DISAGGREGATE FORECASTING MODELS: APPLICATION TO AMEREN UE'S TRANSFORMER USAGE

A Master's Thesis presented to the Faculty of the Graduate School
University of Missouri

In Partial Fulfillment
Of the Requirements for the Degree

Master of Science

---

by
MATTHEW ROMAN

Dr. Wooseung Jang, Thesis Supervisor

DECEMBER 2008

The undersigned, appointed by the Dean of the Graduate School, have hereby examined the Thesis entitled

DISAGGREGATE FORECASTING MODELS:
APPLICATION TO AMEREN UE'S TRANSFORMER USAGE

Presented by Matthew S Roman

A candidate for the degree of Master of Science

And hereby certify that, in their opinion, it is worthy of acceptance

____
Dr. Wooseung Jang

____
Dr. James Noble

____
Dr. Lori Franz

ACKNOWLEDGEMENTS

This research would not have been realized if not for the time and effort put forth by several individuals from both the Industrial Engineering department of the University of Missouri and Ameren UE. I would like to thank Dr. Wooseung Jang for his discerning guidance throughout the course of my thesis; his knowledge and experience benefited every aspect of this work. Na Deng was also helpful by contributing to the initial stages of this research and because of her aid in developing objectives for subsequent work. I also thank Dr. James Noble who was instrumental for his work with Ameren UE making the project, from which this thesis was based, possible. Finally, I would like to thank the many employees of Ameren UE who devoted their time and expertise to provide critical information so as to impart a better understanding of the company, its historical records, and the industry in which it works. For the contributions of all these individuals I am grateful.

Table of Contents

Figures and Tables

Abstract

The importance of accurate forecasts to proper inventory management is a well known and abundantly addressed issue in industry. Maintaining appropriate inventory levels is essential when attempting to maximize potential revenue and customer satisfaction. Within the utilities industry the significance of customer satisfaction is of utmost importance and the ability to predict when and where certain materials will be needed is highly valued. This research was motivated by these requirements and was focused on creating a customized forecasting model which could address the specific needs and demand patterns experienced by Ameren.

Amongst the various materials used during energy delivery, transformers were selected due to their importance and increased lead times from suppliers. The historical transformer usage was attributed to three primary causes: new construction, storm and emergency, and general maintenance. Each of these displayed a distinctive demand pattern, thus a specific forecast was made for each disaggregate segment. Creating an individual forecasting model for each type of demand provided the ability to address the uniqueness within each demand pattern. More specifically, this approach allowed for the input of a forward-looking trend, generated from external factors, during the new construction forecast, the use of a model which followed historical trends within the general maintenance data, and a long-term averaging model which limited outliers found in the storm and emergency demand pattern. These disaggregate forecasts were then added together to create a final aggregate level forecast for the item or group of items being investigated. This model showed up to a 20% improvement of accuracy over more traditional methods when compared using median absolute percent error.

Chapter 1

Introduction

The use of forecasting models during inventory planning is a widely accepted practice throughout a variety of industries. It is also well recognized that the difficulty and importance of forecasting intensifies as usage levels and the number of individual units increase. The generally accepted practice when creating forecasts for many different time-series is to apply a general automatic forecasting model that is easily reinitiated. Amongst these an appropriate form of exponential smoothing is traditionally recommended due to its relatively good performance which has been reaffirmed time and again in practice. Although exponential smoothing tends to be a reliable option when creating forecasting for inventory control, it has difficulties in situations where historical data is not sufficient to predict future usage levels. This often occurs when outside factors have a significant influence on demand.

Another common approach to forecasting within industry is to disaggregate data. This often occurs when it is necessary to create individual forecasts for specific regions, items, or product families. However, creating individual disaggregate forecasts and then adding each to obtain an aggregate level forecast causes overall variance of the model to increase. In addition important information relating to the data can be lost when statistically similar usages are separated simply because they are from separate stores or regions.

The purpose of this research was to create a forecasting tool which could be used to forecast monthly transformer usage for a large utility company in the coming year. It was necessary for the model to be easily reinitiated across many different time-series, thus an exponential smoothing model was an appropriate choice. However, upon examination of the aggregate time-series it was observed that trends in usage levels could quickly change before any backward looking forecasting model could adapt. This lead to the identification of the various causes of transformer usage and the development of a specialized disaggregate model. This model recognized three different causes of demand: new construction, storms and emergencies, and general maintenance. Of these three, new construction was of particular interest due to its dependence on outside factors. A new model was developed for the new construction portion of the data which could incorporate a forward-looking trending variable. Suggestions for this trending variable were based on construction data that was determined to indicate future transformer usage. The benefits of creating disaggregate, demand pattern specific, forecasting models was further utilized when creating individual models for both the storm and emergency and general maintenance portions of the data. Storm and emergency usage is volatile by nature so a model which could better smooth this tendency was selected. The final disaggregate segment, general maintenance, demand pattern behaved much like one would expect from an older inventory item. The general level was rather constant and shifts in trend occurred over long periods of time. For this reason, exponential smoothing was a good forecasting technique for this time-series. The complete model was programmed into Excel to enable automatic reinitiation and to simplify its use by the company.

The specialized disaggregate model was found to outperform more traditional forms of forecasting by being more broadly applicable and more easily customized, through the use of an inputted forward-looking trend. In every tested case the specialized disaggregate model was found by either geometric mean relative absolute error (GMRAE) or median absolute percent error (MdAPE) to have the most accurate forecasted values. This work is intended to aid the inventory planning needs of the previously stated utility company and to help further develop the proper use of disaggregation in forecasting.

Chapter 2

Literature Review

Forecasting has been a heavily researched and important part of inventory management for decades. Over this period of time the frequency and severity of market shifts has escalated leading to an increase in forecasting difficulty and its importance to inventory planning. Due to the vast amount of research done in this field there are many different types of forecasting techniques which can be used. Researchers often outline the importance of model selection and emphasize that forecasting is an art form. There is never only one good solution for any given situation. Because of this, familiarity with the data and its demand pattern are seen as essential during model selection. In industry, a common need is to create forecasts at various levels of aggregation. Statistically aggregate level forecasts tend to be more accurate; however, research has shown that it is essential to forecast disaggregate segments within the data when these are known to have significantly unique usage patterns, Weatherford 2001 [18].

Being an early and intuitive approach to forecasting, exponential smoothing has served as a standard of comparison for many years. Even though it is a relatively simplistic model when compared to other ARMA approaches, exponential smoothing has been shown in many cases to be just as accurate. Increases in model complication do not necessarily lead to increased accuracy, and often these added complications only elevate the cost of the model. In cases where it is necessary to create forecasts for many different time-series a model which can be automatically reinitiated, such as exponential smoothing, is essential.

The literature review below will serve to provide a background of topics related to the work contained within this paper. Special attention has been paid to the development of exponential smoothing and the selection of aggregation level. The work done within both these fields was of significant importance to the construction of this research. An outline of relevant background information related to exponential smoothing is presented and the work done concerning aggregation is discussed to provide the framework for which the addition of this investigation will hopefully contribute. Finally, pertinent research into model evaluation techniques is included to verify their usage during final validation.

2.1 Exponential Smoothing

Originating during World War II as part of Robert G. Brown's work on tracking models, for fire control information, on the location of submarines for the US Navy, exponential smoothing has been a well known and broadly used method for decades. In the 1950s, Brown extended his research to forecast the demand for spare parts in Navy inventory systems. Also, Charles C. Holt worked independently of Brown during the 50s' on an exponential additive model which was published in his original work Holt (1957). Holt's work gained recognition due to Winter's 1960 empirical study and the method became known as the Holt-Winter's forecasting system. Over the years exponential smoothing has become one of the most widely understood and employed types of forecasting, and has served as the standard for comparison during forecasting evaluation in many theoretic and functional applications.

In 1988 Chatfield [5] outlined the process for using Holt-Winter's exponential smoothing because at the time he believed the method to be less known than it deserved and unfairly considered inaccurate despite empirical evidence to the contrary. He notes that exponential smoothing had been shown to perform just as well as more complicated projection methods and even as well as multivariate methods. Chatfield mentions that Holt-Winter's forecasts can be easily improved using subjective judgment and that forecasts are not sacred and should be modified in the light of any external knowledge.

Thomas's 1993 [17] argued that beyond a modest level, additional sophistication is detrimental due to the increased cost of both time and money during their development. However, there may be cases where econometric models (causal models) should be expected to out perform extrapolation methods (exponential smoothing). Thomas hypothesized that econometric methods would outperform extrapolation methods in situations with greater buyer sensitivity to changes in product/market factors; and that extrapolation will be more accurate in cases of less buyer sensitivity to such factors. This hypothesis was confirmed and a recommendation of extrapolation was made for cases where buyers were less sensitive such as in mature product markets and that extrapolation would be less effective in markets where change is intended by management through promotional techniques. Thomas recommends that future research should more explicitly include factors that characterize the product/market situation for which the forecast is being developed and that the evaluation of factors such as major environmental variables may lead to more specific guidelines about the appropriateness of various methods.

The majority of the research into exponential smoothing entails whether or not its selection as a forecasting method is appropriate. This work is often based upon the development of new forecasting procedures, new methods of model evaluation, or during the consideration of various types of time-series. In Chatfield's 1997 work [6] a procedure is outlined on how to best select the most appropriate forecasting model. Chatfield discuss that there are many considerations which must be made while selecting the best model such as the data being analyzed, the expertise available, and the number of series to be forecast and he raises the question, "What is meant by 'best'?". This question is posed to emphasize the importance of clarifying the objective of the forecast and the necessity to understand the context of the problem in order to formulate it properly. A simple automatic univariate procedure is considered best when there are many series to forecast. "There is little overall difference in accuracy between several methods and so it seems sensible to choose a method which is simple, easily interpreted and for which software is readily available." Chatfield recommends an appropriate form of exponential smoothing but mentions there are several good alternatives. Once the purpose of the forecast is clearly defined and an understanding of the data obtained, a time plot is necessary to enable the analyst to look for trend, seasonality, and outliers. The selection of an automatic versus a non-automatic approach is important and relies upon the context of the problem. Chatfield recommends an automatic approach in the case of stock control due to the large number of items to forecast which is in agreement with his 1978 suggestion that a fully automatic version of the method should be used on a computer based system to make routine forecasts without human intervention, Chatfield [7]. During automatic stock control forecasts he also suggests the use of a multiplicative seasonal model for every case [7]. In closing of is 1997 article, Chatfield notes that all forecasts are based on assumptions and a lot can be said for not producing a single point forecast, but rather producing a range of forecasts based on different known assumptions.

In 2006 Gardner [10] provided a more up-to-date and comprehensive review of exponential smoothing techniques and model selection. The issue of whether or not exponential smoothing is a special case of ARIMA modeling was considered resolved and that exponential smoothing methods are optimal for a very general class of state-space models which is broader than the ARIMA class. The formulization for all standard exponential smoothing methods were given and it was noted that equations for the seasonal methods are only valid for forecast horizons which are less than or equal to the length of the seasonal cycle. Gardner also states that there is no longer any excuse for the use of arbitrary parameters in exponential smoothing with the popularity of good search algorithms such as Excel Solver. He also suggests that initial values can be refined simultaneously with the smoothing parameters during the optimization process; however, it is also mentioned that there was little difference in average post-sample accuracy regardless of initial values. It is important to renormalize seasonal indices during Holt-Winter's method and little guidance is given on appropriate parameter selection in the multiplicative seasonal case. It is concluded that much of the theory within exponential smoothing still needs to be validated and the basis for choosing amongst the different approaches to time-series forecasting expanded.

2.2 Aggregation vs. Disaggregation

It is common in practice to have an organization wish to be able to not only generate aggregate level forecasts, but to also create forecasts for individual families and products within the aggregate. This necessity brings about two possible processes for achieving this type of information. The first is known as the top-down strategy, which involves the generation of an aggregate level forecast which is then allocated to the segments of interest in accordance with the ratio of the aggregate which the disaggregate part entails. The second approach, known as a bottom-up strategy, creates individual forecasts for each of the disaggregate segments and then adds these together to create the aggregate level forecast. In general the top-down approach is favored due to decreased variance, lower cost, accuracy during times of stable demand. Conversely, a bottom-up approach is considered essential when it is necessary to capture differences amongst demand patterns.

Dangerfield in 1992 [8] tested 15,000 aggregate series constructed from individual series used during the M-competition using exponential smoothing and two-item families. The statistical background for both the top-down and bottom-up approaches was discussed. Support for the top-down approach is based upon the statistical fact that the variance of the aggregate demand is equal to the sum of the variances of the independent item demands. This means that if you simply add together the forecasts for the individual items the aggregate variance will be quite large due to the aggregate variance being equal to the summation of each of the individual variances. Dangerfield mentions that it is not clear that the top-down approach, even if the independence assumption is correct, will improve the accuracy of individual forecasts. His result of this study found the bottom-up approach to be more accurate in three out of four series tested; additionally no combination of item correlation and/or proportion were found where top-down resulted in lower total MAPE than forecasts developed from individual exponential smoothing models.

In 1999 Bunn [3] asserts, that in practice, the general approach is to forecast each item's data series individually and then aggregate as necessary. This approach is considered to waste peripheral data if the behavior of similar products is not taken into account when producing the individual and grouped forecasts. The author suggests that there tends to be a trade off between the unbiasedness of individual forecasts and the robustness and efficiency from individual and related series. The relationship between these series is often caused by economic considerations, competition, regulation, and weather. The author believed that estimating seasonal components based on aggregated related series should not be dismissed, due to the fact they are revised less frequently than the other parameters (level and trend). Bunn used three different strategies for grouping data: according to business classes, cluster analysis within business classes, and cluster analysis across the whole time-series. He found that seasonal indices based on combined series improved forecast performance and that both methods of classification, business and statistical, offered improvement.

