1994

Adjusting Holt-Winters Exponential Smoothing for External Intervention: A Mathematical Technique for Making Quasi-Judgmental Adjustments For Anticipated Changes

Daniel Wayne Williams

Adjusting Holt-Winters Exponential Smoothing for External Intervention: A Mathematical Technique for Making Quasi-Judgmental Adjustments For Anticipated Changes

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Public Administration at Virginia Commonwealth University.

By

Daniel Wayne Williams
Bachelor of Arts with Honors, University of Virginia, 1975
Masters of Education, University of Virginia, 1979

Director: Ralph S. Hambrick, Jr., Professor,
Department of Public Administration

Virginia Commonwealth University
Richmond, Virginia

April 2, 1994

"Plurality is not to be posited without necessity."

- William of Ockham
circa 1280 - 1349

TABLE OF CONTENTS

Table of Contents
(Continued)

Table of Contents
(Continued)

LIST OF TABLES

LIST OF FIGURES

List of Figures
(Continued)

List of Figures
(Continued)

List of Figures
(Continued)

List of Figures
(Continued)

Abstract

ADJUSTING HOLT-WINTERS EXPONENTIAL SMOOTHING FOR EXTERNAL INTERVENTION: A MATHEMATICAL TECHNIQUE FOR MAKING QUASI-JUDGMENTAL ADJUSTMENTS FOR ANTICIPATED CHANGES

Daniel W. Williams, D.P.A.

Virginia Commonwealth Universtiy, 1994

Public administration data is sometimes extrapolated through exponential smoothing. Sometimes such data may undergo a level shift because of a policy decision. The slope of the curve formed by connecting the periodic observations increases or decreases significantly for a brief period, thereafter returning to a slope similar to the slope preceding the policy change. This discontinuity might be called a ramp or a step. Forecasts made with exponential smoothing immediately before, during, or immediately after the ramp or step may be considerably inaccurate unless adjusted. A technique called adjusted exponential smoothing is proposed to reduce or eliminate the inaccuracy of forecasts made under such circumstances when the ramp or step arises from a planned policy decision. An empirical study is conducted to determine whether the proposed technique constitutes an improvement over other exponential smoothing techniques. The empirical study shows that the proposed technique improves the accuracy of forecasts when planned level shifts subsequently actually occur. Guidelines are provided for using the technique.

CHAPTER 1: INTRODUCTION

In this chapter I will:

- Provide a general introduction to this dissertation.
- Provide an overview of the following chapters.

General Introduction

Forecasting is an integral part of planning and budgeting for many public administration activities. It is used for revenue projections, planning for prisons, budgeting for Medicaid expenditures, and numerous other public administration planning and budgeting activities.

Forecasting techniques can be roughly classified as simple, intermediate, and sophisticated.

- **Simple** techniques can be as simple as assuming that the last observation will also be the next, which is sometimes called the random walk or naive method. Other simple techniques follow the random walk approach after preprocessing data by such methods as deseasonalizing it or adjusting for such factors as inflation, sometimes this sort of preprocessing is called "decomposition." Sometimes a simple trend is added to the data; the trend may be the difference between the last two observations or the ratio of the last observation over the prior observation.

- **Intermediate** techniques include moving averages, exponential smoothing, and the use of more complex decomposition which may include building a model of the process that is to be forecasted.

- **Sophisticated** techniques generally involve use of single or multiple correlation techniques or the use of complex forecasting algorithms such as ARIMA.

These labels do not necessarily capture the richness of the variety of techniques, for example some moving average techniques, like X11 (defined in Appendix I), may be very complex. Empirical studies in forecasting do not demonstrate that sophisticated techniques produce better results than intermediate techniques. Exponential smoothing, in particular, is often cited to be as effective as more sophisticated techniques.

In this dissertation I examine a problem that arises when forecasting with exponential smoothing. Data series sometimes undergo shifts (see Figure 1) that make them difficult to forecast. In data forecasted for public

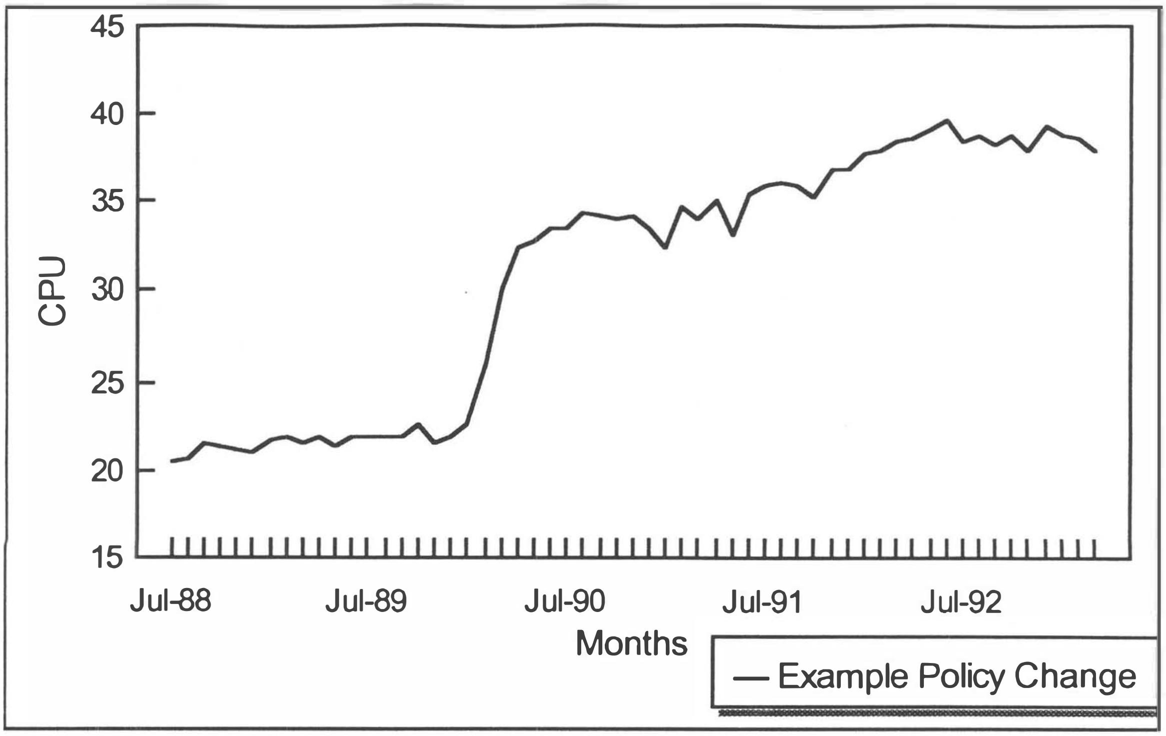

Figure 1

administration purposes, these shifts may arise because of policy changes, so I call them policy changes or policy shifts in this dissertation.

Data forecast for public administration frequently contain policy changes. Decisions made by legislatures often transform into changes in data series. For example, the data series shown in the example is actual average reimbursement per service unit for certain Medicaid expenditures, the level shift reflects a legislative action. The existence of level shifting data in revenue forecasting is documented by Wilpen Gorr.[1]

With techniques that use correlation based mathematics, such as regression, *historical* level shifts generally can be incorporated in the forecast model either by including independent variables that undergo similar level shifts, or by use of dummy variables. This approach is not available to intermediate techniques such as exponential smoothing. Techniques that are available for incorporating historical level shifts are discussed in the literature review.

Additionally, there is no available technique for integrating expectations of *future* policy changes into exponential smoothing models. Such future policy changes, especially those that have little historical precedence, also pose a problem for more sophisticated techniques. Somehow the level shift must be put into the model in order to get it out of the model.

In this dissertation I propose a technique for integrating independently developed estimates of policy changes into exponential smoothing models to forecast through future periods that have policy shifts. This method consists of modifications of an exponential smoothing model to incorporate the first differences of a policy change estimate within the exponential smoothing model.

Two major hypotheses are examined:

1. The proposed technique provides forecasts that are more accurate than are available from other exponential smoothing techniques for the period beginning with the onset of the actual level shifting data.

2. The proposed technique provides forecasts that are more accurate and better fit than are available from other exponential smoothing models for periods of time following the period of a level shift.

These hypotheses are made more precise in Chapter 6.

The first hypothesis is examined through 12 different simulations of policy shifts over 20 data series. The proposed technique, four other exponential smoothing techniques, and the random walk approach are used to project each data series through a period where a level shift is anticipated. Hold out data is adjusted for simulated policy changes. The forecasts are updated through six periods under twelve scenarios of simulated actual level shifts. Forecasts are compared with simulated actuals and errors are calculated for up to fifteen periods subsequent to the six update periods. The errors arising from various techniques are compared across an array of statistics in the manner of forecast competitions.

Descriptive statistics include three measures of Mean Squared Error, three measures of Absolute Percent Error, a measure of range of percent error and a measurement of rank of absolute error. Each of these statistics is computed for each series and then summarized for all twenty series. Four summarizations are provided, the average of the twenty series, the geometric mean of the twenty series, the average ranks of the 11 scenarios for each of the twenty series, and the summed Kruskal-Wallis ranks of the twenty series. Each of these results is ranked among the 11 scenarios. The average ranks and the Kruskal-Wallis ranks are tested for statistical significance through two non-parametric tests, the Kruskal-Wallis analysis an the Analysis of Variance by Rank.

The second hypothesis is examined in a separate forecast comparison. Twenty data series that are known to have undergone previous level shifts are forecast using the proposed technique in model fitting stage, and using four other techniques. Magnitude of historical level shifts is empirically estimated from the data. No simulated policy changes are added to these data. Six updates are completed.

Hold out data is used to examine the accuracy of the forecast as with the first hypothesis. Analysis of results is as with the first hypothesis.

Overview of Chapters

In the **second** chapter I provide background related to forecasting in public administration and background discussion of forecasting in general and exponential smoothing in particular. I describe Holt-Winters exponential smoothing and discuss the relative worth of exponential smoothing as a forecasting technique. In the **third** chapter I provide a more complete discussion of the problem proposed in this introductory chapter, specify the research question for this dissertation, and discuss the need for this study. In the **fourth** chapter I examine the literature to determine what techniques might already exist for addressing the proposed problem through exponential smoothing. In the **fifth** chapter I propose an exponential smoothing solution to the problem. In the **sixth** chapter I examine the literature to determine various models of forecasting research and examine appropriate models for empirical evaluation of proposed forecasting techniques. I provide more precise formulations of the hypotheses. In the **seventh** chapter I define two research projects that are used to examine these hypotheses. In the **eighth** chapter I present the results of the two research projects. In the **ninth** chapter I discuss results and draw conclusions. **Terms** are defined in **Appendix I**. **Appendix II** includes formulas for forecast techniques discussed. **Appendix III** includes certain correlation matrices related to the data series used in the research projects. **Appendix IV** contains tables produced in the data analysis. **Appendix V** contains information concerning the fit of the model in the second research project. **Appendix VI** contains estimated level shifts for the second research project. **Appendix VII** contains formulas for the statistics demonstrated in Appendix IV and certain other error statistics.

Summary

Forecasting is important for planning and budgeting, two integral elements of public administration. Exponential smoothing is a valuable forecasting technique. Level shifting data is difficult to forecast whether through exponential smoothing or other techniques. A method is proposed for forecasting level shifting data. Two major hypotheses concerning this technique are examined through two research projects.

CHAPTER 2: BACKGROUND

In this chapter I:

- Identify some uses of forecasting in public administration.
- Identify some occasions where a level shifting problem arises with data forecasted for public administration. This problem is discussed in more detail in chapter 3.
- Discuss some general ideas regarding the display of forecast data in graphs.
- Discuss forecast techniques in general.
- Discuss the use of a form of exponential smoothing known as Simple Exponential Smoothing (SES) and two variants known as Holt exponential smoothing and Holt-Winters exponential smoothing.
- Briefly explain why exponential smoothing is a useful form of forecasting.

Forecasting in Public Administration

Some uses of forecasting in public administration are:

- Revenue is forecast for budget planning and other uses.[2]

- Entitlement programs forecast enrollment and service usage for budgeting and planning.[a]
- Prison populations are forecast for budgeting and planning.[3]
- Expenditures are forecast for budget planning.[4]
- Special health care populations are forecast for planning and other purposes.[5]

Many of these forecasts are made for budgetary and planning purposes, and are intimately associated with the analytic roles of such public administration professionals as budget or policy analysts. Forecasting is an important analytic tool for these public management and public management support roles. Many techniques are used in the practice of public administration. Publications and entities that focus on public administration have sponsored articles, chapters, or books concerning forecasting.[6]

Exponential smoothing, which is the specific forecasting method studied in this dissertation, has been suggested as a technique that would benefit local governments in budget forecasting.[7] This suggestion rests in part on the simplicity of use. Further, 10% of municipal

---

[a]Known from my personal experience as the budget director for the Department of Medical Assistance Services.

governments us exponential smoothing for revenue forecasting and 7% use it for expenditure forecasting.[8] Analysis of exponential smoothing forecasting models has been accepted as a dissertation topic in public administration.[9]

Governmental forecasting inaccuracy receives heavy scrutiny even when estimates are extremely close.[10] Both revenue and expenditure forecasting receive close attention in evaluating the fiscal status of the states.[11] In recent years, evaluation of expenditure forecasting has specifically focussed on Medicaid programs.[12]

Medicaid Forecasting

In Virginia, Medicaid general fund expenditures account for 13% of all general fund expenditures in the Commonwealth of Virginia for fiscal year 1994[13] and is cited as among the fastest growing components of the Commonwealth's budget.[14] Other states are experiencing similar growth in their Medicaid programs and their Medicaid programs are of similar magnitude within their state budgets.[15] Medicaid expenditures have been projected to climb to 25% of state spending by 1995 (using a different measurement scale which includes federal funds and has these expenditures at about 20% in 1994).[16]

The Medicaid program is also a rapidly growing large component of the federal budget, which currently consumes about 6% of the total federal budget.[17] Federal and state spending on Medicaid overtook other public spending for the poor in the early 1980s[18] and is now the largest federal grant program in state government, amounting to 35% of all federal grants to states in 1991.[19] In 1991, federal spending on Medicaid totalled $52.5 billion.[20] Medicaid is overtaking the currently larger Medicare program which accounts for 10% of the federal budget.[21]

The Medicaid program is generally considered to be uncontrollable. One of the components of the perception this Medicaid is the fact that many states have had difficulty forecasting their programs during a period of significant policy change during the late 1980s.[22] Federal findings show state forecasting errors averaging 18% across the country with Alabama underestimating its federal grant by almost 90% for fiscal year 1991 (not all individual states are reported);[23] however, the report leaves it unclear as to how much error arises from actual forecast error.

The federal government attributes a substantial portion of the budget errors to "Substantial increases in inpatient hospital care[,]...Increase numbers of beneficiaries, some of whom now receive benefits as a result of post-1985 Congressional expansions of eligibility for Medicaid[, and]... a generally unpredicted upturn in acute health care costs."[24] This is a not very clear attribution of a substantial portion of these forecast errors to policy changes, possibly as much as 41% (reporting categories are not adequately clear to show a definite share). A recent Health Care Financing Administration publication shows 39 "major" Medicaid expansions between 1986 and 1990.[25] The states have been complaining about federally driven Medicaid expansions since at least 1988.[26] While exact attributions of magnitude are not possible with this data, it is apparent that it is difficult to forecast **through** periods of policy change.

In this dissertation a technique is developed for forecasting through periods of policy change when externally developed estimates of the policy change are available. I have developed this technique with the forecasting problems of the Medicaid program in mind. It is not unreasonable to expect that such estimates are often available.[27] The policy change problem is discussed in more depth in the next chapter.

Policy Changes Occurring Throughout Governmental Data

As discussed in the next chapter, the same problem arises in the corrections environment. Corrections is another significant area of state budgeting that is thought to be out of control.[28] In the corrections environment this problem may be exacerbated by the fact that policy changes may have extremely long lead times from policy decision to impact on the data series.[a]

While the significance of the policy change problem is not well documented in public administration literature, it is not completely ignored. Wilpen Gorr has written on the significance of tracking public policy factors in governmental MIS systems for the purpose of using them to explain forecast.[29] A portion of his argument is that when public planners forecast data series that have level shifts, user confidence in the forecast depends, in part, on there being adequate understanding of the reasons for those shifts. He cites techniques such as those of Lewandowski and of Makridakis and Carbone (reviewed in a Chapter 4 below) as potential methods for forecasting with data that has such shifts.

---

[a]In a presentation to a forecasting technical panel in July 1993, a Virginia Department of Corrections analyst presented data that showed lead times of five or more years.

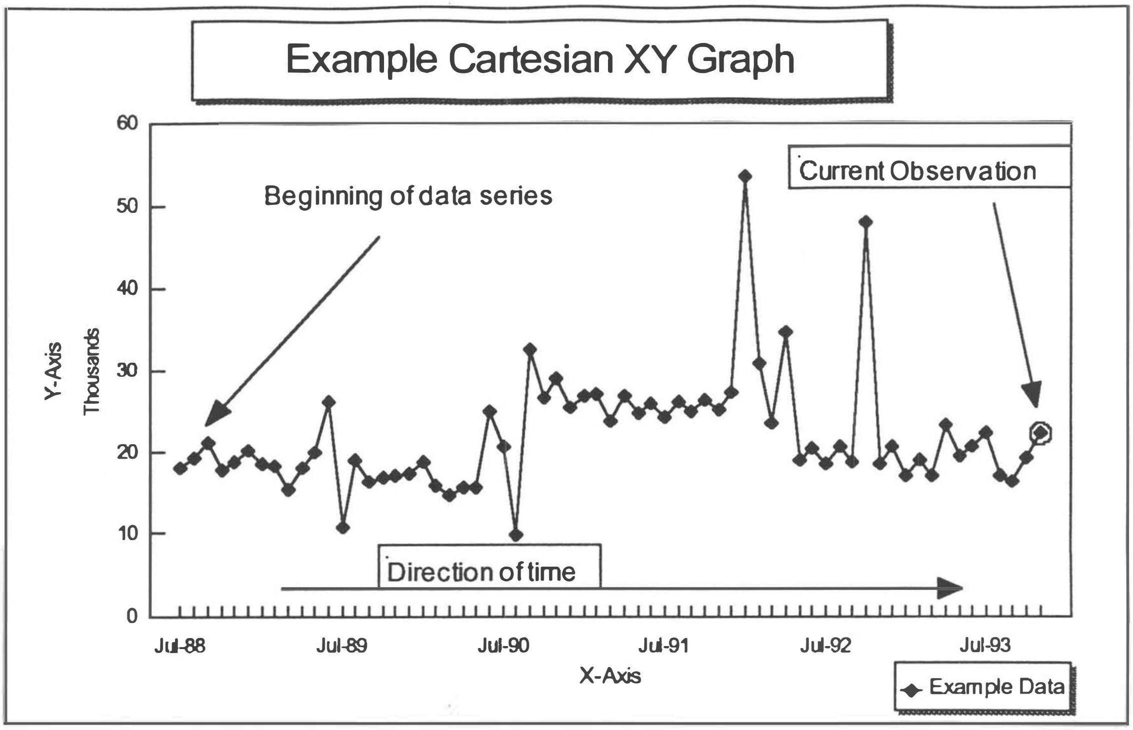

Figure 2
Forecasting Background

Cartesian XY Graph

Forecast data is frequently displayed on a Cartesian XY graph calibrating the X axis in roughly equal time units, and the Y axis in the observation measurements of the data. The data is usually displayed with the first available observation at the Y axis and each subsequent observation displayed one time unit to the right. Each observation is located to the right of its Y measurement and above its X axis time mark. By connecting these observations one obtains a curve that follows this data across time. Figure 2 shows a Cartesian XY graph. Throughout this dissertation XY graphs are used in this manner.

Forecasting Techniques

There are number of forecasting techniques ranging from intuitive or judgmental to highly complex mathematical approaches. In general these techniques rest on an estimate of central tendency, such as an average, or its variate with possible consideration of some predictable variation from central tendency, such as seasonality.

Simple techniques generally treat a recent observation (sometimes called the **naive method**[30] or **random walk**[31]), the average, or a trend added back to one of these as a the forecast. Simple trends include both differences between recent observations (additive trend) and ratios between recent observations (multiplicative trend). (Formulas that demonstrate these trends and other forecasting methods discussed in this dissertation can be found in Appendix II.)

Simple techniques can also involve **data cleaning** such as adjusting data to remove the effects of general inflation by converting nominal dollars to constant dollars.[32] An even simpler technique is the use of a known fixed number as a forecast. Where the forecaster has good reason to believe it is correct, no other forecast can compare with a fixed number, for example, the number of days in any future week in the relevant future can be reliably forecast to be 7.

Intermediate levels of forecast techniques generally include moving averages (defined in Appendix I) and exponential smoothing.[33] There are several forms of exponential smoothing, one that is frequently mentioned in forecasting literature is called Holt or Holt-Winters Exponential Smoothing. This technique forms the basis for most of the discussion in this dissertation.

Two types of techniques that are more **sophisticated** than exponential smoothing are ARIMA (Auto-Regressive Integrated Moving Average) techniques, and correlation based techniques.[34] ARIMA is a complex system of equations and evaluation techniques that are similar to exponential smoothing techniques, in fact, Holt exponential smoothing can be shown to be a special case of ARIMA.[35a] Correlation techniques generally involve determination of a causal relationship between independent and dependent variables through regression models or systems of regression models.[36]

These techniques account for a large share of forecast techniques used in actual practice,[37] and the more sophisticated ones (including exponential smoothing) are commonly included in forecasting texts.[38]

---

[a]Multiplicative Holt-Winters is not reducible to an ARIMA model.

Exponential Smoothing

Exponential smoothing is a method of projecting serial data into the future through a statistical evaluation of forecast errors arising in the exponential smoothing model. Single exponential smoothing (SES) is a moving average that places more weight on recent observations. It is sometimes called an **exponentially weighted moving average**. The weights diminish exponentially as new observations are added to the model, which fact gives this method the **exponential** part of the name exponential smoothing.[39]

The idea of SES is that more recent observations are a better predictor of the future **level** of a series than are older observations. **Level** refers to the central tendency of a data series. Since every individual observation contains some random noise, the latest observation by itself is not the best predictor of the future series. By averaging in older observations, this noise is **smoothed** away. However, by weighting the observations towards the current period, the model still reflects the current period information.

Exponential smoothing provides a summary of this data through a curve that follows the data through the same time period, but has less overall variation. The difference between the model curve and the actual data is the known as residual noise, random variation or **error**. When the observations are exhausted at the current period, SES projects future values as the level calculated from the error calculated with the last actual observation. Figure 3 depicts a forecast made with SES.

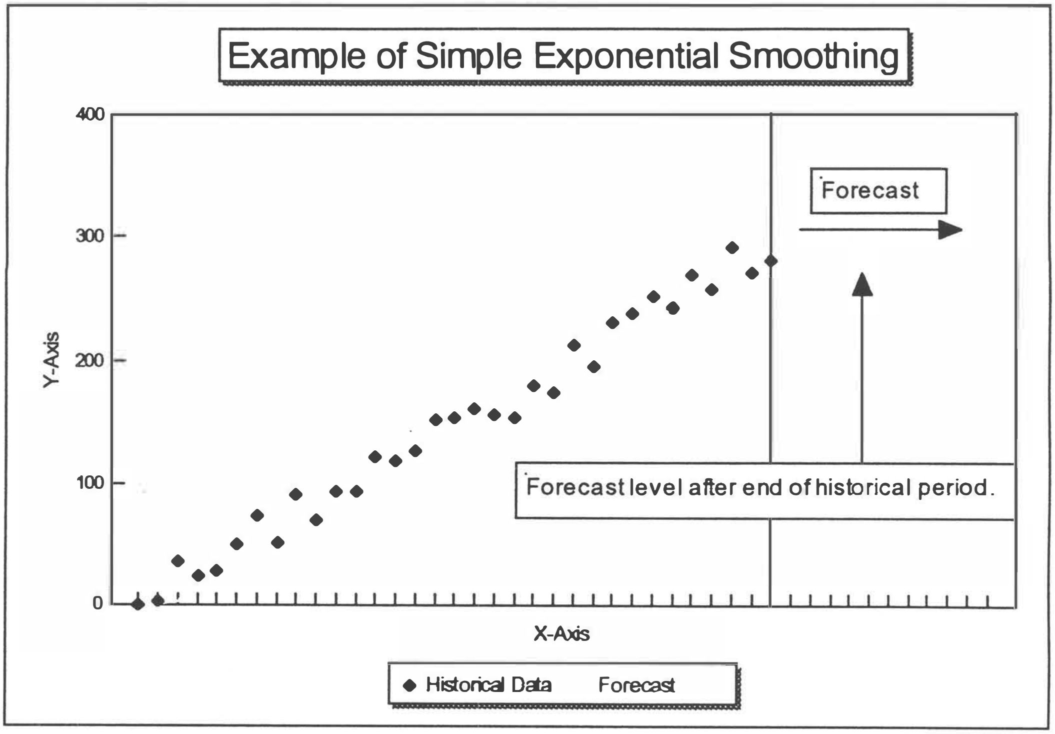

Figure 3

SES is produced by the following formulas:[40a]

$$e_t = F_t - X_t$$

$$F_{t+1} = F_t + \alpha e_t$$

Where,

e_T = Error of observation t,

X_t = Observation t,

F_t = Forecast of observation t,

α is a weight parameter subject to **0** α **1**, and

t is an index of time.

To fit a model, the parameter α is fit to the curve using an optimizing technique. The optimizing technique generally involves reducing Root Mean Squared Error, Mean Absolute Percent Error or some other loss function to a minimum. A loss function is a statistic that represents the cost of error in the forecast.[41] Different types of loss functions treat error differently. For example, loss functions that square errors place greater emphasis on reducing the largest errors, while absolute error loss functions do not.

---

[a]Throughout this dissertation notation is adjusted from that arising in source documents to increase the consistency.

The first forecast (F_0) in the series can be zero, or it can be initialized at some value that the forecaster believes to be close to the initial level of the series. F_0 is the model value for the first observation in the series, that is, at the far left side of the graph; it is **not** the first forecast after the end of the historical data. In the graph the forecast is shown as a line with F_0 beginning at the same point as the first actual observation, X_0.

When α is set at zero, then F does not change from period to period, so it remains the initial value, F_0. When α is set at 1, F for periods beyond the current period are identical to the current period X.

It is apparent in Figure 3 that the average of the actual observations, 147, would be a considerably worse predictor of the future value of the series than the SES prediction of 279. However, it is also apparent that the series is trending towards even higher numbers. SES is not able to capture such a trend. A technique that can capture a trend is Holt Exponential Smoothing.

Holt Exponential Smoothing[a]

Holt exponential smoothing decomposes data into level and **trend**. **Trend** refers to the slope of a data series and is the difference between two successive observations.[42] When the observations are exhausted at the current period, the Holt model projects future values by repeatedly adding back trend to the level to project the next level. It is explicitly defined through formulas[43] which can be found in Appendix II.

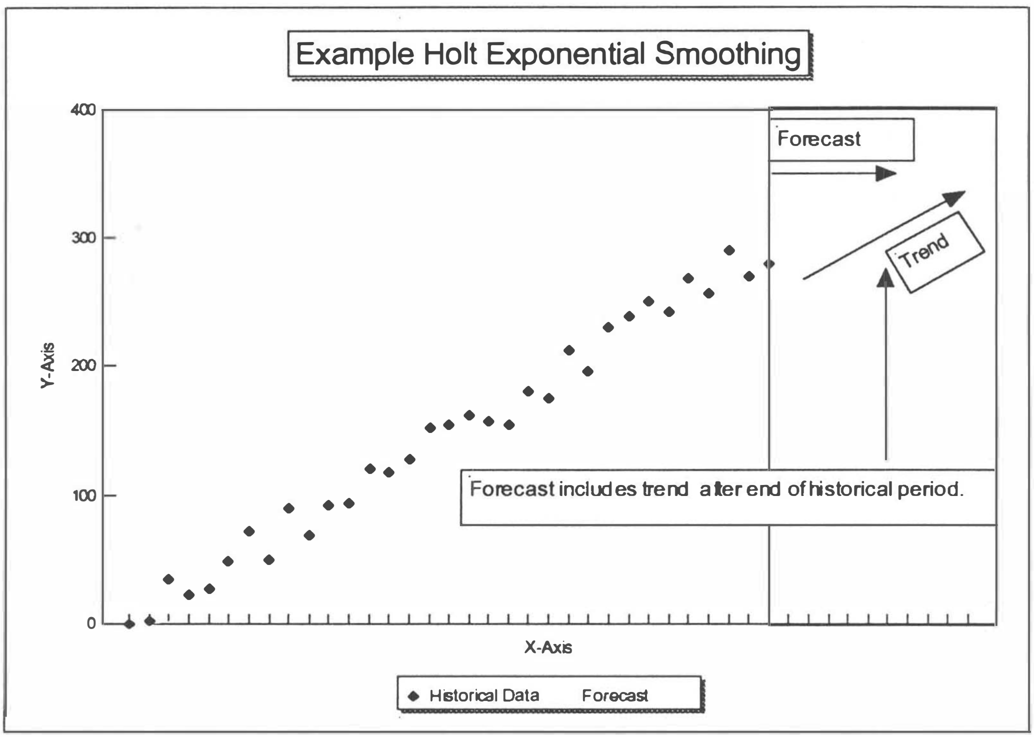

Figure 4

---

[a]Discussion of Holt and Holt-Winters exponential smoothing closely follows Makridakis, Wheelright, *and McGee*, see endnote 12.

Holt exponential smoothing adds a β parameter for trend. Both α and β are optimized subject to the restriction Φ = {α,β), 0 Φ 1. This technique begins with an exponentially weighted moving average, but also adds an observed trend to the extrapolation. When the trend elements of the Holt model are set to neutral values, Holt exponential smoothing is equivalent to SES.

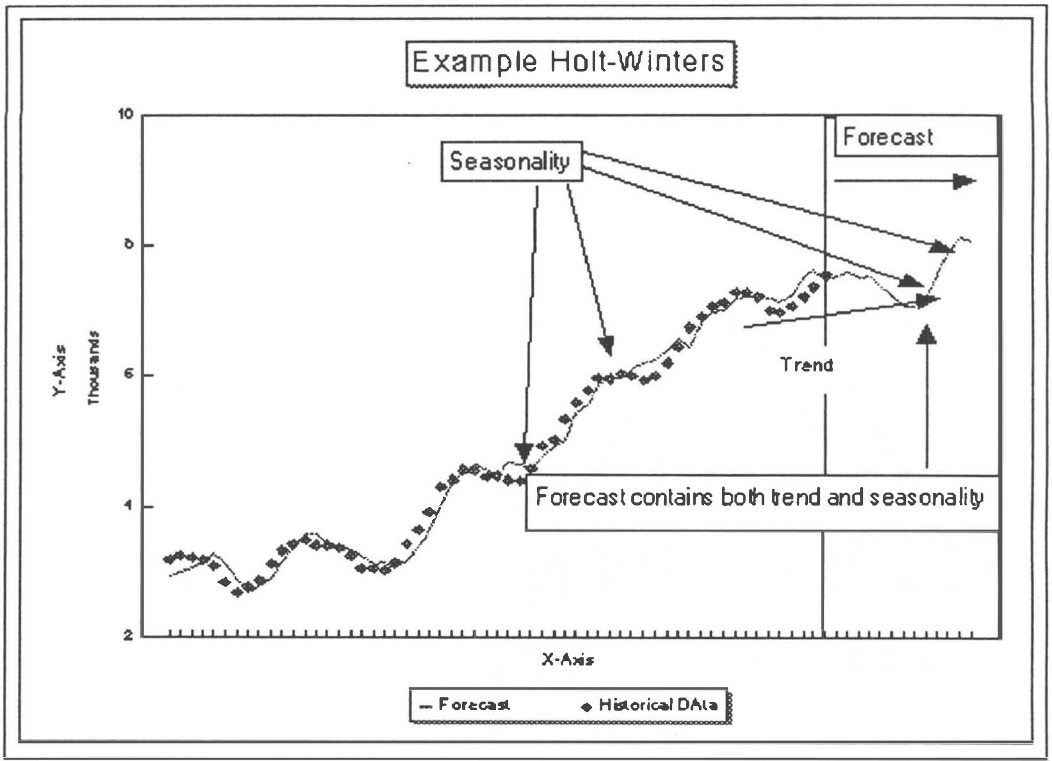

Figure 5

Holt-Winters Seasonal Exponential Smoothing

While Holt allows consideration of trend, it does not help with seasonally fluctuating data. The following graph demonstrates Holt model that also includes a Winters

multiplicative seasonal component.[44] Formulas are in Appendix II.[45]

On the XY graph, Holt-Winters makes the curve less smooth by including expected seasonal variation through a third parameter, γ. The parameters, $\Phi = \{\alpha, \beta, \gamma\}$, are optimized subject to the restriction, $0 \quad \Phi \quad 1$. Where data is significantly seasonal, normal seasonal variation is treated as expected rather than as error, so it does not result in misleading forecast correction with each update. When the seasonality component of Holt-Winters is set to neutral values Holt-Winters is equivalent to Holt.

