TESTING MARKETING HYPOTHESIS

BY

JOHN C. PICKETT
PROFESSOR OF ECONOMICS
UNIVERSITY OF ARKANSAS AT LITTLE ROCK
2801 S. UNIVERSITY AVE.
LITTLE ROCK, AR 72204
FAX: 501.569.8898
PHONE: 501.569.8878
JCPICKETT@UALR.EDU

AND

DAVID P. REILLY
SENIOR VICE-PRESIDENT
AUTOMATIC FORECASTING SYSTEMS, INC.
PO BOX 563
HATBORO, PA 19040
FAX: 215.672.2534
PHONE: 215.675.0652
DAVE@AUTOBOX.COM

ABSTRACT

Avoiding cannibalization of an existing store's sales as a result of opening a new store is a perennial challenge facing a firm's store expansion policies. The research question is to identify the statistical techniques that can be used to determine whether or not a new store cannibalizes the sales of an existing store.

Using OLS regression techniques, the initial conclusion is that new store sales have a positive effect on existing store sales. Examination of a plot of the residuals suggests the OLS model is insufficient, and the presence of autocorrelation leads the analyst to conclude that the OLS model is invalid.

The application of time series techniques concludes (1) that new store sales do not affect existing store sales, (2) that a strong weekly effect is correctly modeled by the inclusion of seasonal pulses for Saturday and Sunday effects, and (3) that the inclusion of nine pulse variables accounts for the effects of the previously identified outliers. These conclusions offer convincing insights into a firm's store expansion policies.

THE RESEARCH QUESTION

Avoiding cannibalization of an existing store's sales as a result of opening a new store is a perennial challenge facing a firm's store expansion policies. Some form of a territorial protection policy is frequently adopted to ensure the continued success of existing stores and to enhance the success of new stores. The press reported that the opening of the second casino in Atlantic City significantly increased the first casino's business. This phenomenon continued until recently as the market has become saturated.

Cannibalization is a broadly defined term. A search on Google will return a long list of examples, papers, and articles addressing this issue. Microsoft's recent release of Vista will cannibalize XP sales, especially since Microsoft plans to adopt a strategy of forced migration to Vista. Pharmaceutical companies release new drugs for which there are no substitutes. Other times, the FDA approves new drugs "me-to drugs" that offer little improvement in the therapeutic value of existing compounds.

Another example of cannibalization occurs when a business opens another outlet in the same area as an existing shop. The two establishments may or may not compete, i.e., the new outlet may or may not cannibalize the sales of the existing outlet. All businesses have a new store opening policy.[1] Businesses maintain sales in units and dollars for all stores and product lines within stores. These data series are constantly monitored to gleam trends and patterns of unit and dollar sales to support the businesses' proactive expansion policies. Given the existence of data for store sales, what statistical techniques can be used to determine whether or not a new store cannibalizes the sales of an existing store?

DATA SERIES

We have data series for two different store locations recording total unit sales by day of the week.[2] See Figures A and B for time series plots of the data and Table 1 for descriptive summary statistics. The line of business is providing supplies to businesses.[3] The data are daily unit sales for 180 consecutive periods, which we call a time series.

A visual examination of the time series plots reveal some anomalous data and detect what appears to be a strong day-of-the-week pattern. A standard bivariate scatter plot (Figure C) is often useful in

exploratory data analysis as a backdrop to standard regression and correlation results. The main body of data stands out against the anomalous points. First, one observes the data in a cluster. Second, the spread of the plot indicates the presence of numerous seemingly extreme values, perhaps caused by unknown promotions, competitor activity, or unpredictable events. The third is the plot of the regression line passing through the middle of the scatter. The usual and customary statistical technique for summarizing the relationship between the two variables, existing store sales vs. new store sales, is by using OLS regression techniques. The regression equation for the data is:

Existing Store Sales = 36,200 + 1.16 New Store Sales
(.0000) (.0538)

The initial interpretation of the regression indicates the intercept term is necessary and the slope is significant at alpha = .0538. The initial conclusion is that new store sales have a positive effect on existing store sales much like the aforementioned Atlantic City casino results. Sales in the old store increase at a rate of 116% of the new store sales reflecting what some might argue was the result of the promotion of the new store. This is sometimes referred to as a branding effect. Now comes the marketing staff to explain why the existing store sales are positively related to the new store sales. One hypothesis is that a marketing campaign to announce the opening of the new store has a spillover effect on the existing store's sales.[4] The regression results support this hypothesis, and an unsuspecting analyst might conclude that existing store sales are positively affected by the new store, and more stores should be considered. However, as we shall see this conclusion is premature and seriously flawed.

But first, examine the plot of the residuals from the OLS equation in Figure D. Inspection reveals two conclusions. First, the residuals are clustered around zero, as they should be. Second, there are a number, perhaps as high as 10 or more extreme values or outliers and an apparent within week predictable pattern. It is well known that the estimated parameters resulting from the application of OLS techniques are sensitive to extreme values and omitted deterministic variables like day-of-the-week effects. The usual fix-up for extreme values is to create a dummy variable for each

outlier identified by inspection. However, in many cases outliers cannot always be easily identified by inspection due to the auto-projective structure and changes in level or trend as we show below.

DETERMINING THE VALIDITY OF THE REGRESSION RESULTS

Those using OLS regression techniques to summarize the relationship of time series data typically do not verify the results by applying various tests to determine the validity of the assumptions underlying OLS techniques. Addressing the validity question of the assumptions simply asks the question "Are the regression results true or do they reflect the impact of omitted series?" A classic case of spurious correlation is the relationship between the number of fireman at a fire (X) and the reported damage (Y). The omitted (concomitant) variable is 'the intensity off the fire' that affects both the Y and X, and its omission results in the incorrect conclusion. Another example often quoted is the significant correlation between the number of churches (X) and the number of crimes (Y) and other measurable social phenomenon.

In this case, "Is it true that existing store sales are statistically related to new store sales?" The technique used to determine whether or not the result is true or false is to test the assumptions underlying the regression analysis.[5] These are the sufficiency conditions to determine the validity of the statistical results. The assumptions are well known, but bear repeating now.[6]

(1) The expected value of the residuals is zero.

