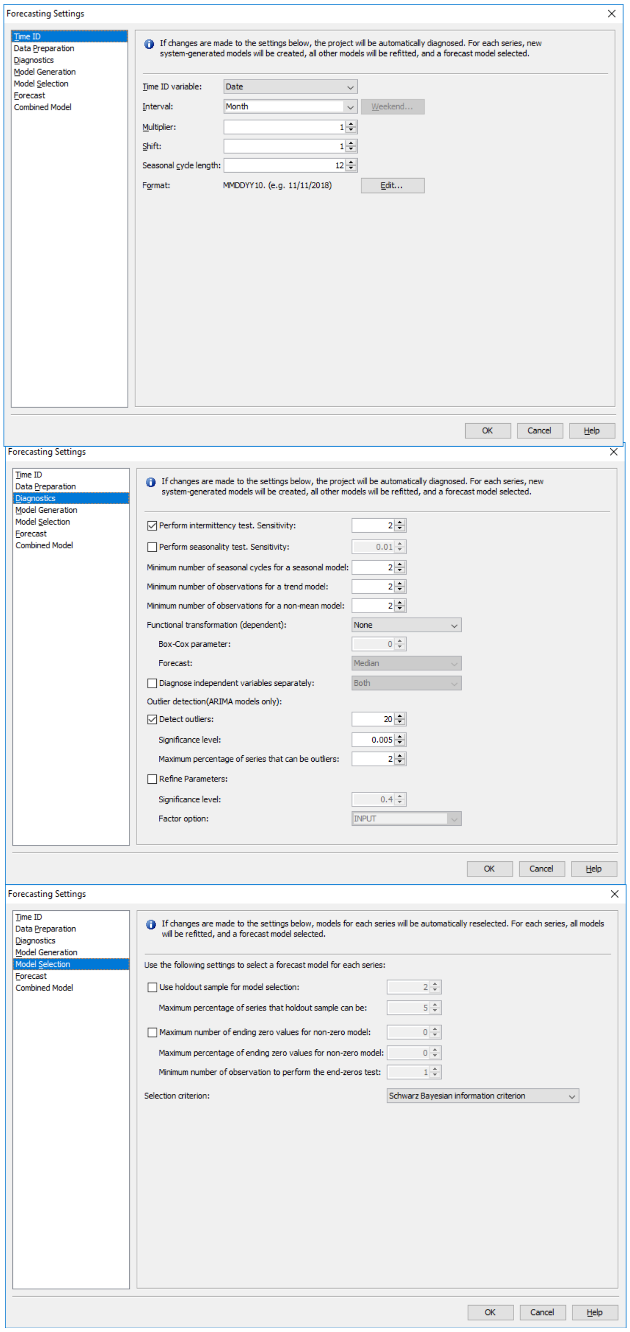
288	4594.3224	87.3783	4423.0641	4765.5807	4593.0000	-1.3224
289	4583.6854	87.3783	4412.4271	4754.9437	4679.0000	95.3146
290	4803.6029	87.3717	4632.3575	4974.8483	4699.0000	-104.6029
291	4904.8630	87.3717	4733.6176	5076.1083	5060.0000	155.1370
292	5191.4431	87.3716	5020.1978	5362.6883	5183.0000	-8.4431
293	5158.3559	87.3716	4987.1107	5329.6011	5066.0000	-92.3559
294	5189.3472	87.3716	5018.1021	5360.5924	5149.0000	-40.3472
295	4939.6799	87.3716	4768.4348	5110.9250	4830.0000	-109.6799
296	5060.5016	87.3716	4889.2565	5231.7467	5200.0000	139.4984
297	5030.6792	87.3455	4859.4853	5201.8732	.	.
298	5100.4522	125.7173	4854.0508	5346.8536	.	.
299	4960.9510	160.2542	4646.8585	5275.0434	.	.
300	4807.7716	193.1108	4429.2814	5186.2618	.	.
301	4830.4302	225.0515	4389.3375	5271.5230	.	.
302	4864.8115	256.4094	4362.2583	5367.3647	.	.
303	5116.6708	287.3432	4553.4885	5679.8530	.	.
304	5242.8543	317.9301	4619.7228	5865.9857	.	.
305	5189.2629	348.2063	4506.7910	5871.7347	.	.
306	5277.1652	378.1869	4535.9326	6018.3979	.	.
307	5059.5939	407.8751	4260.1734	5859.0144	.	.
308	5319.6531	437.2686	4462.6223	6176.6839	.	.

SAS output of predictions after additive outliers and level shifts are controlled for in Model 2.