Predictable Variation

Seasonality is a form of predictable variation. Other techniques are also available for reducing predictable variation, e.g., data can be divided by the number of trading days,[46] before it is entered into the **statistical forecast model**[a] and then readjusted after forecasting through the model to produce a full forecast. Such

---

[a]In this study I use "forecast model" to refer to statistical forecast models which are equations or systems of equations that include parameterized evaluation of forecast error. It is also possible to refer to any set of equations that result in a forecast as a "forecast model." Although either usage may be correct, I use the term to refer to statistical models.

adjusting and explaining involves attempts to reduce the amount of variation that is left as residual noise. The use of trading days is part of the broader approach of data decomposition, where forecasters try to break down a series into simpler component series before forecasting,[47] for example, one may break down a forecast of expenditures for health care services into a forecast of units (services delivered) and a forecast of expenditure per unit, this would be a **multiplicative decomposition**; alternatively, one might break down a forecast of service units into service units delivered to adults and service units delivered to children, this would be an **additive decomposition**.

By decomposing data into simpler series, forecasters have a better opportunity to determine the intuitive reasonableness of forecast projections. This benefit arises because homogeneous data series are more likely to have only a few primary factors generating their trend; thus, incongruous information is more obvious (e.g., a forecast of gross expenditures may be allowed to grow because of "trend" when a forecast of units of service would not be because the forecaster knows that a regulated service capacity limit has recently be exhausted). Thus, decomposition is closely related to causal analysis.

The SES formula presented here and related formulas included in Appendix II are the **error formulation** of these models.[48] This formulation demonstrates the relationship between the determined error (variation between forecast and actual observation for any period t), and the projection for the next period. Specifically, the forecast for the level or trend components for period t + 1 is the forecast for that component for period t plus a **proportion of the error** in that forecast.[49] In the case of seasonality the interpretation is slightly more complicated but essentially the same.

This proportion-of-error provides a common sense interpretation of the parameter restriction, $\Phi = \{\alpha, \beta, \gamma\}$, $0 \quad \Phi \quad 1$, as proportions are naturally limited between none = 0 and all = 1. On this interpretation, error that is highly likely to indicate change in a component should be weighted highly, while error that is likely to be random noise should be given limited weight. Thus, a parameter approaching 1 indicates that, for the specific component, forecast error can generally be taken to mean that there is a change in the data, while a parameter approaching 0 indicates that error is best interpreted as random noise and the forecast performance is not significantly improved by heavily considering the magnitude of the most recent error.

Exponential Smoothing And Sophisticated Techniques

While some of the discussion in later sections and chapters addresses the presences of a discontinuity problem that is the subject of this dissertation when forecasting with sophisticated techniques such as ARIMA, Kalman filters (defined below on page 69), or regression, **the objective of this study is to examine the problem when forecasting with exponential smoothing**. In this dissertation the discussion of both problem and solution focuses on exponential smoothing. This should not be taken to imply that the problem does not exist with other forecasting techniques, nor that it cannot be addressed through those other techniques.

A Valuable Alternative to Sophisticated Techniques

Regardless of potential benefits of sophisticated techniques, exponential smoothing is a valuable forecasting technique. It has some of the advantages of the less sophisticated techniques and some of the advantages of the more sophisticated techniques.

Advantages as a Sophisticated Technique

Like the more sophisticated techniques exponential smoothing is reasonably reliable.[50] Sometimes regression based techniques and ARIMA are assumed to be most reliable forecasting techniques because they are the most sophisticated. However, forecast literature tends to support the view that simpler techniques, **particularly exponential smoothing**, are more reliable.[51] Armstrong has argued that the persistent belief that the most sophisticated techniques are the most reliable does not reflect actual empirical evaluations of such techniques.[52] Some of the specific considerations about sophisticated techniques involve forecast fitting, turning points, and sophistication itself.

Forecast Fitting

It is sometimes thought that techniques that are more effective in fitting data during the sample period also do a better job in forecasting. This has not proved true.[53] Sample period fit is considered an unreliable indicator of forecast accuracy. Consequently, more sophisticated techniques are not *de facto* better than exponential smoothing solely because of any increased effectiveness in sample period fitting.

Turning Points

Turning points are points in a series where a trend shifts. Simple extrapolative techniques like exponential smoothing can discover turning points only after they have occurred. Sometimes sophisticated techniques are thought to be effective in predicting the turning points of economic cycles; however, such techniques are not proved to be effective.[54] The point here is not that simple techniques can predict turning points, but only that more sophisticated ones are not particularly better at it.

Sophistication

A particular problem that arises with sophisticated techniques is that sophistication can lead to error.[55] This problem arises because sophisticated techniques involve a higher risk of confusing forecast noise (unexplained variation) with information. As parameters or other strategies for extracting information from varying data increase, so too does the risk of finding a pattern that appears to be meaningful when it is not.[56] Simpler techniques do not risk as much error of this sort because they do not attempt to explain as much variation.

Advantages as a Simple Technique

In the previous sections I reviewed reasons why simple techniques such as exponential smoothing are often just as accurate as more sophisticated forecasting techniques. In that respect exponential smoothing has the same forecasting benefit as sophisticated techniques. Even when exponential smoothing may not be as accurate than these more sophisticated techniques, it may be better because of its advantages as a simple technique. These advantages include lower cost and lack of dependence on exogenous data.

Lower Cost

A modest gain in forecast accuracy attained by using more sophisticated techniques may not justify the cost in analyst time and skill.[57] Exponential smoothing is fairly easy to learn and to apply.[58] It is, therefore, useful when the forecasting work force is not itself skilled in more complex statistics. Also, it can be applied to a large number of data series with a relatively small amount of work. This advantage is in direct contrast to the need for sophisticated skills and considerable analyst time for applying correlation based techniques and ARIMA techniques.

This advantage should not be thought to imply that exponential smoothing is associated with less capable analysts. Instead, the use of skills associated with fitting sophisticated statistical models may consume time and effort that may be better spent investigating the data generating functions that produce the data series being forecast.[59] In an actual work environment as may arise in public administration, the analyst must allocate an appropriate level of time and effort to various tasks. Exponential smoothing may allow the allocation of less time to model fitting which, in turn, allows the allocation of more time to other tasks.

Endogenous Data

Exponential smoothing does not require the availability of data series and forecasts of data series that can be used as exogenous (independent) variables. Correlation based forecast techniques depend on the availability of forecasts of independent variables which, in the end, must be generated either from macro-economic models, judgement, or extrapolation techniques;[60] or which **may not be available at all.**[61] As Vollmann, Berry, and Whybark put it, "In the first place, in certain instances, we simply have no past data [with which to develop correlation analysis]."[62] These techniques are not appropriate where forecasts of causal data is not available, "The [econometric] approach, even if fundamental for policy analysis, is often inappropriate for very short-term predictions, first of all owing to the lack of the relevant data on the exogenous variables."[63]

The mere fact that independent variables can be correlated with the data that one wants to forecast in a regression model is not sufficient for use of the independent variables in forecasting, since the forecast of the dependent variable can extend only so far into the future as the availability of the independent variable unless the model also provides for a forecast of the independent variable. Extrapolation based techniques such as exponential smoothing do not have this difficulty.

Even where forecasts of independent variables are available, they may have too much variance to be useful for forecasting. Richard Ashley has demonstrated that correlation based forecasts that depend on forecasted independent variables may be particularly inaccurate.[64] He finds that when a regression based forecast depends on a forecasted independent variable where the forecast of the independent variable is subject to significant variance, it is likely to be less accurate than a misspecified forecast (one that ignores an obviously significant independent variable). This finding casts considerable doubt on whether correlation techniques (which include some more complex

ARIMA techniques that are discussed in this dissertation) are likely to be more beneficial in forecasting than simpler extrapolation techniques such as exponential smoothing.

Causation

I am left with the idea that some people simply cannot accept that forecasts that ignore regression and covariance may be better than, or at least as good as, forecasts that rely on extrapolation techniques. This remaining hesitancy undoubtedly relates to the notion of **causation**. Although it a tenant of research design that covariance does not, by itself, imply causation,[65] covariance is commonly assumed to be statistics' most powerful measure of causality. For the moment I will set aside the philosophic problem of induction[66] which is the root of the problem of covariance. In practice causation is demonstrated by logically isolating the relationship (or accounting for all important component causes within the covariance structure), telling a good story as to why there is causality, establishing temporal order, and demonstrating covariance.[67] Regression gets only the last of these. No amount of model fitting and regression diagnostics is an adequate replacement for following all of these steps. In practice, forecasters may frequently find limited amounts of data that are available for forecasting to the horizon they need. When faced with this condition and, perhaps, a false perception that a high correlation coefficient is proof of a good model, forecasters may turn to correlation maximizing techniques such as stepwise regression[68] rather than carefully demonstrating causality. Even if they avoid the correlation maximization error, they may not be afforded the luxury of fully demonstrating causality. As compared with techniques that carry such heavy baggage, it should not be surprising that techniques that rely on much simpler assumptions (essentially, that demographic or economic data does not fluctuate widely over a short period of time) can produce comparable forecasts.

Conclusions Regarding More Sophisticated Techniques

Exponential smoothing is a valuable forecasting technique. In this dissertation it is considered worth further examination and refinement. The possibility that other more sophisticated techniques may provide alternative solutions to the problem that is described in this dissertation is not considered a reason why it is not worth resolving the problem within exponential smoothing. In particular, more sophisticated techniques may be considered to exhibit problems that simpler techniques may avoid. If the problem presented in this dissertation can be resolved or ameliorated without significant loss of simplicity, it will be considered a worthwhile improvement, regardless of whether the solution is the most accurate solution available under ideal conditions for use of sophisticated techniques. Actual forecasting may frequently occur under conditions that are not ideal for use of sophisticated techniques.

Summary

Forecasting is an important analytic technique used by public administrators for numerous public budgeting and planning purposes. Forecasting is an accepted topic of discussion in public administration literature. The level shifting problem identified briefly here an discussed in the next chapter arises in data that is forecast for planning and budgeting in public administration. There are various techniques for forecasting serial data, these range from simple to sophisticated. Exponential Smoothing is at the intermediate level of sophistication. Several versions exist. Research indicates that exponential smoothing may be as accurate in forecasting as more sophisticated techniques. This dissertation focuses on the use of exponential smoothing as a forecast technique. Discussion of problems arising with this technique should not be taken as implying that the same problems do not arise with other techniques.

CHAPTER 3: LEVEL SHIFTING DATA

In this chapter I will:

- Describe a problem that arises with level shifting data series.
- Describe some simple approaches to coping with this level shifting problem.
- Explain why these approaches should fail.
- Specify the research question of this dissertation.
- Show that additional research into the level shifting problem is needed.

Level Shifting Data

A type of data series that is particularly difficult to forecast is one that adjusts upwards or downwards reflecting some external intervention.[69] I generally refer to these external interventions as ~~a~~ **policy changes** or **level shifts**. They may also be known by such terms as **discontinuities, exogenous events, externalities, interruptions, irregularities, outliers, ramps, shifts, steps, transients,** etc.[70*] These shifts constitute a significant source of forecast failure.[71]

---

[*]These terms are fairly generic and may also refer to events in data that are not associated with external interventions.

In this study I am interested in interventions that result in a **permanent level shift** in data. Level shifts involve two nearly parallel slopes (trends) that are connected by a **ramp** of two or more observations including the end points of the ramp. A ramp is a series of observations that has a steeper or less steep slope than the slope of the periods immediately before or immediately afterwards. When the ramp occurs in the slope between just two end points, it may be called a **step**.*

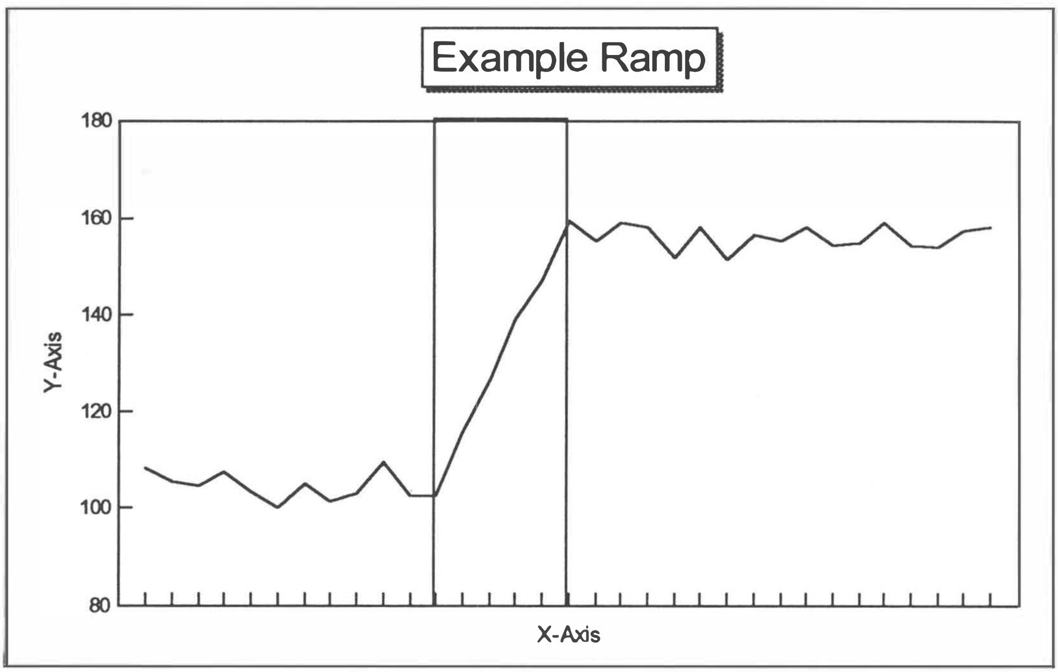

Figure 6

---

*These terms are being defined here, however, they are consistent with uses that are common in forecast literature.

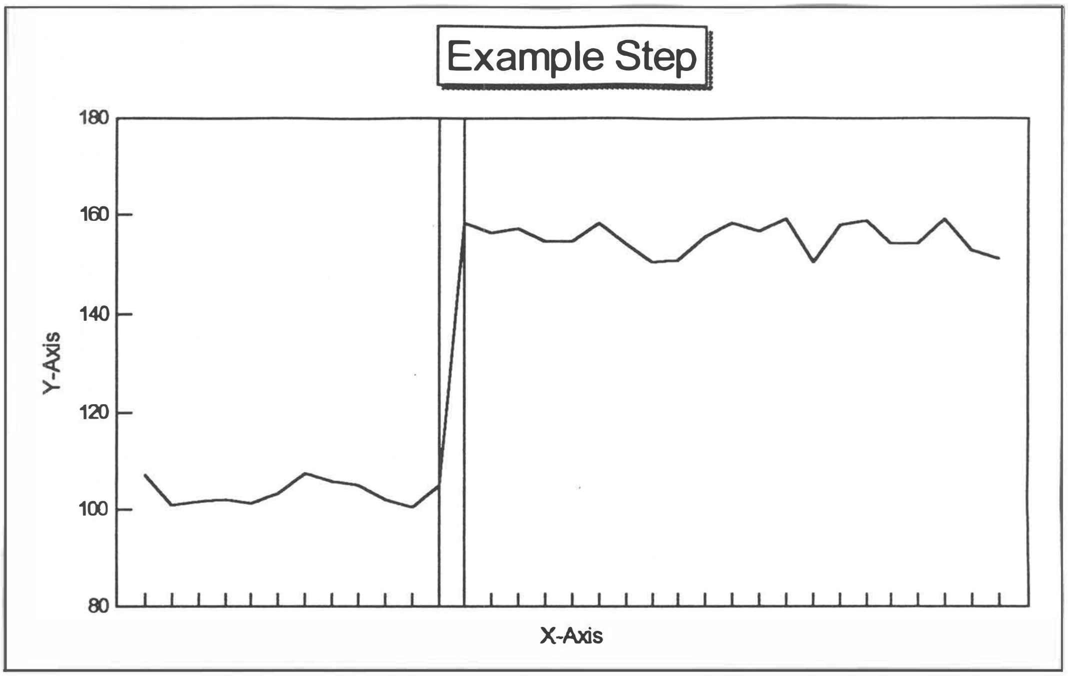

Figure 7

Public administration related data is frequently subject to interventions because of decisions made in the policy making cycle; however, interventions can occur for reasons other than policy making, such as changes in the billing practices of government vendors, addition of significantly large entities to tax or other revenue roles or removal of the same, changes in the items counted in data, re-categorization of data, changes in practices that generate data, etc.

This study concerns level shifts that arise from **planned** policy changes, or other events that are similar to planned policy changes in that they can be **anticipated in advance**. When such events occur, it is possible that the magnitude of such events can be anticipated before they occur. Following is an example of a planned policy intervention.

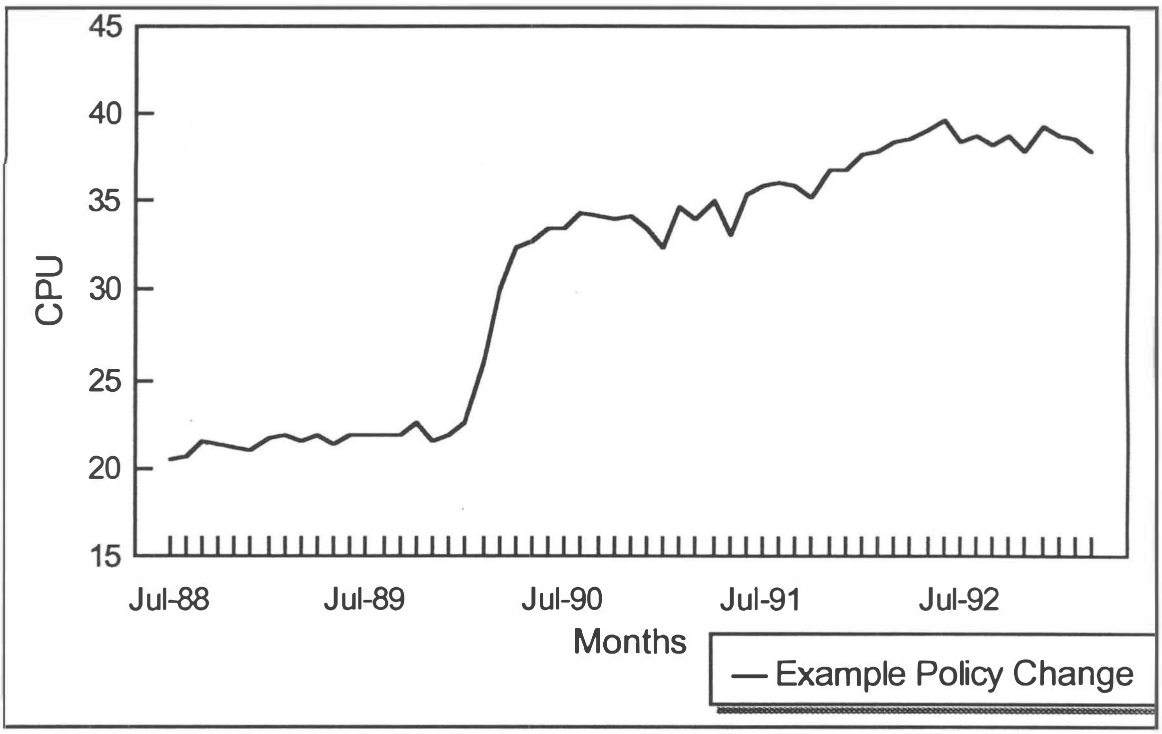

Figure 8

In the Medicaid program each service in a class of services may be subject to a rate ceiling. A public policy intervention may be to raise the ceiling to a new benchmark level. In fact, in 1990 Virginia substantially raised its rate ceilings for physician services.* The impact of this policy change on the data series can be seen in Figure 8. A rate change amounting to over 30 percent of the prior level occurs between the 18th and 23rd periods. The Department of Medical Assistance Services anticipated this level shift in the planning that went into setting the new rates.

Poor Performance

As a user of exponential smoothing,** I have found that this technique performs poorly when data undergo the sorts of adjustment described above. I have found three problems of **reliability** and **accuracy***** associated with the three locations in time where the current period might be in relation to the level shift. The same sorts of problems arise with more sophisticated techniques.[72]

---

*Information regarding the Virginia Medical Assistance Program is known from the researcher's employment with this program for 13 years.

**This discussion reflects my experience as a user of exponential smoothing and is consistent with the proportion-of-error analysis as discussed beginning on page 26. While the general problem of discontinuities is discussed in forecasting literature (see endnote 82), it has not been well analyzed. Thus, there is no precedence for this description of this aspect of the problem.

***In this discussion a "reliable forecast" is one that is not subject to vast variation from one update to the next while an "accurate forecast" is one that turns out to be right.

Reliability and Accuracy

Exponential smoothing forecasts of data that experience level shifts are not reliable, that is, forecasts **fluctuate considerably** as the forecast is **updated**, i.e., new observations are added to the historical observations. There are three phases to this aspect of the forecast problem:

- **Before** the change exponential smoothing models do not reliably **project through** the period of change.

- **During** the shift and immediately afterwards, exponential smoothing models do not **effectively respond** to the change.

- **After** the change exponential smoothing models be **less effective** for a considerable number of updates while waiting for the problems that arose during the shift to clear up, **or** they become **highly volatile** reflecting the undesirable effects of adjusting parameters to let the forecast keep up with the change.

Future Level Shifts

Exponential smoothing forecast models are not particularly good at **forecasting through** future periods during which such planned level shifts are anticipated. Exponential smoothing models contain no information about future level shifts in their historical data, have no other source of information, and have no means of efficiently using information that may be known to the forecaster, but not found in the historical data.* As a result, they do not forecast the level shift. When the actual level shift occurs, data is considerably different than expected, so the model must adjust. Earlier projections are replaced by considerably different later projections.

Concurrent Level Shifts

D. W. Trigg and A. G. Leach describe the **ineffective response** of exponential smoothing models, "With low values of α the forecasting system will take an unacceptably long time to home in to the new level; biased forecasts will occur and will continue for some time."[73] This problem arises because of the proportion-of-error adjustment that has been described. Under ordinary circumstances the forecast has already explained most variation and the remaining variation is noise. Consequently, the model tends

---

***This assertion follows logically from the fact that exponential smoothing models are fit to minimize a loss function that measures historical errors; historical errors are the only source of information for exponential smoothing models.**

to avoid adding back a large proportion of error to the previously existing forecast. However, when the level shifts much more of the error is information. If the model assumes that the error is noise, it will discount the error too much, which leads to inaccurate forecasts.

Past Level Shifts

This **ineffective response** can continue for a long time after the level shift occurs[74] while the parameters continue to sort out the forecast error into noise and information. Alternatively, the forecaster may intervene by raising the α parameter which may allow for more rapid sorting out of noise and information in the level shifting period. However, this may cause a **loss of stability** in the underlying forecast, particularly where the underlying data series is characterized by high variance.

Example Level Shift With Exponential Smoothing

The following graphs show the impact of the previous level shift on a forecast made through a variant of the Holt exponential smoothing model with parameters of $\alpha = 0.25$ and $\beta = 0.01$.

Forecast With Level Shift

The movement of the partial line shows the updated actuals.

Initially the model contains no information about the policy change, it projects the series without the policy change. At this point the forecaster can lump on the estimated impact of the policy change directly onto the results of the statistical forecast model to get a forecast.

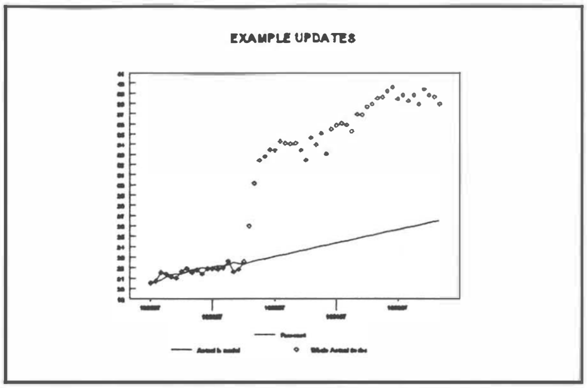

Figure 9

Second Update

As the policy change begins to take effect, the statistical forecast model follows the policy but at a rate that is discounted by the amount of the α parameter. At this point, the forecaster must begin to estimate how much of the policy is "in" before he can lump on the remainder to make a whole forecast.

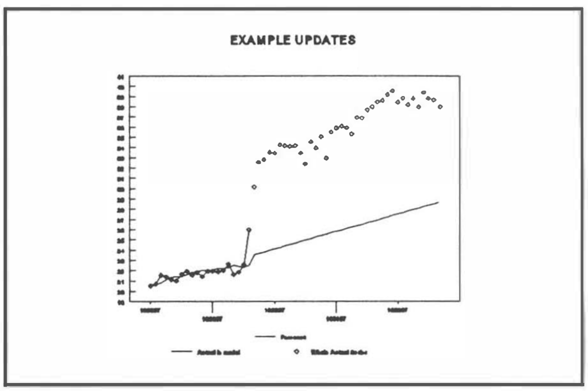

Figure 10

Third Update

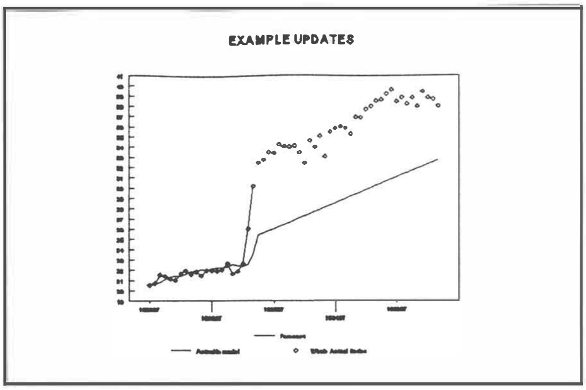

Figure 11

As the policy ramp continues, the trend begins to respond to the policy impact. Here, even if the forecaster accurately estimates how much the policy is "in," he must also factor out how much the statistical forecast model is overestimating before he can make an accurate whole forecast. During this period, it is not possible to simply add the lump sum value of the policy to the prior forecast, because some of the impact of the policy is in the forecast. The data associated with the policy change is beginning to enter into the historical series, the forecast model is adjusting the projected future level and trend - however inaccurately - for this change. If the policy is added back to the forecast, the overall estimate will be too large. If it is not added back to the forecast, the overall estimate will be too small (or large for negative changes). The only option available is judgmentally adjust the lump sum amount to add back to the forecast.

Fourth Update

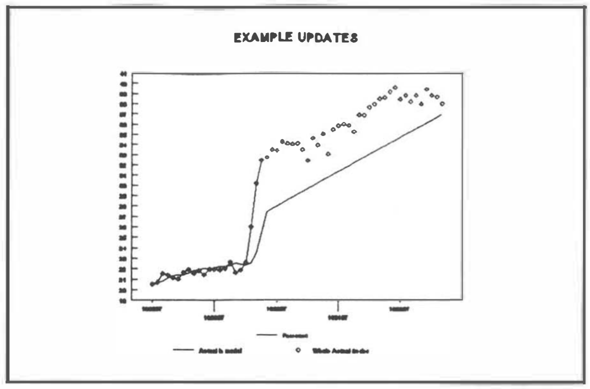

Figure 12

While the policy change is fully in effect, the forecast errors remain large. Both level and trend continue to adjust upwards reflecting the presence of positive forecast errors. At this stage, the forecaster no longer worries about how much the policy change is "in." Instead, his problem is with how much the statistical forecast model is affected by the policy impact on the data.

Fifth Update

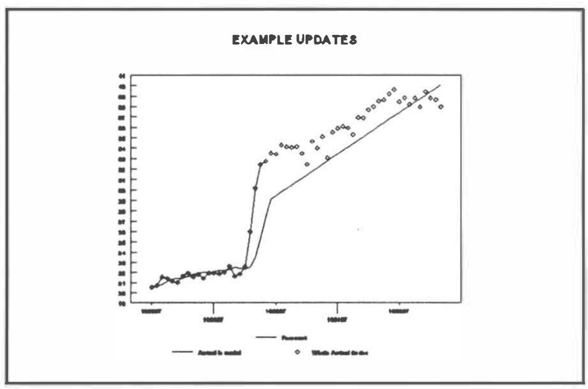

Figure 13

While the data series begins to return to pre-policy change patterns at a higher level, the statistical forecast model continues to adjust upwards due to the large positive errors. (In this actual data series other events led to an shift in the trend just after the rate change. Had this not occurred, the overestimation of the trend would be even more extreme.)

Sixth Update

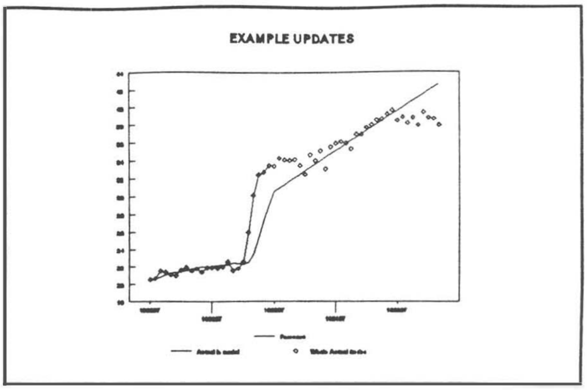

Figure 14

The actual data has returned to the pre-change pattern, but the forecast continues to adjust upwards in both level and trend. The short term forecast is too low, because the level is underestimated. The longer term forecast is too high, because of an overestimation of trend.

Seventh Update

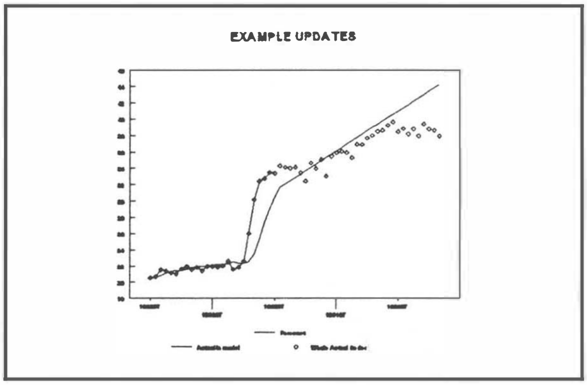

Figure 15

As the positive errors continue to raise the level and trend, the forecast begins to over estimate most of the data series while continuing to underestimate the next few observations.

Eighth Update

Since the short term forecast continues to underestimate the next observation, the level and trend continue to adjust upwards, producing a severe over estimate of the intermediate and longer horizons.

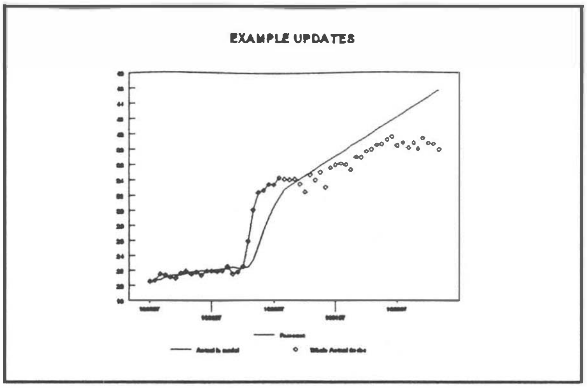

Figure 16

Ninth Update

Ultimately the level catches up with the trend. At this point, the short term forecast may be reasonably accurate for one to three future periods. However, the intermediate forecast is over estimated because of the trend adjustments and the longer term forecast is severely over estimated.

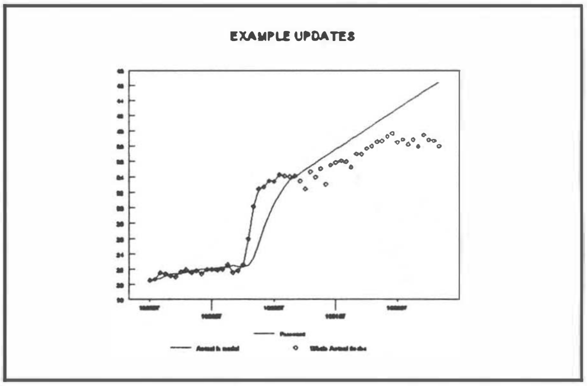

Figure 17

Ineffective Responses to Level Shifting

Following are some of the obvious ways that exponential smoothing models can be used when the data they forecast experience the sort of shift described above (these approaches are presented to clarify the problem, other approaches are reviewed in the next chapter):

Level Left Optimized

The level parameter can be left as optimized before the shift and the forecaster can wait until the forecast model ultimately becomes effective again after many periods of ineffective forecasts.[75] The forecasts will be ineffective for two reasons:

- Since the forecast is optimized at a time when the data does not experience such a massive shift, it will consider a high proportion of the variation in the data as random noise (i.e., the α value will be set at a low number). Thus the information contained in the level shift will be excessively discounted, and the level of the forecast will fail to keep up with the more massive shift associated with the policy intervention. Thus, the near term periods of the forecast will be under estimate the level shift (become biased in the opposite direction of the level shift).

- Since the level will fail to keep up with the shift, the errors will become much larger than usual, and will be highly autocorrelated, i.e., will repeatedly have a positive or negative sign. This result will cause the correctly estimated trend and seasonal factors to over respond to the error and become incorrectly estimated. After a few periods the forecast will severely over estimate the change in trend with respect to the level shift, thus the more distant horizons of the forecast will be over project in the direction of the level shift. Similar, but more complex, confusion will occur with seasonality.

Adjust the Level

The level parameter can be adjusted (increased) to allow the forecast to respond to the new level shifting information.[76] Since the level parameter is a proportion that is multiplied against the error to produce a new level estimate, the shift can be rapidly included in the forecast by setting the α parameter very high. **In this study, this approach, when combined with adding in the lump sum value of**

the expected policy change, will be called the *ad hoc* method.*

As seen in the following graphs, this approach also produces an ineffective model. Although the model responds to the level shift, thereby resolving, or at least mitigating, the difficulties discussed above, it does not restrict is response to that level shift. When the level shift is over, it continues to respond just as rapidly to random noise. Over time the forecast changes significantly from period to period, making specific forecast results unusable as it is difficult to determine which projection to rely upon unless the forecast user is interested only in the next future observation.