$E(\varepsilon_i) = 0$

(2) The variance of the residuals is constant.

$\sigma_i^2 = k$

(3) The residuals are not autocorrelated.

$Cov(\varepsilon_i,\varepsilon_j) = 0$

(4) The residuals are normally distributed.

$\varepsilon \approx NID$

(5) The regressors are independent variables.

$X_i \neq f(\varepsilon_i)$ and $\varepsilon_i \neq f(X_i)$

Failure of any one of the first four assumptions means the functional form is incorrect. Failure of the fifth assumption means the relationship between Y and X is not one-way from right to left. Rather, the failure of the fifth assumption means the relationship between Y and X should be estimated using simultaneous equations.[7]

Testing whether the expected value of the residuals equals zero confirms that the first assumption is satisfied. The second assumption is tested by dividing the residuals into two groups and testing the equality of the two variances. Division of the residuals into two groups followed by testing the equality of the two variances confirms the second assumption is satisfied.[8]

The third assumption is where the application of OLS regression techniques to summarize the relationship of two time series data almost always can be shown to be an incorrect statistical technique. Table 2 shows the autocorrelation and partial autocorrelation coefficients and their t-values for lags 1 – 14. If the Durbin-Watson statistic had been reported, then autocorrelation at lag 1 would not be significant. However, the standard Durbin-Watson statistic only tests for autocorrelation at lag 1 and requires that the mean of the residuals is zero everywhere (assumption 1 above). Table 2 reveals that autocorrelation is present at lag 7, perhaps due to the need to incorporate auto projective structure at this lag or a deterministic variable representing a day-of-the-week effect. Given that the data is daily, the analyst (and his or her readily available software package) might incorrectly assume that the autocorrelation at lag 7 be significant, i.e., all Mondays are related, all Tuesdays are related, etc. Note also that the partial autocorrelation coefficient at lag 7 is statistically significant supporting the idea that some form of augmentation is required in order to meet the Gaussian assumptions. OLS techniques can be used to estimate a lagged value on X so long as the model is an autoregressive form or a specific day-of-the-week effect. If the model form for lag 7 is a moving average, then OLS techniques cannot be used.

The conclusion is that the test for the third assumption fails, which leads the analyst to conclude that the simple model shown above is insufficient and needs augmentation.[9]

The fourth assumption focuses on whether or not the residuals are normally distributed.[10] Figure E is the Anderson-Darling test for normalcy. The p-value < .005 concludes the residuals are normally distributed.

The fifth assumption focuses on determining whether or not the dependence is one-way from right to left-hand side or simultaneously determined. This assumption is usually tested by estimating two OLS regressions. The first is:

$$\text{Residuals} \neq f(\text{New Store Unit Sales})$$

The second is:

$$\text{New Store Unit Sales} \neq f(\text{Residuals})$$

We have estimated both equations and determined that the null hypothesis is not rejected, i.e., that there is no relationship among X and ε in either equation. Hence, the fifth assumption is satisfied, and the linear equation is appropriate.

IDENTIFYING THE VALID TIME SERIES MODEL

The next step in the analysis of the two data series is to tentatively identify potential deficiencies.[11] We are searching for structural variables that are associated with the outliers and the unknown deterministic variables.[12] The manual procedure is often to augment the user specified casual series with six indicator variables reflecting the effects of the days-of-the-week. While the manual procedure is routinely followed, it is flawed by its omission of deterministic variables such as pulses, level shifts, seasonal pulses, and local time trends.

Table 3 shows a typical augmentation to include dummy variables for six of the seven days of the week. Using the data from Table 3, Table 4 reports the augmented model. Note the coefficient on the new store's unit sales declined from 1.16 in the simple OLS regression to .492 in the augmented model

and is not statistically significant. Only three of the six dummy variables for days-of-the-week are statistically significant.

Figure F is a plot of the residuals from the augmented model. Even with the outliers untreated there is a clear suggestion that certain days of the week have a significant effect on the existing store's unit sales. Since days of the week also affect the new store's unit sales, we immediately acknowledge that the 'background variable' induced the spurious correlation between the existing store's unit sales and the new store's unit sales as shown by the alpha of .4030 for the coefficient for the new store's unit sales.

Referring to Figure D shows the plot of the residuals resulting from the initial OLS regression. There are 10 outliers suggesting that the analysis needs to include 10 intervention variables. If we incorporate the 10 indicator variables to account for the most serious anomalies, then Table 5 results. This again suggests that the new store's unit sales are statistically insignificant (alpha = 0.2086).

Inclusion of the 10 intervention variables results in the X variable, new store's unit sales, being not significant as evidenced by its p-value and t-value. The initial hypothesis asserted by the marketing staff is invalid; i.e., that there is a spillover effect resulting from the marketing campaign supporting the opening of the new store to the existing store's unit sales.

Given the tentative identification of the time series model, the next step in the analysis is to begin a step-down analysis to eliminate the statistically insignificant variables, to determine if the parameters are consistent through time, and to model the residuals. After this effort, the final model is shown in Table 6.

The final model results in a large increase in the adjusted R^2 from .025 to .895 and the reduction in the standard error from 3,306 to 994. The necessary and sufficient conditions for a valid equation are satisfied. The necessary condition is that all parameters in the estimated model are statistically significant. The sufficient conditions include the acceptance of all of the five assumptions about the residuals. All of the sufficiency conditions are met and so the model is valid.

CONCLUSION

Given these results, the analyst may conclude the following about the relationship between the existing store unit sales and new store unit sales. New store's unit sales have no statistical effect on the existing store's unit sales. In addition, the analyst may conclude that existing store's unit sales has no identifiable stochastic trend, has a long-term memory structure in terms of shocks (innovations), has a short term memory structure in terms of observations, and appears to have either a model and/or parameter changes at week 38 and day 6, so the final model used only the last 48 observations to fine tune the model. Finally, the analyst did not estimate the model using any power transformations on either the existing or new store's unit sales.