Additional work has been done concerning the area of improving seasonal demand forecasts through the application of aggregate level seasonal data to lower levels of disaggregation. Dekker 2004 [9] discusses how forecasting for individual items has become more difficult as the assortments of items has grown over the years. He cites Dreze 1994 as saying that the average number of stock-keeping units at a supermarket has grown from 6,000 a generation ago to 30,000 today. For this reason he suggests that independently modeling seasonal demand for individual products may no longer be optimal. His work follows the same framework as Bunn 1999 [3] as he states that product aggregation into families of individual products with similar seasonal patterns can be used to determine seasonal indices at the product family level and then apply these indices when making forecasts at the individual product level. This is assumed to be beneficial because family level data tends to be less erratic than product level data which would be advantageous when separating the seasonal pattern from the randomness, resulting in better season indices. In the course of this study Holt-Winter's was found to perform poorly due to demand uncertainty and stochastic seasons, however, using aggregate demand data through product aggregation to calculate seasonal indices at the disaggregate level was considered to improve short term forecasts.

As researchers continued to examine the benefit of both the top-down and bottom-up approaches and more application based research became available, a more specific question arose addressing in what circumstances various levels and types of aggregation are most beneficial. Weatherford 2001 [18] emphasizes the importance of accurate forecasts to the hotel industry. Within this industry forecasting is important when predicting the durations of use. Accurately predicted durations can be utilized by managers to maximize overall revenue during all time periods and not just during periods of high demand. The general approach to forecasting in the hotel industry is to separate demand according to the length of stay (LOS) and rate categories (RC). Some first forecast the aggregate and then separate this forecast down to the disaggregate length of stay and rate categories in accordance to historical probability distributions. Two other common approaches are to forecast the LOS and RC first and then apply probability distributions or simply create a forecast for each of the possible combination of LOS and

RC groups. This raises the question as to which level of aggregation is the most appropriate and accurate approach. Weatherford proposes that individual disaggregate forecasts are essential when it is important to detect distinctions between demand patterns for individual items. The results of his research were that a purely disaggregate forecast strongly outperformed even the best aggregate forecasts. It was recommended to forecasts at the completely disaggregated level and thus create a forecast for each combination of LOS and RC.

Although research has shown the benefits of disaggregation additional work has focused on finding an appropriate balance between aggregate and disaggregate forecasts. Zotteri 2005 [19] addresses the trade off between the differences within the data and the ability to accurately address these differences. The author worked to forecast demand for a European grocery retailer. He hypothesized that, "to accurately forecast demand one needs to estimate the driver of demand fluctuations." The issues on how to cluster demand was addressed since in practice it is most common to cluster demand according to geographic location of corporate structure. A different approach was recommended so as to cluster demand according to the degree of similarity between time-series. Since the purpose of this research was to analyze the effect of aggregation level the same forecasting technique was used at all levels of aggregation. It was argued that clustering stores according to their demand pattern rather than region or size could lead to a better grouping for forecasting. The project found that a cluster analysis which grouped information from stores with similar demand patterns performed slightly better than the aggregate, particularly when forecasting high demand items. Zotteri concludes that even in a specific context, there is no one best way to determine aggregation level and should be further researched by practitioners.

A distinction should be drawn between disaggregation according to retail location or product and disaggregating according to demand pattern. The problems with variance, as noted in Dangerfield 1992 [8], arise in both cases. In the first case, where data is disaggregated based solely on managerial interest, this is a major concern for reliability of aggregate level forecasts. However, when the data is disaggregate according to statistically significant demand patterns, such as in Weatherford 2001 [18], these concerns are lessened due to the benefits received from segmenting the data into structures from which accurate forecasts can more easily be created.

2.3 Evaluation

Although the focus of this paper is not on the many evaluation techniques available for assessing forecast performance, it is still necessary to provide a background of the techniques selected for the purposes of this research. Determining model performance is just as much of an art as creating the forecast itself, and it is clearly of substantial importance when comparing techniques. Below is a selected discussion of previous work within the field of forecast evaluation relevant for the purposes of this paper.

In 2006 Hyndman [11], found it necessary to summarize the many various types of forecast evaluations and provide the literature's, more or less, consensus view of which evaluations are the best and in what situations they are most appropriate. As assessment techniques have evolved over the years, the more traditional and common types of measurement like mean square error (MSE) and root mean square error (RMSE) have come under scrutiny. These are now considered to be overly sensitive to outliers and are, for the most part, no longer recommended when determining forecast accuracy. In cases where forecast performance is compared across different data sets percentage errors tend to me more accurate because they have the advantage of being scale independent. Of these median absolute percentage error (MdAPE) has been selected because of Armstrong and Collopy's [2] finding that when evaluating a moderate to large number of data series GMRAE provided the most robust results. Relative error measures are also considered an improvement to general RMSE or MSE. Of these geometric mean relative absolute error (GMRAE) is recommended by Armstrong 1992 [2] and has been applied in cases such as Weatherford 2001 [18].

Chapter 3

<u>Specialized Disaggregate Forecasting Model</u>

The above discussion outlined work previously done in the areas of exponential smoothing and aggregation level. The review of literature showed the appropriateness of exponential smoothing when creating forecasts for inventory management. Also, the need for further research into the selection of aggregation level and usefulness of case studies was established. Due to the advantages of creating disaggregate forecasts based upon the distinctions of usage within the aggregate, a disaggregate model was created to forecast transformer usage by a major utility company.

3.1 Model Objective

The objective of this effort was to serve as a practical application of previously developed techniques and also, as an extension to current methodologies. In previous work one model was generally applied to all disaggregate segments; in this work, not only was the data disaggregated into statistically significant segments, but also the most appropriate type of forecasting model was applied to each. This approach was a logical development of the research into the possible benefits of disaggregation and was based upon the notion that forecasts should be tailored to the demand pattern shown for each relevant segment of the data. Furthermore, this study can serve as empirical evidence from which further comparisons can be drawn.

The focus was to create a reusable and effective forecasting tool that could be utilized for the everyday inventory control needs of a major utility company. Thus the purpose was not to create a one time forecast but to design a tool that could be easily reimplemented well into the future and across multiple time-series. This required an automatically reinitiated forecast created in an easily understood framework with a software tool that was readily available to the company. As a result, exponential smoothing and averaging techniques were the most appropriate. Furthermore, Microsoft Excel was selected for implementation due to its broad accessibility and the ability to easily refit exponential smoothing parameters through the use of Excel Solver.

The mathematical formulization for each forecasting model is discussed, and the reasoning for their selection is presented. Every model used was eventually programmed into Excel. An explanation of the specific techniques used during implementation is provided in the Chapter 5.

3.2 Formulation

Earlier work has shown disaggregation to be an important tool when statistically unique groups within the aggregate can be defined. Furthermore, it is recognized that disaggregation generally increases the variance at the aggregate level; however, these effects are alleviated due to the benefits of organizing the data in such a way that characteristics specific to each time-series can be exploited to create more accurate disaggregate forecasts. Here this concept is applied to forecast monthly transformer usage by a major utility company for the coming year. Three specific demand patterns were found within the aggregate which were determined by the cause of transformer usage and confirmed through an investigation of each segment's time-series. These were new construction (NC), storm and emergency (SE), and general maintenance (GM). The segments were identified because each displayed particular characteristics which could be utilized while creating a forecasting model. Thus, simply grouping the data into appropriate segments and applying one general forecasting model was not sufficient for creating forecasts due to intrinsic differences between the time-series. Specifically, the new construction demand pattern was significantly determined by external factors which could not be established from historical usage alone. An individual forecasting model was used for each disaggregate segment and then, as in previous work, added together in creation of an aggregate forecast. The specific application of each model and further explanation of their use is presented in Chapter 4.

3.2.1 New Construction

The central concept to the new construction model was that forecasts should be adjusted whenever external information is available and known to affect demand. The transformer usage caused by new construction was heavily dependant on economic trends and the general health of the housing and construction industries. So as to enable the consideration of these external trends, which influenced Ameren's transformer usage, a clear-cut model which utilizes a forward looking user specified input variable was created.

New Construction Model:

$$L_t = (X_{t-12} + X_{t-24})/2 \quad (1)$$

$$y_t^{NC} = L_t (1 + \Delta) \quad (2)$$

Notation of Variables:
L_t = Expected level in period t
X_t = Observed value of the time-series in period t
Δ = Expected change in level (user specified)
y_t^{NC} = Disaggregate level forecast for new construction

This model determines a forecast by averaging the usage during a given month over the past two years (X_t) and adding a trending value delta. The level was established using only two years of data because a trend is not generated over time. Using only two points of data also allowed the model to adapt more quickly to changes in usage while maintaining a more stable level than only one year of historical data which would be more subject to outliers. This mean level was then modified according to the anticipated changes in future demand by the forward-looking input variable (Δ). This variable was the expected change in usage from period (t) to period (t+12). The value was established by examining two indicators that had been determined to be predictors of usage by comparing their levels with respect to new construction transformer usage. It was found that historically the year over year percent change in NC transformer usage lagged that of both St. Louis and national housing permit applications by approximately one year.

The recommended input for NC's trend is the average of these two indicators; please see Chapter 4 and Appendix A for further description. Although there is a recommended outside value for trend, a historical trend can also be inputted into the model which assumes that factors which have influenced usage in the past will continue to do so at the same rate into the future. The new construction model succeeds at incorporating two primary considerations: namely, automatic reinitiation and the consideration of factors outside of the historical record of usage. The trending factor can be determined through the use of Appendix A or historical data alone; however, the consideration of potential eventualities is also helpful due to the ease in which the model can adapt to a new trending input.

3.2.2 Storm and Emergency

The nature of storm and emergency usage was to be sporadic and difficult to predict. The ability to distinguish this usage from the rest of the data allowed for the use of a forecasting model which could better smooth the inherent volatility of this data. An additional benefit of separating this demand pattern was to remove it from the more stable and predictable usage due to general maintenance. Once again it was necessary for the model to be automatically reinitiated and broadly applicable. Also future storm and emergency usage would be somewhat independent of past observations of level making an exponential smoothing model a poor choice; however, observations of seasonal factors from past data could be used to forecast increased usage during the summer months. The model below used the maximum amount of historical data available, five years, and was intended to both limit outliers and reduce instability.

Storm and Emergency Model:

$$\mathrm{X}_{t,a}^{SE},\ldots,\mathrm{X}_{t,n}^{SE} = \mathrm{X}_{s+12(\alpha-1)}^{SE} \quad \begin{Bmatrix} if\ \ t \bmod 12 = 0, & s = 12 \\ else, & s = t \bmod 12 \end{Bmatrix} \tag{3}$$

$$\mathrm{M}_{t,\alpha}^{SE} = 2^{\text{nd}} \text{ largest } \mathrm{X}_{t,a}^{SE},\ldots,\mathrm{X}_{t,n}^{SE} \tag{4}$$

$$\mathrm{Z}_{t,\alpha}^{SE} = \begin{Bmatrix} \mathrm{X}_{t,\alpha}^{SE}, & if\ \ \mathrm{X}_{t,\alpha}^{SE} \le \mathrm{M}_{t,\alpha}^{SE} \\ \mathrm{M}_{t,\alpha}^{SE} & if\ \ \mathrm{X}_{t,\alpha}^{SE} > \mathrm{M}_{t,\alpha}^{SE} \end{Bmatrix} \tag{5}$$

$$y_t^{SE} = \frac{\sum_{\alpha=1}^{n} \mathrm{Z}_{t,\alpha}^{SE}}{n} \qquad \text{s.t.} \quad \begin{matrix} \alpha = \text{int} \\ n \ge 1 \end{matrix} \tag{6}$$

Notation of Variables:
t = Period being forecasted
n = Number of years previous to the year containing t
s = Period of t during the first year
$\mathrm{X}_{s+12(\alpha-1)}^{SE}$ = Observed values of the SE time-series in the same month as t, in years 1 to one year prior to the year containing t
$\mathrm{X}_{t,\alpha}^{SE}$ = Observed value of the SE time-series in the same month as t, during year α
$\mathrm{M}_{t,\alpha}^{SE}$ = Second largest $\mathrm{X}_{t,\alpha}^{SE}$ value found in $\mathrm{X}_{s+12(\alpha-1)}^{SE}$
$\mathrm{Z}_{t,\alpha}^{SE}$ = Observed value of the SE time-series in the same month as t, during year α whenever $\mathrm{X}_{t,\alpha}^{SE} \le \mathrm{M}_{t,\alpha}^{SE}$, otherwise $\mathrm{Z}_{t,\alpha}^{SE}$ = the second largest $\mathrm{X}_{t,\alpha}^{SE}$ value found in $\mathrm{X}_{s+12(\alpha-1)}^{SE}$
y_t^{SE} = Disaggregate level forecast for period t for storm and emergency

The model organized the data into the month that each observation of usage occurred. A maximum allowable value was set equal to the second largest value occurring during a particular month and then an average was taken across the five points of data available, one from each year. Due to no predictable trend or seasonal effect this monthly averaged level became the forecast for the same month in the coming year. It would be beneficial for this model to consider a greater amount of historical data whenever this information becomes available to further increase smoothing of variability.

This process was selected over somewhat simpler approaches such as averaging or simply finding the median for two reasons. During the overall average calculation one outlier could significantly affect the value, and if only two periods (assuming five points of usage) had substantial levels of usage the median would not be an accurate representation of the series. Table 3.1 shows both of the above possibilities occurring within the aggregated location JA family time-series.

Month	Year 1	Year 2	Year 3	Year 4	Year 5	Avg	Med	SDM
Jan	5	4	10	6	96	24	6	7
Dec	7	4	4	105	71	38	7	31

Table 3.1 Comparison of Possible SE Forecasting Techniques

3.2.3 General Maintenance

As opposed to the previous two segments, general maintenance usage could be fully determined by observations of historical usage and a more traditional time-series forecasting model was appropriate. Shifts in trend occurred slowly overtime and displayed the same qualities one would expect to see from older, more stable inventory items. Exponential smoothing's proven performance and applicability to inventory control due to automatic reintiation and adaptability made it the best choice for this data segment. A minimum of two years of data is required to implement the exponential smoothing model. For this project five years of data was used because of the variety of time-series for which the model was applied, thus allowing a maximum number of observations, and the ability of smoothing parameters to appropriately weight recent and historical data.