In the following graphs, the same data as shown in the previous models is forecast with $\alpha = 0.8$ and $\beta = .01$. Selected updates from periods after the level shift are shown to demonstrate the consequence of raising α to allow

---

***I have not found literature that demonstrates the use of the *ad hoc* method; however, it is clearly the simplest approach available. When I have discussed the technique proposed in Chapter 5 with forecasters, I have been asked how it differs from the *ad hoc* technique which is, by implication, potentially adequate to meet the problem.**

the policy change level shift to rapidly come into the model.

Level Shift α = 0.8

When the forecast parameter is raised to reflect the anticipation of a level shift, the forecast catches up with the level shift after a fairly short lag. This reduces, if not eliminates the negative impact on the forecasted trend.

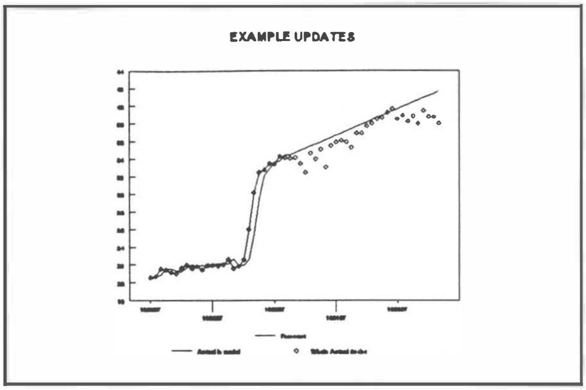

Figure 18

Update with Variation

Afterwards, when the data experiences significant variation that is simply unexplained noise, the forecast level follows the variation just as faithfully as it follows the explained level shift.

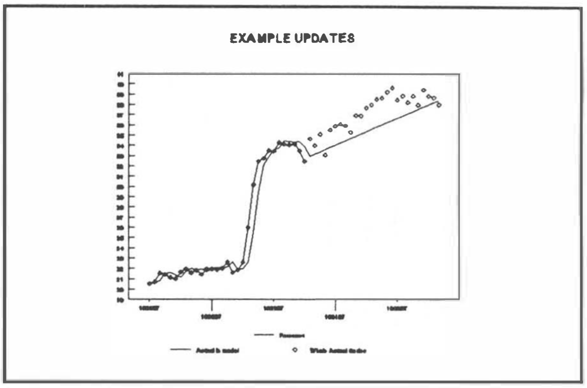

Figure 19

Second Update

This tracking of variation produces a roller coaster effect in the forecast. Sometimes the forecast is down as with the last graph and some times it is up as with this one.

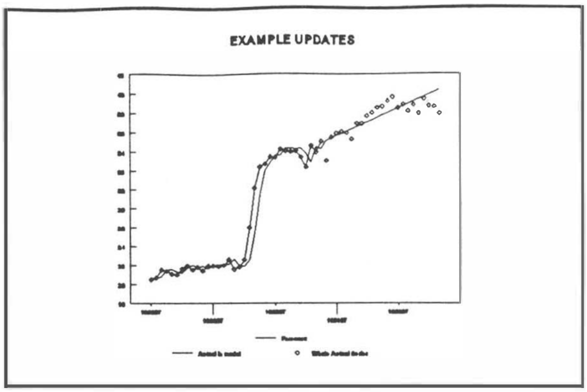

Figure 20

Third Update

The consequence of this roller coaster effect is that each update produces a whole new forecast that is significantly different from the previous forecast update. The user does not know which one represents the anticipated future. The forecast is no longer an abstract summarization of the historical data, it is instead a trend that takes off from a point near the last actual observation.

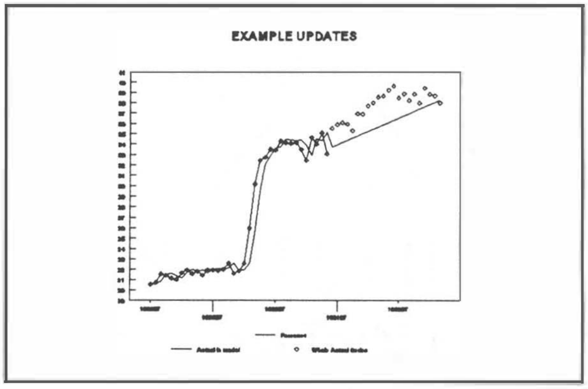

Figure 21

Updating

A topic that is not widely discussed in forecasting literature is the matter of updating, for example while he widely cites studies for almost all other assertions, J. Scott Armstrong's discussion of updating exponential smoothing models is limited to: "*Frequent updating is important for accuracy* [Italics in original]."[77] Many organizations make periodic forecasts of the same data series. When such reforecasting is made with the same or similar models from time to time, such forecasting may be called updating. In my practical forecasting experience, updating is an important aspect of forecasting.

For intermediate and longer forecast horizons, including many horizons of governmental forecasts, there are two contrary needs with forecast updates. First, the forecast should not experience high variability (bounce around a lot), in other words, it should be reliable.[78] If updates are frequent, say monthly, and the forecast frequently changes by a significant amount, say 10 percent of the incremental growth from a current year to a budget year, the forecaster cannot have a lot of faith in the current forecast as compared with the forecast from the most recent previous update. On the next update it might bounce back. Second, the forecast should change when there is a change, in other words, it should be accurate; that, it would seems, is the point of updating.[79]

In part, the problem that is raised in this dissertation is a problem of updating under conditions where one of these two needs may not be met. If the forecaster uses commonly accepted techniques to avoid excessive variability, i.e., optimizes a forecast in the absence of level shifts, the forecast likely will not change when the level shift occurs. On the other hand, if the forecaster adjusts the forecasts to recognize the level shift by raising the level parameter to a high value (or, as will be discussed later, by using an adaptive technique), the forecast may respond not only to the level shift, but also to every other event that might cause noise in the data.

Research Question Specified

In this dissertation I ask: **Can a method be devised to use prior knowledge of policy shifts to improve performance of exponential smoothing forecasts?** I compare the performance of various simple exponential smoothing models and approaches that might be used to forecast through periods of level shifting. I examine whether any of the methods considered is superior. Included among these methods is a method I have developed for use when policy changes are anticipated (see Chapter 5). Two major hypotheses are specified in Chapter 6.

Need for a Study

As a forecast practitioner for a public program I find that policy changes pose significant difficulties for forecasts. When I have information about prospective level shifts, I have found it difficult to effectively use that information. The approach I formerly used was to add back the data in lump and let the forecasts adjust to the data as policies went into place (the *ad hoc* method). This proved unsatisfactory in practice because:

- Forecasts perform particularly poorly when the data series they forecast are undergoing or have recently undergone level shifts.

- It is difficult to account for out year consequences of policy changes when they are added back in lump sums. Typically in forecasting for the Virginia budget one needs forecasts that address the current year and two subsequent years. Lump sums that may reflect mid-year policy start up are seldom of much use for any year but the year that the policy is expected to go into effect.

- It is difficult to work with lump sum amounts when summing the forecast over various periods for various purposes. The Medicaid forecast is summed over one fiscal year for use in state budgeting and another fiscal year for use in federal budgeting. Lump sum policy estimates only add confusion under such circumstances.

- Lump sum changes are particularly difficult to work with when initial assumptions change due to the nature of decisions that are made as the policies are put into place.

- Once lump sum policy changes begin to become part of the data series, there is little information available to guide the judgements necessary to decide how much of the policy is "in."

The practice of increasing the level parameter to allow the policy change in quickly avoids some of the consequences of the problem, although not all (e.g., it does not resolve the question of how much of the policy is "in"). However, it brings its own costs in terms of increased forecast variation (see Figure 18 through Figure 21).

In discussions with other practicing forecasters I have found that they too have difficulty with policy changes. For example, the Virginia Corrections forecast has difficulty with accounting for the impact of new laws that result in prison sentences on their forecast of new commitments.[80]

As discussed by Fred Collopy and J. Scott Armstrong,[81] the problem of ramps and steps is one of the more severe unresolved problems in time series analysis. As discussed above, these problems arise in data that is forecasted by and for the public sector and are significant to planning and budgeting for major public programs. It is, therefore, useful to the practice of public administration to evaluate a technique that may mitigate this problem.

Summary

Level shifting data poses a significant problem for forecasting through exponential smoothing models. It can cause an exponential smoothing model to experience a considerable period of serially correlated errors. Such errors may lead to inappropriate estimates of trend which may continue for some time after level projection errors are corrected. Adjustment of the α parameter may reduce these problems, but it can lead to instability (high variability) in the forecast. When forecasters possess information about prospective level shifts, exponential smoothing models do not make optimal use of that information either in making projections or in minimizing error in the model. With the standard Holt-Winters type exponential smoothing models one cannot simultaneously minimize variability of forecasts and maximize response to anticipated level shifts. In this dissertation various models are examined to determine whether one is more effective than another in projecting level shifting data. One model that is included in this comparison has been developed explicitly for dealing with this sort of data. Recent publications indicate that forecasters consider the family of problems of which it is a member, **discontinuities**, to be one of the more severe problems presently unresolved in time series analysis.

CHAPTER 4: FORECAST LITERATURE CONCERNING LEVEL SHIFTS

In this chapter I will:

- Examine the forecasting literature regarding discontinuities.
- Identify techniques used to identify level shifts.
- Identify techniques used to adapt to level shifts.
- Identify techniques for forecasting where prospective level shifts are anticipated.
- Briefly review literature regarding estimation of prospective level shifts.

Literature Regarding Discontinuities

Fred Collopy and J. Scott Armstrong have recently asserted, "[Time] series forecasting research and practice have largely ignored abrupt changes."[82] They find this particularly mysterious because they find that 92% of forecasters consider this topic to be important in selection of extrapolation methods, ranking it the third most important feature examined.[83] An examination of the literature reveals that things are not as bleak as Collopy and Armstrong assert. Three common approaches to forecasting through periods of discontinuous data series are techniques that:

- Alert the forecaster to suspect that the model is misfit at the point of update,
- Recognize the discontinuity and adapt to it, or
- Assist in fitting a model under conditions of discontinuity.

Alerting the Forecaster to Discontinuities

A tracking signal is a statistic that is used to signal the forecaster that something has occurred near the end of the historical period that may result in the forecaster's wanting to reconsider and replace the current forecast model. Everette S. Gardner, Jr., identifies several tracking statistics including the simple cusum, the backward cusum, the smoothed error tracking signal, and the autocorrelation tracking signal.[84] These tracking signals monitor forecast errors to determine whether the models are in control, that is still reasonably well fit. They demonstrate out of control conditions when they exceed certain critical values. The principal underlying these methods is to establish a ratio related to the error term in the forecast model. Critical values are values at which these ratios indicate that the forecast is out of control, for purposes of this study that would mean it had undergone a level shift. Gardner's findings suggest that all these methods are useful under appropriate conditions. John O.

McClain has compared cusum with the smoothed error tracking signal and found that the smoothed error tracking signal is more effective for identifying out of control conditions quickly.[85]

Lewis W. Coopersmith develops an alternative approach that uses the F-statistic:

> When a continuous model is assumed, a procedure for detecting knots [Coopersmith's term for either level or trend shifts]. . . . involves a search over an interval which is first set small enough so that it is unlikely that more than one knot would occur. The point is determined which maximizes the F-statistic used in testing for a significant change in trend. If the maximum F-statistic is not significant, the evaluation interval is extended and testing is repeated. For discontinuous models that include shifts, the search is extended to determine the point where the F-statistic which tests for the significance of [ARIMA intervention] parameters . . . is maximized. After the knots are determined, robust procedures . . . are used to estimate the linear pieces; the last piece is extended for use in forecasting.[86]

While Coopersmith's approach assumes that the forecast technique is ARIMA oriented, it is readily generalizable to other techniques so long as sums of squared errors can be calculated in comparison between forecast models that do and do not contain appropriate adaptation for discontinuities.

Vollmann, Berry, and Whybark suggest another a tracking signal, Bias divided by Mean Absolute Deviation. They do not clearly define their measure of Bias; however, the context suggests that they intend average error or the smoothed error tracking signal. Thus, as they say, this tracking signal is limited to:

$$-1 \leq \text{Tracking Signal} \leq 1$$

This tracking signal suggests a forecast out of control when it approaches either 1 or -1.[87]

Some of these techniques, e.g., Coopersmith's technique, require considerably more effort than others, e.g., cusum, and may, for that reason, be more appropriate for initial model identification than for a tracking signal functions. All provide the possibility of identification of points where a data series has experienced a discontinuity and can, in principle, be used to identify out-of-control conditions. However, the tracking signal approaches only serve to signal the fact that discontinuities have occurred. The forecaster must still intervene with the forecast model to refit it under conditions of discontinuity. Also, these techniques are not designed to provide for **anticipation** of discontinuities. Problems related to forecasting **through** prospective periods of discontinuity are not addressed by these techniques.

Recognizing Discontinuity and Adapting To It

A natural extension of the tracking signal is the use of some signal in the data to adapt the forecast to the conditions near to the end of the historical period. This approach is sometimes called adaptive forecasting. There are several forms of adaptive forecasting. D. W. Trigg and A. G. Leach pioneered the approach of an adaptive α parameter.[88] The tracking signal modifies the α parameter so that it is **large when the error is unusually large**, i.e., the forecast is out of control, and **small otherwise**. A commonly recognized Trigg-Leach model is as follows:[89]

$$\text{Smoothed Error} = E_t = \Phi e_t + (1-\Phi)E_{t-1}$$

$$\text{Smoothed Absolute Error} = M_t = \Phi|e_t| + (1-\Phi)M_{t-1}$$

$$\alpha_t = \text{Absolute Tracking Signal} = T_t = |E_t/M_t|$$

In the Trigg-Leach model, the α parameter automatically adjusts with every update. Another approach is the Whybark method[90] which adjusts parameters to preset adaptive levels only when tracking signals exceed certain critical values.

Independent Fit

T. M. Williams has improved the Trigg-Leach approach by removing a source of confusion in the underlying Holt-Winters model.[91] Williams' idea is that the parameters of the Holt-Winters model are not independent, so when the α parameter is adjusted by the Trigg-Leach approach, the β and γ parameters are incidentally also adjusted. He develops a revised Holt-Winters model that does not have these interactions, thereby allowing independent adaptive modification of the α parameter. In Holt and Holt-Winters, the level parameter, α, is fit to all three forecast components, level, trend, and seasonality. In the Williams modification, this parameter is not included in the fit of the trend and seasonal components of the model. As a result β and γ are allowed a broader range of possible fitted values. Formulas are shown in Appendix II. Williams uses the Trigg-Leach smoothing signal for calculating the adaptive α parameter,[92*]

Because the parameters $\Phi = \{\alpha,\beta,\gamma\}$ are restricted to values of $0 \leq \Phi \leq 1$ and because Williams modifies Holt and

---

*He actually presents a different formula, but his formula makes no sense (it exponentiates out of control under certain conditions). A careful review of his math and his text shows that he erroneously substituted " $\alpha_t = E_t/\Delta_t$. for "$\alpha_t = |E_t/\Delta_t|$".

Holt-Winters by dropping α from certain multiplications, the result is to increase the effective magnitude of the β or γ parameter by the size of α. In practice α is frequently selected to be quite small, e.g., $\alpha \cong 0.1$, so to retain roughly the same effect in a Williams model, the parameters should be adjusted downwards by a factor of roughly 0.1.[93]

While Williams makes these adjustments to allow for less problematic adaptive forecasts, it is equally reasonable where adaptive forecasting is not in use. Williams' technique has been reinvented by Blyth C. Archibald in 1990.[94] The Williams model is used as a basis of a proposed model in this dissertation and is referred to as the Holt-Williams* or the Holt-Winters-Williams model.

The idea of these adaptive models are that a tracking signal can automatically signal the α parameter to increase or decrease as necessary to keep the forecast in control. Thus, α is given a time index rather than being treated as static across the whole model. The time indexed α is increased when the model appears out of control and decreased when the model is in control. Such increasing or

---

***"Williams" in these models refers to T. M. Williams, not the current researcher.**

decreasing is intended to allow the forecast to remain fairly steady during periods of stability yet respond rapidly when the tracking signal detects instability.

The Williams article provides an extensive bibliography of other adaptive models, including those by Eilon and Elmaleh,[95] Steinar Ekern,[96] Theil and Wage,[97] and Nerlove and Wage,[98] as well as many other citations.[99] Adaptive ARMA (autoregressive moving average models) have also been developed.[100] These articles generally discuss variations of adaptive models, the generalization of adaptive models, and the effectiveness of adaptive models which Ekern in particular questions. It is generally accepted that adaptive models are not satisfactory. Armstrong cites 12 studies that support the view that adaptive forecasting is ineffective.[101]

Autocorrelation

C. Chatfield proposes an exponential smoothing technique where the forecast is adjusted by adding the factor (e_{t-1} * $\varrho_{e,et-1}$), that is the autocorrelation of the errors at time t-1 is multiplied by the error at time t-1 and added back to the forecast.[102] This factor is exponentiated for periods beyond the end of the sample period. This technique is not specifically proposed for the purpose of adaptive forecasting, but it can be seen to be a variation of an adaptive technique. The factor will become large where errors are autocorrelated and small where errors are not autocorrelated. Autocorrelation may arise under other situations, but should certainly arise when the forecast systematically erroneous due to level shifts. Later references to a Holt-Winters-Williams variation of this model in this proposal will label it autocorrelation corrected Holt-Winters-Williams. That model uses the Williams correction to Holt-Winters and also uses the Chatfield autocorrelation correction. This model is demonstrated in Appendix II.

It should be apparent that all adaptive techniques implicitly employ tracking signals. The Trigg-Leach method employs the signal a smoothed error signal.[103] The Whybark method employs information about the standard deviation.[104] Chatfield employs autocorrelation of errors.[105]

Like tracking signals, these techniques are designed solely for dealing with level shifts that are identified retrospectively through the data used to fit or update the model. They have no method of efficiently using information the forecaster may have concerning planned policy changes.

While they may help a forecast catch up to a new level once it is observed in the historical data, they do not particularly help forecast and update through a period of updating.

More Sophisticated Adjusting Models

Kalman Filters

The Kalman[106] filter approach optimizes meta-parameters that allow for the forecast parameters to adjust with the level of variation in the data series. P. J. Harrison and C. F. Stevens have demonstrated that the Kalman filter approach can be generalized to include both correlation based models and time series extrapolation models. Kalman filters are sometimes called state-space models or the use of these models may be called Bayesian forecasting. Kalman filters are not necessarily sensitive to level shifting data, instead they allow self-adjusting parameters with ordinary data.

Harrison and Stevens developed a multi-state model which is specifically designed to allow Kalman filter forecasts to respond to level shifts.[107] This multi-state model allows the forecaster to define multiple Kalman filter models (they recommend four) which are designed to respond to various specific types of data discontinuities (level shift, trend shift, outliers, and no discontinuity). The multiple models are aggregated through an assignment of probability to each of the various states.

Duk Bin Jun and Robert M. Oliver define another variate of the Kalman filter which is designed specifically for level shifting data.[108] This technique adds a dummy variable to the Kalman filter model at the point in time where the level shift is thought to occur. Duk Bin Jun also conducted further analysis concerning statistics that assist with identifying the period of the level shift.[109] This technique assumes that the level shift occurs over a single period. While Jun an Oliver argue that this technique should be better than Trigg-Leach, they do not demonstrate comparative effectiveness.

While the Kalman filters discussed here are variates of exponential smoothing, they are not appropriately classified as simple models. They are mathematically more complex than exponential smoothing, particularly the multi-state model that is most comparable to an adaptive exponential smoothing technique, and they may require more sophistication for model fitting. Also, these are most appropriately classified as adaptive models, they do not provide for anticipation of level shifts.

Non-Gaussian Models

The non-Gaussian forecast model,[110] and closely associated, approaches attempt to distinguish between ordinary variance and level shifts through the use of heavily weighted tails in the probability distribution function surrounding the forecast. In effect, the forecast shifts from one level to another when repeated observations indicate a new mean. The use of non-Gaussian probability density functions allow for smoother transition between level estimates. Non-Gaussian models are, however, another form of adaptive model that provides no opportunity for anticipating change.

ARIMA

ARIMA provides three approaches to accounting for externally driven data shifts. These are the transfer function model, the intervention model, and the multivariate time series model.[111]

- The transfer function model combines the features of a univariate time series ARIMA model with features of a regression model. It adjusts the autoregressive results of an autoregressive univariate model to also take into account the effects of a known causal variable. If this causal variable can be made to change to reflect the policy change, it can be used to include the policy effect in the forecast. If not, then if the input variable changes naturally as the policy changes, it can at least help the forecast keep up with the policy change.

- The intervention model is a special case of the transfer function model that uses a dummy variable. The dummy variable is set at zero for periods during which the policy (or other source of level shift) is not in effect and 1 for periods during which the policy is in effect. Use of dummy variables requires knowledge that non-stationarity* has occurred.[112] Analysis of the data series should reveal the existence of non-stationarity. However, this analysis is a major component of the increased analyst cost for using ARIMA type models.

***"Non-Stationarity" refers to a condition of a time series where the series does not have a constant mean and/or variance Linearly trending data can be induced to be stationary through differencing, so a data series that can be fit to a Holt model is implicitly stationary. For the purposes of this study it is adequate to assume that a data series that can fit to a Winters model is also stationary. There are many reasons why a data series may be non-stationary, level shifts are only one form of non-stationarity.**

- The multivariate ARIMA model is similar to the multiple equation econometric regression model. It simultaneously solves interrelated multiple time series. Transfer function models and multivariate ARIMA models can account for future policy changes by including independent variables that contain anticipated policy changes in the future period, if such variables are also significantly related to historical periods. However, to do so those independent variables must themselves have forecasts that reflect the prospective policy change.

Structural change non-stationarity can be classified into five types, additive outliers, innovational outliers, level changes, transient level shifts, and variance changes. Ruey S. Tsay has developed specific procedures for identifying each of these sorts of non-stationarity in forecast data and identified specific ARIMA models that are appropriate to each.[113] David J. Pack[114] has developed theoretical ARIMA models that allow for modelling any sort of non-stationarity.

The Pack models are designed to provide the forecaster with precise techniques for modelling variation, including any form of non-stationarity or intervention variation that might arise in the **historical** or **sample** data. However, they provide no guidance for forecasting through prospective shifts that are unrelated to historical variation. A related model designed to suppress irrelevant outliers is described by Steven Hillmer. This model prevents the forecast from becoming biased when one time outliers occur.[115] The foregoing discussion is not a comprehensive review of ARIMA modelling; such a review is beyond the scope of this dissertation.

In this study I am interested in level changes which may or may not be preceded by innovational outliers. These are the sorts of data series that reflect onset and permanent change related to external causes. The intervention ARIMA model provides for forecast model fitting in the case that historical data reveals a level shift or other non-stationarity. It is not designed to assist with forecasting through future periods that include anticipated non-stationarity. In fact, with ARIMA modelling it is necessary to **supplement the model fitting procedure** with other **special non-stationarity identifying procedures** to achieve a similar level of effectiveness for dealing with non-stationarity while updating as is available with other models discussed in this section.

Regression models are not discussed separately because their use in forecasting of level shifts is generally parallel to transfer function, intervention, and multivariate ARIMA models. This brings us back to the reason simple methods like exponential smoothing may be better than more complex methods, "The approach . . . is often inappropriate . . . first of all owing to the lack of the relevant data on the exogenous variables."[116]

Summarizing Sophisticated Techniques

The techniques discussed in this section will not be further examined in this study because they are not associated with exponential smoothing techniques. They have been examined to determine whether the problem identified in the previous chapter is readily resolved with other techniques. In general these techniques rest on the assumption that level shifts are identified in the data that is used to fit the forecast model rather than anticipated before the fact. They do not employ information that may be available to the forecaster concerning anticipated policy changes. Even transfer function models require that the independent variable that reflects the prospective level shift must also be correlated with the historical data. While these techniques may provide some solutions to level shifting data, they do not provide so clear a solution as to rule out the potential benefit of identifying an exponential smoothing solution to the level shifting data.

Models Targeting Level Shifts in the Horizon

Adaptive and intervention techniques are available for forecast modelling under conditions of discontinuity when the discontinuity is discovered in the history of the data. In this section I discuss models that anticipate level shifts in the forecast horizon.

An ARIMA Model

Victor M. Guerrero provides another use of ARIMA in modelling level shifting data.[117] He begins with the following problem:

> Since some new economic policies were to be implemented, a structural change on the behavior of IMP was expected and a higher than usual rate of growth of IMP was agreed upon. Then, a future monthly path, consistent with the annual target and with the [available] historical records, as well as tolerance limits for the pay, were needed to determine whether the observed behavior of IMP during the year should be considered accurate.[118]

This problem involves an adjusting the forecast for policy decisions. He reviews other articles and concludes, "[None] of these papers considered the possibility of structural changes during the forecast horizon."[119] He identifies several formulae that can be used as follows:

$$Z_{F,D,V} = Z_F + D_F + V_F$$

Where, Z is the original time series (where this can be understood as either Z_t from formula 1 above or Y_t from formulas 2 or 3 above), D is the deterministic effect of structural change, V is the stochastic effects of structural change and F is the future period.

$$E(Z_{F,D,V}|Z_0) = E(Z_F|Z_0) + D_F$$

This formula can be understood to mean that the expected value of the future series is the expected value of the old series plus the deterministic effect of the policy change.

While it appears that Guerrero is dealing with the problem raised in this study, prospective structural shifts, the actual results of the math he demonstrates allows for estimating the **interim** values of a forecast when a plan of action is assumed to achieve a certain end point. The vectors for the deterministic and stochastic effects must be estimated based on the anticipated value at the end point. The point of his models is to determine how to estimate the interim vectors and their variance so as to be able to track actual interim performance and determine whether actual observations are leading to the anticipated end point. This is significantly different than the problem posed in this study, i.e., having prior knowledge of the deterministic structural effect and needing to combine it with the underlying series to achieve a full forecast. Guerrero's main contribution to the objective of this dissertation is further confirmation of the general absence of studies focussed on "structural changes during the forecast horizon."

When Patterns Change

Spyros Makridakis and Robert Carbone have developed a forecasting approach that is particularly aimed for forecasting when there are pattern changes that occur beyond the period of the historical data.[120] This method distinguishes between short and long term forecasts using adaptive or responsive methods for forecasting the short term while using less responsive methods for forecasting the longer term. These forecasts are combined through weighted averaging with the weight beginning in favor of the responsive technique and shifting to the non-responsive technique. The underlying idea of this approach is that short term fluctuations may not reflect permanent changes and, therefore, should not be allowed to excessively influence the calculation of the longer term forecast. The

Carbone and Makridakis approach allow the forecast to treat these as trend shifts for the short term, thereby allowing the forecast to follow their impact over the short term. However, the technique returns the forecast to the discipline of the more stable trend over the longer term, preventing the longer term forecast from falling completely off track.

This approach tends, implicitly, to support the disposition that arises with many forecasters to prefer setting extremely low forecast parameters.[121] This disposition can be understood to reflect an effort to avoid excess influence of short term fluctuation in forecasting the longer term. Where the forecaster is more concerned about the shortest of the short term or about making a forecast that captures a fluctuation that occurs near the end of the historical period, this disposition to set low parameters disappears.[122] The use of high forecast parameters has the effect of allowing the forecast to respond to new information in much the same way as adaptive forecasting responds to such new information with the difference being that adaptive forecasting adjusts the amount of response such that it responds less where the tracking signal indicates the variation should be counted as noise rather than information. While such biases as a preference for low parameters may reduce forecast fitting success, practical experience may support the use of such strategies since optimal models established within the sample period frequently are not the optimal models in the forecast period.[123]

Repeating Historical Fluctuations

Wilpen L. Gorr has developed a protocol for establishing special event data bases.[124] Such data bases provide for the possibility of retaining factors or other information that can be used in forecasting through periods where events have occurred. These factors might be additive or multiplicative in classic decomposition models, or might be other information. Gorr's articles look at this issue not, primarily, from a forecasting point of view, but from the perspective of information management. From this perspective, it is crucial that special event information should be retained in a manner that allows for use after those who have first hand knowledge of the event have left the organization. The exact use of the information in forecast modelling or other analysis is not necessarily pre-specified. While quantitative information is useful, this approach also focusses on qualitative information that might be used to understand discontinuities in data series.

Gorr's articles point towards the work of Rudolf Lewandowski[125] whose technique, FORSYS, is identified as exhibiting superior performance over longer time periods in a frequently cited forecasting competition.[126] Lewandowski describes a specific additive technique used to remove the effects of special events from forecasts before extrapolating them into the future. In addition, he asserts that such special event information can be used by managers to anticipate the effect of recurrences of the same special events in the future. He describes two uses of this technique:

- Where special events are known to have occurred in the past, he decomposes the data series by determining an additive level shift that adjusts for the discontinuity that occurred in the history of the data.[127]

- Where a special event has occurred in the past and is anticipated to occur again in the future, the magnitude of the past special event is used as a guide to gauging the special event in the future.[128]

This second usage appears to partially address the problem raised in this dissertation. Lewandowski is using the prior temporary level shift as an estimator of the future temporary level shift (period by period). Lewandowski's discussion shows that he uses this estimator as an adjustment factor for the forecast in the future periods where a similar special event is anticipated. He explains this usage as accounting for such activities as a sales promotional campaign.[129]

Gorr also identifies the FUTURCAST software package of R. Carbone and S. Makridakis as containing a multiplicative technique associated with special events.[130] The multiplicative special event factor is similar to a multiplicative seasonal factor of a multiplicative Holt-Winters forecast model. In effect, it estimates a percent change from the underlying base line forecast associated with the special event. This multiplicative factor allows for modelling of future special events by analogy to prior special events in a manner similar to Lewandowski's second use of his special event factor. The analogical use of the multiplicative factor may be more beneficial where the underlying series has changed in magnitude between two occasions of the special event.

While the special event factors of Lewandowski or Carbone and Makridakis may be useful when there is a temporary divergence from a normal condition, or perhaps a cyclical pattern that varies from an underlying linear pattern,[131] they are not the most parsimonious method of representing permanent changes. The difficulty, in the case of permanent changes, is that the special event factor requires maintenance for each future period during which the special event is in effect. If the event is permanently in effect, the factor must be maintained for all future periods. It would be more efficient to permanently adjust the level of the model.

Another Repeating Model

Jose Juan Carreno and Jesus Madinaveitia developed a modified Holt-Winters model that uses an index similar to a seasonal index to forecast the impact of announced price increases on a forecasted time series.[132] This index is a set of multiplicative factors that are computed to reflect the impact of the price increase over a cycle in the demand series. The factors computed during one cycle are taken as prior expectations during the next cycle after being re-selected by the forecaster upon becoming aware of the intention to announce the price increase. The error terms occurring in the cycle adjust the factors for use in the next cycle. They also develop a similar model that is a modification of Brown's double exponential smoothing. They demonstrate significant improvements using either of their techniques as compared with unadjusted double exponential smoothing.[133] This approach is similar in concept to the Lewandowski special event factors and especially the Carbone and Makridakis special event factors, although their citations do not indicate a familiarity with those approaches. Like these other approaches, this model is aimed at occasions of temporary interruption in trend and level rather than permanent adjustments, and it rests on data from similar historical events rather than estimates supplied from external sources.

Conclusions Regarding the State of the Art

While there are many techniques available for forecasting with discontinuous data, the recent survey by Fred Collopy and J. Scott Armstrong suggests that these techniques are not considered satisfactory. Since the late 1960's techniques have been proposed for identifying discontinuity through the characteristics of the forecasted data using tracking signals. Later techniques were proposed that developed these tracking signals into forecast parameters for adapting to discontinuous situations. Some studies have indicated that these adaptive techniques have been less than successful. Other techniques that have been proposed, e.g., Kalman filters, non-Gaussian methods and complex ARIMA models, have moved away from the simplicity of exponential smoothing models. Often these more complex techniques are explained and justified in the literature based on their mathematical properties rather than on an empirical evaluation of their performance in actual forecasts.

These techniques are designed to identify level shifts as they occur in the observed data. They **react** to the level shift rather than forecast through level shifts, i.e., forecasting with a level shift in the prospective period. Certain ARIMA models - intervention models - may be able to forecast through level shifts, however, doing so depends on the availability of an exogenous variable that contains a forecast of the level shift.

In general, techniques do not exist for taking advantage of knowledge a forecaster may have that a level shift will occur in the future. There are a few exponential smoothing techniques that do forecast through future level shifts - actually, temporary interventions - when similar interventions can be found in the history of the data. These suggest a model for forecasting through a permanent level shift where there is information available about the magnitude of that level shift. The suggested technique would allow the forecaster to adjust the forecast projection by the level of an externally supplied estimate of the level shift. This technique will be further developed in the next chapter.

Estimation of Level Shifts

In parts of this dissertation I have suggested that forecasters may have externally supplied information concerning level shifts. Investigation of such methods is not the objective of this dissertation. Nevertheless, in anticipation of a technique that is proposed in the next chapter, it is necessary to establish the credibility of the assumption that such estimates may exist.