The results of the statistical analysis allow the analyst to further conclude that:

1. The statistical identification of the two seasonal pulses (X1 and X2 in Table 6) indicates that the X2 day = 4 is Saturday and X1 day = 5 is Sunday. These seasonal pulses cannot be identified by inspection of Figure D. Existing store's unit sales decline (the estimated coefficients are both negative) on both Saturday and Sunday.
2. There is a strong relationship between all Mondays, Tuesdays … and Sundays for the existing store's unit sales.
3. And most intriguing, asking the firm's marketing staff and existing store's management team to identify exactly what may have occurred during the weeks and days of the weeks that are associated with the interventions (pulses) included in the model. The marketing staff and store management want to repeat those actions and programs that are associated with intervention variables that have positive coefficients and avoid those actions and programs that are associated with intervention variables that have negative coefficients.[13]

REFERENCES

John C. Pickett, David P. Reilly and Robert M. McIntyre (2005), "How to Select a Most Efficient 'OLS" Model for a Time Series Data, *Journal of Business Forecasting*, 24 (Summer), 28 – 32.

FOOTNOTES

[1] Two extremes may be offered. Subway sandwich shops approve new stores within a close distance of existing stores. Observation in almost any community will confirm the numerous Subway sandwich shop locations. In contrast, Lexus dealerships are often placed far away from each other. Arkansas has only one Lexus dealer in the entire State.

[2] Available from www.autobox.com/xkos

[3] Confidentiality of proprietary information prevents the release of the name of the company, the store locations, and the exact weeks in a specific year. However, the data series are actual unit sales.

[4] We shall defer to the marketing experts for the exact techniques suitable to support a new store opening. The preliminary analysis of the data reveals that the existing store sales are positively related to the opening of the new store.

[5] Testing the underlying assumptions goes to time series techniques also.

[6] John C. Pickett, David P. Reilly and Robert M. McIntyre, "How to Select a Most Efficient 'OLS" Model for a Time Series Data, *Journal of Business Forecasting*, Summer 2005, Vol 24, No. 2, pp. 28 – 32.

[7] For time series, the statistical technique is multiple time series estimation procedures.

[8] This is a much simplified test. A complete testing of the second assumption requires dividing the residuals into n groups and testing each pair of variances.

[9] The magnitude of the error is unknown. We simply know that the OLS model is incorrect.

[10] If not, then we cannot use the critical values printed in the normally distributed t tables.

[11] We use AUTOBOX - the most powerful and flexible time series program available from Automatic Forecasting Systems, Inc. It is available at www.autobox.com. It follows the ideas presented here but replaces the human eye with very smart heuristics designed to augment models and redefine models using tests of necessary and sufficiency.

[12] Some outliers can be easily identified by inspection of the plot of the residuals but some cannot as we demonstrate in a following section.

[13] It is highly unlikely that interventions associated with weather, holidays, competitor's marketing efforts, and other uncontrollable events can be avoided, but some might be mitigated if anticipated.

TABLE 1

DESCRIPTIVE STATISTICS

EXISTING STORE		NEW STORE	
Mean	40275.3	Mean	3555.79
Standard Error	3313.5	Standard Error	410.84
Median	28633.5	Median	1793
Mode	#N/A	Mode	1761
Standard Deviation	44455.2	Standard Deviation	5511.99
Sample Variance	2E+09	Sample Variance	3E+07
Kurtosis	10.9201	Kurtosis	24.2297
Skewness	3.03409	Skewness	4.20167
Range	279405	Range	46313
Minimum	3	Minimum	13
Maximum	279408	Maximum	46326
Sum	7249554	Sum	640042
Count	180	Count	180

TABLE 2

AUTOCORRELATAION AND PARTIAL AUTOCORRELATIONS OF RESIDUALS

Lag Value	ACF Value	Stnd. Error	t -Ratio	Chi-Square	X^2 Prob.	PACF Value	Stnd. Error	t -Ratio
1	.123	.075	1.65	2.8	NA	.123	.075	1.65
2	-.039	.076	-.52	3.1	NA	-.055	.075	-.74
3	-.063	.076	-.83	3.8	.0518	-.052	.075	-.69
4	-.121	.076	-1.59	6.5	.0387	-.111	.075	-1.49
5	.024	.077	.31	6.6	.0854	.049	.075	.65
6	.137	.077	1.77	10.1	.0383	.119	.075	1.59
7	.210	.078	2.67	18.5	.0024	.179	.075	2.40
8	.033	.082	.41	18.7	.0047	-.010	.075	-.14
9	-.094	.082	-1.15	20.5	.0048	-.070	.075	-.94
10	-.079	.082	-.96	21.6	.0058	-.025	.075	-.34
11	-.026	.083	-.31	21.7	.0099	.016	.075	.22
12	-.093	.083	-1.12	23.4	.0095	-.131	.075	-1.76
13	.150	.083	1.80	27.8	.0035	.119	.075	1.59
14	.065	.085	.77	28.6	.0045	-.015	.075	-.21

TABLE 3

AUGMENTING THE DATA SERIES TO INCLUDE DUMMY VARIABLES

	EXISTING	NEW						
	STORE	STORE	DAY 1	DAY 2	DAY 3	DAY 4	DAY 5	DAY 6
OBSERVATION NUMBER	UNIT SALES	UNIT SALES	DUMMY	DUMMY	DUMMY	DUMMY	DUMMY	DUMMY
1	23863	1565	1	0	0	0	0	0
2	26100	417	0	1	0	0	0	0
3	31367	1830	0	0	1	0	0	0
4	23664	2183	0	0	0	1	0	0
5	20116	117	0	0	0	0	1	0
6	16399	13	0	0	0	0	0	1
7	20239	2073	0	0	0	0	0	0
A SIMILAR CODING SCHEME WOULD BE APPLIED TO THE SUBSEQUENT OBSERVATIONS								