General Maintenance Model:

$$S_t = \alpha(X_t/I_{t-12}) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{7}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{8}$$

$$I_t = \delta(X_t/S_t) + (1-\delta)I_{t-12} \tag{9}$$

$$y_t^{GM}\ (m) = (S_t + mT_t)I_{t-12+m} \tag{10}$$

Notation of Variables:
α = Smoothing parameter for the level of the series
γ = Smoothing parameter for the trend
δ = Smoothing parameter for seasonal indices
X_t = Observed value of the time-series in period t
S_t = Smoothed level of series, computed after X_t is observed
T_t = Smoothed additive trend at the end of period t
I_t = Smoothed seasonal index at the end of period t
p = Number of periods in the seasonal cycle
m = Number of periods in the forecast lead-time
y_t^{GM} (m) = Disaggregate level forecast for m periods ahead from origin t for general/maintenance

Holt-Winter's triple exponential smoothing was appropriate for this data because it showed the same seasonal tendencies as the previous segments, peaking in the summer months, and the consideration of trend was necessary due to the long period of time this model was going to be implemented and the occurrence of trending should be expected. During triple exponential smoothing three values are used to generate the final forecast. Each of these values is an exponentially weighted average of past and present computations; as the model progress from period to period the weighted importance of previous calculations decreases exponentially. First, the level is updated by deseasonalizeing the most recent period's data, next a trend is found by computing the change is this level, and finally the seasonal index is determined by dividing the current periods observed usage by the previously calculated level. A forecast is made by projecting out these values m periods ahead. This is done by multiplying the most recent trend by m periods and adding this to the newest level. This value is then multiplied by the known seasonal index for the month being forecasted.

3.2.4 Aggregate

Once a forecast was completed for each of the three disaggregate segments generating an aggregate level forecast was a relatively simple process of adding each of the disaggregate forecasts. As mentioned previously this meant that the variance of each segment was also added together causing a general increase in total variance when compared to aggregate forecasting models. Although it should be expected that in some cases performance would be negatively affected; the benefit of creating demand pattern specific forecasting models for groups of statistically similar data outweighed any general case increases invariance. This benefit is particularly evident in the case of the forward looking trend within the new construction model. The ability of this model to segment data and apply suitable forecasting models was fundamental in its ability to create functional and precise forecasts.

Aggregate Model:

$$Y_t = y_t^{NC} + y_t^{SE} + y_t^{GM} \tag{11}$$

Chapter 4

Ameren Case Study

4.1 Company Background

AmerenUE has enjoyed growth throughout eastern Missouri and areas of Illinois during the past 100 years primarily through acquisitions and mergers. During this time Ameren has grown to be the largest investor owned electric and gas utility in Missouri and in 2007 was in the top third in market capitalization amongst the nation's utility companies. AmerenUE, as it exists today, was created in 1997 by the merger of Union Electric Company and Central Illinois Public Service Company. Ameren grew further through its 2003 acquisition of Central Illinois Light Company and 2004 acquisition of Illinois Power Company. Ameren serves approximately 2.3 million electric customers, 1.2 million of which are in the St. Louis area. These customers currently enjoy rates which are nearly 40% below the national average and the second lowest of any metropolitan region in the nation. The company prides itself on financial strength, low rates, cost containment, and customer service. Ameren works to ensure reliable low-cost service to all of its customers by continuously improving systems to ensure company efficiencies into the future.

4.2 Project Objectives

The intent of this project was to create a precise and dynamic forecasting model which can easily be applied to several time-series at varying levels of aggregation. This forecasting tool will be utilized by Ameren to help insure that the correct items are in the correct locations whenever they are required. The ability to properly predict transformer usage will aid in sustaining a continuous transfer of power and thus help insure overall customer satisfaction. The approach here was to exploit current research in addition to developing a new general methodology while creating a company specific tool which would be used well into the future.

4.3 Data Provided by Ameren

The original inventory data provided by Ameren consisted of a Microsoft Access database file containing inventory records for all issues and receipts for four primary warehouses from January 2003 until December 2007. Included in this document were the dates and quantity of each transaction along with a brief description, unit price, and several inventory control values for every item. The entries were made according to stock numbers and were contained within six tables, one for each year and an additional table containing the descriptions and unit prices.

Further information was provided from the DOJM database in the form of a Microsoft Excel file which contained transformer usage history. Usage values were at the monthly level from January 2003 until July 2008. The database was separable by distribution number, operating center, activity code, transformer type, and date. Supplementary forms were also provided which explained operating center, transformer, and activity codes used within Excel.

4.3.1 Inventory Database

All transformers issued from inventory were totaled for each month over the five years of data that was provided. The values found within the inventory database were relatively complete. However, there was no way of separating the data according to the reasons items were used. This was a significant drawback to the inventory database because forecasts at the aggregate level were unsatisfactory due to the knowledge that several unrelated causes of transformer usage were not being considered. A more powerful approach would consider the varying causes and significant types of demand patterns within the aggregate.

4.3.2 DOJM Database

The DOJM database, for the purposes of this model, was much more informative than the issues and receipts alone from the inventory database. The flexibility provided by the capability to separate not only incoming and outgoing items, but also activity codes allowed for disaggregation of the historical monthly totals into three subcategories: new construction (NC), storm and emergency (SE), and general maintenance (GM). The three subcategories were distinctive from one another in that they each reflected notably unique causes of transformer usage. The aggregate monthly totals were calculated along with the monthly totals for each disaggregate part. Although the DOJM database was of great use, it unfortunately had lower usage levels in almost every month when compared to those contained in the inventory database. Usage values appearing in the DOJM database were considered incomplete due to reporting error and delay caused by the additional detail recorded in the database.

4.4 Desired Data

In order to create a forecasting model which is as accurate as possible, it is exceptionally important to use historical data that is both accurate and appropriate for projection into the future. The most desirable data in the case of this study would include the most precise record of usage with respect to both amount and period of use. Additionally this data would need to be separable into the three categories of interest. Regrettably neither database individually could provide the level of effectiveness required.

4.5 Data Extraction

So as to benefit from the unique and significant advantages of each database a method was developed which utilizes the more complete inventory data while separating it in accordance with the historical records shown within the DOJM database. In order to do this it was first necessary to determine whether or not the DOJM database represented a large enough portion of the inventory data to provide relevant information pertaining to the ratios of each disaggregate segment. During periods in which the DOJM database did not sufficiently denote the inventory data it was desirable to use a standardized ratio for that particular month of the year. Once an appropriate ratio was determined for each period, they could then be applied to the more complete inventory data so as to disaggregate it into the three segments of interest.

4.5.1 Step One: Comparison of Quantities

Transformer usage is documented by both the inventory and DOJM databases. While the historical records within these databases are fairly similar, there remains a significant amount of discrepancy between them. The inventory database in almost all periods contained higher quantity values than the DOJM database and it was necessary to preserve this more complete record of usage. Validating the usage of DOJM's ratios upon the inventory database required that a minimum level of the transformer usage shown in the inventory records also be displayed in the DOJM database. Figure 4.1 compares the inventory and DOJM reported monthly usage levels over the provided 5 years of data.

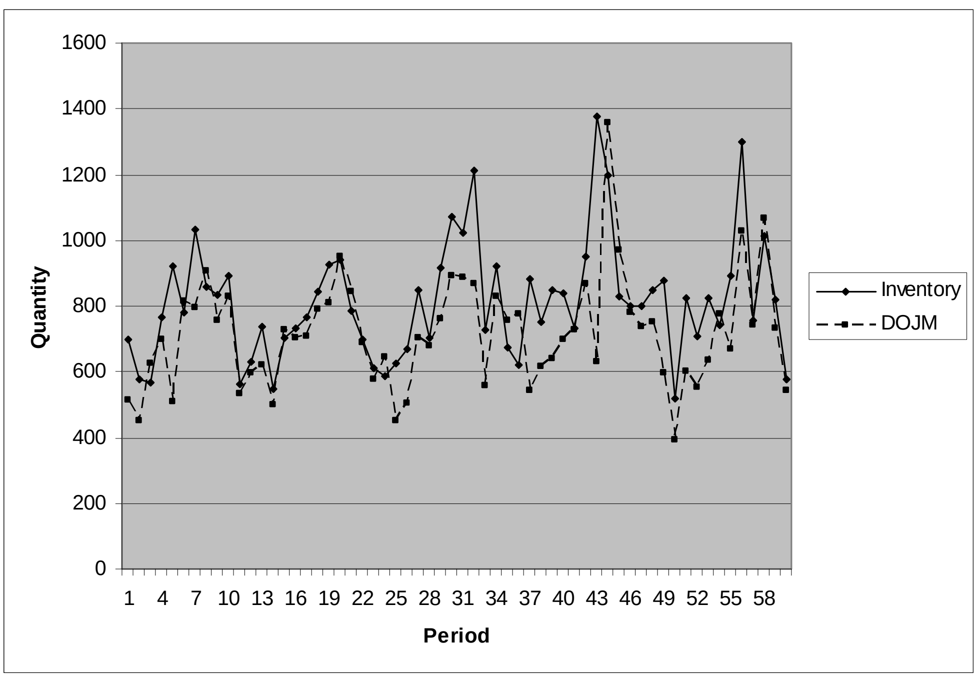

Figure 4.1 Comparison of Database Usages

It was established that the DOJM database must contain a usage level greater than or equal to 70% of what was found in the inventory records. In periods where there was not sufficient usage exhibited by DOJM it was reasoned that the ratios would be unreliable due to certain segments being under represented within the total usage and thus causing an inaccurate calculation of the ratios for that period. Table 4.1 presents the percentage of the inventory usage that was accounted for within DOJM during each period while insufficient levels of DOJM usage are marked with an asterisk.

Period	Inventory	DOJM	Ratio		Period	Inventory	DOJM	Ratio
1	696	515	74%		31	1021	888	87%
2	579	449	78%		32	1214	866	71%
3	568	624	110%		33	726	558	77%
4	766	698	91%		34	919	830	90%
5	923	508	55%*		35	675	756	112%
6	782	815	104%		36	619	778	126%
7	1031	796	77%		37	882	542	61%*
8	856	906	106%		38	753	615	82%
9	834	754	90%		39	850	641	75%
10	893	829	93%		40	840	698	83%
11	563	531	94%		41	734	728	99%
12	631	595	94%		42	950	870	92%
13	739	619	84%		43	1376	632	46%*
14	549	498	91%		44	1196	1360	114%
15	703	725	103%		45	829	969	117%
16	731	702	96%		46	800	781	98%
17	768	707	92%		47	802	739	92%
18	846	790	93%		48	849	750	88%
19	927	811	87%		49	877	595	68%*
20	941	949	101%		50	519	393	76%
21	787	843	107%		51	824	601	73%
22	698	688	99%		52	708	554	78%
23	613	579	94%		53	822	634	77%
24	586	646	110%		54	743	775	104%
25	624	453	73%		55	894	670	75%
26	671	506	75%		56	1301	1030	79%
27	850	705	83%		57	754	741	98%
28	701	678	97%		58	1011	1067	106%
29	914	762	83%		59	820	734	90%
30	1072	894	83%		60	576	544	94%

Table 4.1 Percentile Comparison of Database Usage Levels

4.5.2 Step Two: Calculation of Standard Ratios

As shown in Table 4.1, the DOJM database had insufficient levels of usage in periods 5, 37, 43, and 49. During these periods the ratios shown by DOJM are not reliable enough to be applied to the inventory usage. Table 4.2 displays the ratios of each segment found in DOJM for every period over the five years of data, once again insufficient periods are denoted.

Month	2003 NC:SE:GM	2004 NC:SE:GM	2005 NC:SE:GM	2006 NC:SE:GM	2007 NC:SE:GM
Jan	45 : 5 : 50	58 : 2 : 40	59 : 5 : 36	71 : 3 : 26*	50 : 29 : 21*
Feb	60 : 13 : 27	62 : 3 : 35	65 : 3 : 32	71 : 4 : 25	59 : 15 : 26
Mar	60 : 3 : 37	58 : 3 : 39	63 : 3 : 34	63 : 8 : 29	62 : 10 : 28
Apr	51 : 3 : 46	49 : 3 : 48	57 : 9 : 34	63 : 16 : 21	59 : 10 : 31
May	53 : 12 : 35*	60 : 8 : 32	59 : 3 : 38	63 : 13 : 24	55 : 9 : 36
Jun	45 : 30 : 25	45 : 14 : 41	53 : 12 : 35	51 : 16 : 33	54 : 18 : 28
Jul	54 : 10 : 36	49 : 12 : 39	58 : 9 : 33	48 : 30 : 22*	52 : 11 : 37
Aug	51 : 5 : 44	61 : 9 : 30	59 : 14 : 27	40 : 39 : 21	40 : 33 : 27
Sep	60 : 3 : 37	64 : 2 : 34	62 : 7 : 31	55 : 22 : 23	56 : 11 : 33
Oct	63 : 2 : 35	65 : 3 : 32	75 : 2 : 23	69 : 7 : 24	50 : 9 : 41
Nov	71 : 2 : 27	68 : 2 : 30	71 : 5 : 24	62 : 7 : 31	47 : 9 : 44
Dec	67 : 3 : 30	60 : 3 : 37	69 : 2 : 29	53 : 29 : 18	41 : 26 : 33

Table 4.2 DOJM: Monthly Segmented Ratios

In these cases a standard ratio must be determined based upon DOJM records for the same month but in different years that did have sufficiently high levels of usage. Thus a standard ratio was generated for each month, which could be used in low periods occurring during that month. The monthly standard ratios were computed by averaging each segment's (NC, SE, and GM) ratio across the five points of data we had for that month. If a period of low usage occurred within a month, the periods were eliminated from the average so as not to be reflected in the standard ratio.