Forecast literature suggests a few techniques for projecting relatively new things into the future. Techniques that can be borrowed from new-product forecasting includes subjective estimates, analogy, consumer-based testing, extrapolation of early sales, and diffusion models.[134] **Extrapolation** generally refers to the use of time series techniques. **Analogy** may refer to purely subjective analogy, i.e., reasoning from one case to another and borrowing information from the source case. Alternatively, it may refer to use of mathematical techniques that rest, in part, on analogy between a new case and old cases, e.g., regression or diffusion models. **Subjective estimates** generally refer to the use of expert or management guesses. **Consumer-based testing** has its analogy in policy making under the guise of pilot projects. Generally, this approach involves trying something out on small scale before going large scale. It can also involve surveying people's interests.

Available cost oriented techniques include learning curve models[135] (which are sometimes included in the class of diffusion models), econometric techniques, and engineering estimates.[136] **Learning curve** models and their relatives require considerable data from the **new** series for fitting,[137] therefore, they are of little value in providing forecasts of anticipated new series or level shifts in old series, although they could be used heuristically for subjective analogical models (there is no literature that suggests that they are used in this manner).

Econometric techniques can be taken to refer to correlation based techniques, that is, regression. Where appropriate, these techniques may provide for adequate estimation of policy impacts. The use of an econometric technique for estimating a policy impact does not guarantee the availability of an econometric technique for forecasting. Cross sectional models may reasonably estimate the impact of a policy, through analogy with other entities that have implemented similar policies, without providing a reliable estimate of the previously existing data generating function. Thus, they may provide only the incremental impact of the policy.

Armstrong, *et. al.*, provide no further explanation of the sort of thing they mean when they say that changes in costs can be estimated through **engineering estimates**, except that they characterize these techniques as "judgmental."[138] However, it appears that they are referring to the use of techniques which focus on costing out actual component cost generating activities, building up the overall cost from these components.[139] In actually performing cost estimation functions for a government program, I frequently find this approach to be the method of choice for costing out proposed changes in governmental services. This approach may not necessarily reflect the precision of engineering studies applied in industrial settings. However, the conceptual structure is similar with a focus on:

- Identifying the actual cost generating activities or units,
- Estimating the quantity of these units,

- o Determining reasonable estimates of cost associated with these units, and
- o Accounting for such factors as:
  - o start-up time,
  - o special start-up costs,
  - o collateral costs,
  - o offsetting savings, and
  - o time frame conversions between accrual of liabilities and cash transactions.

The use of such costing out procedures frequently rests on a combination of use of planned activities (decision maker intentions), market information (current and projected price information), and analogy (information regarding service utilization, etc., borrowed from existing programs). This approach is similar in concept to the idea of decomposition, focussing attention on individual components of cost rather than sophistication of estimation technique. Some forecasting literature supports the view that understanding the process may be more important than use of sophisticated techniques.[140]

This review in not a thorough review of the techniques used to estimate prospective policy shifts. It is intended solely to show that it is credible to conclude that such techniques may exist and may provide reasonably accurate estimates. Actual results arising from the technique proposed in the next chapter may depend on the which techniques are actually used and how reliable their results may be.

Summary

The problem of level shifts is recognized as significant by a large number of forecasting practitioners. Existing techniques include those that identify level shifts (out of control conditions) through tracking signals, those that incorporate tracking signal into the estimation of the α parameter (or other parameters), those that provide other methods for incorporating historically identified level shifts into forecast models, and those that use historical level shifts in analogy for anticipating new level shifts. Three models that might provide some guidance for further development are the Lewandowski additive model, the Carbone-Makridakis multiplicative model and the Carreno-Madinaveitia multiplicative model. Forecast literature supports the view that there may be techniques available to estimate policy changes although they are not serial estimates of whole data series.

CHAPTER 5: A MODIFICATION OF EXPONENTIAL SMOOTHING

In this chapter I will:

- Propose a modification of Holt-Winters-Williams exponential smoothing that might provide a specific solution to this problem.
- Provide a theoretical justification of this solution.
- Specify some limitations of the proposed solution.

Need for a Technique

The techniques discussed in the last chapter allow the forecaster to identify level shifts occurring in historical data and to **react** to them. The reaction may be to refit a model based on a tracking signal that indicates that the model is no longer reliably fit, or it may be to use a tracking signal or another similar statistic to fit a more complex model. In any case it is still a reaction. Even the best technique for reacting to a level shift only follows the data as the data changes. In some spheres it is thought that proactive approaches to future problems are better than reactions, even good reactions.

When a policy decision is made, the data can be expected to change even before it actually changes. Only the undocumented *ad hoc* technique is available for including anticipated changes in the forecast projection unless the anticipated change is simply a repetition of a previous temporary level shift. A technique that allows the forecast to include any anticipated changes may provide a more realistic forecast.

In this chapter I propose a technique that allows the forecaster to **prospectively anticipate** a level shift so that the model does not need to react to it. In other words, the technique includes the level shift in the forecast projection.

Techniques that react to level shifts do not allow the forecaster to take advantage of all of the information that is available to them, particularly information that may have been developed for the purpose of supporting policy decisions that lead to level shifts. The technique proposed in this chapter is particularly designed to take advantage of externally supplied information that can be used to anticipate the effect of a policy change. I anticipate that by taking advantage of this information, a more accurate forecast can be made. The study that is described in a chapter 7 compares the technique proposed in this chapter with some of the simple approaches for reacting to a level shift identified in the previous chapter.

A Proposed Exponential Smoothing Solution

The *ad hoc* method suggests that a policy change may be included in a forecast model by adding the anticipated value of the policy change to the forecast produced into the forecast model. The Lewandowski method suggests a similar addition when a historical fluctuation is expected to repeat. When such additions are lumped onto the forecast produced by the exponential smoothing model they may provide for a more accurate ultimate forecast. However, they do not correct for problems that may arise within the exponential smoothing model itself. Also, these techniques require that the adjustment be added to **each** projected observation produced from the exponential smoothing model. While this may be a suitable approach where a temporary level shift is anticipated over a short period of time, it presents more difficulty where the level shift is long lasting and where multiple level shifts may arise over time.

To address these difficulties, I propose the following modification to the Holt-Williams model:*

---

*Numbering of these formulae continue the same series as the Holt-Winters-Williams formulae.

1. e_t = Error at time t = $X_t - F_t'$
2. F_t' = Adjusted Forecast at time t = $F_t + P_t$
3. F_t = Initial Forecast at time t = $S_{t-1} + B_{t-1}$
4. S_t = Level at time t = $F_t' + \alpha e_t$
5. B_t = Trend at time t = $B_{t-1} + \beta e_t$
6. A_t = Adjustment factor at time t = $P_t - P_{t-1}$
7. P = A periodic estimate of a policy in a vector: $(\ldots,0,0,0,a,b,c,\ldots,n,n,n,\ldots)$ where, a, b, c, ..., n all have the same sign, and $|a| < |b| < |c| < \ldots |n|$.

Other constraints are as with Holt-Williams as described in Appendix II.

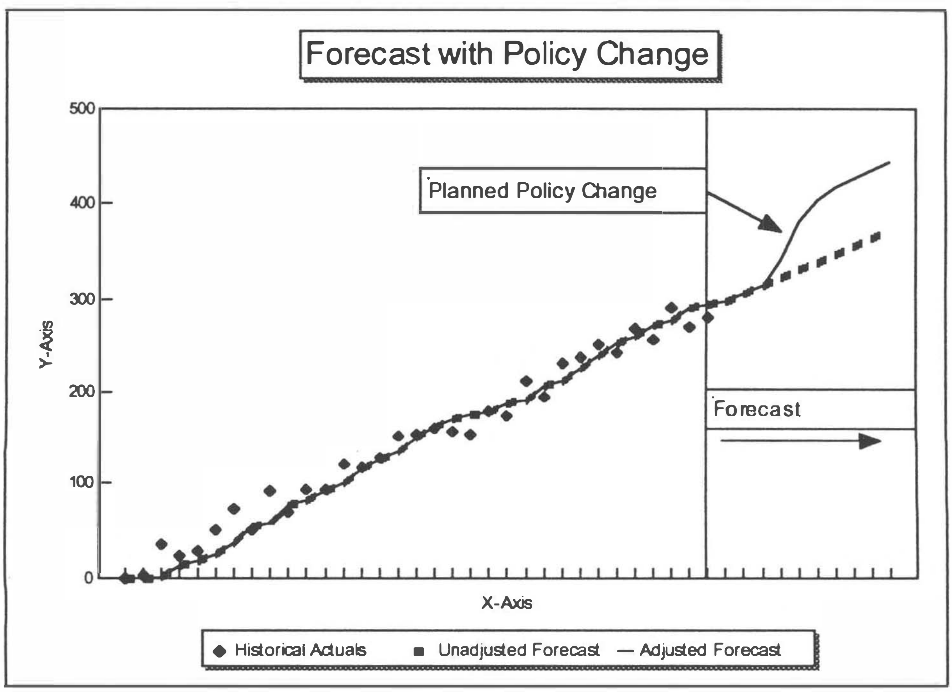

Figure 22

These three comments serve to clarify these formulas somewhat:

- The second restriction on the vector P is for practical considerations only, to avoid use of this level shifting technique in cases where the change under way is actually a trend shift or a seasonality change. When the user has knowledge that a complex level shift is under way, as when a level adjustment is expected for a specific time period only, this restriction can be removed.

- Formula 2. can be restated as:

  $F_{t+m}' = \quad S_t + B_t * \sum_{t+1,\ t+m}(P_t)$

  Then the expression $\sum_{t+1,\ t+m}(P_t)$ can be simplified to:

  $A_{t+m} - A_t$

  This formulation shows that the adjustments included in the forecast model at time t are equal to full level adjustment that would be added in the *ad hoc* adjustment at time t.

- By subscripting P as P_j where j is an index that is associated with various policies and summing appropriately, this formula can be generalized to account for multiple policy adjustments. In this case formula 6 is modified to:

A_t = Adjustment factor at time t = $\sum(P_{jt} - P_{jt-1})$

These formulas modify Holt-Williams; however, Holt-Williams is very similar to SES, Holt, Winters, Holt-Winters, and Holt-Winters-Williams. With appropriate substitution of these other basic models, this policy adjustment model can also modify these other exponential smoothing models.

For ease of reference, I sometimes refer to this technique as **adjusted** Holt-Williams (or **adjusted** Holt-Winters-Williams). It may also be called the **differences** technique because it adjusts a policy-adjustment-free-forecast for the **level** impact of policies by **adding the first differences** of an estimate of the policy change to the unadjusted forecast. Since it adds these first differences directly into the forecast model, it permanently shifts the forecast upwards (or downwards in the case of negative differences). Thus, the adjusted forecast permanently includes the policy change.

Making an adjustment in this manner results in the following expected consequences:

1. When the policy occurs at approximately the time expected in approximately the size expected, the forecast is ready for the change and requires no further adjustment. The forecast does not develop large errors at the time of the policy change, so it is not necessary to correct any of the forecast parameters. Problems associated with large forecast errors do not materialize.

2. When the technique is used to empirically fit historical level changes that are known to have occurred, exponential smoothing parameters can be better fit to the remainder of the series.*

3. When the policy fails to occur at approximately the time expected in approximately the size expected, the forecast error increases. This increased error alerts the forecaster to the fact that the policy change has

*This usage should be limited to cases where the forecaster knows that there has been a level shift and knows why the level shift occurred. If a data series has periodic level shifts that are unexplained, the forecast parameter needs to reflect this so that future occurrences arising for the same unexplained reasons will not be ignored.

not occurred, leading him to follow up with the people who are responsible for the policy implementation in order to determine what sort of change is required in the adjustment. It also alerts management of the implementation failure, leading to management use of forecast information.

4. When the forecast is generated through the proposed formulas, the whole forecast including the policy change component is generated without additional manipulation. So when results are tabulated, no adjustments are required. Likewise, when the forecast is used as input to other more complex forecasts, the impact is automatically carried forward to those forecasts. This is particularly beneficial when the forecast is generated in an automated environment, where other approaches may require manual intervention. It can also be very beneficial where various summation periods are required for different reporting purposes.

5. When, prior to the change in the policy, the forecaster learns of revised assumptions about timing or magnitude, the forecast assumptions can be adjusted by revising the same assumptions within the forecast. For example, the forecaster can shift the policy change

forwards or backwards in time, or can increase or decrease the magnitude of the adjustment. Consequences for all future periods and all summation periods are automatically adjusted. Such adjustments can also be made when, empirically, it is demonstrated that the policy change has impacts other than those prospectively anticipated.

The empirical research in this dissertation examines the first (the larger study) and second (the smaller study) of these expected consequences and finds incidentally relevant information for the third. The fourth and fifth consequences are logical in nature and do not require additional research to demonstrate their accuracy.

Theoretical Rationale

The proposed technique augments the use of **information** in forecasting. Quantitative forecasting involves efforts to extract information from sample data (the historical period) that can be effectively generalized to the out-of-sample data (extrapolated into the future). A difficulty with sophisticated techniques is that they sometimes confuse random or unexplained variation for information.[141] The proposed technique deals with the problem of information about the future in a different way. It **decomposes** the forecasting problem into separate problems of forecasting the underlying process and forecasting a prospective change in the process, allows for **quasi-judgmental forecasts** of the prospective change relying on policy maker **intentions**, and finally **reintegrates** the results into a complete forecast.

Decomposition for Efficient Use of Information

J. Scott Armstrong argues that it is particularly helpful to decompose a problem to help analyze it.[142] Armstrong provides numerous citations that show that decomposition improves forecasting. He argues:

> Decomposition has a number of advantages. It allows the forecast to use information in a more efficient manner. It helps to spread the risk; errors in one part of the problem may be offset by errors in another part. It allows the researcher to split the problem among different members of a research team. It makes it possible for expert advice to be obtained on each part. Finally, it permits the use of different methods on different parts of the problem.[143]

This argument cites several specific advantages of decomposition that are directly related to the proposed technique. These include:

- The ability to split the problem up among different members of the research team. The proposed technique allows the forecaster to benefit from analyses completed by staff who have developed specific policy cost analyses.

- The ability to obtain expert advice on each part. The forecaster can separately seek out information about each part of the forecast (at least prospectively) and use the advantage of that information in making a forecast.

- The ability to use different methods for different parts of the problem. The proposed technique is specifically oriented to using different methods in making the forecast, while integrating the results into the most effective combined forecast.

By breaking down the initial forecasting problem into forecasting of the underlying process and estimating the policy change, the proposed technique allows the forecaster to use appropriate techniques and information for each component of the forecasting problem, rather than forcing the problem to fit the technique.

Judgmental Adjustments

The use of externally supplied estimates of policy changes in the proposed technique is very similar to the use of judgmental forecasting based on policy maker intentions. Armstrong cites numerous studies that show intentions provide good predictions where matters are important.[144]

Practicing forecasters commonly advocate adjusting forecasts to account for externally available information.[145] Such accounting is also the underlying purpose of decomposition techniques in general. For example, forecasters frequently adjust data to take into account trading days[146] precisely because they anticipate that this external factor will lead to predictable variation in the data being forecasted.

Nevertheless, forecasting studies generally show that judgmental adjustments of forecasts do not improve forecast quality.[147] Still forecasters persist in believing that use of knowledge about the data series, particularly about future states of the data series, helps in forecasting the data series. Vollmann, Berry, and Whybark offer a typical discussion where they suggest that when a forecaster has knowledge of external information he must chose between adjusting the forecast and adjusting the forecast model.[148]

Don Miller has conjectured that it is more effective to use judgmental adjustments when the forecaster or other expert commits to the adjustment before the forecast is made.[149] The rationale for this conjecture is as follows: When judgmental adjustments are committed to before the forecast is made, they relate to the information underlying the adjustment itself. This use restricts the role of judgmental adjustments to that of including more information in the forecast and may improve the forecast. However, when judgmental adjustments are made as the last stage of the forecasting, they are used to force the overall forecast to the forecaster's subjective estimates. The second use substitutes a subjective forecast for an extrapolation forecast with an associated loss of accuracy.

The proposed technique may bridge between these competing views on the advisability of including judgmental adjustments in forecasts. It provides for clear quantified forecast adjustments that rely on prior existence of quantitative estimates of the factors that are expected to lead to predictable variation in the data series. By relying on the prior existence of quantitative estimates, the technique also follows the normative logic of Miller's conjecture that the commitment should be made before the unadjusted forecast is known.

Integration

The math that allows for integration of the component forecasts into one model is the heart of the proposed technique. For the sort of problem for which this technique has been developed some future event irreversibly combines component data series into a single indistinguishable one. Before they merge, one can forecast them separately and add the results. Exponential smoothing can be used for the series that has a reasonably long data history, but the other component is not in that history so it must be estimated in some other way. Neither component forecast alone accounts for the future expectation. Together, they make a more reasonable forecast.[150] After the data merges, there is only one series to forecast. So, the first differences of the policy change estimate are used to permanently adjust the level of the exponential smoothing forecast model, thereby aggregating the component forecasts into an integrated whole just as the data series itself will be integrated.

This process is a formalization of the *ad hoc* method that forecasters might otherwise use when faced with policy changes. Integrating forecasts that arise from various sources may be considered an improvement over these separate forecasts.[151] When the policy change is known to occur far in the future, the forecaster could simply add the consequences of the policy change to a forecast derived from a statistical forecast model (the *ad hoc* technique). Ignoring any dynamic aspects of the forecast, that approach would have the same result as the proposed technique. The summation of the periodic first differences used in the proposed technique gets the forecasted data series to the same level as would be attained through the *ad hoc* technique.

The difference between the two techniques is not the estimate that it produces, but the consequences of updating with actual data. With the *ad hoc* technique, the data associated with the policy change leads to large forecast errors which throw the statistical forecast model into confusion while data associated with failure to experience the policy change does nothing at all. Neither of these assist either the forecaster or the manager who uses the forecast. In the proposed **integrated** technique, the occurrence of the policy change minimally confuses the forecast (the confusion is less as the estimate of the impact of the policy change is better), while failure to experience the policy change throws the statistical forecast model into confusion. These consequences help both the forecaster and the manager who uses the forecast regardless of whether the policy change occurs or not. The benefit arises because the integrated technique can be expected to have much smaller errors within the statistical forecast model than the *ad hoc* method.

Limiting Factors For the Technique

There are factors that may interfere with anticipated benefits of the proposed technique:

Independent Components

In part the anticipated benefits arise from the assumption that the two series that are to be merged are independent. If they are not independent, the combination of the series is not best accomplished through simple addition. A lack of independence might arise when the main effect of a policy change is to change the level of a data series, but an additional effect is to also change the trend of that series. It is likely that this assumption of independence will be violated with actual policy changes. An important factor to examine is whether the technique is robust to violations of this assumption. In the empirical study, two simulated policy changes reflect a correct estimate (scenarios 1 and 8) and two include both a level change and a trend change (scenarios 2 and 9). The purpose of including these scenarios is to look at one sort of violation of the assumption of independence.

Accurate Component Estimates

It is assumed that the two component series are estimated through techniques that are reasonably reliable. There is a higher risk that this will not be true for the estimate of the policy change A policy estimate may be substantially inaccurate in several ways:

- It may substantially over or under estimate the impact of the policy. Where it underestimates the impact of the policy it still should result in reducing the overall bias of the forecast, that is, it should reduce the size of the error that arises from using a model that contains no information at all about the policy by some portion of the difference between the unadjusted model and the true data generating process. When it overestimate the impact of the policy, there is no guarantee that the policy-adjusted model will be more accurate than the unadjusted model. In the study there are two scenarios that simulate underestimated policy changes (scenarios 3 and 10) and two scenarios that simulate overestimated policy changes (scenarios 4 and 11).

- It may incorrectly estimate the nature of the impact, that is, it may adjust the forecast to reflect a new level and minimal impact on trend when just the opposite occurs. In the study two scenarios simulate a trend change instead of a level change (scenarios 5 and 12), one simulates no change at all (scenario 6) and one simulates a variance change instead of a level change (scenario 7).

- It may place the policy at the wrong point in time, either before or after the actual impact occurs. This possibility is not examined in the empirical study.

- It may predict the opposite of the actual impact, that is, increase when decrease occurs or vice versa. This possibility is not examined in the empirical study.

- It may incorrectly account for the timing and length of the ramp of the level shift, either by not accounting for it at all (assuming a step) when, in fact, phase up, or by assuming too short or too long a ramp. This possibility is not examined in the empirical study.

Trend, Seasonality, and Variance Shifts

The technique only addresses the case of a level shift. Trend, seasonality, and variance shifts are not adjusted by this technique. It is likely that a variation of the technique would be available for trend shifts.* I am not aware of a modification of the technique to adjust the forecast for seasonality or variance shifts.

Competing Shifts

The proposed technique may lead the forecaster to overlook a trend shift, seasonality shift, or other unexplained level shift that occurs at the time of the expected level shift. The large errors that may have caused the forecast to adjust for either of these alternative types of changes will be lost because the forecast is already adjusted to reduce the size of these errors.

---

*With the following modifications this technique can also adjust for trend shifts:

3.5	B_t	=	Trend at time t	=	$B_{t-1} + \gamma e_t + K_t$
8.5	K_t	=	Trend Adjustment Factor	=	$P_t + P_{t-2} - 2P_{t-1}$

This places the second differences of the policy estimate into the trend. It is likely that the forecaster would have to make a judgement that the chief problem he/she has with a particular policy is its impact on trend or its impact on a policy and, then select which to modify. Modifying both may be too aggressive.

The forecaster is not relieved of this problem because he fails to anticipate a policy change. If the policy change is not included in the model through the proposed technique and the forecast begins to perform poorly, the forecaster will have to investigate the phenomena in order to decide what sort of remedial action is required. Upon finding an expected policy change it will be reasonable for the forecaster to assume that a change occurring at the time it is expected and in the order of magnitude expected is the change that is expected. While this will occasionally result in erroneous results, it is less likely than the opposite assumption. Thus, the same result occurs whether the policy change is accounted for in the model or not. In either case, additional experience **may** correct the error after more updates; however, there is no guarantee it will.

Shifts that are Too Small for Significance

The changes that this technique imports into the forecast may be so small as to be of little consequence to the forecast performance. This may be the case particularly when the data series that is modified by the policy change is subject to wide variation in the first place, or when the estimate of the prospective policy change is subject to wide variation. So, the technique may import a risk of error without significantly improving the forecast. This is undoubtedly true for some applications. A forecast that suggests an expenditure of $500 million a year may be insignificantly impacted by a policy change that adjusts it by $500,000 a year, since the confidence interval around the $500 million forecast may be much larger than that. Yet, ignoring the change would appear to bias the forecast, assuming it was previously unbiased, since the estimated impact of the program change implies that the expenditure will be greater than the previously estimated central tendency. So, as insignificant as the item is for the effectiveness of the forecasting model, ignoring it creates an underestimate.

More importantly, the forecaster should not overlook **practical** forecasting consequences when deciding technical factors. The acquisition of the funding for the policy that cost $500,000 may not be inconsequential. By systematically ignoring all such policy changes, it may well be that a program creates unnecessary budgetary turmoil for itself by first seeking budgetary approval for policy impacts, then later seeking funds to support slight unexpected growth in the forecast, when the growth in the forecast is merely the forecast model's recognition of the policy that was previously ignored.

Explaining the Forecast

The use of this technique may **reduce** the forecaster's ability to explain the forecast model to the forecast user. When a forecast is adjusted through the *ad hoc* method, the forecaster can tell the user the exact size of the policy adjustment. When the adjustment is incorporated within the forecast through the proposed technique and entered as a component into a more complex forecast, the exact size of the policy adjustment is no longer easy to state.

Summary

When forecasters have information about anticipated policy changes, it may be effective to include that information in the statistical forecast model. A technique is defined to allow for a permanent inclusion of such information. This technique makes sense because it simply allows the forecaster to include information in the statistical forecast model. This is a natural extension of the decomposition approach to forecasting.

CHAPTER 6: VALIDATING FORECAST TECHNIQUES

In this chapter I will:

- Identify the criteria that are available for comparing forecast techniques.
- Specify two major hypotheses (5 constituent hypotheses) that are examined through empirical research which is discussed in the final three chapters of this dissertation.
- Identify the various types of forecast evaluations.
- Discuss the generally accepted approach to evaluating forecast accuracy.
- Identify the range of statistics that are available for evaluating forecast accuracy.

Forecast Criteria

To determine whether forecasts made using the proposed technique are better, it is important to consider the meaning of **better**. This term has been used with several meanings among forecasters. Roughly, these include:

- Better fit in the sample period.[152]
- More accurate in the forecast period.[153]
- Lack of bias.[154]
- Less expensive to use.[155]

- Easier to use.[156]
- Easier to understand.[157]
- Not having systematic errors.[158]
- Containing more information.[159]
- Providing more useful information to a forecast user.[160]

Also, I proposed two criteria in chapter 3, accuracy and reliability. These criteria, beginning with the two I proposed, are discussed below.

Accuracy

There is considerable dispute in the literature regarding which statistic, if any, reliably measures the relative accuracy of various forecasts.[161] In addition, studies indicate that forecast confidence intervals calculated by standard statistical formulas are unduly narrow[162] which suggests that statistical comparisons between forecasts may not be reliable. Nevertheless, severely inaccurate forecasts would seem to be pointless. So, forecast accuracy is considered a criterion of forecast model adequacy. Later in this chapter methods for evaluating accuracy are reviewed more thoroughly.

Reliability

Forecasting literature does not address the issue of reliability as I have raised it. I use **reliability** to refer to a lack of update driven fluctuation. Forecasts of more than one period ahead are of little value when they change significantly with every update. If the current forecast and the last one are both the best made at the time, but the current one is 10% more (or less) than the last one, and if the forecast fluctuates this much with every update, how does the user know which one is right? In the empirical analysis a statistic is included to look at this issue.

Better Fit

It appears that the **better fit in the sample period** criterion is a proxy for the **more accurate in the forecast period** criterion. However, forecast literature does not support the view that the one implies the other.[163] In this study accuracy will be measured more directly by looking at the actual consequences of actual forecasts. Better fit will not be a criterion of forecast model adequacy.

Lack of Bias

In estimation, bias arises when the expected value of an estimator is not equal to the parameter it is used to estimate.[164] It is, therefore, related to accuracy. In this study, concerns of bias will be subsumed under concerns of accuracy. Nevertheless, Mean Deviation (a measure of bias) will be considered in fitting of forecast models.

Less Expensive to Use

As has been cited elsewhere in this study (page 30), it is commonly believe that exponential smoothing is a relatively inexpensive forecasting technique. This view rests on the fact that exponential smoothing calls on relatively little data and can be taught to staff who have little statistical or other high cost analytic skills. It also requires only a moderate amount of computer time and calls on formulas that are available in numerous computer programs and which formulas are fairly easy to recreate when off-the-shelf software is not available or desired. In the usual case, interpretation of the results of exponential smoothing models is not considered difficult. I have not been able identify research that demonstrates these assertions, but I assume from personal experience that they are accurate. Another element of forecast cost is the cost of wrong decisions due to forecast inaccuracy; however, forecast accuracy is a separate criterion for this study.

If exponential smoothing is an inexpensive forecasting technique, and if the proposed technique uses only information that is on hand requiring only a small amount of effort to put in a usable format, the proposed technique should remain inexpensive. The question whether the proposed technique is less expensive than unadjusted exponential smoothing rests on whether the value gained through improved accuracy or other value added is greater than the increased effort required for application of the technique. I am not able to operationalize this question and do not propose evaluating it question in this study.

Easier to Use

It is widely held that exponential smoothing is a relatively easy to use forecasting technique. Reasons for this view are not widely discussed but, I assume, are associated with the relatively simple math and the relative ease of interpreting output in the usual case. The proposed technique is a relatively simple extension of exponential smoothing, requiring primarily that the forecaster grasp the mathematical operationalization of a first difference. Consequently, if exponential smoothing is relatively easy to use, so too is the proposed technique.

Also, the proposed technique provides for several efficiencies compared with its primary alternative, the *ad hoc* approach. These are discussed beginning on page 97.

One of the more important of these efficiencies is that, as compared with the *ad hoc* technique, this technique has the advantage that it is "more seamless," i.e., it allows for the computerized forecast model to contain a level shift. A forecast practitioner who wants to accomplish the same results without the proposed technique makes a two stage forecast, first he uses a computerized forecast model to produce an initial forecast, then he takes the results of the model and manually adds the *ad hoc* adjustment. With each forecast update, the same two stage adjustment must be made. Also, when more than one forecast horizon is reported, the *ad hoc* adjustment must be included for each reported horizon. The proposed technique minimizes the number of adjustments made by permanently including the adjustment in the forecast level and contains the adjustments in the computerized forecast model, thereby eliminating manual adjustments.

Easier to Understand

An easy to understand inaccurate forecast is of little value. The ease of understanding criterion is, in the first place, related to forecasts that are anticipated to be accurate. At this point, the accuracy of the proposed technique has not been evaluated. Thus, evaluation of the ease of understanding criterion would be premature unless it can be included in the study of accuracy with minimal additional impact.

I anticipate that, in fact, the proposed technique will be somewhat less easy for the user to understand than is *ad hoc* method. Under the *ad hoc* method, the user can see the forecast from the statistical forecast model, see the policy impact, and see the combination of these. In the proposed technique, the policy change is mysteriously absorbed into the statistical forecast model.

In the discussion above it is suggested that this *ad hoc* method sometimes results in significant inaccuracies. It is, therefore, important to separate the issue of accuracy from the issue of understanding. If the study of accuracy demonstrates significant gains, the issue of understanding may become more significant for a separate study.

There are two issues of ease of understanding:
(1) Whether the technique as described in this study can be understood by forecasters and end users with relative ease.
(2) If not, whether an alternative description can be articulated that will make the technique accessible to

forecasters and end users. A negative finding on the first issue does not settle the matter.

Absence of Systematic Errors

In his discussion of this issue, H. O. Stekler is clearly interested in the issues of bias and accuracy.[165] These are issues previously discussed above under separate topical headings. In general, a systematic error is an error that indicates the failure to take into account an important explanatory variable. In causal or econometric type forecasting, this would generally mean that the model is missing an important explanatory variable. In Holt-Winters-Williams models it may indicate a need to analyze the decomposition of the data into the series that are being forecast. For example a periodic up and down cycle in monthly data may indicate that the data is inherently weekly in nature with a special weekly end date, e.g., Fridays, such that in months that have five Fridays (roughly one a quarter) the series bumps up, in other months it bumps down. This sort of decomposition analysis is another approach to dealing with explanatory variables.

The proposed technique is developed specifically to remove certain sorts of systematic error, it is designed to remove the serial correlation that arises over a period of time during which an exponential smoothing forecast model is behind the curve on a level shifting policy change. Consequently, the evaluation of the relative accuracy of the technique is also an implicit evaluation of its effectiveness in removing systematic error from the forecast.

Containing More Information

The proposed technique is specifically designed to incorporate information known by the forecaster into the statistical forecast model. With respect to the overall forecast model, including non-statistical equations, the forecast made in advance of the date of the policy change using the proposed technique does not contain more information that a forecast using the *ad hoc* method. However, it is proposed that the statistical model in the proposed technique does contain more information than the **statistical model** used in the *ad hoc* technique. The difference in the statistical models is that the proposed technique uses the anticipation of a change to adjust the number that is to be compared with the forecast error. As the policy change unfolds, the **errors** computed in the alternative statistical forecast models will be different. In one case the statistical error reflects the anticipation of a change, in the other it does not. In the proposed technique the statistical error indicates a need to adjust (and adjusts to the degree that the parameters allow) in the case that the policy change fails to occur. It continues without significant adjustment if the policy change occurs as expected. In the *ad hoc* model, the reverse occurs.

While this difference is solely a difference in where the information is stored before the policy change occurs (within or outside of the statistical model), it becomes a **difference in information** as the model is updated **while policy change is going into effect**. The difference is found in the error term of the statistical forecast model. Assume that the estimate of the policy change is reasonably accurate. In this case, the user of the proposed technique continues to have available a statistical forecast that is not affected by abnormal errors. The user of the *ad hoc* technique, on the other hand, has a statistical forecast that is not working because the statistical forecast model is affected by abnormal errors. Once the policy change is in effect and in the historical period of the data, the *ad hoc* technique contains the information about the policy change within the statistical forecast model in the α parameter where it is at risk of confusing other unexplained variation with policy changes while the proposed technique contains the information within a component of the model that is targeted specifically to the specific policy change and has no risk of confusing unexplained variation with the policy change.

Moving to the assumption that the policy change does not take place, the user of the proposed technique is free to change the forecast model to exclude an unexperienced adjustment, and would be ill advised to do otherwise. In that case the forecast would contain the same information as that used in the *ad hoc* method.

Consequently, it appears that the proposed technique produces a forecast model that contains at least as much information as the alternative technique, that it includes more information in the statistical forecast model under certain circumstances. While the discussion above asserts such benefits based on the proposed rationale, an empirical evaluation of the accuracy of the proposed technique also constitutes the evaluation of this information strategy.

Providing More Useful Information to the User

It is possible that the proposed technique can be further developed for use in evaluating or testing the occurrence of the anticipated level shift, thereby providing useful information to the user. Such testing would consist of considering a statistic that would behave differently between models that have both anticipated and actual policy changes and models that have anticipated policy changes that do not materialize. Development and evaluation of this statistic is beyond the scope of this dissertation

Summary of Criteria

In this study the proposed technique is evaluated to determine whether it meets the criterion that it is more accurate and reliable than some comparative forecast in the forecast period.