TABLE 4

THE OLS MODEL INCLUDING DUMMY VARIABLES

MODEL COMPONENT	LAG	COEFFICIENT	STANDARD ERROR	p - VALUE	t -VALUE
CONSTANT		57200	8630	.0000	6.62
NEW STORE SALES	0	.492	.586	.4030	.84
DAY 1 DUMMY	0	-8770	11600	.4489	-.76
DAY 2 DUMMY	0	-4570	11500	.6924	-.40
DAY 3 DUMMY	0	-11600	11500	.3185	-1.01
DAY 4 DUMMY	0	-24700	11500	.0335	-2.14
DAY 5 DUMMY	0	-36200	11600	.0022	-3.11
DAY 6 DUMMY	0	-44700	11900	.0002	-3.78

TABLE 5

TENTATIVELY IDENTIFIED TIME SERIES MODEL PARAMETERS

MODEL COMPONENT		LAG (Week/Day)	COEFF	STANDARD ERROR	p VALUE	t VALUE
CONSTANT			.301E+05	.200E+04	.0000	15.08
INPUT SERIES X1	NEWSTORE SALES	0	.399	.316	.2086	1.26
INPUT SERIES X2	PULSE@63	9/7	.248E+06	.221E+05	.0000	11.25
INPUT SERIES X3	PULSE@135	20/2	.234E+06	.221E+05	.0000	10.63
INPUT SERIES X4	PULSE@134	20/1	.187E+06	.221E+05	.0000	8.47
INPUT SERIES X5	PULSE@115	17/3	.188E+06	.221E+05	.0000	8.51
INPUT SERIES X6	PULSE@128	19/2	.170E+06	.229E+05	.0000	7.42
INPUT SERIES X7	PULSE@89	13/5	.120E+06	.221E+05	.0000	5.43
INPUT SERIES X8	PULSE@92	14/1	.119E+06	.224E+05	.0000	5.33
INPUT SERIES X9	PULSE@24	4/3	.106E+06	.221E+05	.0000	4.80
INPUT SERIES X10	PULSE@70	10/7	.105E+06	.222E+05	.0000	4.74
INPUT SERIES X11	PULSE@11	2/4	.987E+05	.221E+05	.0000	4.47

TABLE 6

FINAL TIME SERIES MODEL

MODEL COMPONENT	LAG (Week/Day)	COEFF	STANDARD ERROR	p VALUE	t VALUE
1CONSTANT		.236E+05	.117E+04	.0000	20.20
2Autoregressive-Factor	7	.502E-01	.148E-01	.0018	3.39
INPUT SERIES X1 SEASONALPULSE@14	40/5	.178E+05	.229E+04	.0000	-7.78
INPUT SERIES X2 SEASONALPULSE@13	40/4	-.161E+05	.213E+04	.0000	-7.54
INPUT SERIES X3 PULSE@45	45/1	.285E+05	.446E+04	.0000	6.38
INPUT SERIES X4 PULSE@15	40/6	-.237E+05	.446E+04	.0000	-5.33
INPUT SERIES X5 PULSE@43	44/6	.236E+05	.448E+04	.0000	5.26
INPUT SERIES X6 PULSE@25	42/2	.224E+05	.444E+04	.0000	5.04
INPUT SERIES X7 PULSE@30	42/7	.217E+05	.443E+04	.0000	4.89
INPUT SERIES X8 PULSE@8	39/6	-.156E+05	.445E+04	.0013	-3.50
INPUT SERIES X9 PULSE@17	41/1	-.120E+05	.443E+04	.0107	-2.70
INPUT SERIES X10 PULSE@44	44/7	.103E+05	.446E+04	.0270	2.31
INPUT SERIES X11 PULSE@36	43/6	-.102E+05	.447E+04	.0281	-2.29

The final model may be rewritten in a more conventional form as:

Existing Store Unit Sales$_t$ (Y) = 23600.
$+[X1_t][(-17787.)]$
$+[X2_t][(-16069.)]$
$+[X3_t][(+28479.)]$
$+[X4_t][(-23731.)]$
$+[X5_t][(+23553.)]$
$+[X6_t][(+22365.)]$
$+[X7_t][(+21670.)]$
$+[X8_t][(-15570.)]$
$+[X9_t][(-11959.)]$
$+[X10_t][(+10292.)]$
$+[X11_t][(-10227.)]$
$+[(1 - .0502B^7)]^{-1}[A_t]$