4.5.3 Step Three: Determine the Standardized Ratios for Insufficient Periods

Due to each month being distinct from year to year it was appropriate to still utilize the ratios shown during each period even if that particular period was determined to be of low level. However, these ratios could not be simply applied as they are in acceptable periods for the reasons described above. Due to this, the standard ratios were calculated and then applied in the form of a weighted average. The ratios found in these low level months were maintained but given a weight equal to the percentage of inventory usage contained in the DOJM database for that period (ρ_t^{DOJM}). The standard ratios were given the remaining weight $(1-\rho_t^{DOJM})$ and added to the ratios found in the low level period. The calculations for the four periods of insufficient data are shown below.

Notation of Variables:
θ_{AVG}^{JAN} = Standard ratios for the month of January
θ_{AVG}^{MAY} = Standard ratios for the month of May
θ_{AVG}^{JUL} = Standard ratios for the month of July
ρ_t^{DOJM} = Percentage of the Inventory Database shown in the DOJM database during period t
NC_t^{DOJM} = Ratio shown for new construction in the DOJM database during period t
SE_t^{DOJM} = Ratio shown for storms and emergencies in the DOJM database during period t
GM_t^{DOJM} = Ratio shown for general maintenance in the DOJM database during period t
Θ_t = Standardized ratios for period t

Standardized Ratios for Period 5

$$\theta_{AVG}^{MAY} = \left(\frac{60+59+63+55}{4} : \frac{8+3+13+9}{4} : \frac{32+38+24+36}{4}\right) = (59:8:33)$$

$$\Theta_5 = \rho_5^{DOJM}(NC_t^{DOJM} : SE_t^{DOJM} : GM_t^{DOJM}) + (1-\rho_5^{DOJM})(\theta_{AVG}^{MAY})$$

$$\Theta_5 = 56:10:34$$

Standardized Ratios for Period 37

$$\theta_{AVG}^{JAN} = \left(\frac{45+58+59}{3} : \frac{4+2+6}{3} : \frac{50+40+36}{3} \right) = (54{:}4{:}42)$$

$$\Theta_{37} = \rho_{37}^{DOJM} (NC_t^{DOJM} : SE_t^{DOJM} : GM_t^{DOJM}) + (1 - \rho_{37}^{DOJM})(\theta_{AVG}^{JAN})$$

$$\Theta_{37} = 64:4:32$$

Standardized Ratios for Period 43

$$\theta_{AVG}^{JUL} = \left(\frac{54+49+58+52}{4} : \frac{11+11+9+11}{4} : \frac{36+39+33+37}{4} \right) = (53{:}11{:}36)$$

$$\Theta_{43} = \rho_{43}^{DOJM} (NC_t^{DOJM} : SE_t^{DOJM} : GM_t^{DOJM}) + (1 - \rho_{43}^{DOJM})(\theta_{AVG}^{JUL})$$

$$\Theta_{43} = 51:19:30$$

Standardized Ratios for Period 49

$$\theta_{AVG}^{JAN} = \left(\frac{45+58+59}{3} : \frac{4+2+6}{3} : \frac{50+40+36}{3} \right) = (54{:}4{:}42)$$

$$\Theta_{49} = \rho_{49}^{DOJM} (NC_t^{DOJM} : SE_t^{DOJM} : GM_t^{DOJM}) + (1 - \rho_{49}^{DOJM})(\theta_{AVG}^{JAN})$$

$$\Theta_{49} = 51:21:28$$

Now that the low level periods had been properly standardized in accordance with the calculated standard ratios segmented monthly ratios were once again generated. Now standardized, these ratios were able to be applied to the usage levels shown by the inventory database. The updated monthly segmented ratios are shown in Table 4.3.

Month	**2003** NC:SE:GM	**2004** NC:SE:GM	**2005** NC:SE:GM	**2006** NC:SE:GM	**2007** NC:SE:GM
Jan	45 : 5 : 50	58 : 2 : 40	59 : 5 : 36	64 : 4 : 32*	51 : 21 : 28*
Feb	60 : 13 : 27	62 : 3 : 35	65 : 3 : 32	71 : 4 : 25	59 : 15 : 26
Mar	60 : 3 : 37	58 : 3 : 39	63 : 3 : 34	63 : 8 : 29	62 : 10 : 28
Apr	51 : 3 : 46	49 : 3 : 48	57 : 9 : 34	63 : 16 : 21	59 : 10 : 31
May	56 : 10 : 34*	60 : 8 : 32	59 : 3 : 38	63 : 13 : 24	55 : 9 : 36
Jun	45 : 30 : 25	45 : 14 : 41	53 : 12 : 35	51 : 16 : 33	54 : 18 : 28
Jul	54 : 10 : 36	49 : 12 : 39	58 : 9 : 33	51 : 19 : 30*	52 : 11 : 37
Aug	51 : 5 : 44	61 : 9 : 30	59 : 14 : 27	40 : 39 : 21	40 : 33 : 27
Sep	60 : 3 : 37	64 : 2 : 34	62 : 7 : 31	55 : 22 : 23	56 : 11 : 33
Oct	63 : 2 : 35	65 : 3 : 32	75 : 2 : 23	69 : 7 : 24	50 : 9 : 41
Nov	71 : 2 : 27	68 : 2 : 30	71 : 5 : 24	62 : 7 : 31	47 : 9 : 44
Dec	67 : 3 : 30	60 : 3 : 37	69 : 2 : 29	53 : 29 : 18	41 : 26 : 33

Table 4.3 Standardized: Monthly Segmented Ratios

4.5.4 Step Four: Segmentation of Inventory Database

Now that a standardized ratio existed for each period, these ratios could now be simply applied to the data contained in the inventory database. This was done through a process of causing the summation of each period's ratios to equal one and then multiplying the aggregate level inventory data by each segments ratio, thus creating a usage level for each of the three disaggregate levels (NC, SE, and GM) which summed to the aggregate. Figure 4.2 shows the usage levels, as determined from the inventory data and the application of the calculated standardized ratios, for the aggregate and disaggregate levels.

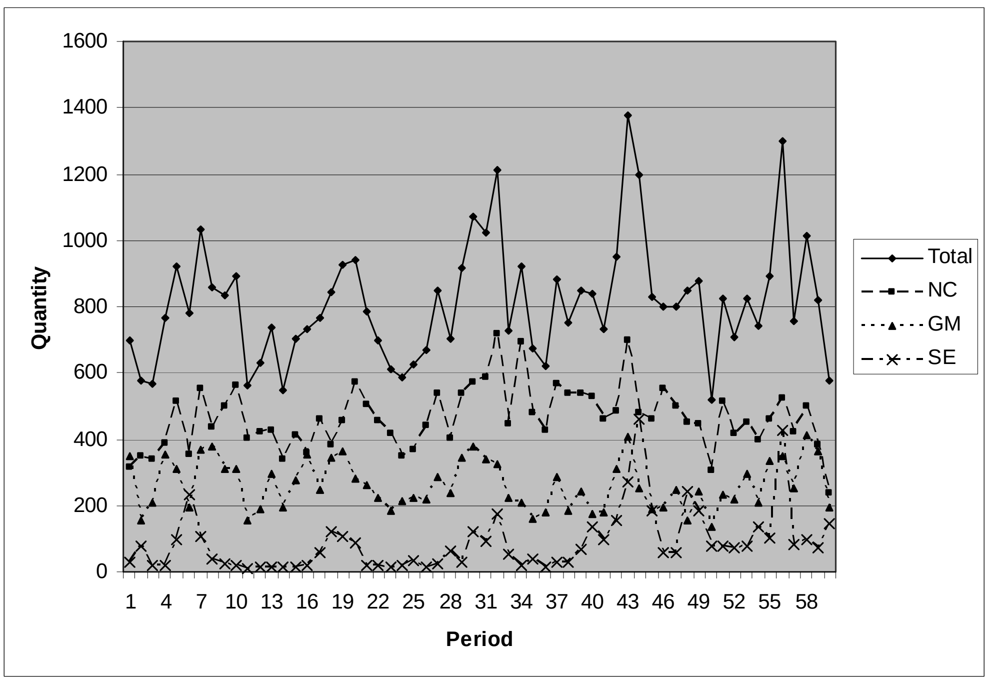

Figure 4.2 Segmented Data from Inventory Database

This methodology successfully presented the provided data in a way that was both accurate and appropriate for future forecasting purposes. The data now incorporated the most beneficial parts of each database, completeness (inventory database) and disaggregation (DOJM database). In the following sections the significance of disaggregating the data will be further discussed and its relevance to the selection of the most appropriate forecasting model for each segment defined.

4.6 Disaggregate Models

Now that transformer usage due to new construction, storm and emergency, and general maintenance was identified, individual forecasts which were most appropriate for each distinctive demand pattern could be created. Initially transformer usage caused by new construction was of particular interest due to its heavy dependence on economic trends and the general health of the housing and construction industries. It was also clear that outside factors had a significant impact on usage caused by storms and emergencies and that predicting weather conditions a year into the future would be highly suspect of error. Additionally, historical records showed significant outliers within this segment of the data whose difficulty of forecasting renders their use within a backward looking model unreasonable. The final segment, general maintenance, differed from the other two segments because no outside factor caused significant trending in the historical record. For this reason a purely backward looking model was justified when forecasting the final segment.

4.6.1 New Construction Forecasting Model

The demand shown in the new construction data did not follow a pattern which would be best forecast by a typical time-series model. Several years of continued upward trend could easily turn into a downward slope at anytime. Figure 4.3 shows the historical data clearly trending up from 2004-2006 and abruptly turning into a downward trend in 2007.

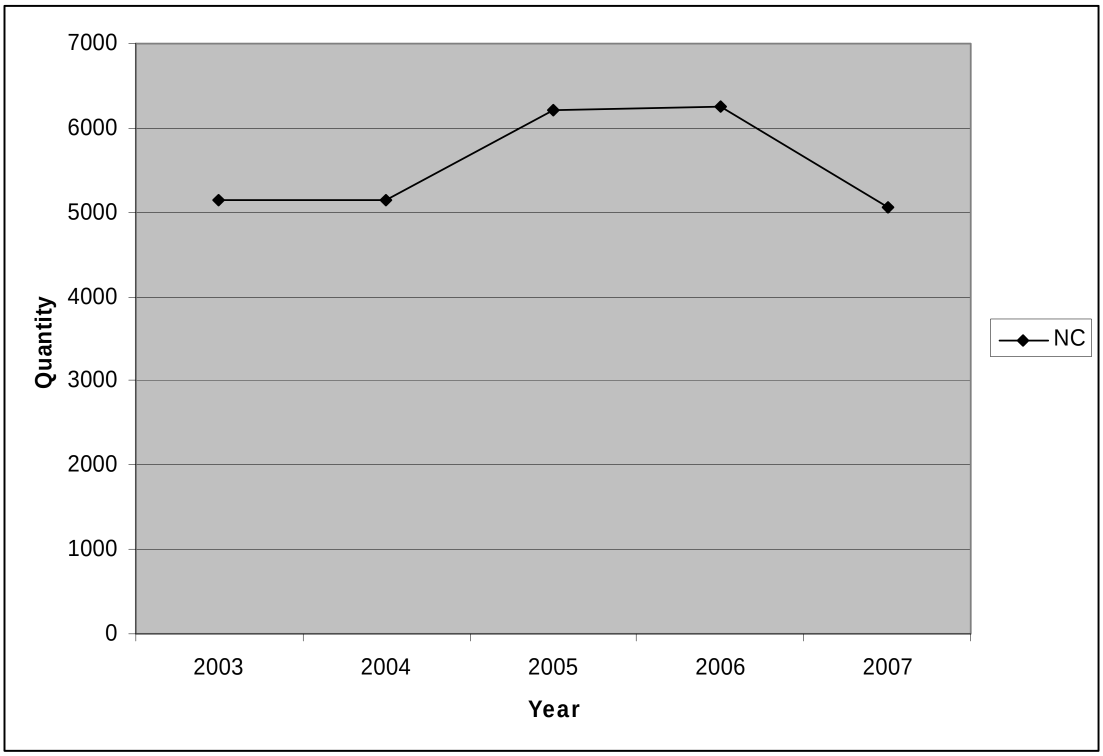

Figure 4.3 New Construction Yearly Usage

Common forecasting techniques such as moving average and exponential smoothing models would not be able to predict this sudden shift in trend leading to inappropriate projected trends and highly inaccurate forecasting results. It would be advantageous to allow for a trend to be inputted into the model, so as to consider currently known conditions and predictable changes in the future, rather than simply being projected based upon past observations.

So as to enable the consideration of outside factors which clearly influence the trend of Ameren's transformer usage, influences which could not be contained in historical data alone were considered. The influence of future construction and construction indicators on Ameren's NC usage was heavily considered and two primary indicators of trend determined. The sources for the economic information provided were the Federal Reserve Bank of St. Louis and the U.S. Census Bureau. The indicators used were both U.S and St. Louis' annual private housing permits. The complete document of provided resources can be viewed in Appendix A. From these resources it was observed that NC transformer usage lagged approximately one year behind both indicators. For this reason the average year over year percent change in the indicators is the recommend forward-looking trend input. It is important to note that the goal of these resources was to provide lasting information which could be relied upon to be maintained in the future. Commercial construction was a sought after indicator during this research; however, this information was extremely hard to obtain and not all commercial construction requires transformers (i.e. bridges, highways, and waste management). Additionally, an observation of housing permits shows them to be a strong indicator of trend and the use of further data not as essential. An experienced and knowledgeable technician, provided with the recommended future trend, could then easily input the user specified trend directly into the model.