Hypotheses to Examine

In this dissertation I propose that a certain sort of problem, level shifting, arises with serial data that may be forecasted. In the ensuing discussion I examine various approaches to forecasting data that experiences level shifts. Considerable attention is given to data that has level shifts during the forecast horizon. I discuss various exponential smoothing techniques that may be used to forecast through periods of level shifting data. I define a specific forecasting technique that incorporates independently developed estimates of policy changes into exponential smoothing models. I discuss several sorts of problems that may arise with the proposed technique. This discussion leads to the following hypotheses.

Accuracy when Forecasting Through Periods of Policy Change

There should be a significant variation in performance of various exponential smoothing techniques in forecasting through periods of level shifting policy changes. Some methods perform should better than others. Where anticipated policy changes materialize, Holt-Winters, Holt-Winters-Williams, adaptive Holt-Winters-Williams, and autocorrelation corrected Holt-Winters-Williams* used alone or in combination with the *ad hoc* method (herein, the alternative techniques)** should not perform as well as the proposed technique when the forecaster possesses reasonably accurate information about the prospective level shifting policy change. Where anticipated policy changes do not materialize, only the models that completely ignore the anticipated policy change should perform as well as or

---

*These terms are used here to reflect a class of models: SES, Holt-Williams, and multiplicative Holt-Winters-Williams. Thus, it is asserted that adjusted-SES is more effective than SES or adaptive SES, adjusted-Holt-Williams is more effective than Holt-Williams or adaptive Holt-Williams, and adjusted-Holt-Winters-Williams is more effective than Holt-Winters-Williams or adaptive Holt-Winters-Williams.

**These techniques represent a reasonably broad range of those techniques that preserve the simplicity advantage of exponential smoothing.

accurate information about the prospective level shifting policy change. Where anticipated policy changes do not materialize, only the models that completely ignore the anticipated policy change should perform as well as or better than the proposed method, that is, the alternative methods may perform better than the proposed method only when the *ad hoc* adjustment is not used **and** the anticipated policy change does not materialize. These expectations lead to these hypotheses:

> HYPOTHESIS 1a: The alternative techniques and the proposed technique are not equally accurate in forecasting through periods where policy shifts are anticipated.
>
> HYPOTHESIS 1b: The proposed technique is more accurate than the alternative techniques when used to forecast through periods where policy shifts are anticipated and such policy changes materialize.
>
> HYPOTHESIS 1c: The proposed technique is more accurate than the subset of the alternative techniques that include use of the *ad hoc* method when used to forecast through periods where policy shifts are anticipated and such policy changes fail to materialize.

Forecasting with Data that has Historical Policy Changes

There should be a significant variation in the performance of forecasting models that are fit across periods of level shifting data. Fitting sample period data series that have **explained** level shifts (policy changes) in the sample period through the proposed technique should produce more accurate forecasts than fitting such historical

HYPOTHESIS 2a: The alternative techniques and the proposed technique are not equally accurate when used to fit data that has had a level shift in the historical period.

HYPOTHESIS 2b: The proposed technique is more accurate than the alternative techniques when used to fit data that has had a level shift in the historical period.

Evaluating Forecast Methods

The hypotheses address whether, and under what circumstances, forecasting with the proposed technique provide better results than would occur in its absences. What constitutes a reasonable study of such a question? Forecasting literature suggests the following research designs related to forecast techniques:

1. Examination of the mathematical validity of a proposed technique.[166]

2. Examination of a mathematical model through the use of simulated data.[167]

3. Forecast competitions, generally involving the use of a hand full of techniques applied to the same data series.[168] The most significant of these over the recent past have included the M-Competition and the M2-Competition, both of which involved comparison *of major*

forecasting techniques by highly recognized members of the forecasting community.[169]

4. Review of the use of a technique on one or more data series known by the researcher.[170]

5. Examination of cross sectional data, e.g., looking at forecasting accuracy for techniques that are frequently used by practitioners.[171]

6. Reexamination of published data and results.[172]

7. Survey research, e.g., into forecaster satisfaction with methods.[173]

8. Forecast technique clarification through the development of protocols for use of a technique.[174]

9. Study of psycho-social elements of forecasting.[175]

No single study could attempt to pursue all these approaches to forecast adequacy. Instead, an actual study should be comparable to one or several of these approaches and should be designed to resolve specific questions raised through specific hypotheses. The study design discussed in the next chapter reflects several of the types of studies cited above. It includes a forecast competition (see bullet 3), using data that is known to the researcher (bullet 4), and simulated policy changes (bullet 2). It is used to determine under what circumstances the proposed technique might be effective (bullet 8). Other designs mentioned above are not included in this study.

N-Period Ahead Evaluations

Two aspects of an forecast update are the **trace** of each forecast update and the **n-period ahead point** of repeated forecasts.[176] The trace is the forecast for periods t+m through t+m+l where t is a time index, m is the number of time periods before the period of interest, and l is the number of time periods in the period of interest. The trace, therefore, is a vector of forecasts:

$$F_{t+m},\ F_{t+m+1},\ F_{t+m+2},\ .\ .\ .\ ,\ F_{t+m+l}$$

The n-period ahead point of repeated forecast updates is the forecast at the observation at t_j+n, where t is the index of the last actual observation and updates by an increment of 1 with each addition of 1 observation to the history of the data, j is the index of the updates, and n is the number of periods from t to the observation measured.

The n-period ahead point of repeated forecast updates moves to a later point in time by the number of additional actual observations added to the history with each update. There is one point observation from each jth update and it is located one period later in time. These are further demonstrated in the following table:

Table 1 Trace and N-Periods

	Period						
Update	1	2	3	4	5	6	7
1	**F1,1**	F1,2	F1,3	F1,4	**F1,5**	F1,6	F1,7
2	A2,0	**F2,1**	F2,2	F2,3	F2,4	**F2,5**	F2,6
3	A3,-1	A3,0	**F3,1**	F3,2	F3,3	F3,4	**F3,5**
4	A4,-2	A4,-1	A4,0	**F4,1**	F4,2	F4,3	F4,4
5	A5,-3	A5,-2	A5,-1	A5,0	**F5,1**	F5,2	F5,3
6	A6,-4	A6,-3	A6,-2	A6,-1	A6,0	**F6,1**	F6,2
7	A7,-5	A7,-4	A7,-3	A7,-2	A7,-1	A7,0	**F7,1**

F = Forecast, A = Actual
First Subscript (i) = Update (row)
Second Subscript (j) = Periods before (negative) or after (positive) the current observation (Ai,0 = Current Period)
i = Subscript of the **Trace** (row)
j = Subscript of the **N-Period Ahead Forecast** (diagonal from left top to right bottom)

The most commonly accepted forecast evaluation design is to compare forecasts based on the n-period ahead point observation.[177] It is acknowledged that evaluation of n-period ahead point accuracy may not fully evaluate trace accuracy. This limitation arises because of the difference between two of the major factors of a forecast, slope and level.

The difficulty of evaluating only the n-period ahead observation is that two different forecasts may not be distinguishable at n-periods ahead. In evaluating a forecast at n-periods ahead, one is evaluating the **level** of the forecast **at that observation point**. If two forecasts have different levels at the end of the historical period of a data series and also have different slopes they may **intersect** at a certain point in the future, thereby having no difference in level at that particular point. If this point is near the n-period ahead observations, it may be difficult to distinguish between these two forecasts at n-periods ahead; they could be confused for two forecasts that had the level at the end of the historical period and the same slope thereafter, i.e., identical forecasts. It is reasonable to assume that if the trace of one of these forecasts is similar to the trace of the actual data series as it unfolds, this forecast is the better forecast. However, the n-period ahead evaluation may not reveal this difference.

Unfortunately, observations of a forecast at various periods ahead are not independent of each other, so the errors from the actual data series members are highly correlated. This correlation invalidates the use of such statistical procedures as ANOVA, t-tests, and regression coefficients when working with forecast traces. Consequently, generally accepted forecast evaluation designs do not provide for the evaluation of the trace of comparative forecasts.[178]

Instead of comparing traces, the practice is to evaluate forecasts at various horizons (n-periods ahead) to determine which technique is more effective for forecasts **at each horizon**. For example, this approach is the technique used in a very widely recognized study known as the M-Competition and a recent follow on to that study known as the M2-Competition.[179] Although this view is not commonly articulated, it can be argued that a forecast technique that produces effective forecasts **at various horizons** reasonably must have a trace that is similar to the trace of the actual data series.

Statistical Evaluation of Forecasts*

Various statistics have been proposed for evaluating forecast accuracy. In general these are descriptive statistics. The general practice with forecast competitions is to demonstrate numerous tables of descriptive statistics

---

***Notation in formulae presented in this section may be altered from the original for consistency within this discussion.**

and to discuss these qualitatively. Statistics that have been proposed for, or used in, forecast evaluation are reviewed below.

Descriptive Statistics

Following is a review of various **descriptive** statistics proposed for forecast evaluation.

M-Competitions

Two common forecast statistics are Mean Squared Error (MSE) and Mean Absolute Percent Error (MAPE).[180] These statistics have been presented in both the M-Competition and the M2-Competition and are commonly cited in other forecast literature. Other statistics presented in the M-Competition include Median MAPE (MdMAPE), Median Absolute Percent Error MdAPE), Average Rank (between various methods used) and comparative performance to Naive 1 (no change) or Naive 2 (seasonally adjusted Naive 1).[181] The comparative performance statistics simply show the number of times that Naive 1 or Naive 2 method out performs the alternative technique. Other comparative performance statistics are also presented. Makridakis, et. al. do not show why these are preferable statistics.

The M2-Competition presents similar statistics include average MAPE, ranking of all series, percent of time better than Naive 2, difference of MAPE between Naive 2 and other methods, Mean Percent Error (MPE), median MAPE divided by MAPE of Naive 1 and median MAPE divided by MAPE of Naive 2.[182] The authors of this study remark, "This paper has provided many tables (some complain *too* many) and used several accuracy measures to report and compare results. [Italics in original]"[183]

Spyros Makridakis and Michele Hibon have published another similar study in which they reported similar statistics as well as reporting Theil's U-Coefficients.[184] Theil's U-Coefficient is a statistic that compares the actual one step ahead forecasts produced in a forecast model with the forecast sometimes known as Naive 1, a no change forecast. A U-Coefficient less than 1 indicates that the proposed technique is an improvement over Naive 1. However, the statistic's distribution is unknown so statistical significance cannot be established.[185]

Robert McLaughlin

Robert McLaughlin[186] defines a statistic that he calls the standardized realization percent (SR). He recommends this statistic for the purpose of easy communication with management. The standardized realization percent is a variant of MAPE.

Benito Flores

Benito Flores provides a general review of forecast statistics and identifies defines a large number including[187*] Mean Error (ME), Mean Absolute Deviation (MAD), Root Mean Squared Error (RMSE). These are absolute statistics, that is, their magnitude is dependent on the magnitude of the original data series. MSE and RMSE are sometimes used as loss functions in optimizing forecasts, but they place more weight on larger errors. Under some circumstances this might not be desirable.

He also defines Percent Error (PE), Mean Percent Error[**] (MPE), and a symmetrically adjusted MAPE (SMAPE) statistic as follows.[188] He points out that MAPE is biased in favor of underestimation because of a bias in the ratio calculation when dividing by the actual. For example with

---

[*]Flores definitions of MSE and MAPE have been discussed above.

[**]This formula is a correction of Flores published number where he calculates the "* 100" against the average PE_i. He has already included this component of the standard percentage formula in calculating PE_i.

two equally large errors, e.g., predicted 50 and actual 100 versus predicted 100 and actual 50, the first will be:

$$(|50-100|/100)*100 = 50\%,$$

While the second will be:

$$(|100-50|/50)*100 = 100\%$$

These statistics (MPE, MAPE, and SMAPE) are dimensionless in that they are divided by the data series or some other quantity that is in the same order of magnitude as the data series. Because of this dimensionless quality, comparison between forecasts and communication with management is enhanced.

Armstrong and Collopy

In a recent article Armstrong and Collopy define a large number of error measures, including the following which they recommend or are used in the calculation of those they recommend:[189] Median APE (MdAPE), Relative Absolute Error (RAE), Winsorized RAE (WRAE), Geometric Winsorized RAE (GMWRAE), Median RAE (MdRAE). They recommend the use of GMWRAE for calibrating parameters, MdRAE for selecting among a small number of forecasts, and MdAPE for selecting among a larger number of forecasts. They recommend against RMSE for use in generalizing about the accuracy of forecasts because of its low reliability. The recommend against the use of MAPE because of its bias in favor of low forecasts.

Although their discussion does not recommend them they also define other statistics including Cumulative RAE, Geometric Cumulative RAE, Median Cumulative RAE, an aggregate RMSE for multiple series, a generalized Theil's U which they call Theil's U2, a Geometric U2, Percent Better, and a statistic that summarizes six of these other statistics which they label Consensus Rank. They do not define inferential tests of statistical significance for various error measures.

Patrick A Thompson

Patrick A. Thompson[190] proposes the use of the log mean squared error ratio for comparison of forecasts. This statistic is defined as follows:

> Let m_{ij} denotes the mean squared forecast error of techniques j on series i. For this series, define the log mean squared error ratio as $lmr_{it} = \log(m_{i0}/m_{ij})$, where m_{i0} is the mean [squared] forecast error of some benchmark technique. Computed with the benchmark MSE in the numerator, a positive LMR indicates that technique j had a smaller forecast MSE on this series than the benchmark.[191]

Thompson goes on to argue that is permissible to average LMR across series. The benchmark that Thompson uses is Naive 2.

Robert Fildes

Several of the statistics defined by Armstrong and Collopy involve the use of Geometric calibrations to summarize across multiple series. Robert Fildes defines a statistic he calls the Geometric Root Mean Squared Error across time periods which he then uses in a ratio to compare between two different forecast techniques and a more generalized statistic called The Relative Geometric Root Mean Squared Error ratio across time periods (TRGRMSE).[192] Fildes also defines an array of other statistics. Ultimately he concludes that TRGRMSE or a relative Median APE are preferable statistics for comparing methods. He objects to MAPE because it is ineffective when actual observations are near zero, which is similar to other objections discussed above.

Inferential Statistics

In addition to these descriptive statistics the following inferential statistics have been proposed:

Benito Flores

In another article Flores discusses the use of the Wilcoxon test of paired data to compare forecasting methods.[193] For n>10, the resulting quantity is compared with the normal distribution for statistical significance. For lesser n, the T_+ distribution is known. In his analysis of this statistic, Flores recommends its use **in combination with other approaches** to comparing forecasts, but does not give it unqualified support.[194] It should be noted that the Wilcoxon test is only appropriate for comparing two forecasts with each other.

Kolb and Stekler

Kolb and Stekler[195] propose a method of comparing forecasts by testing whether the differences in the mean squared error of the forecasts is statistically significant. They define a regression model that compares these means. It should be noted that this regression requires a reasonably large number of observations to provide an adequate N. Use of multiple observations from the same forecast trace encounters the covariance problem discussed above.

H. O. Stekler

H. O. Stekler identifies several inferential statistics in a general review of forecast evaluation statistics:[196]

The **Analysis of Variance by rank** (here in, Rank ANOVA) computes relative ranks of forecasts for each trial, sums those ranks for each model, and compares summed ranks using a chi-squared goodness of fit test. This test is also sometimes called the Friedman test. The **Kruskal-Wallis** test is similar to the Rank ANOVA except that it ranks all observations in one series rather than by trial. The Kruskal-Wallis test also tests significance through chi square. Because of the ranking technique, the Kruskal-Wallis test is not valid when different trials naturally differ in magnitude.

The **Percent Better** is the number of occasions where A out performs B divided by the total number of trials. Where $n>40$, this can be tested for significance using $Z_0=(n_1-n/2)/(n/4)$ where n_1 is the number of times the first method is better and n is the total number of trials.

Stekler defines another test which is aimed at determining whether on forecast contains more information than another. Calculate the regression:

$$A_t = bP_{xt} + (1-b)P_{yt}$$

Where A is the actual and P is the predicted for two different methods. Consider the null hypothesis b = 1. If the test is significant, it is implied that one forecast contains more information than the other. Again, the independence of observations problem lurks in the background.

Stekler discusses several other statistics which will be omitted from this discussion. One, the Wilcoxon test, is discussed above.

Optimization vs. Model Comparison

Before leaving this discussion of statistics, I want to emphasize the difference between optimizing a single model and comparing various models or methods. Many of these statistics can be used for either of these purposes; however, they may not perform equally well for each. This review is aimed at finding a suitable method of **comparing results** from various forecast models. Articles do not show considerable fault with traditional approaches to optimization. Results from recent analyses suggest that most approaches to optimization have similar results.[197]

Selection of Approach for Comparison

These articles show that there is no consensus as to which statistics are adequate for comparison of forecast models. Several of the proposed statistics provided for possible test of significance; however, there is no clear consensus on which, if any, of these tests is appropriate. Some may require more observations than may be available. Others may have less than clear benefit in practice. There is no evidence that any particular statistical test is generally accepted as an appropriate test for forecast accuracy or comparative forecast value. This study follows the traditional approach of displaying and discussing several descriptive statistics. Tables are used to display a variety of statistics summarizing forecast performance. This approach reflects the practice of major studies including one that has been published within the past year.[198]

As an exploratory element of this dissertation three non-parametric statistics are included for evaluation of the descriptive statistics. These statistics are the Rank ANOVA proposed by Stekler, the Kruskal-Wallis statistic also suggested by Stekler, and the multiple treatment comparison statistic for the Kruskal-Wallis test which is used to identify which series are different when the Kruskal-Wallis statistic is significant.[199] These statistics are used to evaluate the relative ranks of various treatments.

It should be understood that these statistics are not generally accepted as a basis for testing hypotheses about forecast methods, thus, use in this dissertation is exploratory, that is, other analyses will be conducted regardless of the statistical significance of these tests. Further, the Kruskal-Wallis test is sensitive to disparity of magnitude among different series, that is, when the series that are compared are significantly different in magnitude in the first place the Kruskal-Wallis test will likely test insignificant regardless of the variation of effectiveness of the treatments (forecast methods). Thus, the Kruskal-Wallis test is not valid for use with statistics that retain the original magnitude of the data (i.e., the Mean Squared Error statistics).

These tests are applied as follows: The descriptive statistics selected in a later portion of this chapter are arrayed using various series of data as trials and various forecast modelling techniques as treatments. The descriptive statistics are then ranked using either the Rank ANOVA or Kruskal-Wallis ranking rules. These ranks are then compared using the Rank ANOVA and Kruskal-Wallis statistics.

Any statistics selected that are heavily influenced by the magnitude of the original data series will be compared only through the Rank ANOVA test as the Kruskal-Wallis test is invalid. The Rank ANOVA and Kruskal-Wallis tests are selected because they allow comparison of more than two series, but they do not require the more complex assumptions associated with parametric statistics. This part of this dissertation is included as a **trial of the possible benefit of these non-parametric inferential statistics in comparison of forecast models**. As discussed in the earlier paragraph, the primary design is a qualitative review based on display of descriptive statistics. For all Rank ANOVA and Kruskal-Wallis tests the null hypothesis is:

$$H_0: \text{Statistic}_1 \equiv \text{Statistic}_2 \equiv \ldots \equiv \text{Statistic}_n$$

The alternative is that at least one statistic is not equal to the others. Results will be compared with $\alpha = .05$ level of significance. However, as these tests are being included to explore their value for this sort of analysis, results will be considered and discussed even if significance is not attained. If the null is rejected, the multiple series comparison analysis associated with the Kruskal-Wallis statistic is conducted to determine which specific model is distinct. The use of these tests is further discussed in the next chapter.

Selected Statistics

Based on the previous review the following factors are considered in selecting the descriptive statistics that are compared in analysis of the empirical research:

- RMSE and MAPE are subject to biases and should be use with caution if at all, nevertheless, these two concepts form the basis of most statistics actually proposed, so it may be desirable to represent each in the selected statistics. Two possible statistics are GRMSE and LMR.

- Variations of MAPE or APE that avoid the biases of MAPE are generally considered good. Two possible variations are MdAPE and SMAPE.

- Relative measures should be compared, absolute measures do not have comparative meaning. Some possible relative measures are LMR, Theil's U, GRMSE, RAE, and RMdAPE.

- Aggregation across multiple series may be improved through Geometric means.

The following statistics are selected for use in this analysis.

1. Geometric Root Mean Squared Error, GRMSE.
2. Log Mean Squared Error Ratio, LMR (as compared with Naive 2).
3. Symmetrical Mean Absolute Percent Error, SMAPE.
4. Median Absolute Percent Error, MdMAPE.
5. Average Rank.

In addition, because these statistics do not address the reliability criterion discussed in the second section of this chapter, the following statistic is added.

6. Range of percent error = largest positive percent error minus largest negative percent error.

Also, because they are frequently cited in forecast literature, Mean Absolute Percent Error (MAPE) and Root Mean Squared Error (RMSE) will be displayed.

In tables presented in chapter 8 and Appendix IV each of these statistics is aggregated across multiple trials through four different techniques: simple average, geometric mean, average rank, and Kruskal-Wallis rank sum. The average rank and Kruskal-Wallis rank sums relate to the associated inferential statistics.

Summary

There are numerous criteria for evaluating forecast models. The proposed technique can be expected to meet criteria related to low cost and ease of use. An empirical study is conducted to evaluate whether the proposed technique meets criteria related to accuracy and reliability. Two major hypotheses (5 constituent hypotheses) are specified. Three constituent hypotheses concern accuracy with prospective policy changes. Two constituent hypotheses concern accuracy with retrospective policy changes. Accuracy and reliability are measured through an array of statistics focussing on squared error, absolute error, and variation in error. Statistics are aggregated across multiple series using geometric averaging and other techniques.

CHAPTER 7: TWO RESEARCH PROJECTS (METHODOLOGY)

In this chapter I:

- Generally describe two proposed research projects.
- Explain the methodology of a research project for analysis of the first major hypothesis proposed above.
- Explain the methodology of a research project for analysis of the second major hypothesis proposed above.

To examine the first major hypothesis (hypotheses 1a, 1b, and 1c as specified on page **126**), six forecast models were built for each of 20 real data series from the Virginia Medical Assistance Program. Two variant forecasts were extracted from five of the models, resulting in a total of eleven forecasts. The forecasts were made with the assumption that the data series would undergo specific policy changes in the horizon period. Various simulated policy adjustments were added to the data series reflecting accurate and inaccurate policy change assumptions. The forecasts were updated through six updates. Accuracy is evaluated for certain periods in the 15 periods subsequent to the end of the six update periods.

To examine the second major hypothesis (hypotheses 2a and 2b), six forecasts models were built for each of 20 real data series that have undergone real level shifts in the observed (historical) period. All these data series are also from the Virginia Medical Assistance Program. Accuracy is evaluated for certain periods in the 15 periods subsequent to the end of the six update periods.

Selection of Scope of the Study

Each study includes 20 data series in order to balance between two objectives. The first is that the study be sufficiently small that it can be completed within the scope of a dissertation. The second objective is that the study should be sufficiently large and general to eliminate the realistic possibility that the findings arise because of chance selection of data series. This need leads to the selection of a larger number of series, even more than 20 may be desirable. For this same reason, different types of data were used (units, cost per unit, enrolled Medicaid eligibles, and gross dollar amounts). Also, data reflecting different origination dates (July 1988 and July 1987) were used.

In the literature review I did not find any standard number of forecasts for empirical evaluation of a technique. Actual empirical evaluations ranged in size from 1 to 1001 and from one organization to many. Limitation of the scope of this study to 20 series and to series originating from the one organization reflect a reasonable compromise to complete the project. As will be seen in further discussion below, 12 different scenarios were compared for 7 updates and a second study is conducted using 20 additional series which were fit to 6 models and updated for 7 periods. The total number of forecasts that were made (excluding those for model fitting) will be 19,320 as shown in the following table:

Table 2 Number of Forecasts

Study	Series	Updates	Models	Scenarios	Total
1	20	7	11	12	18,480
2	20	7	6	1	840
Total					19,320

It seems likely that 20 series updated across 7 periods should constitute an reasonable test of each method for each specific scenario.

The First Major Hypothesis

To examine the first major hypothesis I compared forecasts made with six types of forecast models. The six types of models are Holt-Winters, Holt-Winters-Williams, adaptive Holt-Winters-Williams, autocorrelation corrected Holt-Winters-Williams, Naive 2, and adjusted Holt-Winters-

Williams (the proposed technique). Only one forecast was made using the proposed technique; however, two forecasts were made using the other five techniques: one that uses the *ad hoc* method of adding on a lump amount for the policy change, and one that simply looks for the forecast to catch up while completely ignoring any information that may be available from the estimate of the policy change. In all cases except with Naive 2, Holt-Winters refers to a **selection between four possible models: SES, Holt** (i.e., a model with trend calculated in the manner of Holt)**, Winters** (i.e., a model with multiplicative seasonality calculated in the manner of Winters), or **Holt-Winters.** For the Naive 2 model Naive 1 was the alternative non-seasonal model. Forecasts were made of data series with which I am familiar having forecast this data for a government program. The unadjusted version of Naive 2 (i.e., the one without the *ad hoc* adjustment) was used as the benchmark model for LMR statistic. In fact, in this study seasonal models were not developed. This is explained in further discussion.

Selection of Models to Compare

The proposed technique, the naive model with and without the *ad hoc* adaptation, and four alternative methods with and without the *ad hoc* adaptation are compared. *The* inclusion of the proposed technique is obvious. The naive model is included as a base line for the LMR statistic. The *ad hoc* adaption is included because it appears to be the most obvious technique that a forecaster might use in the case that he had externally supplied information and did not have a forecast model that could integrate this information into the projection, which is the condition I believe forecasters would normally be in. The four selected models are included to represent a reasonable range of possible techniques that forecasters might actually have available to forecast using exponential smoothing when they anticipate policy changes. The Holt-Winters model is included because it is the most common standard exponential smoothing model. The Holt-Winters-Williams model is included because the arguments provided by T. M. Williams suggest that it may be possible to optimize it beyond the degree that the standard Holt-Winters model can be optimized. The adaptive Holt-Winters-Williams model is included as a representative of the adaptive technique using simple methods. The Chatfield autocorrelation corrected model is included as an alternative approach to adaptive modelling, still within the realm of simple approaches.

Obvious models not included within the trial include Kalman filter approaches and intervention based ARIMA models. These are excluded because of the underlying assumption of the study, i.e., that forecasting with simple methods is worthwhile and worth improving even if such forecasting may be less accurate than may be achieved with more complex methods.

The Data

The data are 20 monthly level (i.e., having one observation value per month) data series selected from data series used in budget forecasting for the Department of Medical Assistance Services. Series that have obvious problems unrelated to those under examination in this study, e.g., those with observations with the value of zero, were excluded from the selection. Nevertheless, a variety of series, reflecting a variety of actual forecast conditions, were selected. These series consist of monthly level units, expenditure per unit, and gross dollar amounts for various service categories running from July 1988 through September 1993, and monthly level enrollment data running from July 1987 through September 1992. The following graphs show these data series after certain preprocessing discussed below.

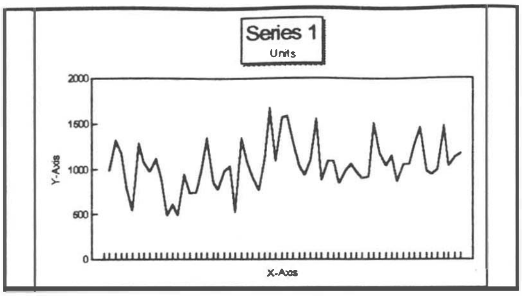

Figure 23

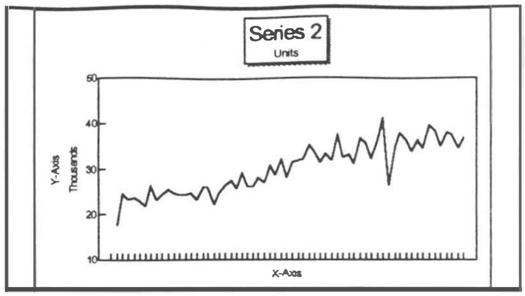

Figure 24

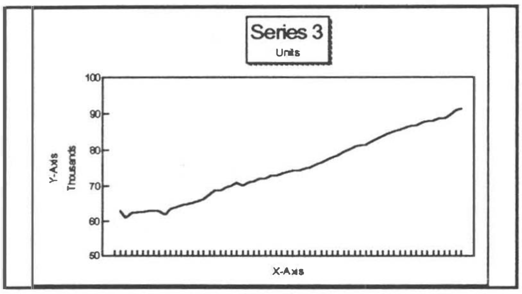

Figure 25

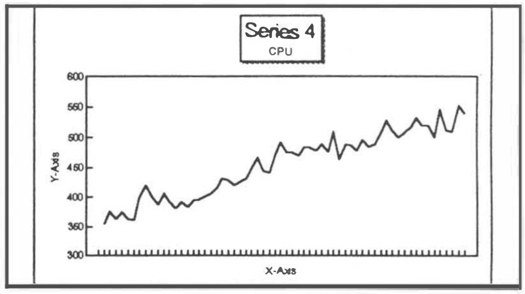

Figure 26

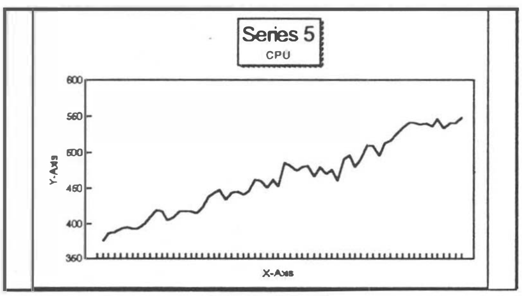

Figure 27

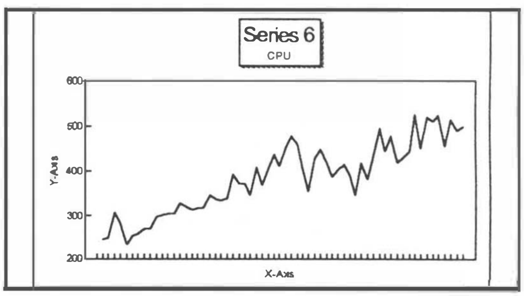

Figure 28

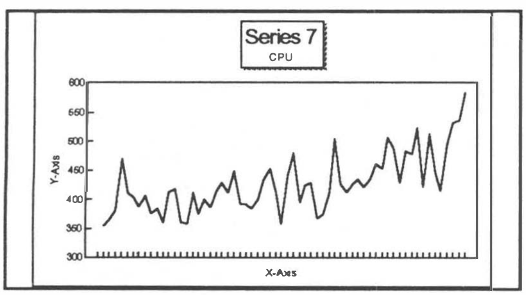

Figure 29

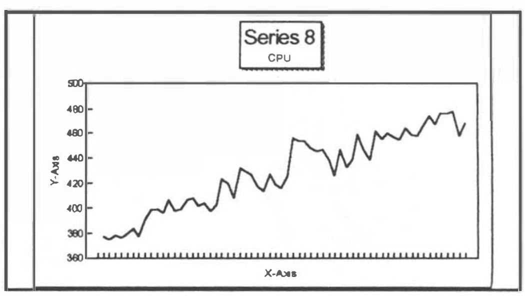

Figure 30

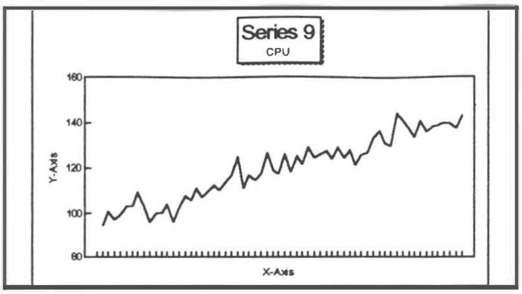

Figure 31

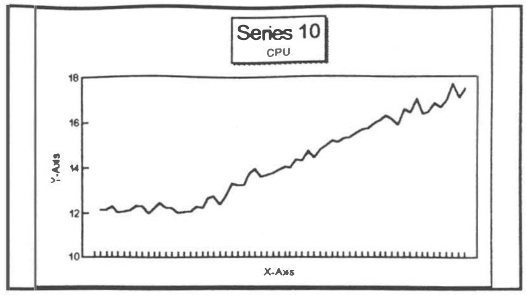

Figure 32

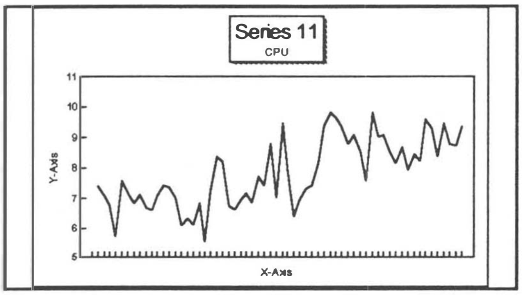

Figure 33

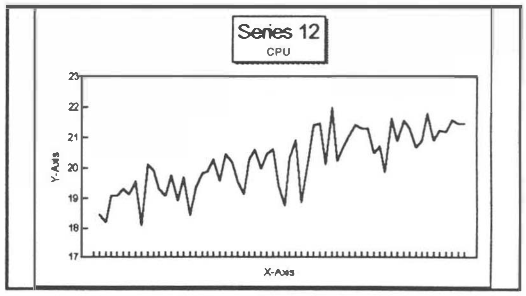

Figure 34

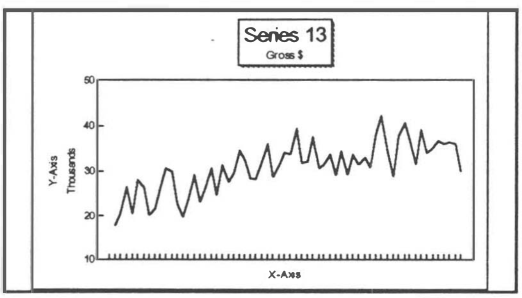

Figure 35

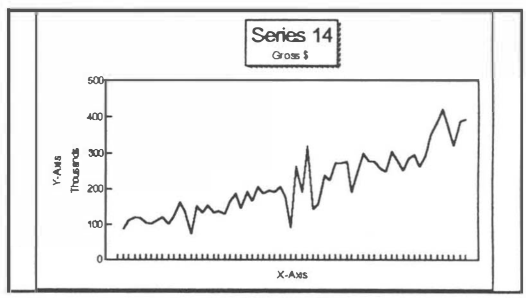

Figure 36

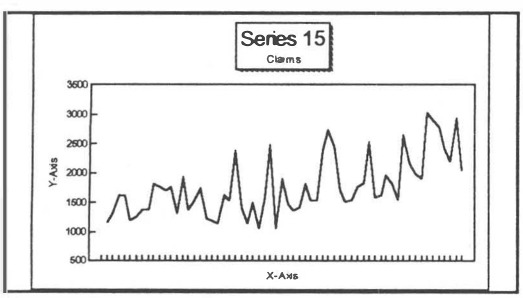

Figure 37

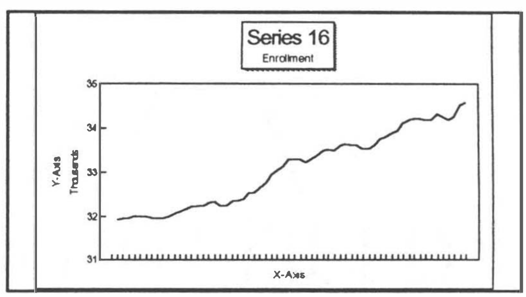

Figure 38

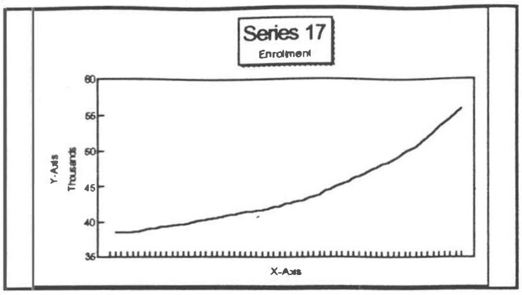

Figure 39

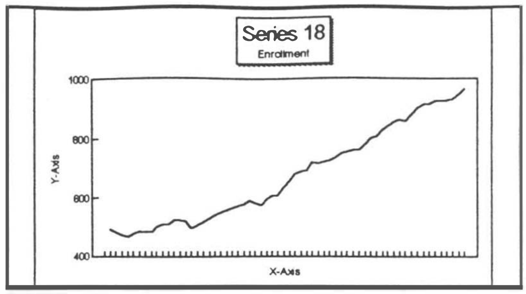

Figure 40

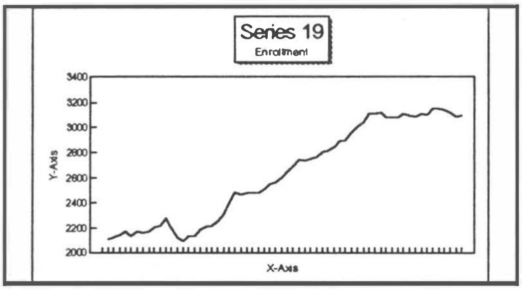

Figure 41

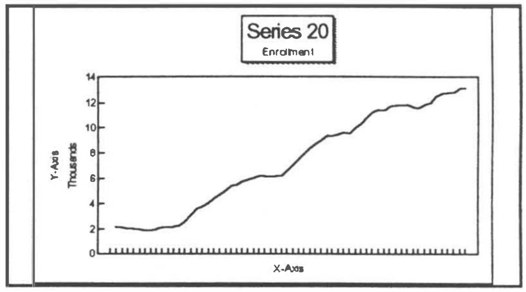

Figure 42

Independence

Because all series originate from a single organization they may render the trial susceptible to some unique condition arising within that organization. As there is a variety of types of data arising within the organization and as the data reaches the end of its fitting period on two different dates that are a year apart, it is unlikely that data would all reflect an undetected aberrance that would render the trials irrelevant to other data. Nevertheless, a correlation matrix was computed to determine whether the series are unduly interdependent. It should be noted that since many of these data series are trending in time they can be expected to reflect some correlation,[200] therefore, correlation matrices were calculated for the first differences of the data instead of the raw data. The data considered in these correlation matrices were the first differences of the pre-processed data described below. Even so, the use of data from a single organization should be considered a limitation that may justify additional trials at a later date. A correlation matrix and a squared correlation matrix are shown in Appendix III.