Adjusted R^2 = .895
Standard Error = 994

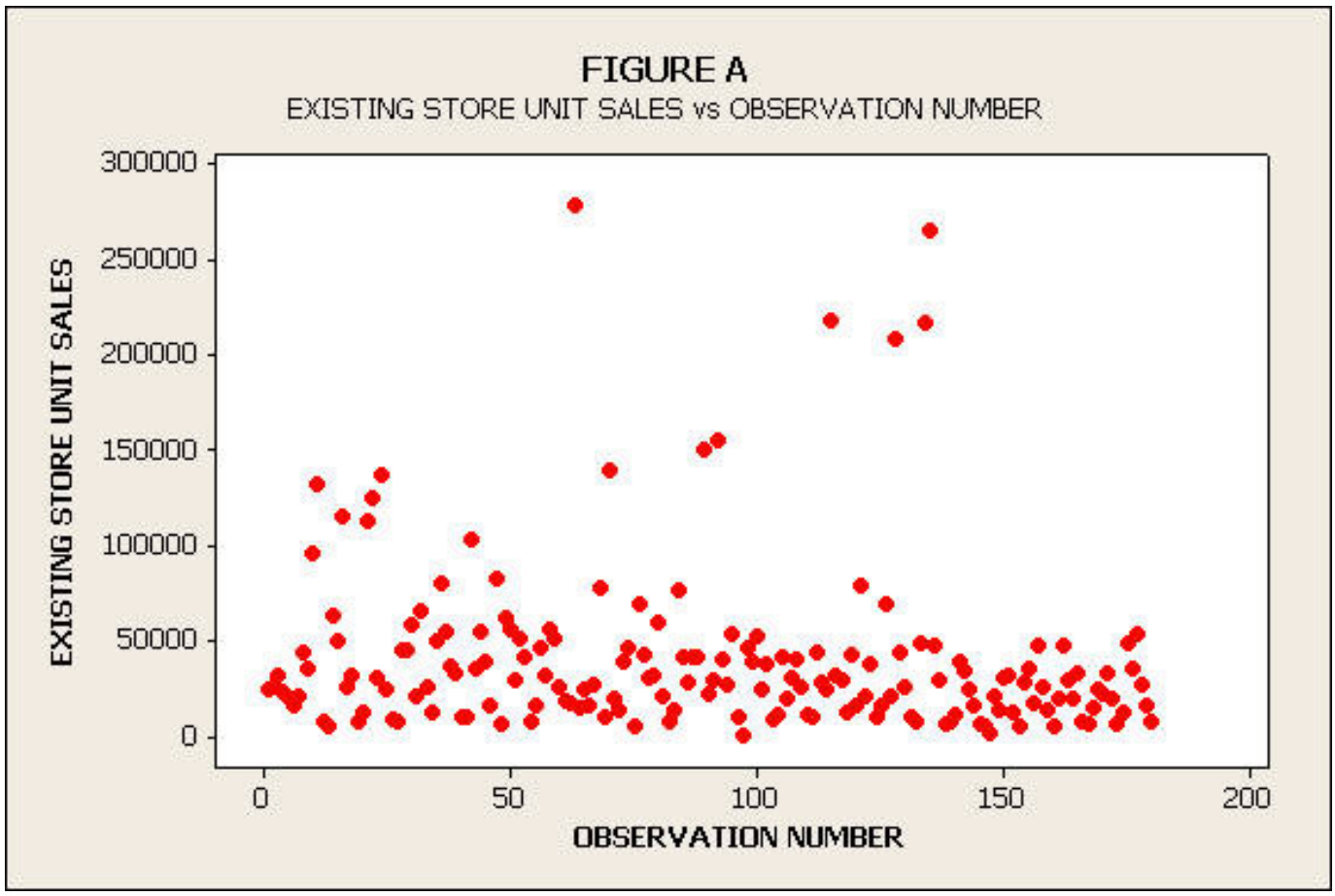


FIGURE A
EXISTING STORE UNIT SALES vs OBSERVATION NUMBER
EXISTING STORE UNIT SALES
300000
250000
200000
150000
100000
50000
0
0
50
100
150
200
OBSERVATION NUMBER

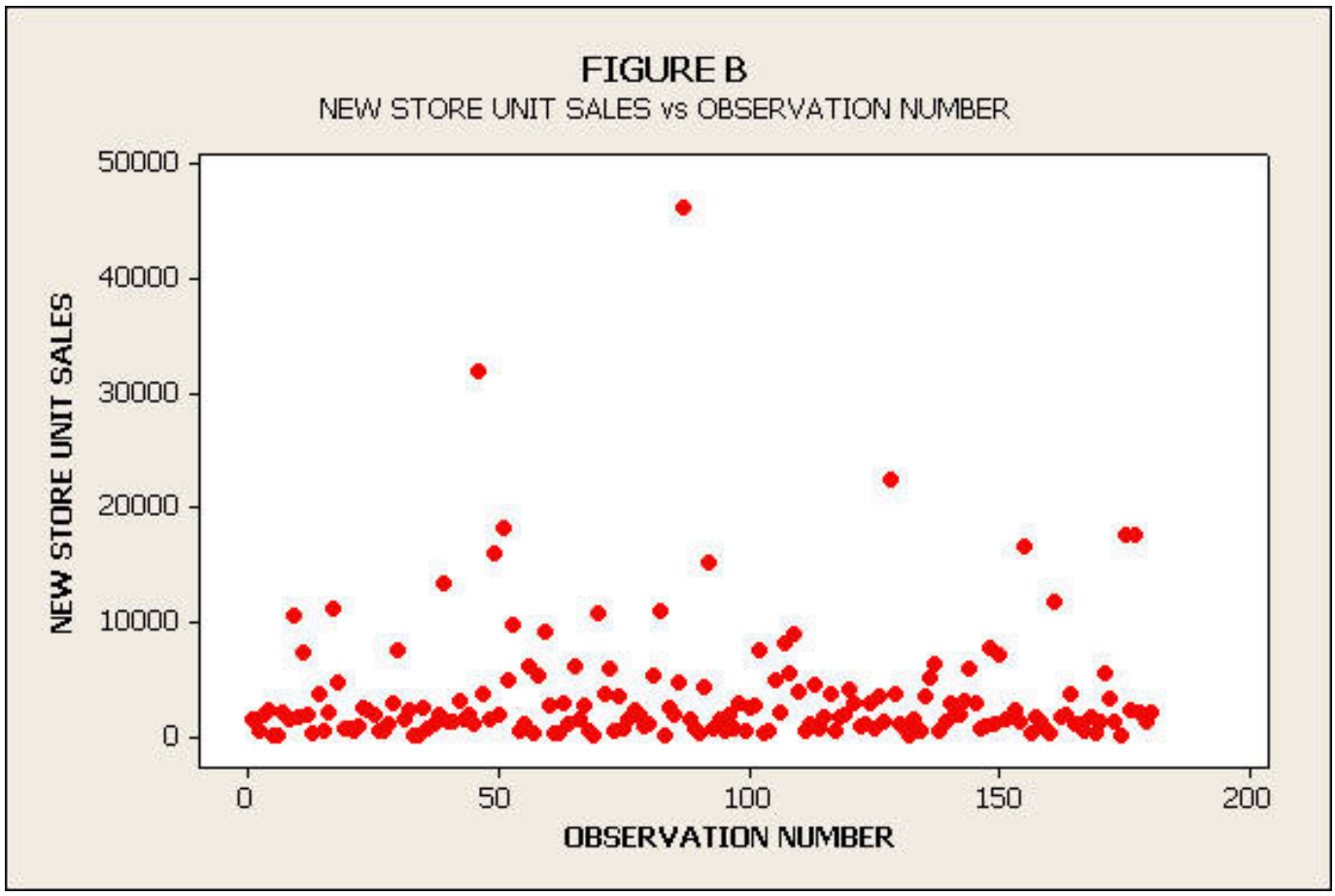


FIGURE B
NEW STORE UNIT SALES vs OBSERVATION NUMBER
NEW STORE UNIT SALES
50000
40000
30000
20000
10000
0
0
50
100
150
200
OBSERVATION NUMBER