Maximum Likelihood Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag	Variable	Shift
MA1,1	0.89884	0.08051	11.16	<.0001	1	New_Multi_Family_Fcst	0
MA1,2	-0.06973	0.06547	-1.07	0.2868	2	New_Multi_Family_Fcst	0
MA2,1	0.89164	0.08484	10.51	<.0001	12	New_Multi_Family_Fcst	0
AR1,1	0.93403	0.05090	18.35	<.0001	1	New_Multi_Family_Fcst	0
AR2,1	0.40277	0.10028	4.02	<.0001	12	New_Multi_Family_Fcst	0
NUM1	-391.85423	86.37109	-4.54	<.0001	0	LS288	0
NUM2	287.00568	75.99876	3.78	0.0002	0	LS185	0
NUM3	-194.15039	53.39789	-3.64	0.0003	0	AO280	0
NUM4	-258.56842	75.36281	-3.43	0.0006	0	LS162	0
NUM5	-169.24245	52.95658	-3.20	0.0014	0	AO212	0
NUM6	-209.59331	76.04797	-2.76	0.0059	0	LS275	0
NUM7	169.69168	53.02881	3.20	0.0014	0	AO21	0
NUM8	173.58606	53.00459	3.27	0.0011	0	AO53	0
NUM9	-213.29355	75.43830	-2.83	0.0047	0	LS198	0
NUM10	166.57066	53.09498	3.14	0.0017	0	AO259	0
NUM11	163.72769	52.98423	3.09	0.0020	0	AO67	0
NUM12	148.45096	52.92735	2.80	0.0050	0	AO50	0
NUM13	144.75758	53.17914	2.72	0.0065	0	AO251	0

Additive outliers and level shift summary for Model 2.

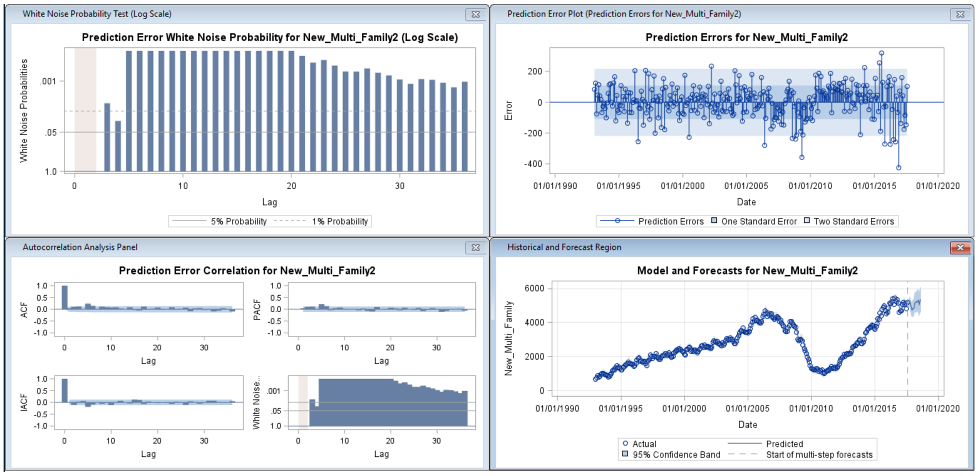
Appendix C - SAS Forecast Studio Model Exhibits



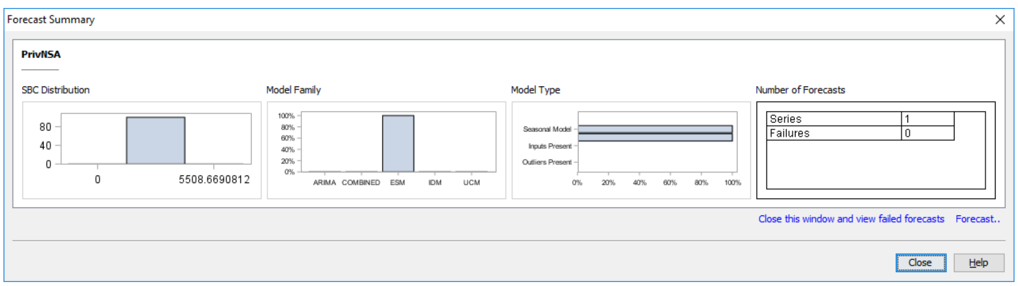
Model 3 parameter settings.