Inevitably, the calculation of a correlation matrix led to selection of the series from the larger universe on the basis of independence. I used the following procedure to select uncorrelated data series from the larger set of all series available for analysis: I calculated the correlation matrix and the squared correlation coefficient of all the data series available. I then identified the two series that had the highest squared correlation coefficient. I eliminated the one of these two that had the highest average squared correlation coefficient. I repeated the process until only 20 series remained. An unexpected side effect of selecting data out was that seasonal series, which are somewhat more highly correlated with each other, were essentially eliminated from the data selection. This result had the effect of eliminating consideration of the Winters models.

Level Shifts

Graphs of the data series were visually inspected for level shifts. Visual inspection was augmented through my prior knowledge of the data series. Data series that experience a prior level shift were excluded from this portion of the analysis. These series are excluded to avoid confusion between reasons for performance results. The second study discussed below evaluates the application of the technique in series that have historical level shifts.

Other Data Restrictions

Data series that include any of the following were not included in this analysis:

- Data that has zero valued observations.
- Data that includes wide swings in variation.
- Data that includes frequent trend shifts.

These data are excluded because it is my experience that such data are difficult to forecast with any technique. Results showing a preference for one technique rather than another with respect to the anticipation of policy adjustments when forecasting through these sorts of data series could not be ruled out as spurious.

Another restriction on the data series is that they were all the same length, 63 observations including 15 hold out observations, although some originate in July 1987 while others originate in July 1986. There is nothing special about the numbers 63 and 15. It was the number of observations that were available for the majority of the series included at the time that this study was completed. The population based series are longer than this and are restricted to 63 to be comparable in length to the other series. These series are allocated to segments as follows: the first 24 periods are allocated to initialization (discussed below), the next 18 periods are allocated to model fitting, the next 6 periods are allocated to simulation *of level shifts*, and the last 15 periods are allocated to *ex ante* model evaluation. This seems to be a reasonable allocation of the available observations.

Hold Out Data

Twenty one months of data were held out from the forecasts for use in simulated monthly updating and for evaluation of the errors.

Model Selection

One model was selected from each of the five types of models for each data series based on optimizing loss functions (producing one forecast for the proposed technique and two forecasts each for the alternative methods). Optimization was based on a grid of possible parameter selections. The grid of parameters is:

Table 3 Grid of Parameters

α =	0.05	0.1	0.2	0.4	0.8			
β =	0.0	0.001	0.005	0.01	0.1	0.2	0.3	0.5
γ =	0.0	0.05	0.1	0.3	0.5			

The settings β = 0.0 and γ = 0.0 represent Winters and Holt. Where both occur, the model is SES. In these cases initialization of trend, seasonality, or both (discussed below) is disregarded. For the non-adaptive models the a fixed α = 0.8 was assumed to be optimal as discussed below.

The α values are selected because they represent a wide range of possible values. Discussion in previous chapters shows my rationale for keeping β and γ fairly low, which is that the Williams adjustment to Holt exponential smoothing raises the implicit values of β and γ by something approaching a factor of 10. The values considered actually represent a broader range than I usually consider while forecasting to avoid overlooking legitimate models. As discussed in chapter 9, I considered a separate set of models for one scenario when I began to suspect that the large β values led to problems; however, the results were not affected by this extra scenario. The lower limits (other than zero) for β and γ are selected based on my experience that values below these tend to have little impact on actual forecast models.

The adaptive model follows the same practice for β and γ and calculates α following the Trigg-Leach formula while using the Williams adjustment to Holt-Winters. In this model the α grid values were used to fit ϕ.

The loss functions that were optimized were SMAPE, RMSE, SMPE. MAPE and RMSE were included as they are commonly used loss functions that measure overall accuracy of forecast. MPE was included to help eliminate models that are consistently erroneous with the same sign as it is assumed that consistently high or consistently low forecasts are particularly undesirable.

I set control limits for SMAPE and SMPE, then minimize RMSE. The control limits were 25% for SMAPE and ± 5% for SMPE. I have found no literature that supports specific control limits for MAPE or MPE (the symmetrical versions of these statistics are not widely known). The specified control limits are proposed based on two reasons. First, symmetry corrected percent error and absolute percent error are likely to be smaller than their non-symmetry corrected versions because the non-symmetry corrected versions are significantly affected by the occasional error where the forecast is intensely higher than the actual. With such small actuals as the denominator in the percent calculation, the percent becomes very large. When the mean of the actual and the forecast is used as the denominator, the calculated percent declines immensely. It is, therefore, assumed that reasonably low percents can be set. Second, reasonably low percents are needed because forecasts that exceed such control limits are likely to be of little value. The specific percents selected are arbitrary and represent my judgement of the point where forecasts begin to significantly lose value due to the various forms of inaccuracy.

The selected model was to be the model with the lowest RMSE that also meets the control limit criteria for the SMAPE and SMPE loss functions, if no model meets these criteria, the model selected was to be the one with the lowest RMSE without regard to other criteria. In fact most selected models meet all criteria.

Model Initialization*

I normally initialize level, trend, and seasonality of exponential smoothing forecast models using the technique described below which I first learned from Don Miller. Forecast literature does not demonstrate a terrific advantage in model initialization.[201] Nevertheless, there is likewise no evidence that initializing causes harm. Under such circumstances, it would appear that the most important consideration is in consistency of practice in comparing models.

I use this initialization process because it seems likely that uninitialized exponential smoothing models may have biased parameters. This bias arises because uninitialized models are actually initialized at zero for level and trend and 1 for seasonality. These initial values guarantee high error values for the first few observations. With guaranteed high error values for the first few observations, uninitialized models are likely to require large (responsive) parameters to self-initialize. As a

---

***This initialization process is similar to one described to me by Don Miller.**

consequence, when the user fits the model to high parameters, it is difficult to know how much the failure to initialize contributed to this fit. The technique described bellow seems to offer a reasonable alternative to these extremely unlikely initial values.

The first twenty four observations are linearlized by subtracting the mean of the first 12 observations from the mean of the second 12 and dividing by 12. The slope so computed is the initial trend. This trend is backed off the mean of the first 12 observations by 6.5 to obtain the initial level. It is then be added back to the initial level for twenty four iterations to establish an estimated deseasonalized series for the first two years. While initialization of seasonality was discussed in the proposal, because of the incidental elimination of seasonal series discussed above, seasonality was not a factor in the model fitting.

Exclusion from Loss Functions

Because the first twenty four observations were used to initialize the forecast, they were excluded from the calculations of the loss function in model fitting.

Trading Days*

Because this data is known sometimes to be affected by trading day information, (Fridays of the month or total days of the prior month) the data series was adjusted (divided) by these factors where it approved to reduce unexplained variation. Use of trading day corrections is recommended by Armstrong.[202]

Data Editing

J. Scott Armstrong advises that data series should be cleaned of erroneous observations and irrelevant outliers.[203] In this study the data series was **cleaned** following Armstrong's advice. Because I am familiar with the series, I am aware that some of the series may undergo a low month followed by a high month, or vice versa. These high/low or low/high events are transitory and reflect short term external events. These were adjusted by averaging the two observations. Certain extreme outliers were edited out of the series by replacing them average of the preceding and following periods. I visually identification of these observations bu inspecting graphs before forecasting data or simulating policy shifts.

***This trading day analysis is similar to one described to me by Don Miller.**

Simplification with Pre-processed Data

Pre-processed data was used to generate the input data for the trials instead of using the preprocessing math to build more complex models that reversed the preprocessing stage as the last step. This was the principal I applied to all data preprocessing in both experiments, e.g., editing outliers or removing weekly variation from monthly level data. This principal was used to simplify assumptions about the end result, i.e., to allow the assumption that the errors of the results arose from the exponential smoothing models rather than from the combination of the exponential smoothing models and the reversal of the preprocessing. Since these forecasts were not generated for practical use, it was not essential to produce a final forecast that was comparable to the original raw series. Forecasters who make practical forecasts are not afforded this convenience.

The Simulated Policy Adjustments

All **models** were developed as if they had an expected level shift occurring beginning on the first month of the forecast period. The prospective level shift is anticipated at 30% of the average of the data series in the last 12 months of the historical period and is expected to phase in (follow a ramp) over 3 months in equal increments. Both positive and negative level shifts were considered for each series. For the model using the proposed technique, this level shift was incorporated into the forecast. For the other techniques, the statistical model was updated without inclusion of policy adjusted information. However, for those models two forecasts were calculated from the data. One forecast included only the information from the statistical forecast. The other included an *ad hoc* adjustment added onto the forecast at the monthly level to increase the forecast to a level similar to that of the proposed technique.

For each series the following conditions were simulated by adjusting the **hold out data**. These conditions allow for consideration of some of the potential limitations of the technique discussed on page 106.

Scenario 1: A level shift occurs exactly as anticipated, beginning on the anticipated date and phasing in over 3 months and attaining 100% of the anticipated amount.

Scenario 2: Each level shift occurs as with scenario 1. In addition, a trend shift of 10% of the average first differences of the 6 periods prior to the level shift is added to the data.

Scenario 3: The level change occurs as with scenario 1 except that it attains 25% of the anticipated policy change.

Scenario 4: The level change occurs as with scenario 1 except that it attains 200% of the anticipated policy change.

Scenario 5: A positive trend shift occurs beginning in the month of the anticipated level shift phase in date and attaining a slope that is 50% of the slope of planned ramp.

Scenario 6: No change is added to the data series.

Scenario 7: No level shift or trend shift is added to the data series; however the variation of the data series is be increased by 100%. This increase is calculated by determining the difference between the observation and a 3 period moving average. That difference is multiplied by 2 and then added back to the original moving average to create a new observation at 100% greater variation.

Scenario 8: A negative level shift occurs exactly as anticipated.

Scenario 9: Each level shift occurs as with scenario 8. In addition, a trend shift of 10% of the average first differences of the 6 periods prior to the level shift is added to the data.

Scenario 10: The level change occurs as with scenario 8 except that it attains 25% of the anticipated policy change.

Scenario 11: The level change occurs as with scenario 8 except that it attains 200% of the anticipated policy change.

Scenario 12: A negative trend shift occurs beginning in the month of the anticipated level shift phase in date and attaining a slope that is 25% of the slope of planned ramp. The 25% trend shift for negative cases is set so as to avoid having the forecast series attain a level below zero in the forecast horizon.

Simulated Scenarios Explained

These adjustments constitute twelve different scenarios that may arise when data series are anticipated to have level shifts. Scenarios 1 through 5 involve planned and actual positive shifts. Scenarios 6 and 7 are planned positive shifts with no actual positive shift. Scenarios 8 through 12 are planned and actual negative shifts.

Scenarios 1 and 8 represents the situation for which the proposed technique is designed. Consistent failure in this scenario would suggest that the technique is of little value. Scenarios 2 and 9 represents the condition for which the technique should be robust. While the policy adjustment is going into place the series also undergoes a significant trend adjustment. The technique should not severely reduce the exponential smoothing model's ability to respond to the trend adjustment. In any case, the model should be expected to perform at least as well as the alternative models.

Scenarios 3, 4, 10, and 11 reflect problems that the forecaster may frequently face. It is not clear which technique should be most accurate under these conditions. Also, these are scenarios for which the forecaster may want to be alerted to estimation failure through a signal such as the smoothed error tracking signal.

Scenarios 5 and 12 should pose a significant problem for the proposed technique since the model can be expected to ignore the trend shift over the first few periods, treating it as a level shift instead. This suggests that the alternative methods may perform better when such a shift occurs. However, the scenario may also pose a significant problem for the alternative methods.

Scenarios 6 and 7 challenge the proposed technique (and the alternative techniques) to perform well in the absence of the expected change. Scenario 7 introduces new confusing information that may cause problems with all of the techniques.

Updating

All forecasts were updated by adding one period of new data to the historical series and recalculating the projections. Parameters were not adjusted during the updating process. For the standard Holt-Winters and Holt-Winters-Williams models, the α parameter was raised to 0.8 before any updating occurs to allow for the anticipated level shift, also the β and γ parameters were changed to the optimal β and γ parameters for the subset of $\alpha = 0.8$ models tested. This rise in the α parameter reflects Armstrong's recommendations following the previous advice of Brown.[204]

Six updates were completed for each series. Each update was recorded with forecasts through the end of the original 21 period horizon. The last fifteen observations, adjusted to reflect each simulated policy adjustment, were available solely for evaluation of the techniques.

Statistical Evaluation

Summarized statistics specified in the previous chapter are displayed in tables for **horizons 1, 5, 10, and 15**. Models that have lower values for all statistics except LMR are considered to perform better. For LMR, the higher value reflects better performance. All statistics except the range of percent error are measurements of accuracy. The range of percent error statistic is a measurement of reliability. Because of the extensive nature of these statistics, the tables displaying them are placed in Appendix IV. That Appendix also includes tables that display the results of the Rank ANOVA and Kruskal-Wallis tests.

In the absence of consensus among forecasting experts concerning what constitutes unequivocal success in a forecast competition, the results are discussed qualitatively rather than specifically compared to a definite standard for acceptance or rejection of hypotheses. The qualitative discussion addresses tendencies of particular techniques to rank as more or less accurate than other techniques, for example the tendency of the proposed technique, or the proposed technique and the *ad hoc* adjusted techniques to rank as more or less accurate than the forecasts made with alternative techniques in which no adjustment is made for policy changes. The discussion also addresses the relative ranks within particular scenarios, e.g., whether the proposed technique increases the risk of forecast error when in actuality the policy leads to a significant increase in variance rather than a level change, which may allow a potential user of the technique to evaluate whether the technique increases or decreases potential forecast accuracy with respect to the particular likely outcomes anticipated for a specific planned policy change.

While qualitative results are discussed, I also examine the results through Rank ANOVA and Kruskal-Wallis, two non-parametric test of rank order. However, because of the small sample size, 20 forecast series, and the lack of general consensus on the applicability, these non-parametric tests are not do not definitively evaluate the hypotheses. As the analysis does not include definitive inferential tests, the results should be considered to add to the overall discussion of techniques appropriate to forecasting discontinuous data.

The Second Major Hypothesis

To evaluate the second major hypothesis (hypotheses 2a and 2b) I made forecasts of 20 data series that have had level shifts during the historical period. The six models used to forecast series in the first analyses were used to forecast the series in this analyses.

The Data

The data are 20 series selected from the Department of Medical Assistance Services. Only series that have a level shift before the beginning of the hold out data are used. These data are shown in the following graphs:

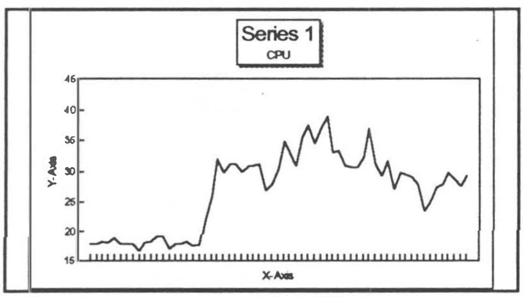

Figure 43

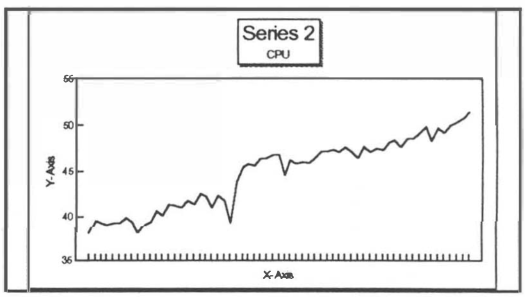

Figure 44

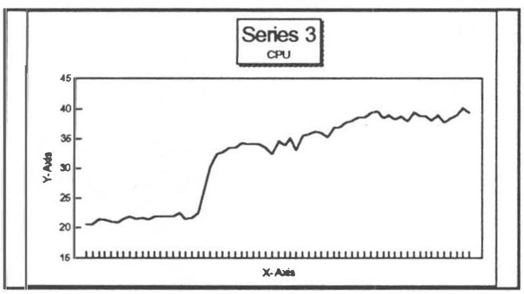

Figure 45

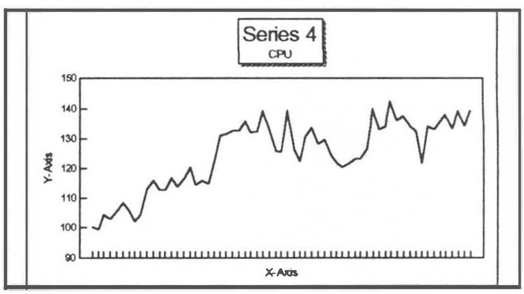

Figure 46

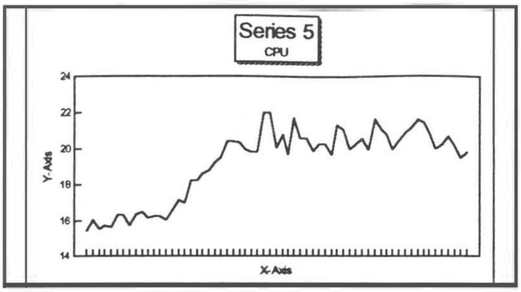

Figure 47

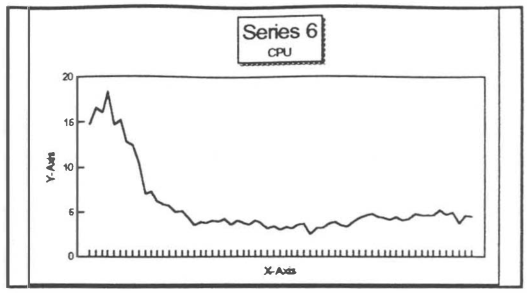

Figure 48

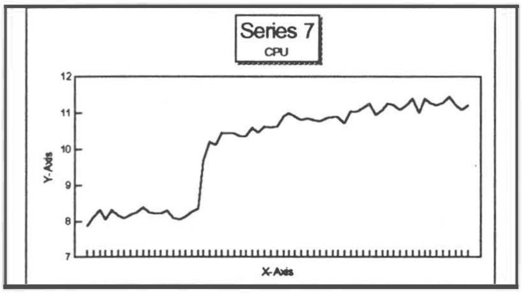

Figure 49

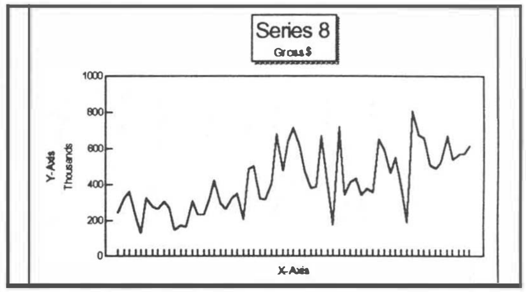

Figure 50

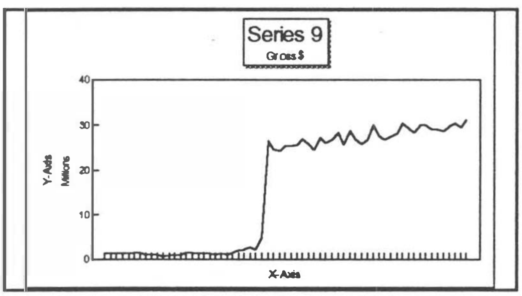

Figure 51

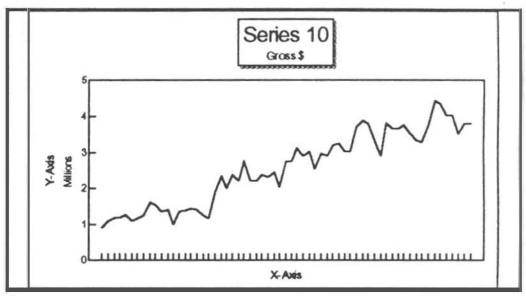

Figure 52

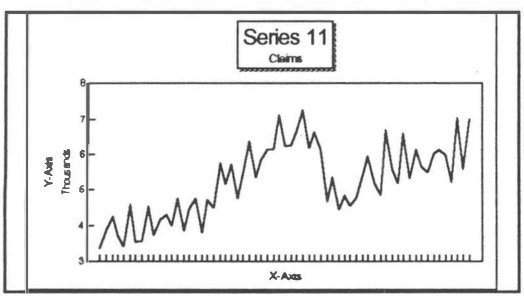

Figure 53

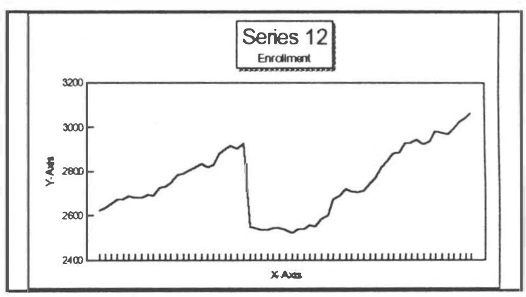

Figure 54

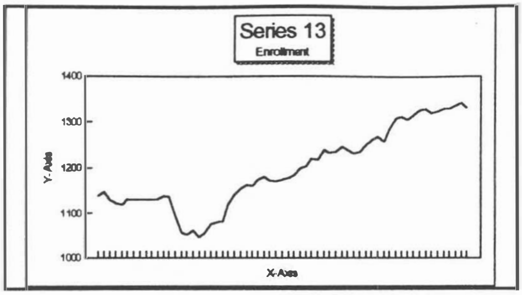

Figure 55

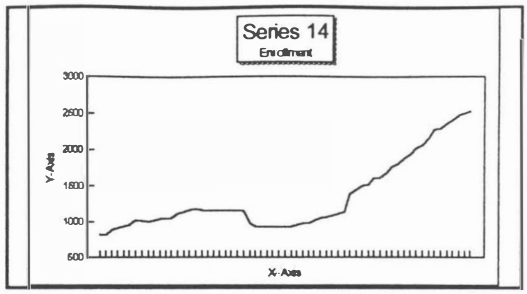

Figure 56

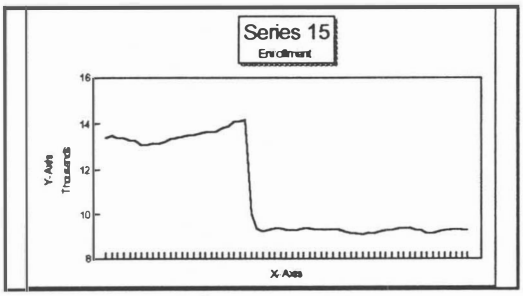

Figure 57

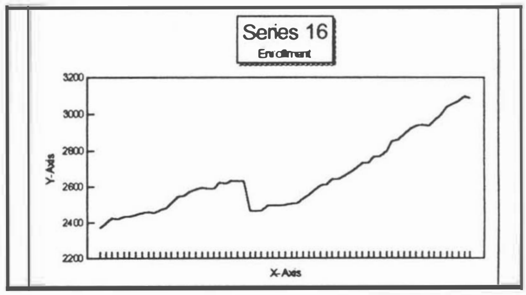

Figure 58

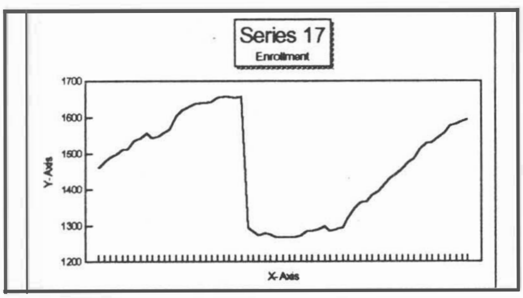

Figure 59

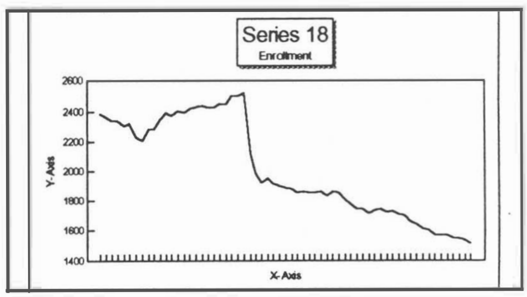

Figure 60

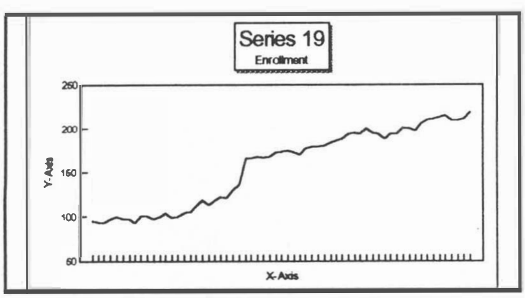

Figure 61

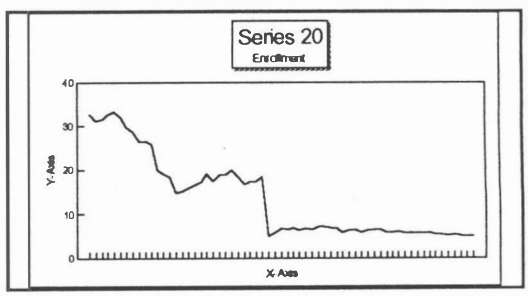

Figure 62

Independence

Correlation matrices were calculated as with the first study. These are also demonstrated in Appendix III. Because this series included level shifting data, the level shifts added confusion to the correlation analysis. The purpose of the correlation analysis was to validate that the data were not correlated in general. Large level shifts arising at the same time would tend to cause these series to be spuriously correlated, while large level shifts arising at different points in time could mask actual correlation in the data. Consequently, the level shifts are averaged out of the first differences before calculation of the correlation matrices. There is some correlation between some of series (R-squared $\cong$ 0.3); however, in general the series are independent. The universe from which these series were drawn did not contain a sufficient number of level shifting series to reduce all correlations further. Two of the 20 selected series appeared to be seasonal. To avoid confounding the results (since all other series were non-seasonal) I deseasonalized these series and treated the deseasonalized data as the pre-processed data as described above.

Forecast Procedures

Initialization and other forecast procedures were as with the first analysis except as follows:

1. There were no artificial level shifts or scenarios. Only the actual data series were forecast.

2. Where the level shift occurs before the end of the initialization period, judgmental (eyeball) values were set for initial trend and level.

3. The level shifting element of the proposed technique was used to fit the historical level shifts in a manner similar to an intervention variable. The historical level shifts were "estimated" by identifying outlier. These first differences were treated as the retrospective level shift estimators. A table of these is shown in Appendix VI.

Updating and Statistical Analysis

The six models were updated for the same six periods as with the first analysis. Forecasts and errors were calculated for **horizons 1, 5, 10 and 15** as with the first analysis. The statistical analysis is conducted in the same manner as with the first analysis. In the next two chapters this second study is referred to as scenario 13. As discussed there, two variants of this study were conducted. The second variant is called scenario 13b. Related tables with those headings are shown in Appendix IV.

Limitations of the Research Projects

This study should be viewed as an exploratory analysis of a new type of forecast model. This model is a hybrid of an exponential smoothing model and simpler estimation approaches. Most results are displayed in descriptive statistics. While non-parametric inferential statistics are presented, they are not widely used for comparing forecast models and are included in part for consideration of whether they may be useful in this sort of comparison. Also, both the number of series evaluated, 20 for each of two projects, and their origins from a single organization limit the generalizability of this study. On the other hand, correlation matrices presented in Appendix III show that the series are independent, thus, it is reasonable to tentatively consider generalizing results. Overall, however, this study provides a preliminary analysis of the proposed technique. Promising results suggest the need for additional study.

Summary

Two research projects were undertaken. In the first project, 20 data series were forecast over 15 horizon periods using 5 different techniques with two variants of five of the techniques (for a total comparison of 11 models). Twelve policy change scenarios will be tested. The series will be updated for 6 periods. Six summary statistics will be calculated for each horizon to evaluate the comparative effectiveness of the 9 models. In the second project, 20 data series that have undergone a level shift in the historical period were forecast over 15 horizon periods using six different techniques. The series were updated for 6 periods. Six summary statistics were calculated for each horizon to evaluate the comparative effectiveness of the six models. Because of cited limitations, results should be considered preliminary.

CHAPTER 8: PRESENTATION AND ANALYSIS OF THE DATA

In this chapter I will:

- Describe the research and the layout of the statistical tables presented here.
- Describe the results of each of the 12 scenarios associated with the first empirical study.
- Describe the results of the 1 scenario associated with the second empirical study.

Results of Statistical Analysis

The statistical analysis is conducted primarily through development of one table of twenty trials for each horizon for each scenario for each descriptive statistic. These tables are then summarized through the average, the geometric mean, the average rank, and the Kruskal-Wallis rank sums. As discussed below, some summary data is not appropriate for some statistics. The summarized data is presented in tables in Appendix IV. Following is a descripiton of these tables.

The Tables

Tables demonstrating summary information from the 12 scenarios of the first study and two variates of the second study run eight pages in length each, for a total of 104 pages, so they are not displayed in the text of this dissertation. These tables are included as Appendix IV. For convenience of reference I label the 11 models generated in the study as follows (The models that are marked with an asterisk are not produced for the second study which is labeled scenario 13):

Adjust The proposed method(or **Adjusted**).