The complete forecast for the new construction disaggregated segment is created in two simple steps. First, a general level for each month over the year being forecasted is computed by averaging the value observed for that month during the past two years. Fewer years were used in this calculation of level than in the other segments because a trend is not established over the years and more recent data is the most relevant. This value is set equal to that period's level, L(t). Next a relative trend, T(t), is determined by multiplying the previously calculated level by either the user specified trend or, whenever this value is not available, a default trend based solely on historical data. This default trend is equal to the moving average of the year over year change in usage of the previous twelve months. Finally these two values, level and trend, are added together in the creation of the final forecast for each period. It should be noted that for the creation of the default or historical trend it is necessary to update usage levels every month. However, in the circumstance where this is not possible the most recent observed twelve months of usage are used during the historical trend calculation and that trend is carried through the forecasted year. A graph of historical levels and a forecast with a trend of -25% is shown in Figure 4.4.

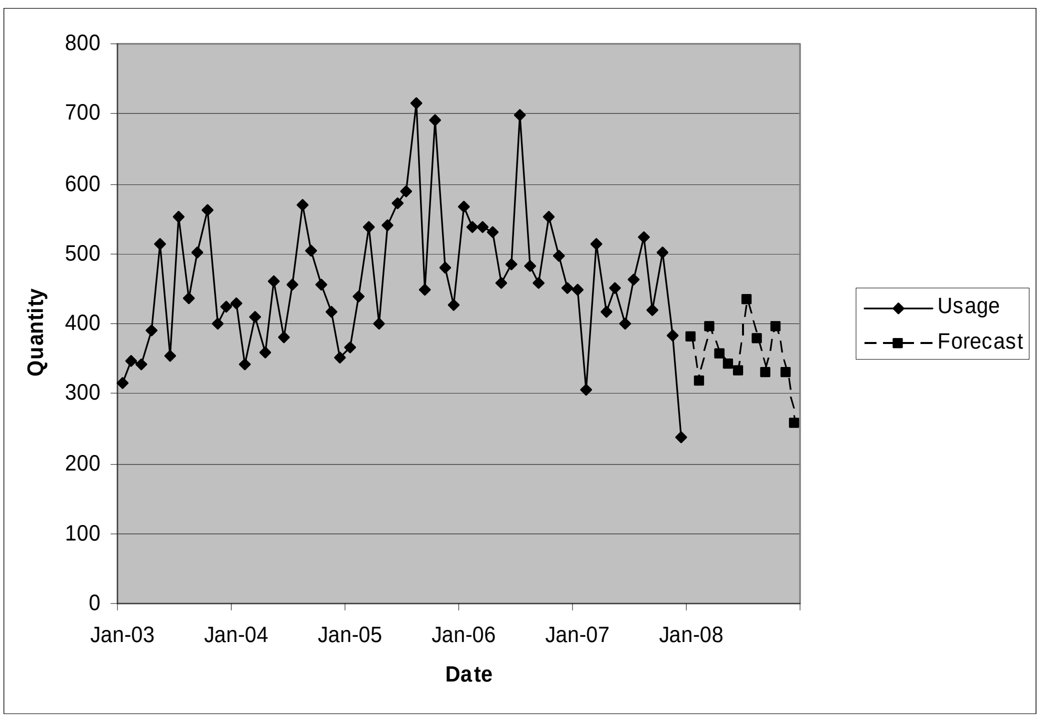

Figure 4.4 New Construction Forecast

4.6.2 Storm and Emergency Forecasting Model

Weather and emergency related usage is inevitably difficult to forecast. For this reason it has been separated from the aggregate data so as to both limit its influence on the general historical based model (i.e. general maintenance) and to allow for the application of a forecasting model which naturally inhibits the occurrence of outliers. Most usage values within the storm and emergency data occur between 0 and 100; yet, there are several occurrences of usage greater than 200 and some as high as 450. These outliers are impossible to predict and should not be assumed to be a reoccurring event. Nevertheless, it should be clarified that the data did display cyclical properties or more correctly season effects. Outages are most often caused by heat. It has been observed by the company that transformers were much more likely to fail after several consecutive days where the temperature does not go below eighty degrees. Due to this, storm and emergency usage levels generally peak during the summer months; however, it would not be reasonable to assume this summer's data would be an accurate predictor of the next. Due to the unique volatility within this data it was appropriate to separate it from the more consistent demand patterns and create a simple forecast which controls substantial outliers. The sixty observations of usage due to storms and emergencies during the five years of data can be seen in Figure 4.5.

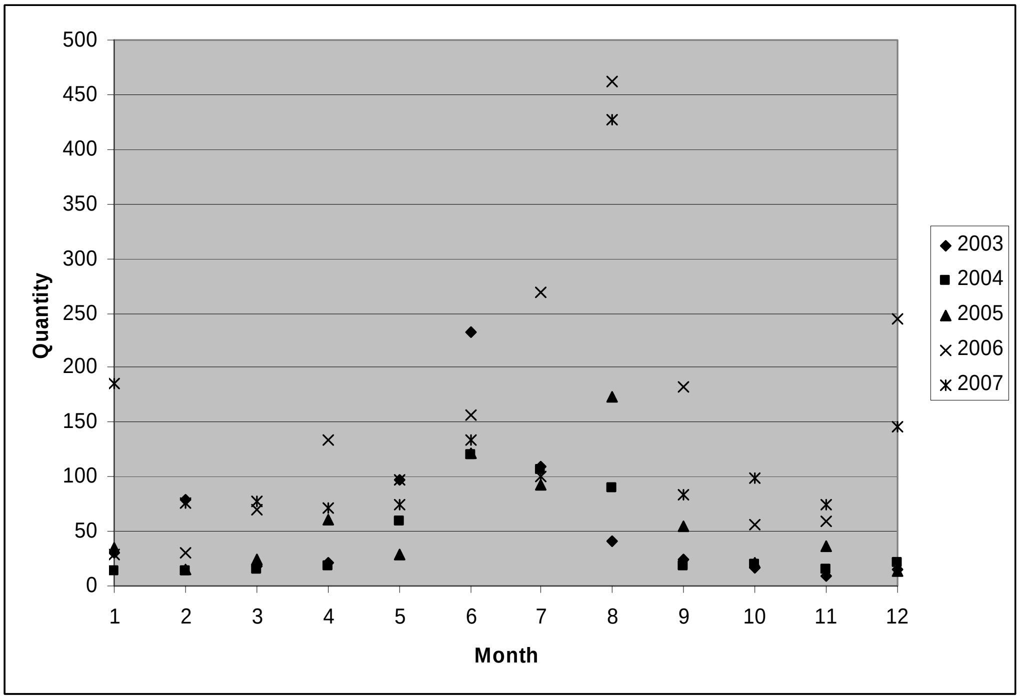

Figure 4.5 Storm and Emergency Monthly Usages

To accomplish this, a model was created which considers all five years of historical data in order to better smooth out volatility once the values are ultimately averaged. Trending the effects of weather over time would be an inappropriate approach because several years of increasingly warm summers is not a reasonable cause of the following summer's temperature to increase. Thus, no trend was to be considered from year to year and the model was entirely dependent on the determined level. The level for each month was an average of the usage shown during that particular month over the past five years. As previously mentioned it was advantageous to automatically limit the number and severity of outliers considered within the average; so prior to its calculation a maximum level (M) was set equal to the second highest value observed for the month being forecast. Thus twelve values for M were determined, one for each month. As a result each month's observed maximum value was lowered to the previously defined M and an average was finally calculated based upon the three values less than M and the two values equal to M. This average was set equal to the forecasted level and, with the absence of a trend, equal to the final forecasted value for that month. Figure 4.6 shows the next year forecast for storm and emergency usage.

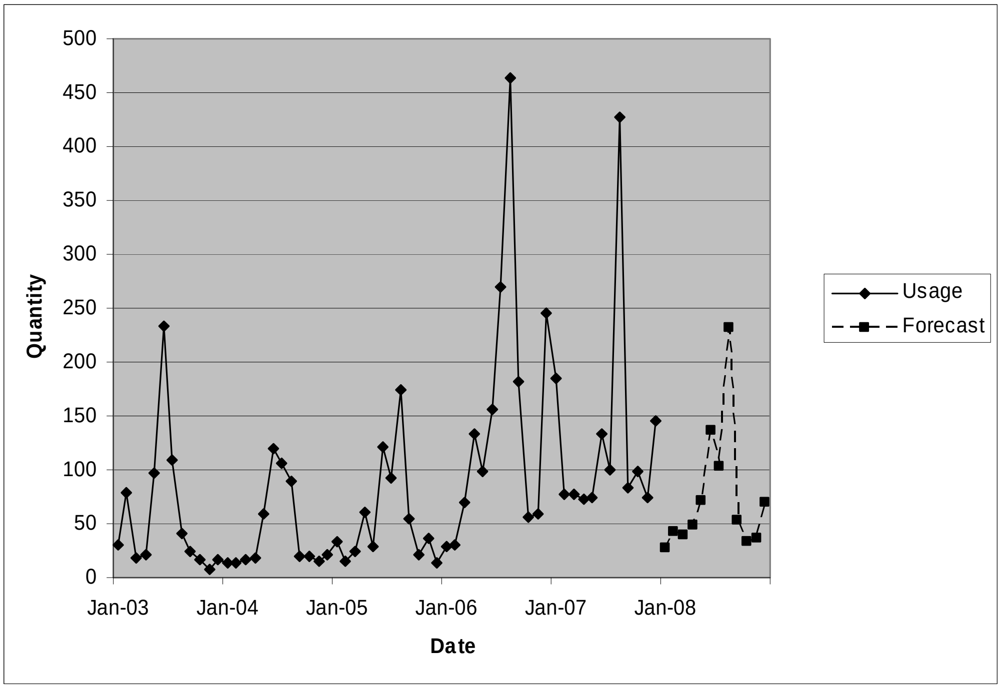

Figure 4.6 Storm and Emergency Forecast

4.6.3 General Maintenance Forecasting Model

After evaluating the historical data, general maintenance usage was found to be independent of outside influences and was determined to be well predicted by past levels of demand. For this disaggregated segment of the data a traditional time-series forecasting model would be most appropriate, and a triple exponential smoothing model was selected due to the data displaying seasonal effects and to allow for better adaption to gradual trending. This trending was significantly different than that found in the new construction data in that it consisted of a slow change over several years. This was caused by general maintenance issues being more insulated from sudden shifts in the general economy than new construction usage. This type of trending is well handled by conventional models.

The five years of previous data was utilized to create initial values for the level, S(t), trend, T(t), and the first twelve months of seasonal effects, I(t). Once these were calculated the model could be activated. The smoothed level, S(t), is updated by dividing the observed usage value for the current month by the seasonal effect for that month in the previous year, thus removing any seasonal component in the level. This current period deseasonalized value is used in a weighted average calculation between itself and the summation of the previous month's level and trend. The purpose is to weight both the currently observed level and all previous levels. The current month is given a weight equal to the smoothing parameter alpha while the past months' weight is equal to 1-alpha. As a new month is added, the weighted importance of each previous month decreases exponentially. Next, the smoothed trend, T(t), was updated based upon the same exponential weighted average method using a new smoothing parameter gamma. However, now the values being weighted are the current trend, which is equal to the current months smoothed level minus the previous months, and the previously calculated trends. The smoothed seasonal component was updated in the same fashion. Once again it uses a new smoothing component, delta, and the values being weighted are the current month's seasonal effect and all previous seasonal effect. Excel solver was used to solve the optimal values for each smoothing parameter, alpha, gamma, and delta. These values created a best-fit line for the historical data which limits error and could be projected into the future to create forecasts. This line was then projected twelve months into the future. The forecasted value for each month was equal to the current months smoothed level plus the current smoothed trend after it has been multiplied by the number of months projected into the future. The sum of these is then multiplied by the smoothed seasonal effect last calculated for the month being forecasted.

This method ultimately creates a twelve month ahead forecast which considers both seasonal and trending effects. It is recommended that new usage levels be inputted once a year. This is because the further into the future projections are made the less reliable they become. The next forecast for demand due to general maintenance can be seen in Figure 4.7.

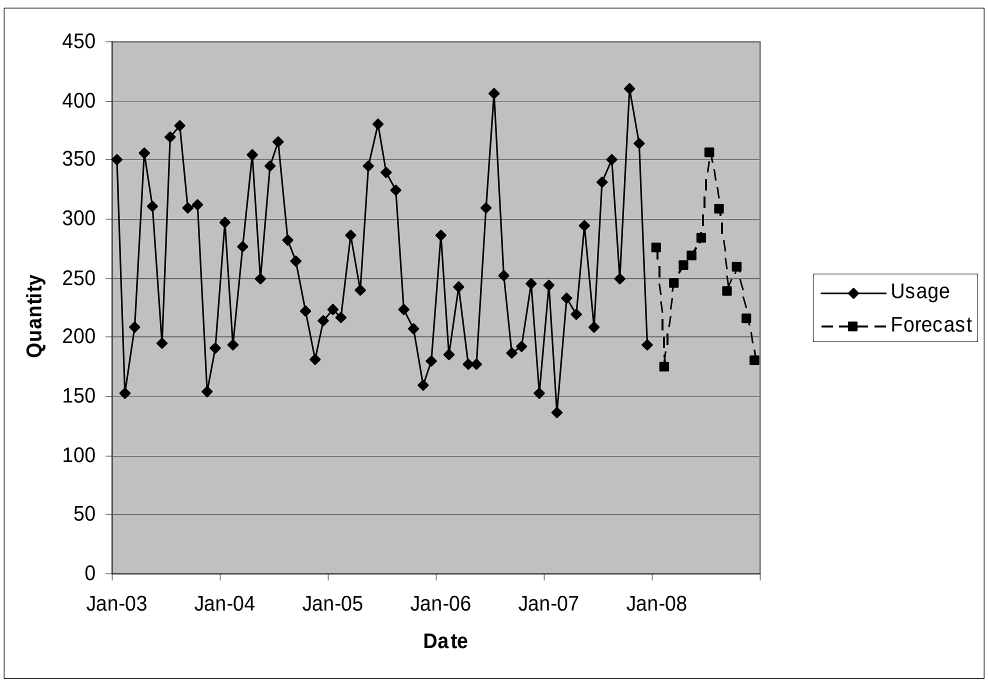

Figure 4.7 General Maintenance Forecast

4.7 Aggregate Model

The primary objective of the project was to ultimately create a monthly demand forecast for transformer usage in the coming year. The central focus was obviously placed on the overall number of transformers used, regardless of the cause of usage. As described above, creating the most accurate aggregate level forecast in the end meant disaggregating usages based upon key demand patterns and using the most appropriate forecast for each case. Once this was accomplished the monthly aggregate level forecast of interest could easily be created by adding together each month's disaggregate forecast. Each disaggregate forecast and their summation, the final aggregate forecast, is shown in Figure 4.8.