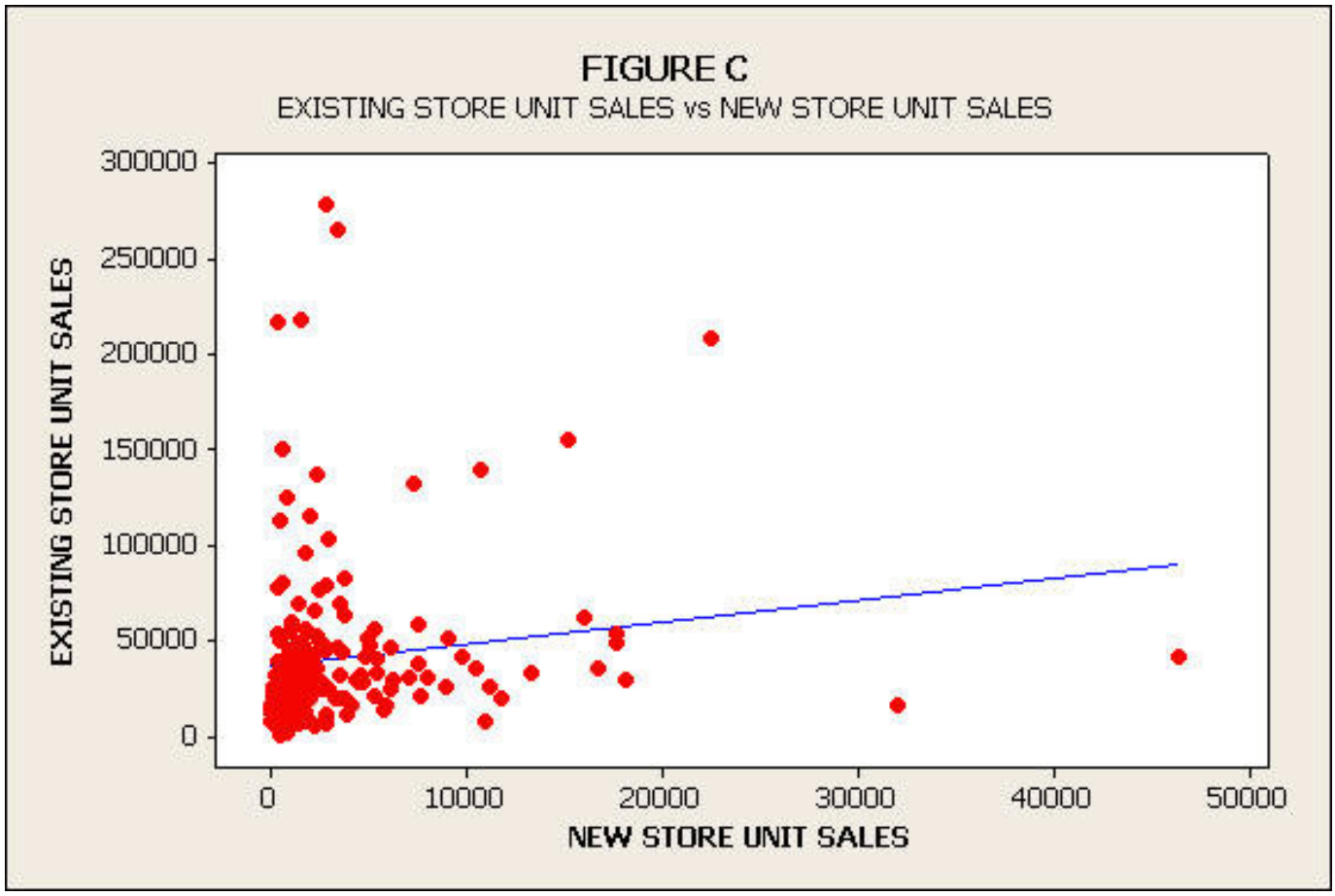


FIGURE C
EXISTING STORE UNIT SALES vs NEW STORE UNIT SALES
EXISTING STORE UNIT SALES
300000
250000
200000
150000
100000
50000
0
0
10000
20000
30000
40000
50000
NEW STORE UNIT SALES