HWW Holt-Winters-Williams.

HW Holt-Winters.

Adapt Holt-Winters-Williams with an adaptive α parameter (or **Adaptive**).

Auto Holt-Winters-Williams with the Chatfield autocorrelation correction.

HWW* Holt-Winters-Williams with an *ad hoc* level shift.

HW* Holt-Winters with an *ad hoc* level shift.

Adapt* Holt-Winters-Williams with an adaptive α parameter and an *ad hoc* level shift (or **Adaptive***).

Auto* Holt-Winters-Williams with the Chatfield autocorrelation correction and an *ad hoc* level shift.

Each table contains four sub-tables for horizons 1, 5, 10 and 15. It summarizes one statistic, e.g., SMAPE, over 20 trials. Trial-by-trial results are not displayed due to the magnitude of information. The different lines of each sub-table reflect different ways of summarizing the statistic. The tables are number X-Y where X is the scenario number and Y is the table number. For each horizon the following information is reported in tables X-1 through X-8:

- The average value of the statistic across twenty trials for each of the eleven alternative models. That is, the basic statistic (error, squared error, absolute error, etc.) is calculated for each of 7 updates for each of 20 trials. It is summarized (averaged, summed, ranked, or aggregated in whatever the appropriate manner for the particular table) to one statistic for each model for each trial. Then the 20 trials are summarized to one average.

- The rank of the average values among the eleven models.

- The geometric mean of the statistic across twenty trials for each of the eleven models.

- The rank of the geometric mean of the statistic among the eleven models.

- The average rank of the statistic by series across twenty trials for each of eleven models, i.e., the rank among the eleven models for each trial, averaged.

- The rank of the average rank among the eleven models.

- The rank sum as calculated in the Kruskal-Wallis statistic (where all observations among the twenty models are ranked in one set).

- The rank of the Kruskal-Wallis Rank Sum.

- The number of models that are statistically distinguished from the reported model in the case that the Kruskal-Wallis statistic is significant. The Kruskal-Wallis statistics are reported in Tables X-9 through X-16. If these statistics are significant, the number reported in this table shows how many of the other models has a statistically significant difference. If the Kruskal-Wallis statistic is not significant, the number reported in this table is irrelevant. When the Kruskal-Wallis statistic is significant and the number reported in this table is 10 (5 for Scenario 13), **all** the other models are statistically distinct from this model.

For most of the statistics reported in tables X-1 through X-8 lower values are superior to higher values. This is not true for the Log Mean Squared Error Ratio, where the higher value is the superior result (in the tables, the ranks reflect this fact). Also, the last row of data for each sub-table implies no superiority for either higher or lower numbers, instead, it shows how many of the other models (models in other columns) are statistically distinct from the reported model when the Kruskal-Wallis statistic is significant. If this number is small compared with the number of models reported, the other models with similar Kruskal-Wallis Rank Sums cannot be statistically distinguished from the reported model. Due to the quantity of statistics calculated for this study, individual pairwise comparisons are not reported.

For the Root Mean Squared Error and the Geometric Root Mean Squared Error certain results are not valid because problems associated with aggregating over data that is not comparable in magnitude. The Geometric Mean is not a valid mean for the Log Mean Error Ratio because negative logs imply that the mean error ratio is less than one, so the technique is worse than the comparison naive model. Invalid information is not reported.

Tables X-9 through X-16 show the results of two non-parametric tests for significant differences of the ranks of the reported statistic in . Thus, in Table X-10 the Rank ANOVA and Kruskal-Wallis statistics for the ranks of the SMAPEs for the twenty trials are shown for horizons 1, 5, 10, and 15. These statistics are reported under the Chi Square Column as they are compared with the Chi Square distribution for determination of significance. The next column reports the degrees of freedom and the last column reports the actual level of significance attained. Traditionally, p values below α levels of 0.05 are considered indication of statistical significance.

Description of the Data Collection and Analysis

I did the following (not necessarily in this sequence) to generate each table (in this example I discuss scenario 1, horizon 1, table 1-2, SMAPE): I fit the model with the observations ending in December 1991 (or December 1990 if the data originated from population based series). I collected the error for observation for January 1992 (or

1991 as may be). I updated the model for the simulated actual for the January observation and collected the error for February, and continued through the seventh update (counting the model fitting observation). I divided each of the seven errors by the average of the observation and the model projection for each of these seven periods to produce a SMAPE for the period (or followed other procedures to produce the relevant periodic statistic for the other tables). I carried the average of the seven SMAPEs to a table for comparison with the other 10 models and the other 19 trials. The average across all twenty trials is shown on the first row (by the label "Average") on Table 1-2, and for the proposed technique is 3.66%. The rank of the averages is shown on the next line, etc. The table of Rank ANOVA and Kruskal-Wallis statistics shows a Chi Squared value of 75.61 for the Rank ANOVA (p value 0.0000) and 42.55 for the Kruskal-Wallis statistic (p value 0.0000) for the SMAPE statistic and Horizon 1. These p values are, of course, significant at the $\alpha = 0.05$ level.

The two non-parametric tests were applied as follows:

- An array of the descriptive statistics were calculated for each trial.

- Ranks were then calculated among the descriptive statistics. For Rank ANOVA, each trial was ranked separately. For the Kruskal-Wallis statistic a table of all the results for the twenty trials was ranked from lowest to highest.

- The Kruskal-Wallis statistic was reported only for relatively dimensionless statistics (percents, ranks, LMRs, etc.).

- Both the Kruskal-Wallis statistic and the Rank ANOVA statistics were compared with the Chi Square distribution for tests of statistical significance.

Material That is Presented

In this overview certain **general results** are observed. I **illustrate** the forecasts made with the various techniques under the various conditions with graphs. These graphs are a **visual guide** to the variation in both the trials and the results; however, they do not necessarily demonstrate the consequences of the various techniques for all twenty data the series. For the first 12 scenarios, these graphs are taken from the application of the techniques to Series 4. Several series are demonstrated for scenario 13.

The scenarios presented below are keyed to initial list of scenarios. To maintain data integrity, this key was not changed although scenarios are grouped in a different order for this discussion. Also, the term **unadjusted models** will be used to collectively refer to the five models that do not take the anticipated level shift into account.

Scenarios 1 through 12 Discussed

The first twelve scenarios address hypothesis 1a:

> The alternative techniques and the proposed technique are not equally accurate in forecasting through periods where policy shifts are anticipated.

The following eight scenarios also address hypothesis 1b:

> The proposed technique is more accurate than the alternative techniques when used to forecast through periods where policy shifts are anticipated and such policy changes materialize.

In the first four scenarios presented, the **simulated actual level shift is equal to the planned level shift**.

Scenario 1: Level Shift as Expected

In scenario 1 the simulated actual level shift is the same as the planned level shift included in the forecast model. In the following graphs, this can be seen with the large level shift at the same point as the beg*inning of the* updating period. To read these graphs, notice that the bumpy line is the actual data, while the first update has the longest forecast line. Forecasts that are near the actual data are more **accurate** than otherwise, while those that are also in a tight pack are more **reliable**.

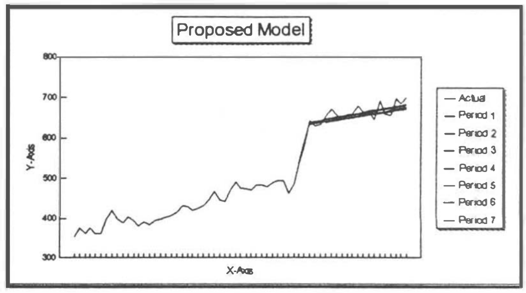

Figure 63

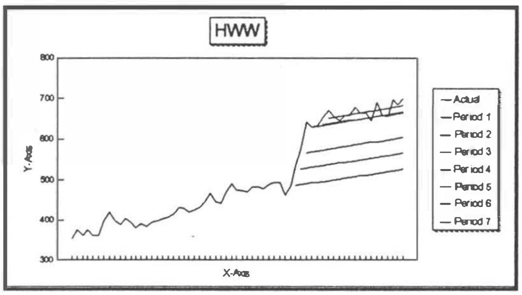

Figure 64

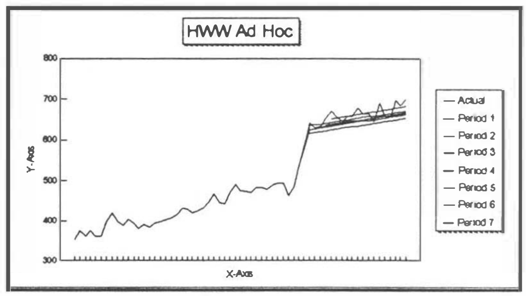

Figure 65

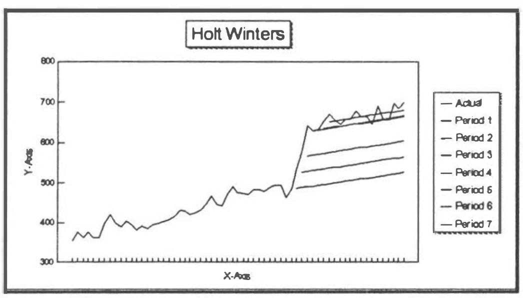

Figure 66

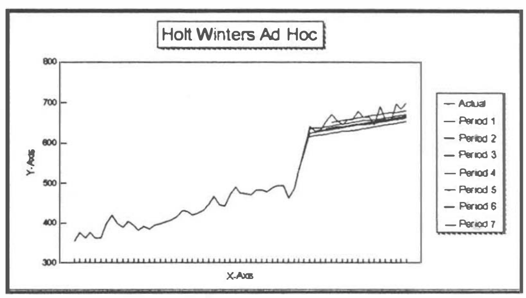

Figure 67

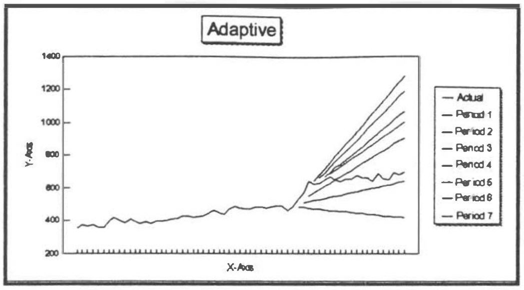

Figure 68

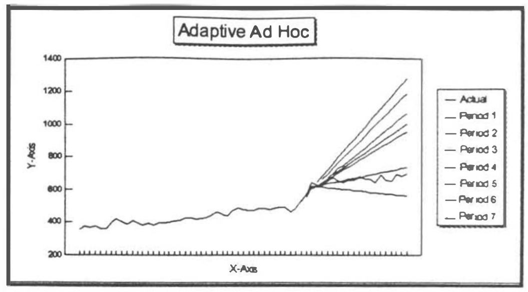

Figure 69

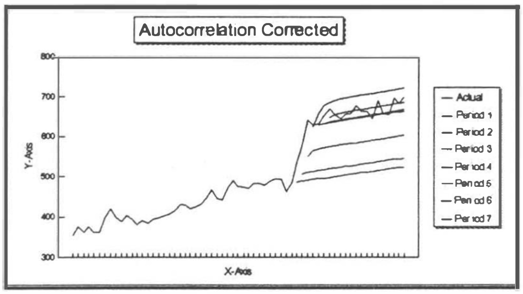

Figure 70

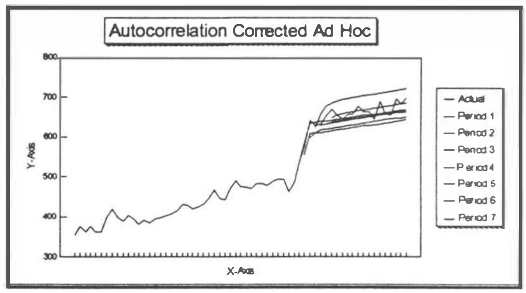

Figure 71

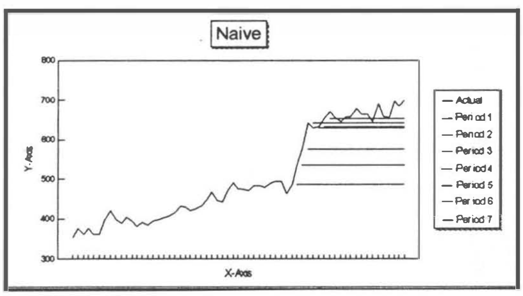

Figure 72

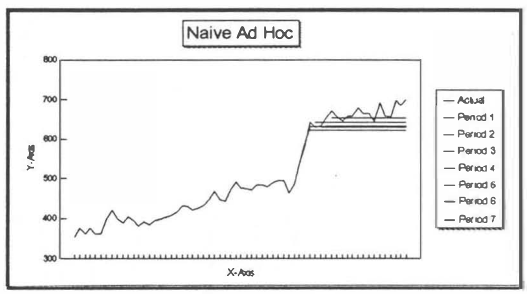

Figure 73

These graphs show a significant variation in the results for the various models. Nevertheless, the specific variation is relevant only for the **example** series. Tables 1-1 through 1-16 summarize the results for all 20 examined series:

- When the planned level shift actually occurs, the proposed method ranks superior to all other models. This result holds for each of the four tested horizons and for all of the reported statistics.

- The Kruskal-Wallis statistic and the Rank ANOVA show statistical significance for these results with p values ranging from 0.0000 to 0.0248.

- The Kruskal-Wallis multiple series comparisons shows that the proposed method can be statistically distinguished from all other methods for all comparisons for which it is valid (i.e., where this result is reported).

- The five other models that take the prospective policy change into consideration, the ad hoc models (marked with an asterisk), consistently perform better than the five **unadjusted models**. However, patterns of results (rank order of performance between these five models) are not consistent across the various statistics.

- The five **unadjusted models** are the worst performing models. Among these, the adaptive, autocorrelation corrected, and Holt-Winters models are frequently the worst performing models, with the Holt-Winters-Williams and naive models performing somewhat better.

- Among the models in which the prospective change is not taken into consideration, the naive model is more effective than the alternative models; however, the *ad hoc* naive model is not superior to the other *ad hoc* models.

Scenario 8, Negative Level Shift as Anticipated

Scenario 8 is like Scenario 1, except that both the expected and actual level shift are a negative shift rather than a positive shift.

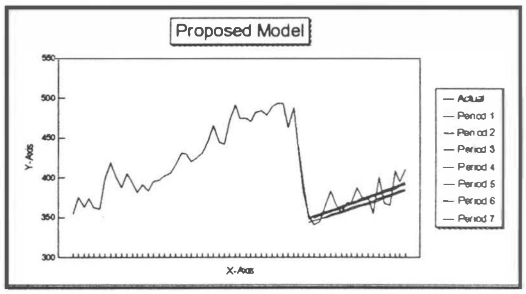

Figure 74

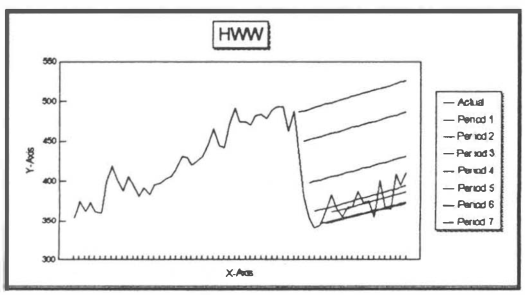

Figure 75

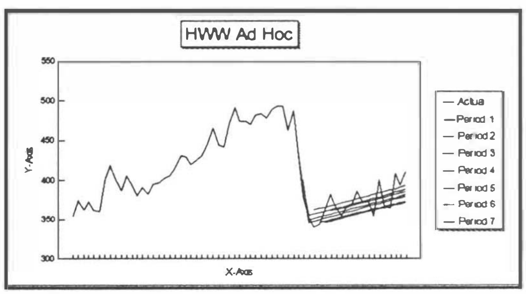

Figure 76

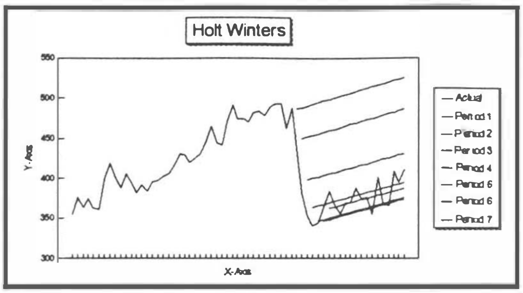

Figure 77

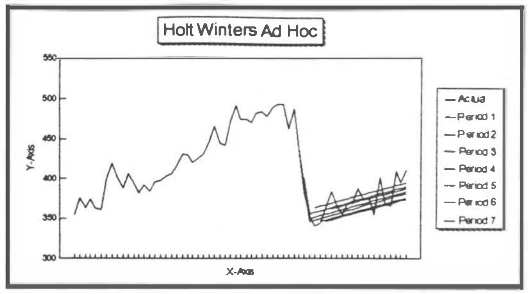

Figure 78

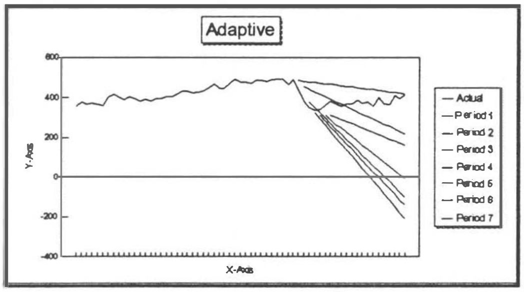

Figure 79

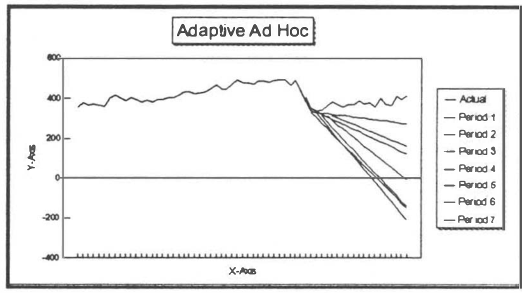

Figure 80

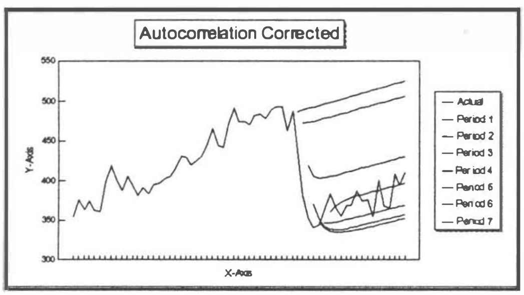

Figure 81

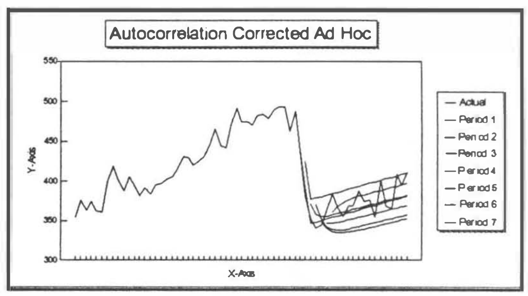

Figure 82

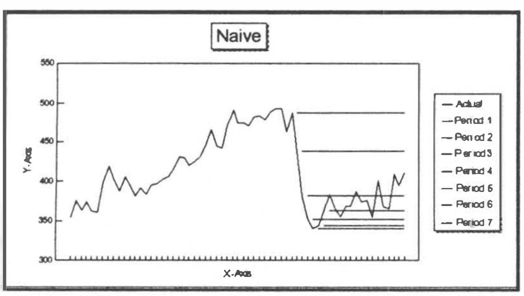

Figure 83

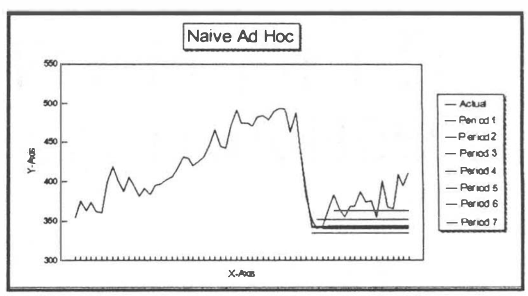

Figure 84

Following are the observed results:

- The proposed technique outperforms all other techniques across all reported horizons and all reported statistics.

- The *ad hoc* methods generally outperformed the unadjusted models.

- The naive method generally outperformed the other techniques that do not include an anticipated level shift and frequently outperformed the *ad hoc* methods.

- The Kruskal-Wallis test is significant for all statistics and all horizons.

- The proposed method can be distinguished from all other models on the Kruskal-Wallis related multiple treatment comparison analysis in all analyses for which the Kruskal-Wallis test is valid.

- The Rank ANOVA test is significant for all statistics for horizon 1 and for all statistics except for the average rank of the absolute error for horizon 5. It is not significant for horizons 10 or 15.

Scenario 2: Level and Trend Shift

In scenario 2, the simulated actual level shift is equal to the planned level shift; however, an unanticipated trend shift also occurs. The simulated trend shift is a 10% increase of the trend over the past 12 periods. It is apparent from the graphs that this change in slope may not result in a significant change in the forecast level.

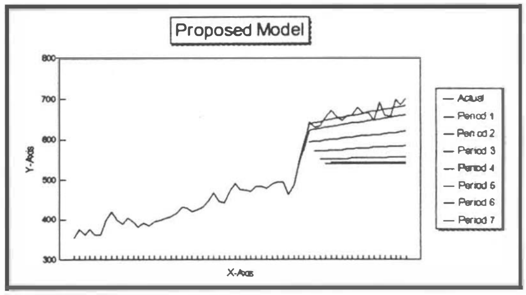

Figure 85

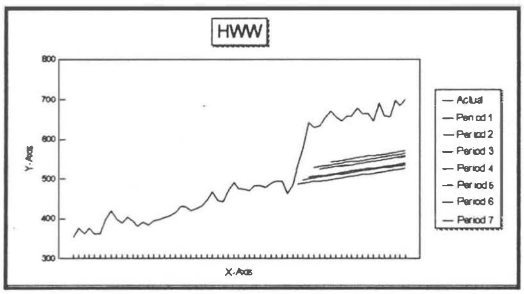

Figure 86

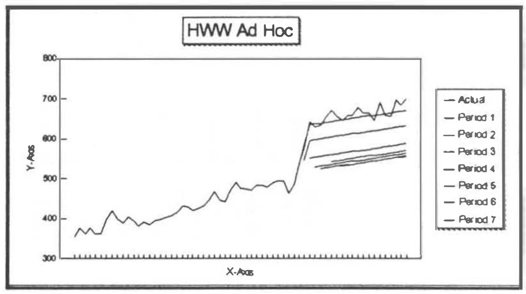

Figure 87

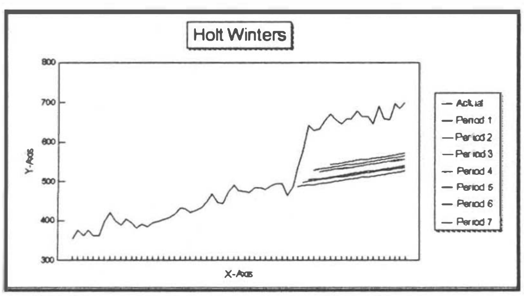

Figure 88

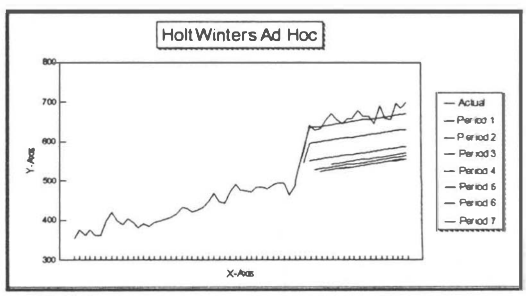

Figure 89

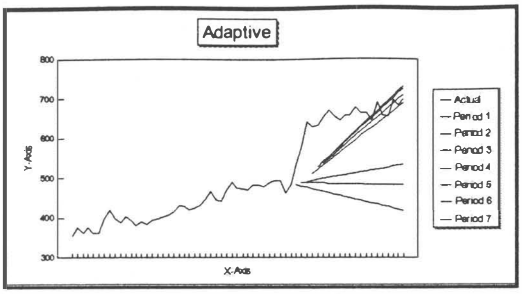

Figure 90

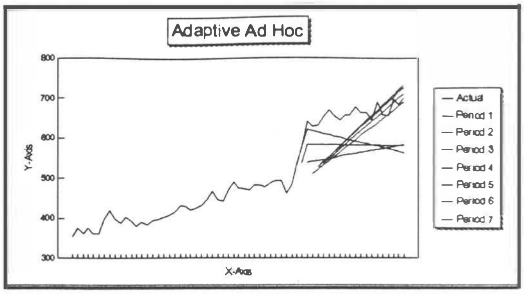

Figure 91

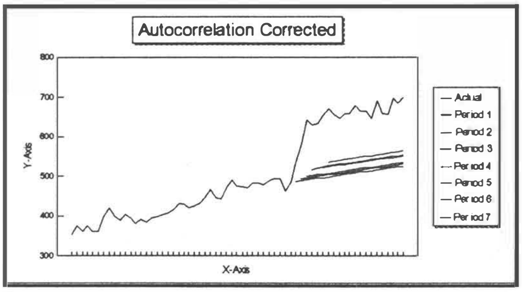

Figure 92

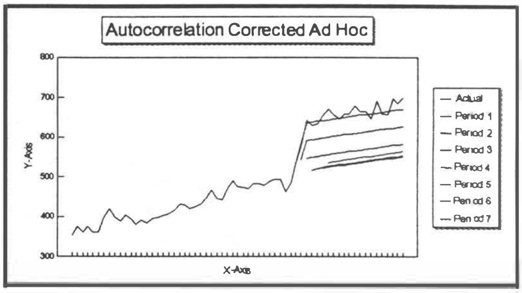

Figure 93

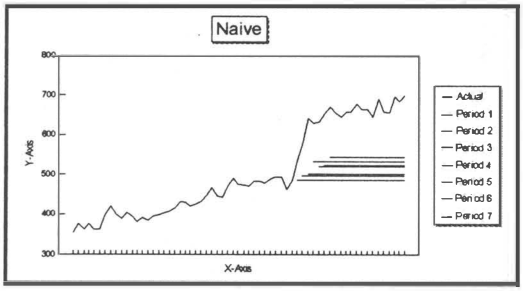

Figure 94

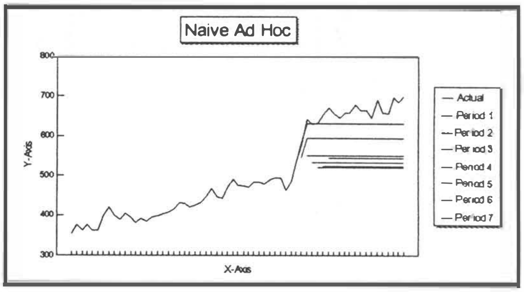

Figure 95

The results of this scenario are similar to those of scenario 1. These include:

- The proposed method ranks superior to all other models for all reported statistics.

- The *ad hoc* techniques consistently rank superior to the unadjusted models.

- Among the five unadjusted models, the least effective models are the Holt-Winters, adaptive, and autocorrelation corrected models.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid.

- The proposed method can be distinguished from all other models on the Kruskal-Wallis related multiple treatment comparison analysis in all analyses for which the Kruskal-Wallis test is valid.

- The Rank ANOVA test is generally significant at the $\alpha = 0.05$ level of significance for all statistics and horizons except it is not significant for horizons greater than 1 for the range of percent error measurement.

Scenario 9, Negative Level and Trend Shift

Scenario 9 is like scenario 2 except that both the expected and actual level shifts are negative rather than positive. As with the positive version, the trend shift is small in the example graph.

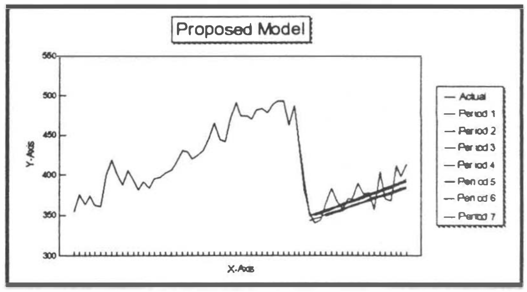

Figure 96

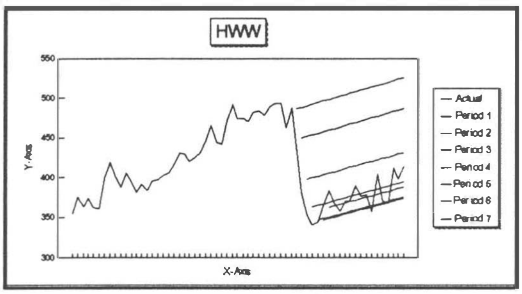

Figure 97

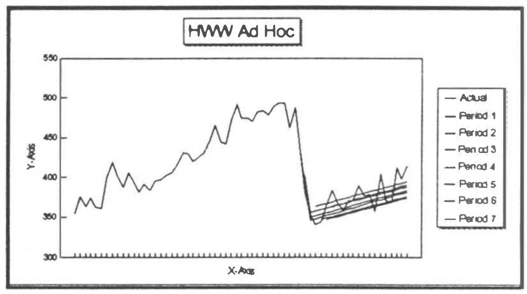

Figure 98

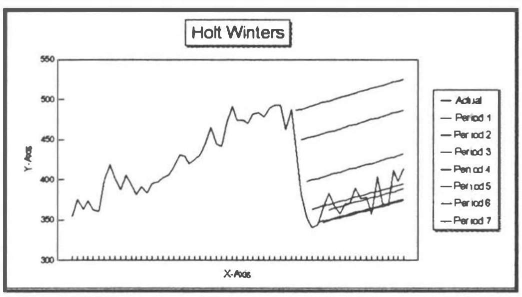

Figure 99

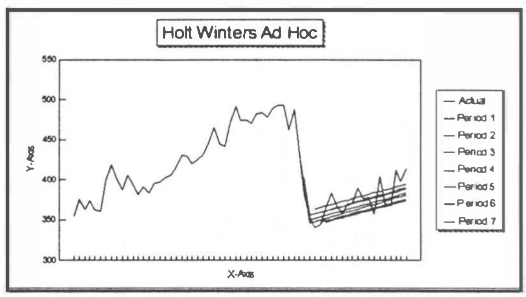

Figure 100

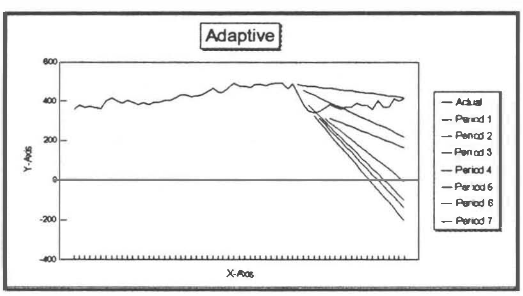

Figure 101

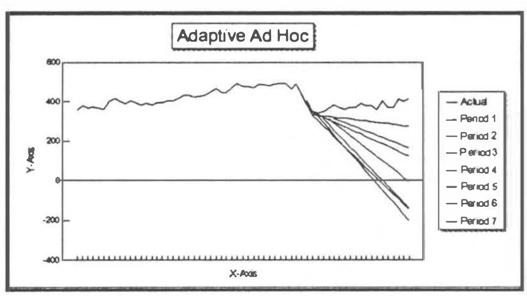

Figure 102

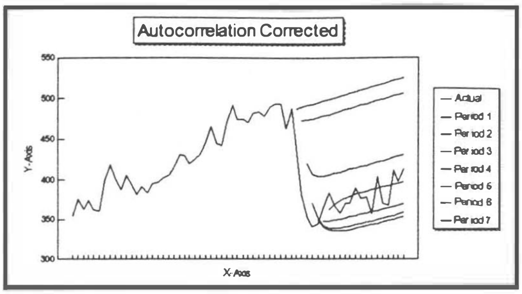

Figure 103

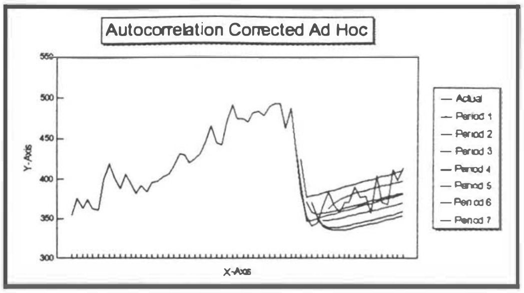

Figure 104

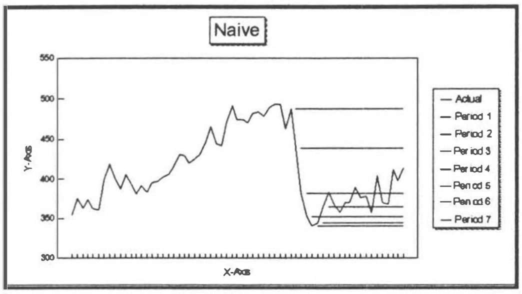

Figure 105

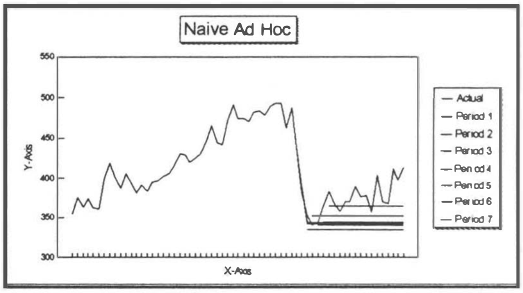

Figure 106

Observed results are as follows:

- The proposed technique is the most effective method across all reported horizons and all reported statistics.