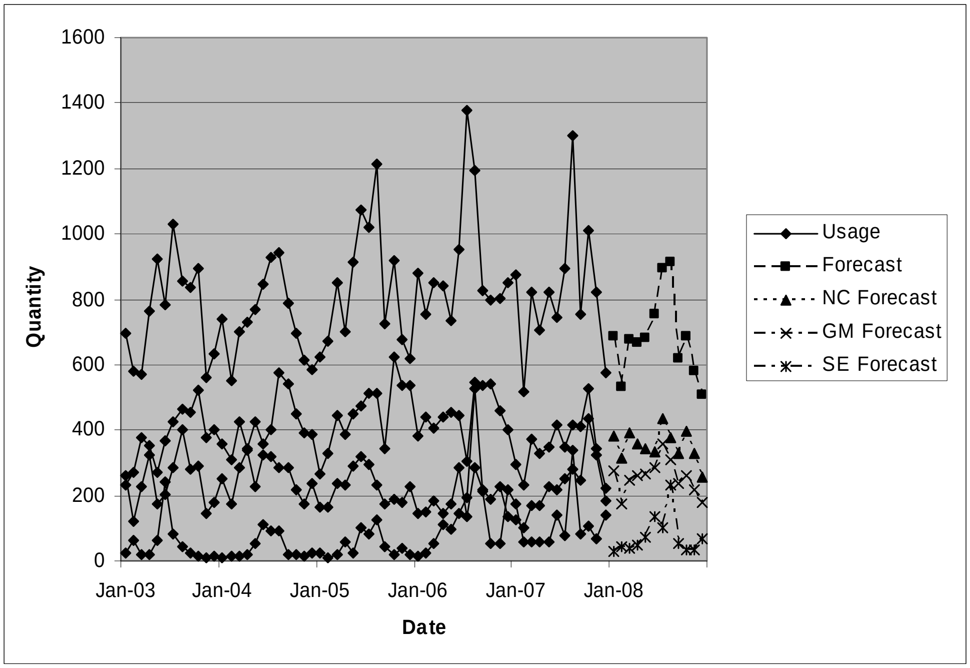

Figure 4.8 Total Combined Forecast

4.7.1 Validation

Validating a forecasting model is often considered as much of an art form as creating the model itself and great amount of research can be done on the varying techniques used during model evaluation. However, for the functions of this paper, model assessment has been limited to MdAPE and GMRAE due to Armstrong and Collopy's recommendations and MSE and MAE due to their greater level of familiarity. As recommended by Armstrong and Collopy (1992) [2] the geometric mean of the relative absolute error (GMRAE) was used when making comparisons between our specialized disaggregate model (SDM), triple exponential smoothing (ES), and simply averaging each month's observed values for the previous years. These forecasting techniques, ES and monthly averaging, were selected for comparison purposes due to both their popularity of use and their simplicity. As mentioned in Chapter 3, because of the large number of time-series needing to be forecasted, a simple technique which could relatively easily be reinitiated was most appropriate and thus the most suitable for comparison.

The benchmark used for comparison during the GMRAE calculation was the random walk where the forecasted value is equal to the last observed value. The value for each of the twelve months being forecasted was found and then divided by the value obtained from the random walk for that month. The geometric average of these twelve months of ratios was then found. This process was repeated for each of the three forecasting methods being evaluated and comparisons were made as to which performed best on several individual time-series at varying levels of usage. The ratios shown in Table 4.4 represent each model's ability to outperform the random walk. A value less than one means the model was more accurate than the random walk, one means they were equivalent, and anything greater than one was less accurate than the random walk, moreover, the lower the value, the more accurate the model. Table 4.4 also compares SDM's forecasts to the exponential smoothing and monthly averaging forecasts using the previously mentioned MdAPE, MAE, and MSE. Each of the cases tested represents a different level of usage varying from all locations and an aggregation of all items down to a particular location (Mexico) and product family (MR). This was done so that the accuracy of the model could be tested across the various levels of usage for which it will be used.

Model	Location	Item	Trend	GMRAE	MdAPE	MAE	MSE
SDM	All	Aggregate	-15%	0.57	16%	141	30377
ES	All	Aggregate	N/A	0.48	25%	168	46920
Mon. Avg	All	Aggregate	N/A	0.54	17%	125	20120
SDM	Mexico	Aggregate	-15%	0.43	18%	26	1374
ES	Mexico	Aggregate	N/A	0.51	25%	29	1522
Mon. Avg	Mexico	Aggregate	N/A	0.71	39%	35	1872
SDM	Cape	Aggregate	-15%	0.82	27%	17	401
ES	Cape	Aggregate	N/A	0.89	44%	24	925
Mon. Avg	Cape	Aggregate	N/A	0.89	26%	18	421
SDM	All	MR	-15%	0.36	17%	32	1456
ES	All	MR	N/A	0.47	20%	36	1791
Mon. Avg	All	MR	N/A	0.46	20%	34	1528
SDM	Mexico	MR	-15%	0.47	27%	6	64
ES	Mexico	MR	N/A	0.73	42%	8	76
Mon. Avg	Mexico	MR	N/A	0.67	42%	7	70

Table 4.4 Comparison of Model Performance

Only the specialized disaggregate model allowed for the input of a forward looking trend. In all of the above cases a trend of -15% was used due to the evaluation of outside indicators of future construction (See Appendix A). Since a disaggregate model was used, the matter of offsetting errors should also be considered during this analysis. This is a well known issue with disaggregate models and occurs when certain disaggregate segments are overestimated while others are underestimated yielding an overall accurate aggregate forecast from less accurate disaggregate forecasts. While this should be acknowledged, the specialized disaggregate model outperformed the comparison models by either GMRAE or MdAPE in every case tested.

Figure 4.9 displays the Mexico aggregated item forecast for each model compared. This comparison illustrates one of the primary advantages of the SD model which is its ability to consider outside factors and thus predict future trends before they become apparent in the usage data. It can be seen that both ES and monthly averaging tended to overestimate usage due to negative trending occurring during 2007 that could not be obtained by exclusively using historical data. According to the MdAPE evaluation for this case, the SD model was 7% more accurate than ES and over 20% more accurate than monthly averaging.

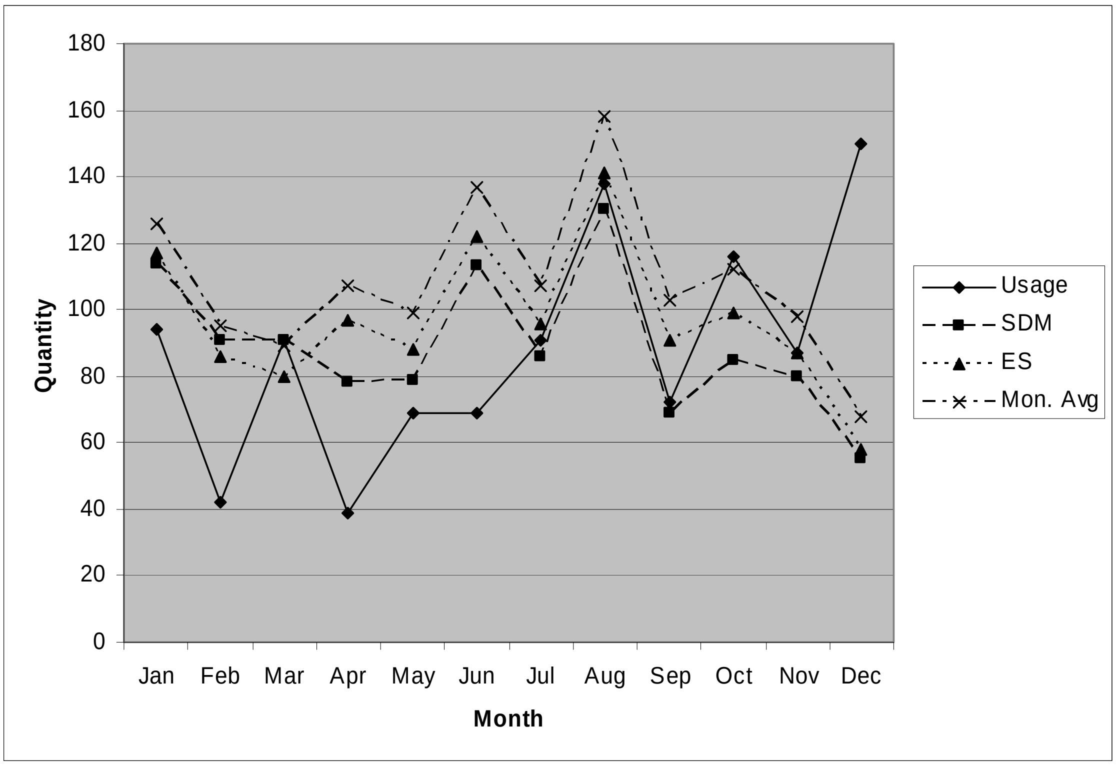

Figure 4.9 Comparison of Forecasting Models: Forecasts for 2007

Due to the disaggregation of the data during the specialized disaggregate model some cases should be expected to underperform because of the addition of variances that occur when the forecasts are aggregated. Nevertheless, the SD model performed the best overall which is significant when forecasting a large number of time-series. As shown in the case of exponential smoothing, which according to GMRAE performed well at the aggregate level but very poorly at lower levels, broad application is important and performance on one time-series does not predict overall model performance.

Additionally, it is recommended to take advantage of the ease of generating forecasts with the SD model and test several scenarios for the future trend. Due to its flexibility in forecasting different time-series and the ability to quickly create “what if” scenarios the specialized disaggregate model was considered to have performed better than both triple exponential smoothing and monthly averaging.

Chapter 5

Specialized Disaggregate Model: Excel Implementation

An Excel based model incorporating the techniques described in previous chapters was utilized to directly implement the developed model in a user friendly and practical configuration. A full description of the model is presented below to better explain how the data was processed during application and to allow future users to more easily update necessary historical data and model parameters. Please note the model described in this chapter is the actual forecasting model, thus in appropriate cases all five years of historical data were considered; however for validation purposes the same model was used but 2007 data was withheld to draw conclusions on model performance.

5.1 Overview

Upon opening the model the user input screen is displayed. This screen is grouped into two primary sections: model parameters and model view. Three input variables (location, item, and trend) are displayed within model parameters (Figure 5.1). Each of these is used as input to the model and should be specified by the user. The variables are inputted via a drop down menu from which possible entries can be selected. There are four possible choices for location, one for each of the three primary warehousing locations and an aggregate of all locations. This allows the user to select a location of interest but also determine general aggregate usage. Eleven choices are provided from the item drop down menu. This includes seven transformers which represent the majority of usage from each of the locations, three product families which can be run on similar manufacturing lines, and an aggregate of all transformers. It should be noted that in all cases a selection of aggregate not only is an aggregation of the other choices but all locations and/or transformers. The third input variable is a forward looking trending variable which is used during the new construction forecast. The intent of this input is to predict future trends within the new construction usage which could not be determined from historical observation alone. Guidance for its selection can be found in Chapter 4 and Appendix A.

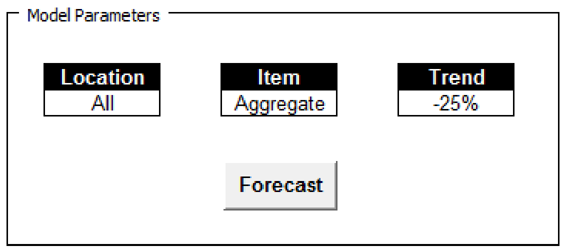

Figure 5.1 Model Parameters

A button labeled “Forecast” is used to run Excel Solver within the general maintenance model. Solver is used to recalculate optimal smoothing parameters whenever new input variables are selected. The “Forecast” button also unhides the general maintenance models, for referencing purposes within visual basic, and rehides once Excel Solver has completed. This means that the detail view must be selected after running the forecasts in order to view the general maintenance model.

The model view section (Figure 5.2) displayed on the user input screen controls which tabs are displayed within the workbook.

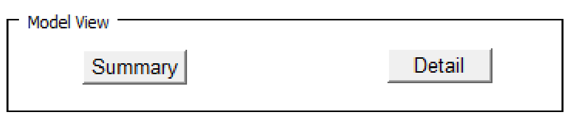

Figure 5.2 Model View

Two view settings are available: summary and detail. The summary view simply displays the input and output tabs for a user who does not wish to view the inner workings of the model. The detail view displays all tabs and is provided to allow for the updating of historical demand and investigation of model formulization. To run this model macros must be enabled and a correct Excel Solver path defined. For fixes to common errors please see Appendix B.

While in summary view five tabs are displayed; the previously described user input tab, colored white, and four others colored red (Figure 5.3). The red tabs are used to display the output of the model.