SBC = 2754.334541

Generated ARIMA Model (LEAF_0)



Model 3 Validity Tests



Model 3 Forecast Summary

Smoothing Model (LEAF_1)

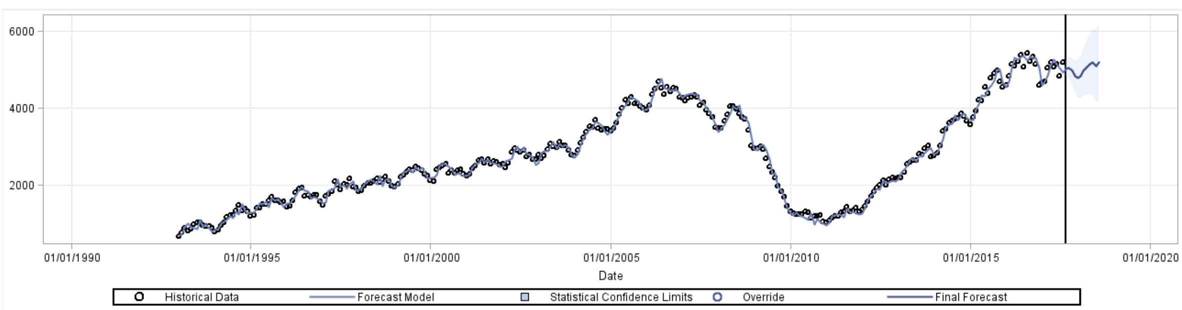
Name: LEAF_0
Description: "ARIMA: New_Multi_Family2 ~ P = (12) D = (1) Q = (12) NOINT"
Details: "ARIMA: New_Multi_Family2 ~ P = (12) D = (1) Q = (12) NOINT"
Model family: ARIMA
Model type: GENERALARIMA
Source: HPFDIAGNOSE **Intercept:** None
Forecast variable: New_Multi_Family2
Delay: 0
Differencing: (1)
P: (12)
Q: (12)

Estimation Options

Method: CLS
Convergence criterion: 0.001
Number of iterations: 50
Delta: 0.001
Singularity criterion: 1E-7
Grid value: 0.005
Restrict parameters to stable values: Yes
NOLS: 0

Subset ARIMA Model (LEAF_0)

Name: LEAF_1
Description: "Winters Method (Additive)"
Details: "Winters Method (Additive)"
Model family: ESM
Model type: ESM
Source: HPFDIAGNOSE**Model code:** ADDWINTERS
Selection code: SBC
Transform: NONE
Forecast option: MEAN
Estimation Options
  Component: LEVEL
  Lower: 0.001
  Upper: 0.999
  Component: TREND
  Lower: 0.001
  Upper: 0.999
  Component: SEASON
  Lower: 0.001
  Upper: 0.999
  Component: DAMPING
  Lower: 0.001
  Upper: 0.999



Active series

	10/01/2017	11/01/2017	12/01/2017	01/01/2018	02/01/2018	03/01/2018	04/01/2018	05/01/2018	06/01/2018	07/01/2018	08/01/2018
Historical Data	.	.	.	.	.	.	.	.	.	.	.
Forecast Model	**5054.8281962**	**4983.7168292**	**4841.4273109**	**4783.1092642**	**4825.0034233**	**4973.096096**	**5077.1102971**	**5147.2117899**	**5191.5177681**	**5108.0774031**	**5205.0351588**
Override											
Final Forecast	**5054.8281962**	**4983.7168292**	**4841.4273109**	**4783.1092642**	**4825.0034233**	**4973.096096**	**5077.1102971**	**5147.2117899**	**5191.5177681**	**5108.0774031**	**5205.0351588**

Model 3 Results

Appendix D – Original Data Set Variable List

Date	Drug store	Auxiliary building
Total Private Construction 1	Building supply store	Amusement and recreation
Residential (inc.Improvements)2	Date	Theme/amusement park
New single family	Other stores	Sports
New multi-family	Warehouse	Date
Nonresidential	General commercial	Fitness
Lodging	Mini-storage	Performance/meeting center
Office	Health Care	Social center
General	Hospital	Movie theater/studio
Date	Medical building	Transportation
Financial	Special care	Air
Commercial (inc. Farm)	Date	Land
Automotive	Educational	Communication
Sales	Preschool	Power (inc. Gas and Oil)
Service/parts	Primary/secondary	Date
Parking	Higher education	Electric
Food/beverage	Instructional	Manufacturing
Food	Dormitory	Food/beverage/tobacco
Date	Sports/recreation	Chemical
Dining/drinking	Other educational	Plastic/rubber
Multi-retail	Date	Nonmetallic mineral
General merchandise	Gallery/museum	Fabricated metal
Shopping center	Religious	Computer/electronic/electrical
Shopping mall	House of worship	Transportation equipment
Other commercial	Other religious	