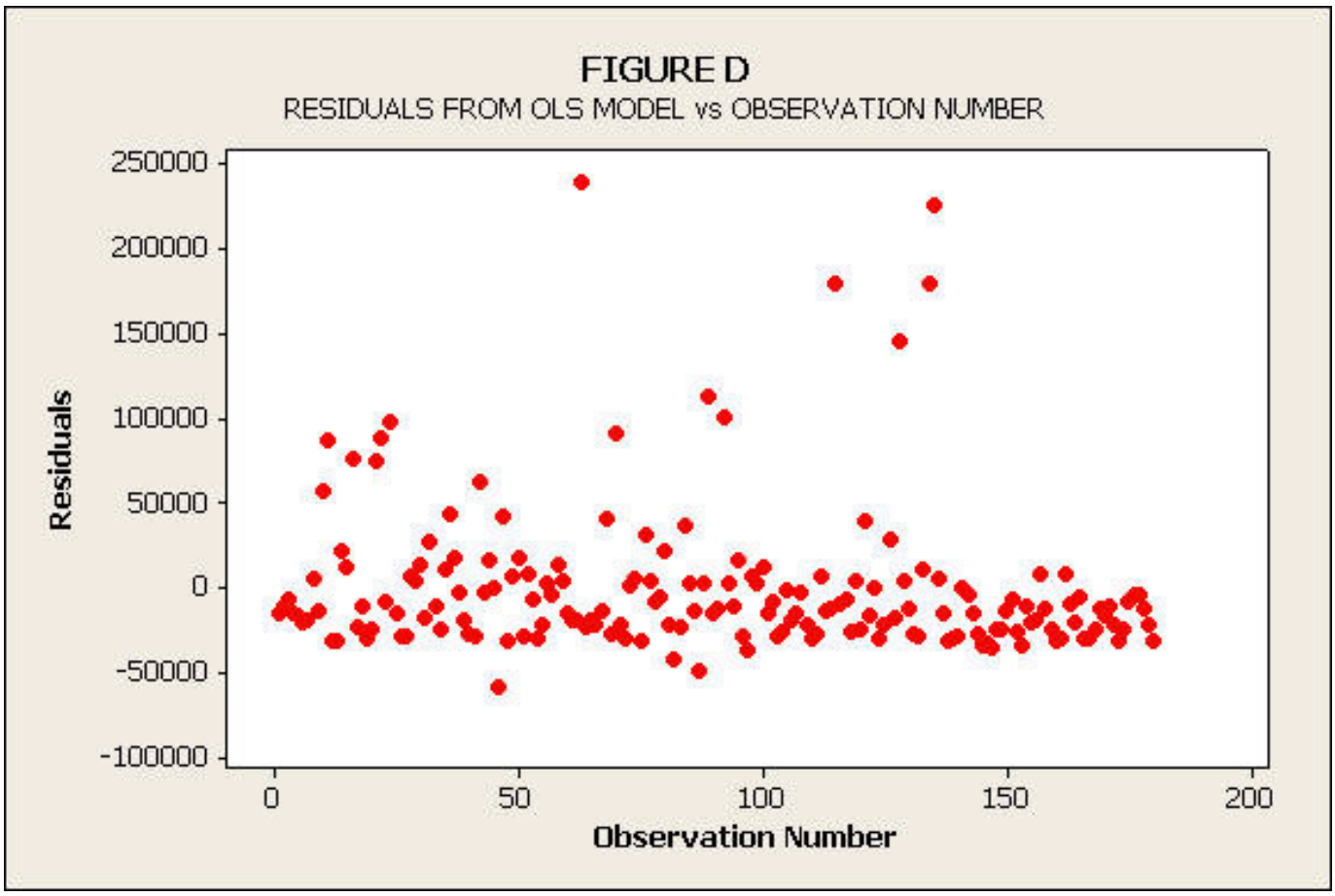
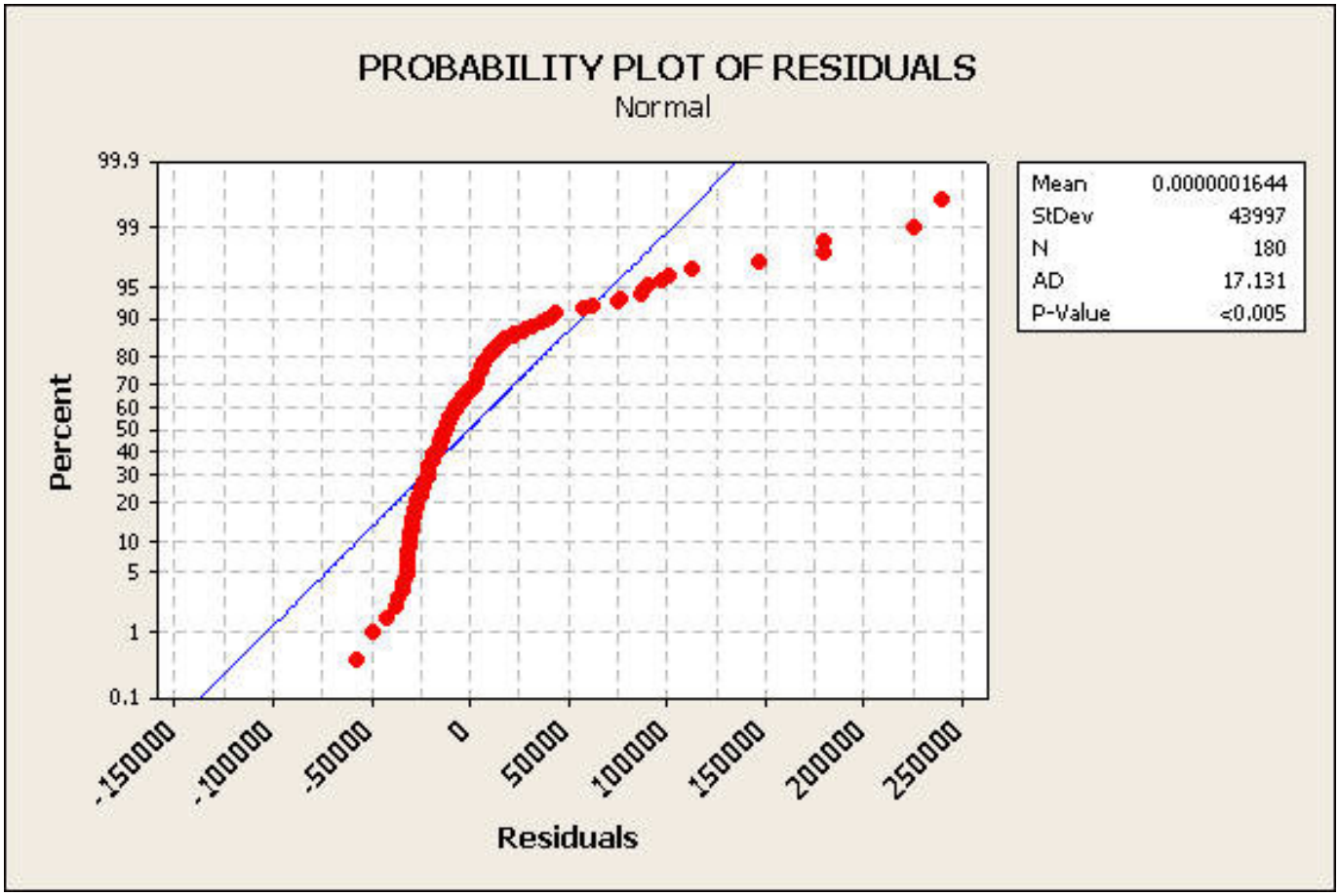


FIGURE D
RESIDUALS FROM OLS MODEL vs OBSERVATION NUMBER
Residuals
250000
200000
150000
100000
50000
0
-50000
-100000
0
50
100
150
200
Observation Number