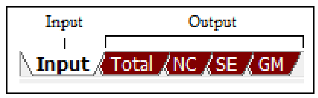

Figure 5.3 Tabs: Summary View

First amongst these is the tab labeled "Total" which shows the three disaggregate forecasts and their combined aggregate forecast. Also displayed are the actual forecasted values and their corresponding months. Similar information is shown on the other three output tabs, one for each disaggregate segment. When detail view is chosen 59 tabs are shown. These are organized into three color coated groups: inputs (white), modeling and output (red), and historical time-series data (blue). Further discussion of these groups is provided below.

5.1.1 Input Tabs

The input tabs are what drive the model by determining what time-series will be used as input and how the data will be disaggregated into the three usage segments of interest (new construction, storm and emergency, and general maintenance). A detailed description of the reasoning and equations used to separate the data can be found in Chapter 4; here it is only necessary to describe how Excel is used to accomplish the necessary effect. Five tabs are used for input purposes, beginning with the user specified model parameters and eventually calculating the input historical usage levels for each of the forecasting models. Detailed discussion of number manipulation within the workbook is contained in section 5.2.

5.1.2 Modeling and Output Tabs

The modeling tabs utilize the usage levels determined by the input tabs to compute the necessary forecasted values. A tab exists for each of the four forecasting models (aggregate, NC, SE, and GM) and four more to display the results of each of these models. It should also be mentioned that since the forecasts dynamically change according to the user inputs it was necessary to create tabs where aggregate level data could be continually stored. In the case of the general maintenance model, this data is employed when usage levels of the user specified time-series are excessively low to calculate seasonal parameters as recommended by Bunn (1999) [3]. Additionally, new construction aggregate data is used when zero values cause errors during the calculation of the historical trend for low usage time-series. Figure 5.4 displays and an example of a disaggregate level output.

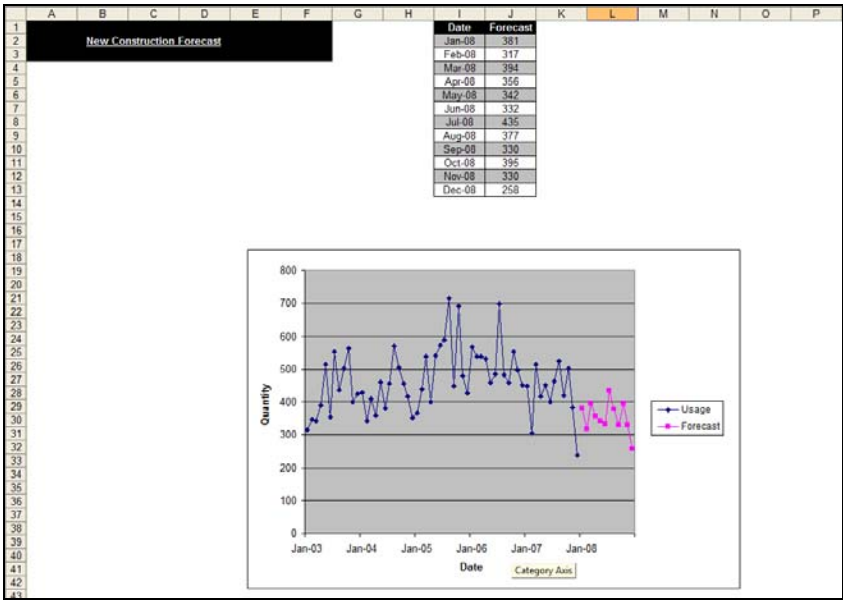

Figure 5.4 Disaggregate Output

5.1.3 Historical Time-Series Data Tabs

The majority of tabs are used to house the many different time-series which can be analyzed within this workbook. They provide the source of data from which all forecasts are made. All possible combinations of user specified location and item are considered. This data was mined from both Microsoft Access and Excel databases. For efficiency reasons all Excel data was exported into Access to more easily separate usages and insert zeros into months of no usage. Appropriate queries were run and the data eventually extracted and contained within the historical data tabs in this model. Macros were used within Access to help automate this process, but further interface programming is recommended to allow communication with the company's primary database. Please note that all historical data tabs must be formatted exactly the same for the model to work correctly. The necessary format is shown in Figure 5.5.

	A	B	C	D	E	F	G
1	**Date**	**Period**	**Inventory**	**DOJM**	**NC**	**SE**	**GM**
2	Jan-03	1	696	515	234	22	259
3	Feb-03	2	579	449	269	61	119
4	Mar-03	3	568	624	375	20	229
71	Oct-08	70					
72	Nov-08	71					
73	Dec-08	72					

Figure 5.5 Required Formatting for Historical Data

5.2 Detailed Execution

In case of future modification of the existing model it is necessary to describe more in detail how data flow throughout the workbook. The previous section describes in general how the model works and why it is organized the way it is. This section will explain how data is taken from the source tabs, organized within the input tabs, and a forecasts generated and displayed. It is recommended for any user who wishes to operate this model to read Chapter 4 so as to have a better understanding of the reasoning behind the data manipulation and model selections.

5.2.1 Data Selection and Extraction

Once the parameters are selected in the user input tab, the specified data is automatically pulled from the source worksheet and inputted into the "Inputs_Dummy" tab. This is done through use of Excel's indirect command referencing the information contained in the tab defined in "Input!M27". Aggregate Inventory and DOJM database data is taken from this sheet along with the new construction, storm and emergency, and general maintenance usage disaggregated from DOJM. This information is then used in the "Separate_Inv" tab to disaggregate inventory usage levels by NC, SE, and GM according to the weighted percentages shown in the DOJM database. Particular rules for separating the data are followed by the model, which have been outlined in Chapter 4. It is also necessary to devote input worksheets to aggregate data that is used when certain computations can not be made within the model due to low levels of usage.

5.2.2 Forecasting Models

For all disaggregate models the data is organized into columns for each necessary component of the model. In all cases this means observed values (X(t)), level (S(t)), and forecast (y(t)). The new construction and general maintenance models have a trending column (T(t)), and general maintenance uses additional seasonal indexes (I(t)). In all models it was necessary to reformat the forecasted values so as to make suitable graphs. This means that the data was placed into columns with dates depending on the source data and organized side by side with past observations. The use of #N/A allowed points on the graph to begin and end in the middle of the times-series, thus producing an accurate looking projection from historical usage.

As previously mentioned the new construction model calculates a level by averaging the observed usage for the month being forecasted during the previous two years. Next a trend is computed by considering the user input within the model parameters. The percent change specified is multiplied by the level and becomes the trend for that period. The model recognizes when the user selects a historical trend and utilizes a year over year percent change moving average during the trending calculation. A forecast is then made by adding the level and trend components and is limited to being greater than or equal to zero. The “NC_Model” and “NC_Aggregate_Model” work the same way, but, as the name suggests, the “NC_Aggregate_Model” always uses aggregate data. (This is also the case with the “GM_Aggregate_Model”.)

In the case of the storm and emergency model the calculated level (S(t)) becomes the forecasted value (y(t)). The level is calculated by organizing the five years of data into the month of usage and limiting each month’s highest usage to the second highest observed during that month. The level is then just the average of those five values. In Excel, first the each month’s highest value is found (Max), then the second highest value is found (M) and used to replace the usage of the highest (Max). Once Max is replace by M the five data point are then averaged. All of this manipulation is accomplished by implementing Excel’s IF command and the methodology can be followed once a simple understanding of this function is obtained.

To operate the general maintenance triple exponential smoothing model initial values for level, trend, and season have to be first calculated. The initialization process automatically implemented follows Kalekar (2004) [12] and NIST SEMATECH [15]. Once the model has been initialized, the model is driven by the triple exponential smoothing equations during the five years of historical data. It is then projected out to created forecasts for the sixth year. The smoothing parameters are fit using years three to year five so as to avoid using the first two years of initialization during the computation of the best fit line.

The specialized disaggregate model is completed by adding all the individual forecasts into one final aggregate level forecast. The forecasts are organized into columns within the “Combined_Model” tab for displaying purposes. These values are then added together and rounded to the nearest integer value, creating the final forecast.

Figure 5.6 shows an example of what is displayed within the "Total" tab for the aggregate forecast.

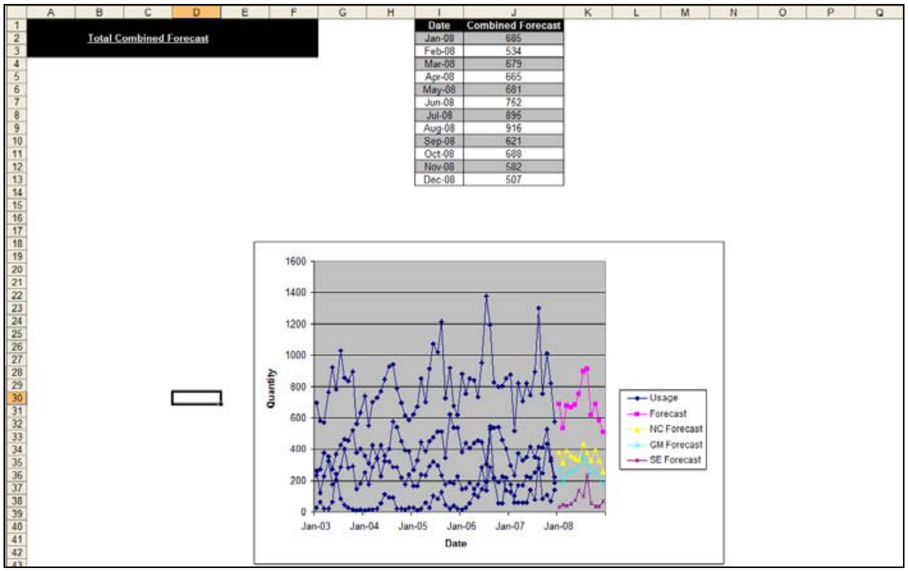

Figure 5.6 Aggregate Output

5.3 Requirements for Continued Use

This model was created to require modest amounts additional maintenance. By simply updating the historical data source tabs it will be able to create accurate and useful forecasts well into the future. Monthly updating is recommended; however, not necessary and additional years of historical data used during modeling would be beneficial, particularly to the storm and emergency forecasts. Updating the new construction trending component is also recommended whenever additional information pertaining to outside factors becomes available. Further in the future it may be necessary to reevaluate what locations and items are of most interest thus necessitating the addition of source tabs. At the very least, usage levels and trending components should be updated once a year to ensure reliable forecast in the future.

Chapter 6

Conclusions and Extensions

The goal of this work was to contribute to the existing body of research which has recognized the importance of defining statistically similar disaggregate segments and the necessity to further explore this approach in practice. Moreover, this was not only an exercise in the application of a new forecasting methodology, but also part of an effort to create a useable and sustainable forecasting tool for a large utility company. The approach used combined known forecasting techniques in a new general methodology and created a company specific forecasting tool.

This work created a forecasting model which was used to forecast monthly transformer usage for the coming year. The technique recognized the causes of demand within the aggregate and utilized this information to disaggregate and forecast demand according to the specific attributes within each class of time-series. The three classes of demand (new construction, storm and emergency, and general maintenance) all had unique properties and thus each warranted a different type of forecast model. New construction was a segment of significant interest because its trend could be predicted by shifts in outside factors that could not be realized through examination of historical data. This led to the creation of a new construction model which incorporated a user specified trend. This trend is vital to exploiting the advantages of this specialized disaggregate model because it allows the user to shift the trend of the forecast against that of historical record, consequently predicting an upward or downward trend which would not be predicted by traditional time-series forecasts.

During this investigation the importance of defining unique demand patterns within the aggregate was confirmed. Moreover, the benefits of a demand pattern specific disaggregate forecast was explored and a confirmation of its accuracy and usability when compared to more traditional forecasting methods found. The forecasting tool and methodology can be immediately implemented and its advantages realized, as future researchers continue to contribute to the work and improve upon its approach.

There are a number of extensions to this work which could significantly improve the accuracy of this methodology and usefulness of the forecasting tool. Further developing the selection of the new construction trend would be very interesting. Possibilities include furthering the data available during trend selection and varying this trend according to the level of usage within new construction. While the former would be a simple excise of finding additional indicators of transformer usage, the later has the potential to greatly improve this methodology by applying the concept of demand pattern specific forecasts to an element of the model which is fundamental to its performance. A methodology which automatically selects the trend which is most appropriate for the time-series being forecast could potentially be a very powerful addition to the specialized disaggregate model.

Another extension would be to offset forecasts according to the cost associated with either under or over estimating demand. If it is determined to be very costly to the company if a stock outage of a particular transformer occurs this could be considered by the model and possibly necessitate an increase in forecasted values so as to cause forecasting error to arise due to overestimation. The converse of this idea can be used if a certain item is extremely costly to house. If the balance between customer satisfaction and storage costs could be quantified in particular cases, forecasts could then be adjusted to appropriately evaluate and adapt to these costs.

Finally, for this particular case it would be beneficial to improve frontend and backend communication so as to fully automate the updating of past data. The advantages of a full ERP system would be valuable here and the need to more easily update usage levels is recognized. By improving this interface the model would become self-sustaining in all aspects except for the selection of the forward looking trend. The ability of the model to automatically generate new construction trends based on outside data would also be very valuable; however, it is recommended that these trends still be evaluated by the user to insure their appropriateness in varying and specific cases.

Appendix A

Historical Indicators of Future Construction and Transformer Usage

Table A.1 displays both St. Louis and national annual housing permits from 2002-2007 (National starts and permits are in thousands of units). The year over year percent changes are graphed in Figure A.1 showing that the NC Usage lags approximately one year behind the two indicators. The average of these indicators for the current year is used as the forward-looking trend within the NC model.

Year	NC Usage	StL Private Housing Permit	US Private Housing Permits
2002	-	13738	1,748
2003	5141	14276	1889
2004	5135	15945	2070
2005	6203	15038	2155
2006	6258	11933	1839
2007	5062	10372	1398.4

Table A.1 St. Louis and National Housing Permits

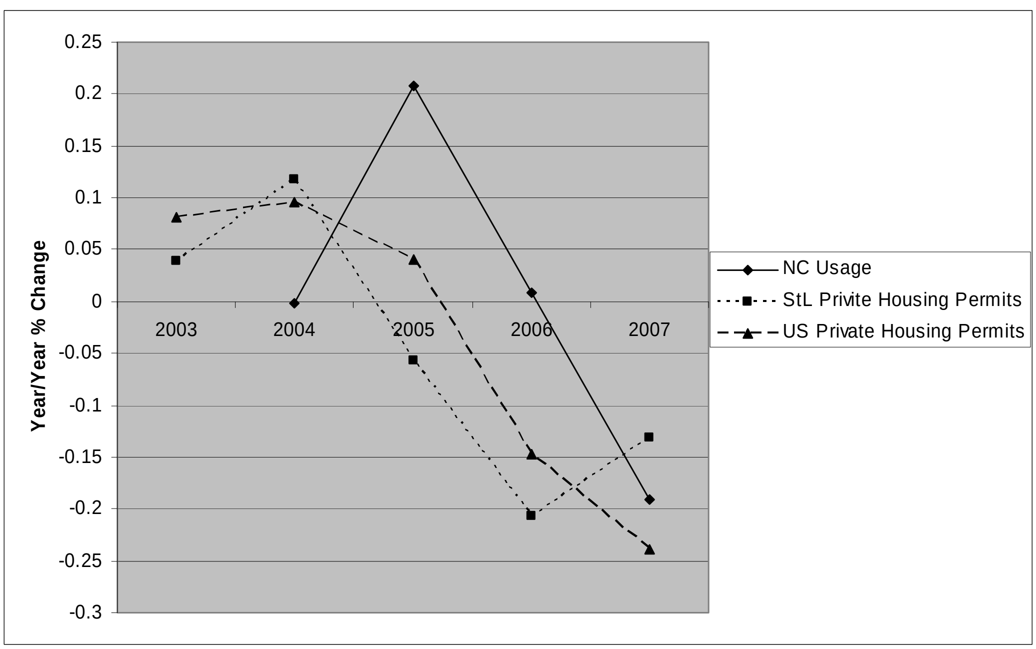

Figure A.1 Indicators 2003-2007 vs. NC Usage 2004-2007

Figure A.2 displays both indicators year over year change from 2003-2007 (the final point was averaged for the expected trend in NC usage for 2008 forecasts).

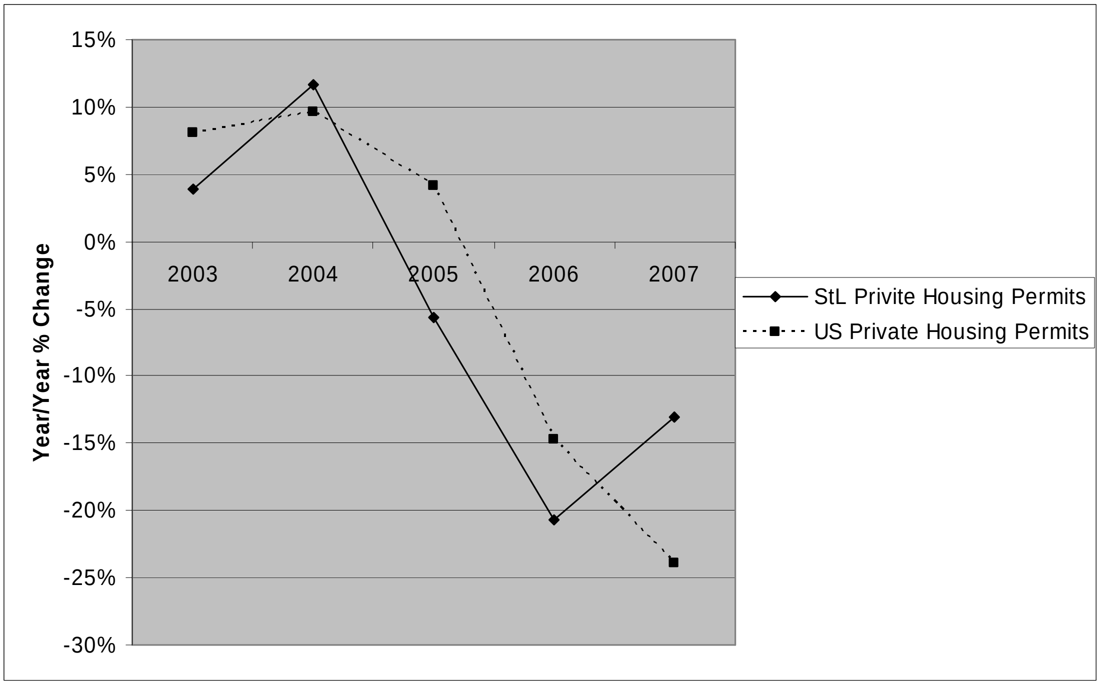

Figure A.2 Indicators 2003-2007: Year Over Year Percent Change

Appendix B

Referencing Solver.xla

Whenever an Excel file is opened which uses Excel Solver within a Macro, the Solver.xla location is read from that file even if it is incorrect. It is common for the location of Solver.xla to vary from computer to computer so each time the saved Excel file is opened on a new computer it should be expected that the path will need to be redefined.

To check if the location is correct (within Excel) go to Tools>Macro>Visual Basic Editor, this opens the visual basic editor. Inside the Visual Basic Editor go to Tools>References, do this will open a dialog box of references. Figure B.1 [1] shows an incorrect pointer to Solver.xla.

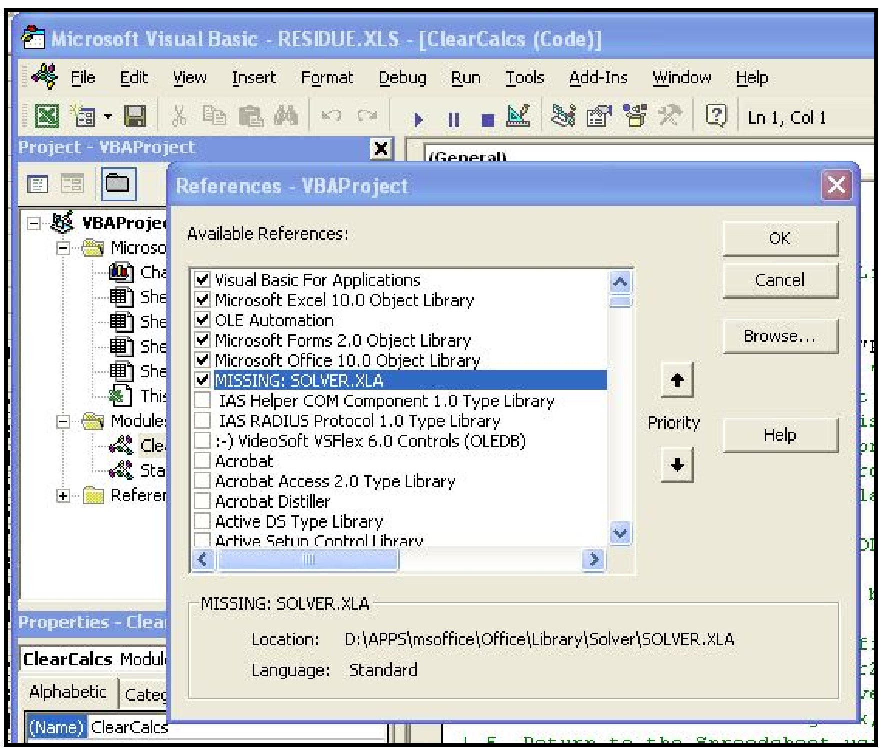

Figure B.1 Illustration of Incorrect Solver.xla Reference Pointer

The highlighted entry indicates that the location has not been found. Notice that the system tried to find the location within a subdirectory of "D:\APPS\...", this means that if a subsequent Excel file is opened that has a reference to the correct location that location will be loaded into memory. This can be used to easily fix the reference error. Return to the Excel file and go to Tools>Add-Ins and check Solver Add-in. (If the box is already checked uncheck it, click ok, and then return to the Add-Ins dialog box and recheck it. This reloads its location into memory.) Now return to the Visual Basic Editor and reopen the Reference dialog box and select Solver as shown in Figure B.2 [1].

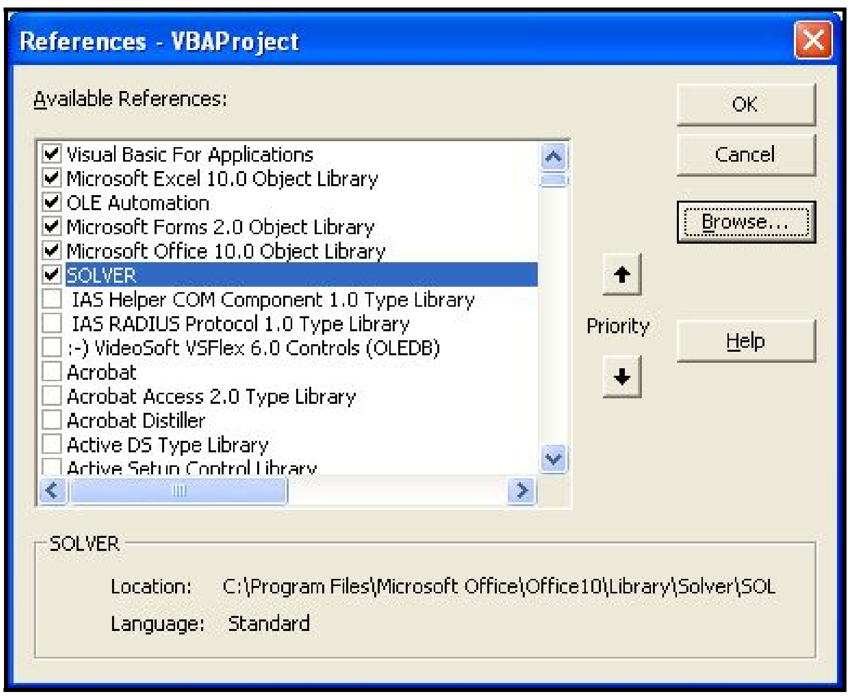

Figure B.2 Selection of Solver

This should complete the fix for providing the location for use of Solver by a Macro. Note that if within the Visual Basic Editor if References is grayed out attempt to click the stop button on the tool bar. Also be sure to save want to the new computer once the location is define or it will be necessary to repeat the above steps.

Appendix Sources

Appendix A

1) http://research.stlouisfed.org/fred2/series/STLBPPRIV/downloaddata?cid=324

2) http://www.census.gov/const/www/permitsindex.html

Appendix B

1) http://www.egr.msu.edu/~lira/supp/macro.htm

Bibliography

1) Archibald, Blyth C., and Anne B. Koehler. "Normalization of seasonal factors in Winters' methods." International Journal of Forecasting 19 (2003): 143-148.

2) Armstrong, J. Scott, and Fred Collopy. “Error Measures For Generalizing About Forecasting Methods: Empirical Comparison.” International Journal of Forecasting 8 (1992): 69-80.

3) Bunn, Derek W., and Angelos I. Vassilopoulos. "Comparison of seasonal estimation methods in multi-item short-term forecasting." International Journal of Forecasting 15 (1999): 431-443.

4) Chan, Chi Kin, Brian G. Kingsman, and H. Wong. "The value of combining forecasts in inventory management -- a case study in banking." European Journal of Operational Research 117 (1999): 199-210.

5) Chatfield, Chris, and Mohammad Yar. "Holt-Winters forecasting: some practical issues." The Statistician 37 (1988): 129-140.

6) Chatfield, Chris. "Forecasting in the 1990s." The Statistician 46 (1997): 461-473.

7) Chatfield, Chris. "The Holt-Winters Forecasting Procedure." Applied Statistics 27 (1978): 264-279.

8) Dangerfield, Byron J., and John S. Morris. "Top-down or bottom-up: Aggregate versus disaggregate extrapolations." International Journal of Forecasting 8 (1992): 233-241.

9) Dekker, Mark, Karel Van Donselaar, and Pim Ouwehand. "How to use aggregation and combined forecasting to improve seasonal demand forecasts." International Journal of Production economics 90 (2004): 151-167.

10) Gardner, Jr., Everette S. "Exponential smoothing: The state of the art---Part II." International Journal of Forecasting 22 (2006): 637-666.

11) Hyndman, Rob J., and Anne B. Koehler. "Another look at measures of forecast accuracy." International Journal of Forecasting 22 (2006): 679-688.

12) Kalekar, Prajakta S. Time series Forecasting using Holt-Winters Exponential Smoothing. 6 Dec. 2004. Kanwal Rekhi School of Information Technology. 3 Feb. 2008.

13) Miller, Don M., and Dan Williams. "Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy." International Journal of Forecasting 19 (2003): 669-684.

14) Ord, Keith. "Charles Holt's report on exponentially weighted moving averages: and introduction and appreciation." International Journal of Forecasting 20 (2004): 1-3.

15) Prins, Jack. Engineers Statistics Handbook. United States of America. Commerce Department. NIST--Commerce Department Technical Admistration. 1 June 2003. 3 Feb. 2008 <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm>.

16) Sanders, Nada R., and Karl B. Manrodt. "Forecasting Practices in US Corporations: Survey Results." Interfaces 24 (1994): 92-100.

17) Thomas, Robert J. "Method and Situational Factors in Sales Forecast Accuracy." Journal of Forecasting 12 (1993): 69-77.

18) Weatherford, Lawrence R., Sheryl E. Kimes, and Darren A. Scott. "Forecasting for Hotel REvenue Management." Cornell Hotel and Restaurant Administration Quarterly 42 (2001): 53-64.

19) Zotteri, Giulio, Matteo Klachschmidt, and Frederico Caniato. "The impact of aggregation level on forecasting performance." International Journal of Production Economics 93-94 (2005): 479-491.