FIGURE E


PROBABILITY PLOT OF RESIDUALS
Normal
Percent
99.9
99
95
90
80
70
60
50
40
30
20
10
5
1
0.1
-150000
-100000
-50000
0
50000
100000
150000
200000
250000
Residuals
Mean 0.0000001644
StDev 43997
N 180
AD 17.131
P-Value <0.005

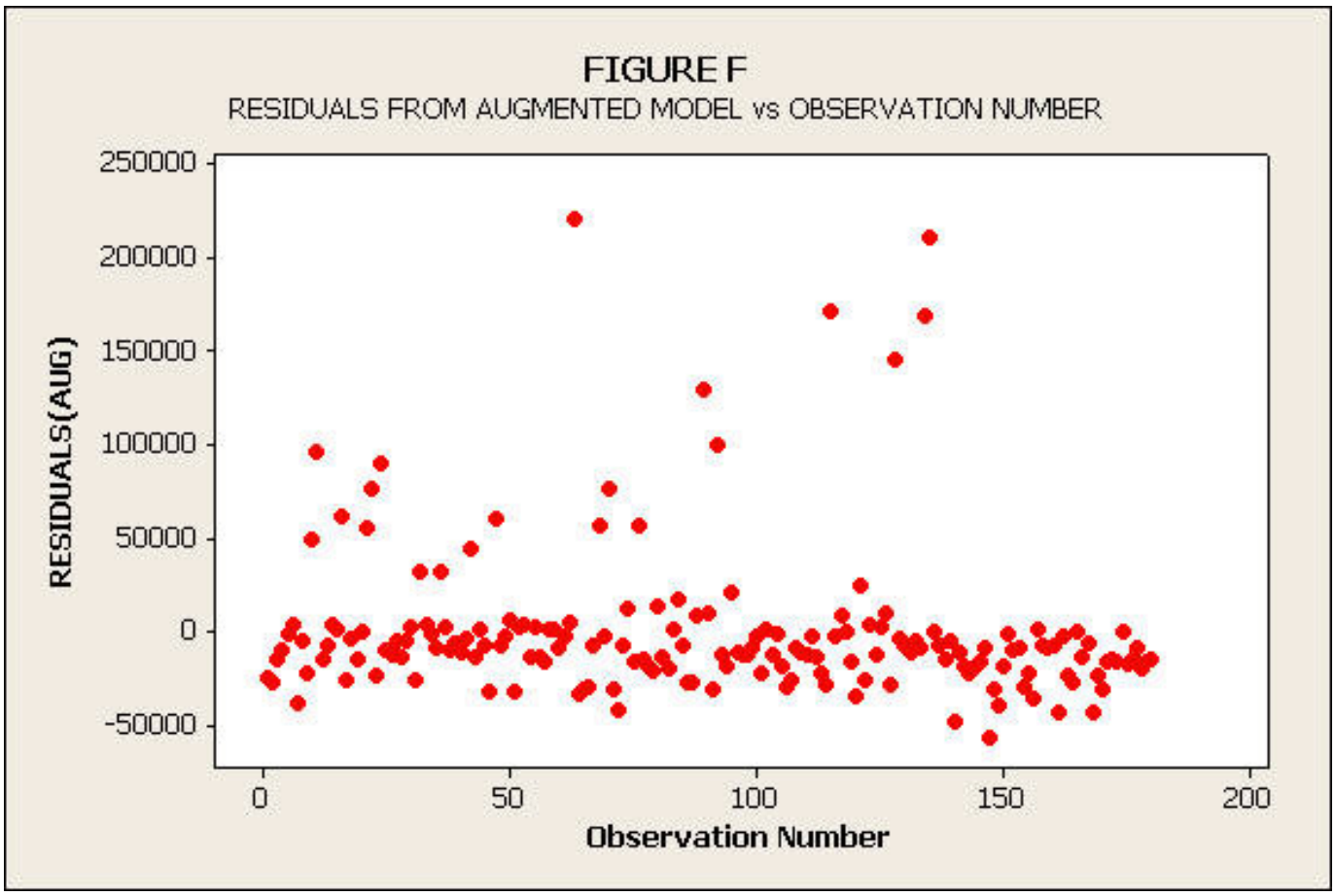


FIGURE F
RESIDUALS FROM AUGMENTED MODEL vs OBSERVATION NUMBER
RESIDUALS(AUG)
250000
200000
150000
100000
50000
0
-50000
0
50
100
150
200
Observation Number

[1] Two extremes may be offered. Subway sandwich shops approve new stores within a close distance of existing stores. Observation in almost any community will confirm the numerous Subway sandwich shop locations. In contrast, Lexus dealerships are often placed far away from each other. Arkansas has only one Lexus dealer in the entire State.

[2] Available from www.autobox.com/xkos

[3] Confidentiality of proprietary information prevents the release of the name of the company, the store locations, and the exact weeks in a specific year. However, the data series are actual unit sales.

[4] We shall defer to the marketing experts for the exact techniques suitable to support a new store opening. The preliminary analysis of the data reveals that the existing store sales are positively related to the opening of the new store.

[5] Testing the underlying assumptions goes to time series techniques also.

[6] John C. Pickett, David P. Reilly and Robert M. McIntyre, "How to Select a Most Efficient 'OLS" Model for a Time Series Data, *Journal of Business Forecasting*, Summer 2005, Vol 24, No. 2, pp. 28 – 32.

[7] For time series, the statistical technique is multiple time series estimation procedures.

[8] This is a much simplified test. A complete testing of the second assumption requires dividing the residuals into n groups and testing each pair of variances.

[9] The magnitude of the error is unknown. We simply know that the OLS model is incorrect.

[10] If not, then we cannot use the critical values printed in the normally distributed t tables.

[11] We use AUTOBOX - the most powerful and flexible time series program available from Automatic Forecasting Systems, Inc. It is available at www.autobox.com. It follows the ideas presented here but replaces the human eye with very smart heuristics designed to augment models and redefine models using tests of necessary and sufficiency.

[12] Some outliers can be easily identified by inspection of the plot of the residuals but some cannot as we demonstrate in a following section.

[13] It is highly unlikely that interventions associated with weather, holidays, competitor's marketing efforts, and other uncontrollable events can be avoided, but some might be mitigated if anticipated.