- In general the *ad hoc* techniques rank superior to the unadjusted models.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid.

- The proposed method can be distinguished from all other models on the Kruskal-Wallis related multiple treatment comparison analysis in all analyses for which the Kruskal-Wallis test is valid.

- The Rank ANOVA test is significant at the α = 0.05 level of significance for horizon 1, has mixed results at horizon 5 although it is usually significant, and generally is not significant at the α = 0.05 level of significance for horizons 10 and 15. It is not significant at horizon 5 for the average rank of absolute error and for the median absolute percent error.

When the Level Shift is Larger or Smaller

In the next four models, *the simulated actual level shift occurs when planned, but is significantly different in magnitude from the planned level shift.*

Scenario 3, Level shift at 25% of Anticipation

In scenario 3 a simulated level shift occurs at the anticipated time; however, it is only 25% as large as anticipated.

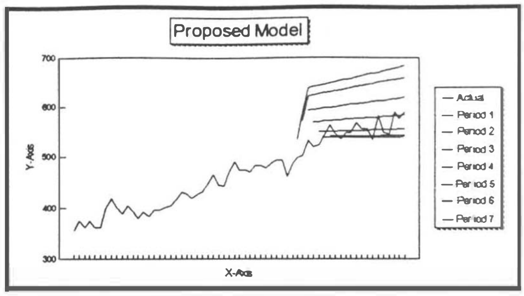

Figure 107

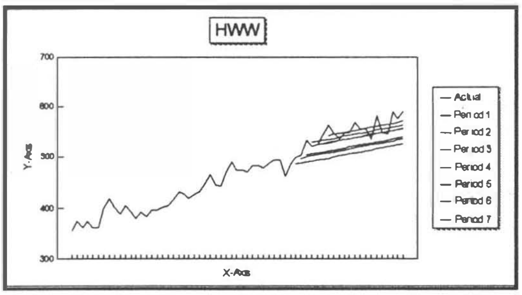

Figure 108

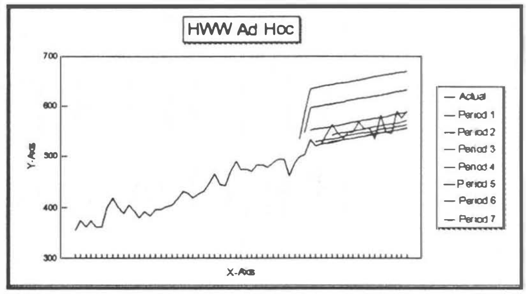

Figure 109

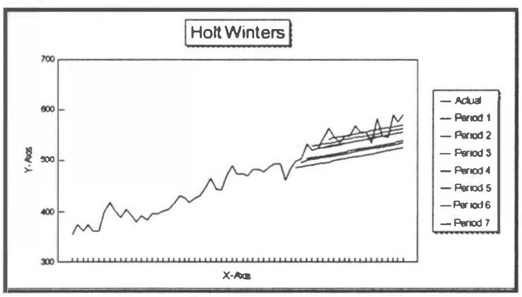

Figure 110

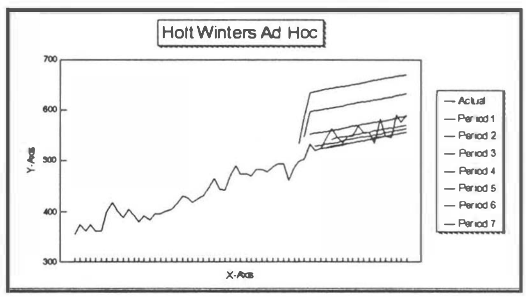

Figure 111

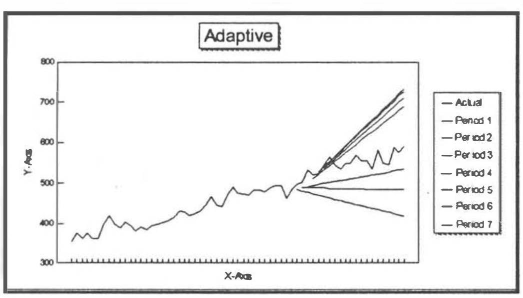

Figure 112

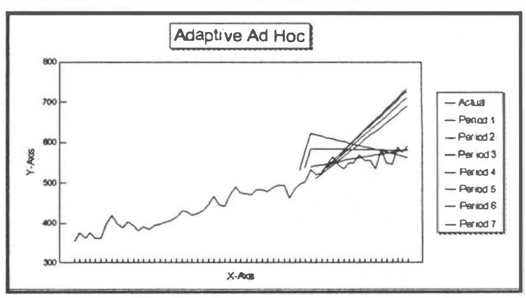

Figure 113

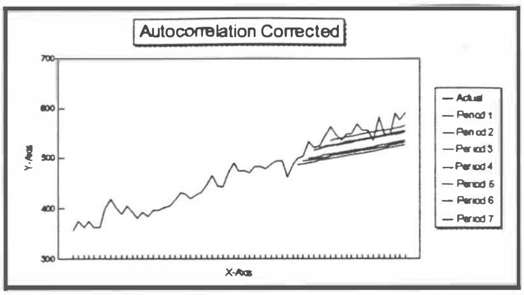

Figure 114

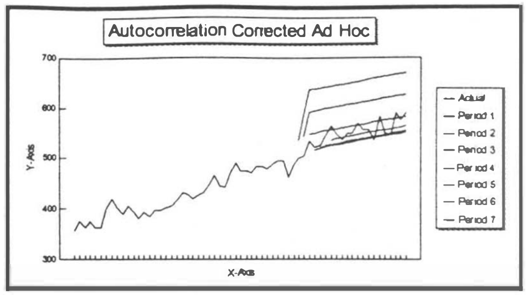

Figure 115

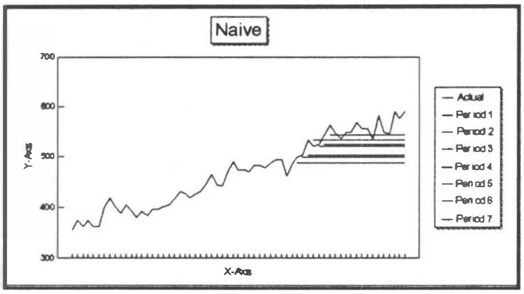

Figure 116

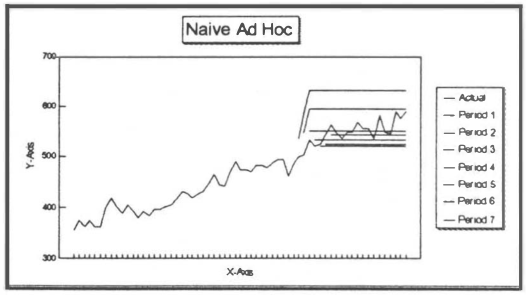

Figure 117

Results for this model are as follows:

- The proposed technique consistently ranks tenth out of the eleven series, only the *ad hoc* Holt-Winters-Williams method is less effective.

- The unadjusted models generally outperform the others.

- The autocorrelation corrected model tends to perform the best.

- Both the Rank ANOVA and the Kruskal-Wallis test significant at the $\alpha = 0.05$ level of significance for horizon 1, both test insignificant at horizon 15.

- In general the Rank ANOVA tests insignificant at the $\alpha = 0.05$ level of significance for horizons 5 and 10 and the Kruskal-Wallis tests significant at those horizons; however, the reader should consult the tables.

- Where the Kruskal-Wallis statistic tests significant, the proposed method is distinguishable from the other techniques with the multiple treatment comparison test, so to is the aforementioned *ad hoc* Holt-Winters-Williams model. However, the autocorrelation corrected model generally is not distinguishable from some or all of the other unadjusted models.

Scenario 10, 25% Level shift, Negative

Scenario 10 is a negative version of scenario 3. The expected level shift is negative. The simulated actual level shift is 25% of the expected level shift.

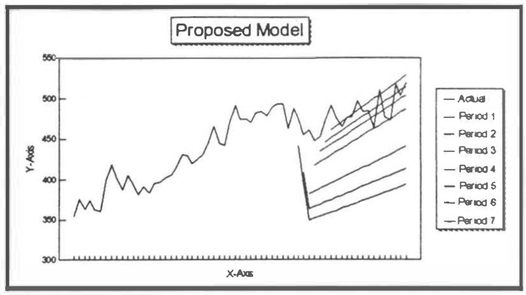

Figure 118

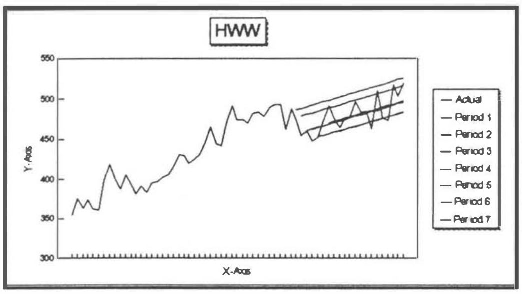

Figure 119

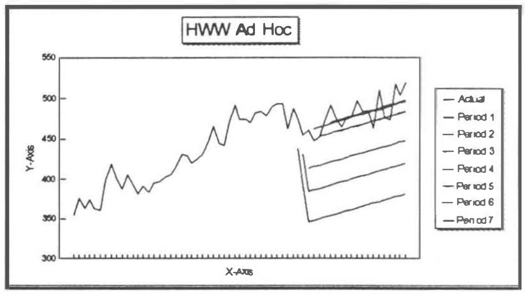

Figure 120

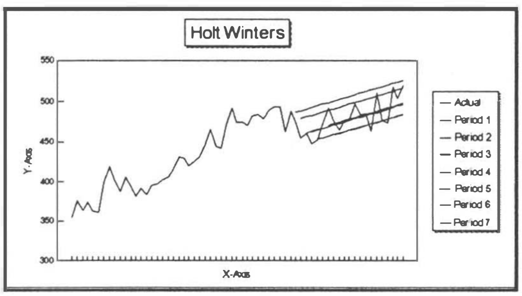

Figure 121

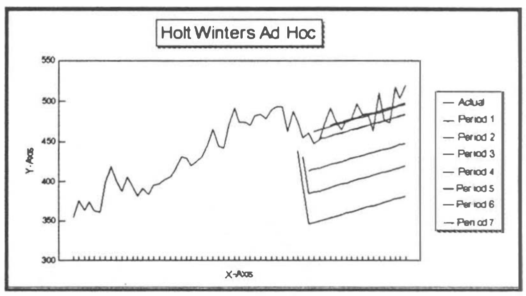

Figure 122

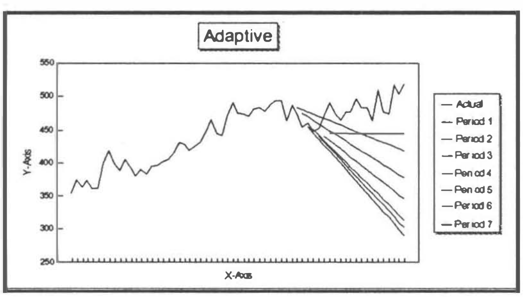

Figure 123

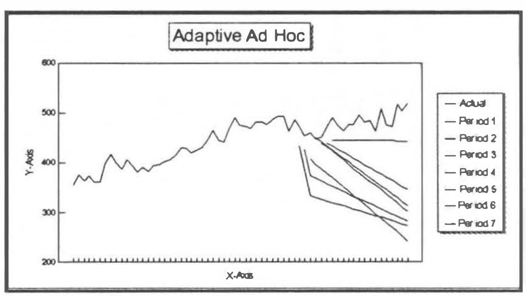

Figure 124

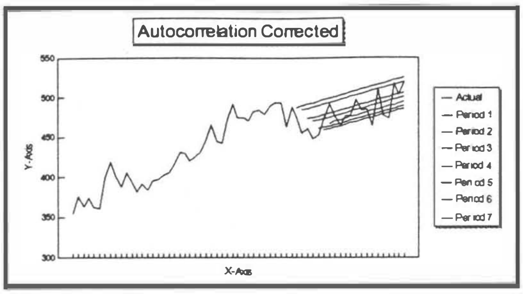

Figure 125

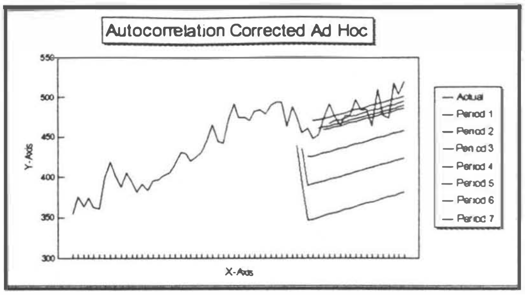

Figure 126

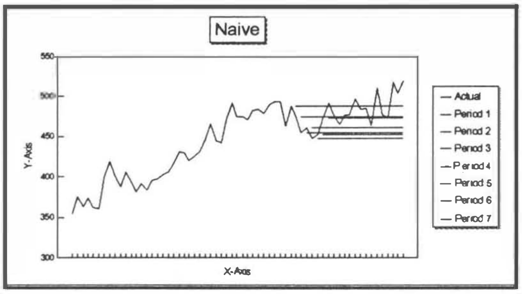

Figure 127

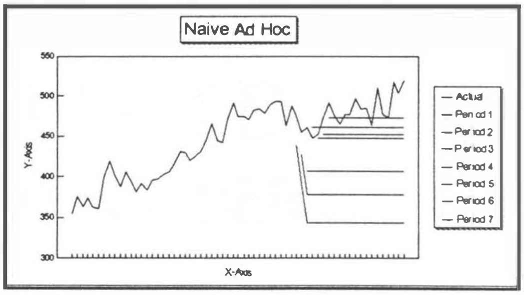

Figure 128

Following are observed results:

- The unadjusted models are more effective than the others.

- In general, for the more distant reported horizons and most reported statistics the autocorrelation corrected model appears most effective. This result is not entirely consistent.

- The Holt-Winters-Williams model is the least effective model for all reported horizons and all reported statistics.

- The proposed technique is generally the next least effective model for all reported horizons and all reported statistics.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid.

- The Rank ANOVA test is significant at the $\alpha = 0.05$ level of significance for all statistics and all horizons.

- The multiple series comparison analysis shows the proposed technique and the Holt-Winters-Williams technique to be distinguishable from the other techniques. In general, the autocorrelation corrected model is not distinguishable from some or all of the other models that do not take the planned level shift into account.

Scenario 4, 200% Level Shift

In scenario 4, the simulated level shift occurs at the anticipated time, however, it is 200% of the anticipated level shift.

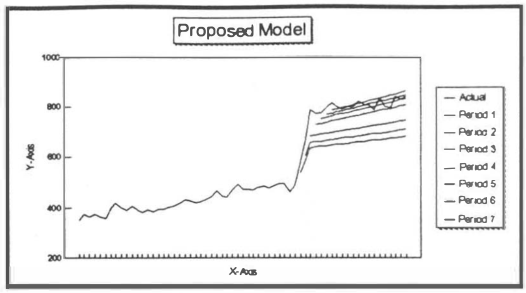

Figure 129

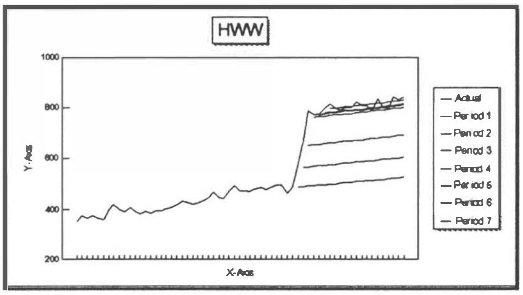

Figure 130

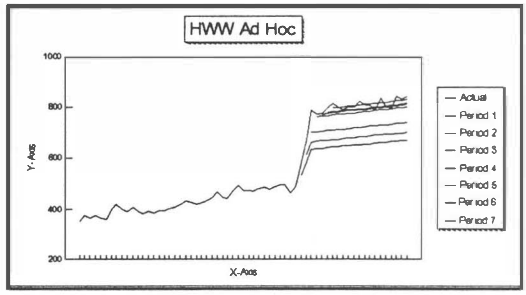

Figure 131

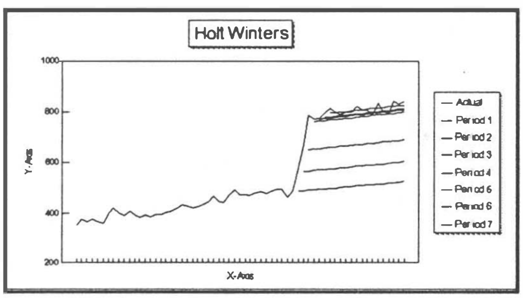

Figure 132

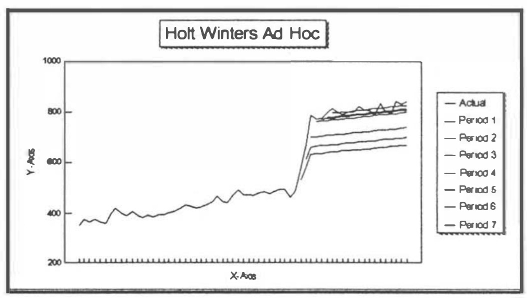

Figure 133

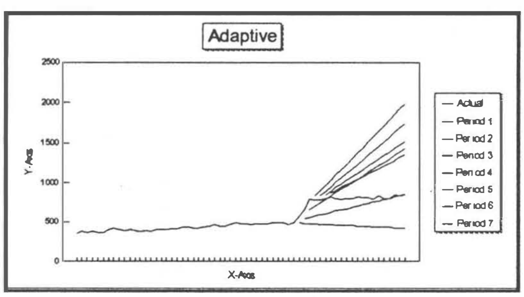

Figure 134

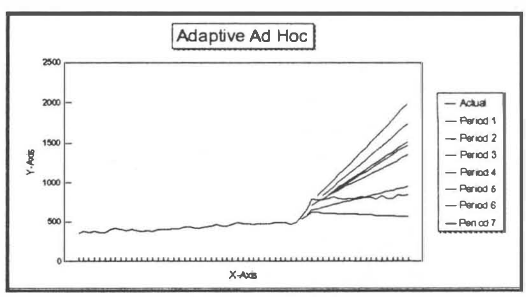

Figure 135

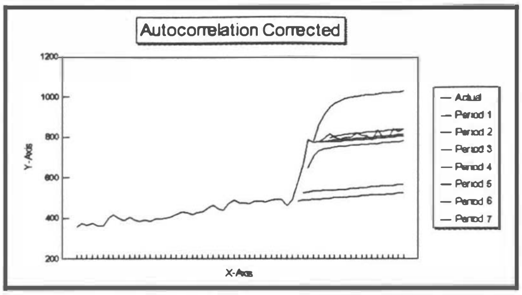

Figure 136

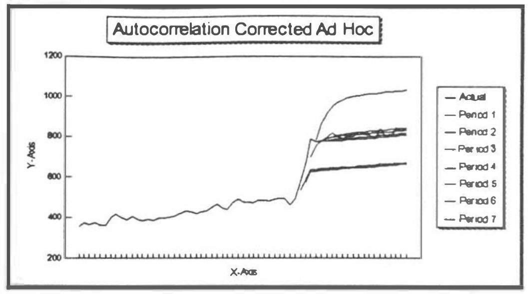

Figure 137

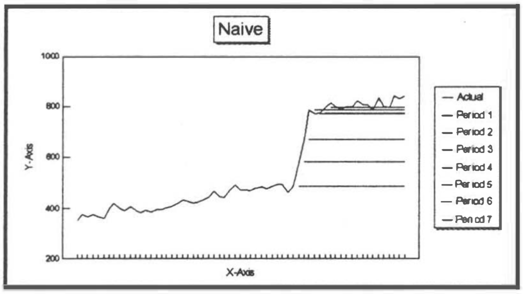

Figure 138

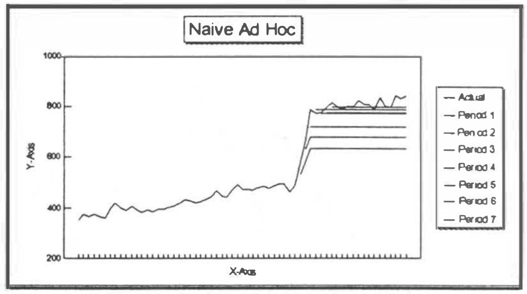

Figure 139

Following are the observed results:

- The *ad hoc* model using the Holt-Winters-Williams technique has superior results for all horizons and most statistics.

- The proposed technique generally performs among the better performing models, with ranks ranging from 1 to 6 and frequently ranking 2.

- The unadjusted naive model has superior results with the median absolute percent error for all horizons.

- In general, but not without exceptions, the models that take the anticipated level shift into account perform better than the unadjusted models. The exceptions are that the *ad hoc* autocorrelation corrected model is a fairly poor performer while the unadjusted naive model is a fairly good performer.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid except for the log mean squared error ratio at horizons 10 and 15.

- The Rank ANOVA tests significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid except for horizon 10 for the log mean squared error ratio and the geometric root mean squared error, and all horizons for the median absolute percent error.

- All models test significantly different from each other, sometimes excepting their nearest alternative by rank or rarely their two nearest alternatives, for all horizons and all statistics where the Kruskal-Wallis statistic tested significant.

Scenario 11, Negative 200% Level Shift

Scenario 11 is the negative equivalent to scenario 4. Both the anticipated and simulated actual level shifts are negative. However, the simulated actual level shift is twice as large as anticipated.

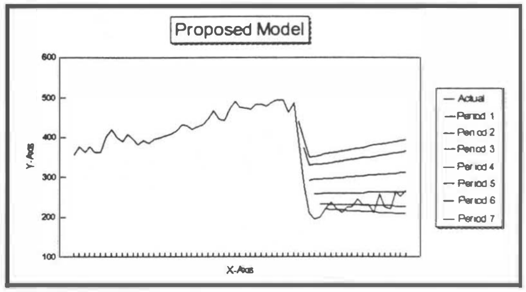

Figure 140

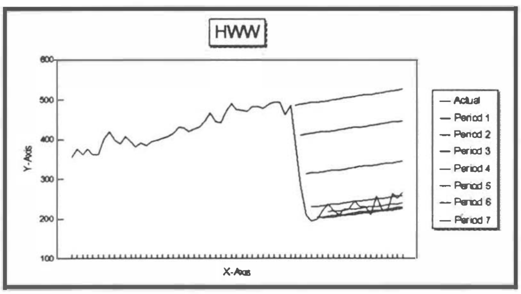

Figure 141

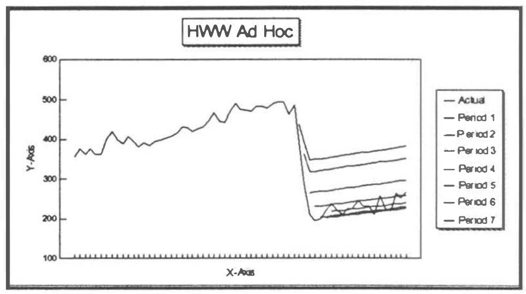

Figure 142

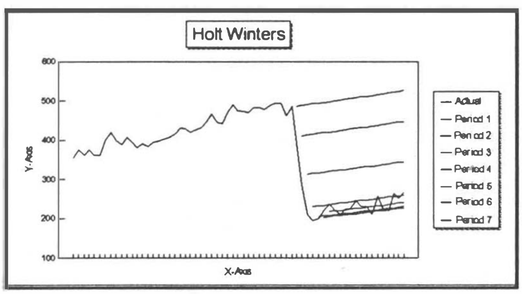

Figure 143

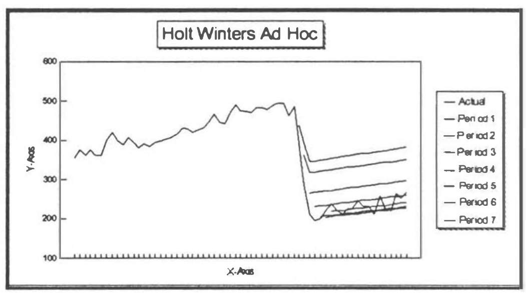

Figure 144

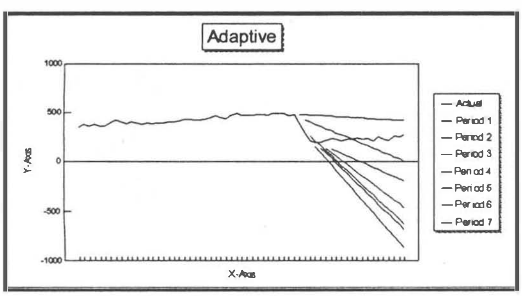

Figure 145

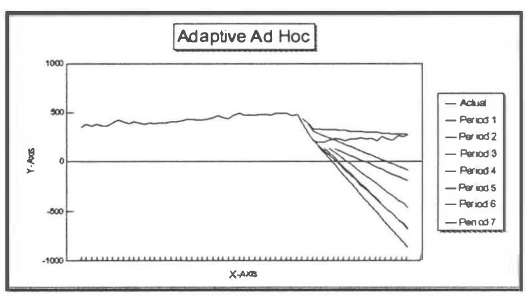

Figure 146

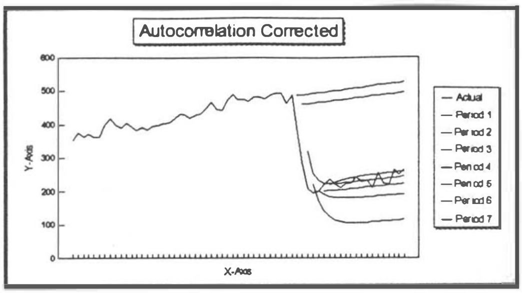

Figure 147

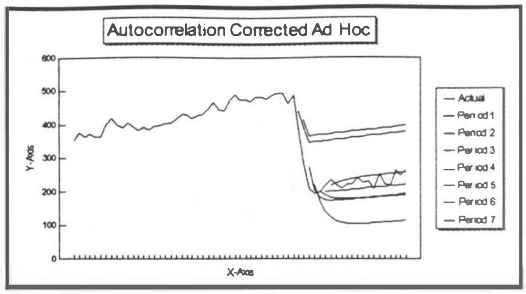

Figure 148

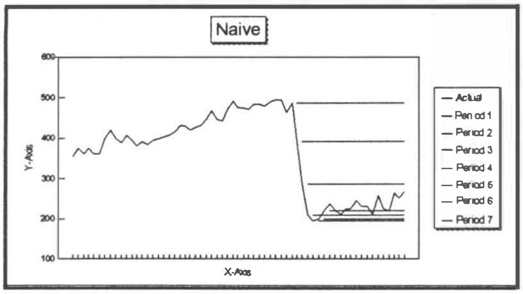

Figure 149

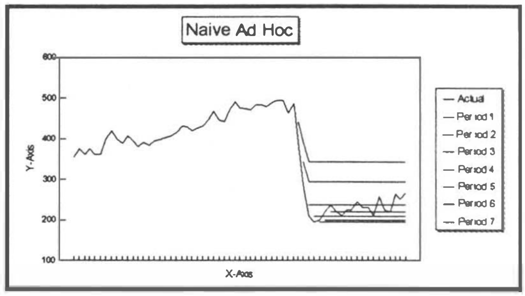

Figure 150

Following are the observed results:

- The *ad hoc* Holt-Winters-Williams technique was clearly superior to all other models.

- The next four models tended to cluster in ranking with the rank order dependent on the statistic presented. These were: the proposed technique, *ad hoc* Holt-Winters, *ad hoc* adaptive, and the unadjusted naive technique. Of these, the proposed technique tended to be rank the best, but not consistently.

- The *ad hoc* autocorrelation corrected model performed very poorly as compared with other techniques that took the anticipated level shift into account.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid with the following exceptions at horizon 1, range of percent error; at horizon 10, log mean squared error ratio and range of percent error; and at horizon 15, symmetrical mean absolute percent error, log mean squared error ratio, and median absolute percent error.

- The Rank ANOVA test is significant at the $\alpha = 0.05$ level of significance except as follows: at horizon 1, range of percent error; at horizon 10, median absolute percent error; and at horizon 15, the test is only significant for root mean squared error, range of percent error, and mean absolute percent error.

- The models are significantly different from 7 to 10 of their alternatives in the multiple treament comparison.

When the Planned Shift Fails to Materialize

The next four scenarios address Hypothesis 1c

> The proposed technique is more accurate than the subset of the alternative techniques that include use of the *ad hoc* method when used to forecast through periods where policy shifts are anticipated and such policy changes fail to materialize.

In these scenarios the proposed level shift does not occur at all. Instead, the anticipated level shift is replaced by a simulated trend or variance shift or, in one case, no simulated data at all is added to the data.

Scenario 5, Trend Shift

In scenario 5 the planned level shift does not occur. Instead, a trend shift is simulated at 25% of the magnitude of the first period of the planned level shift.

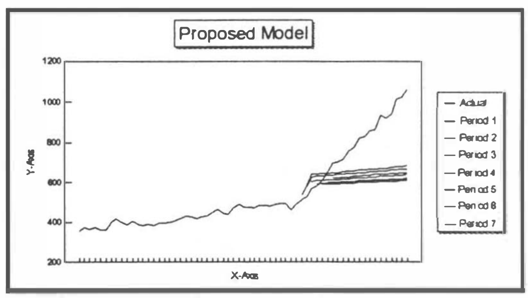

Figure 151

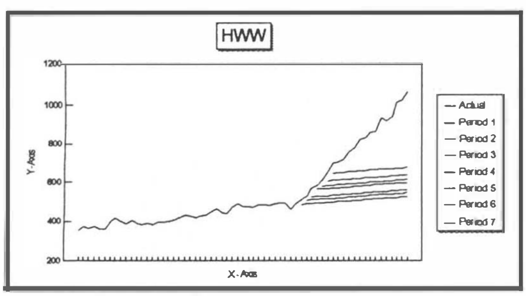

Figure 152

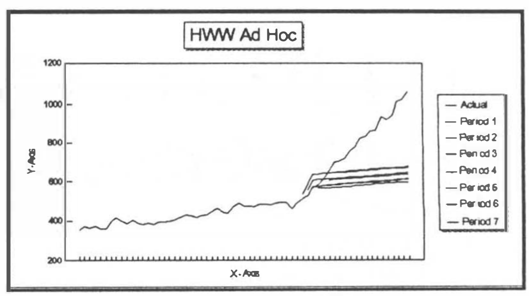

Figure 153

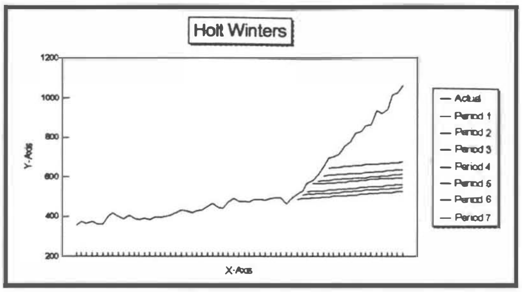

Figure 154

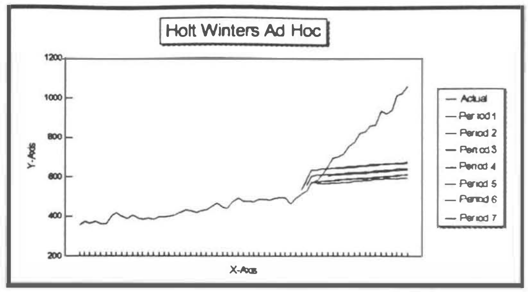

Figure 155

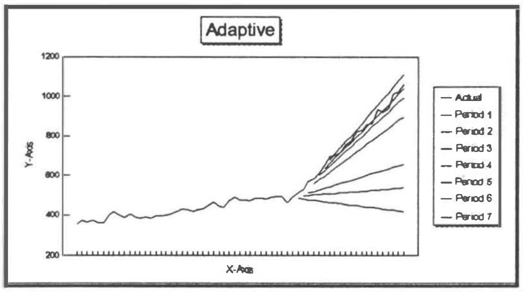

Figure 156

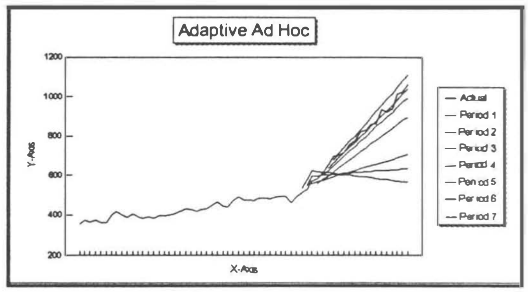

Figure 157

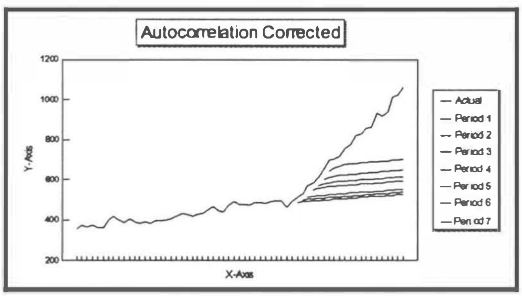

Figure 158

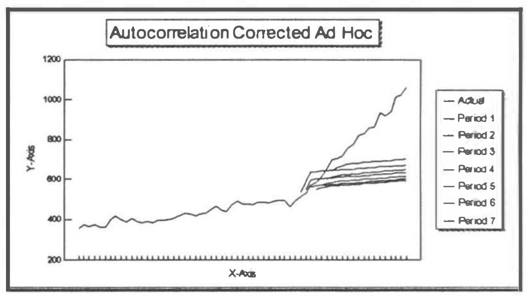

Figure 159

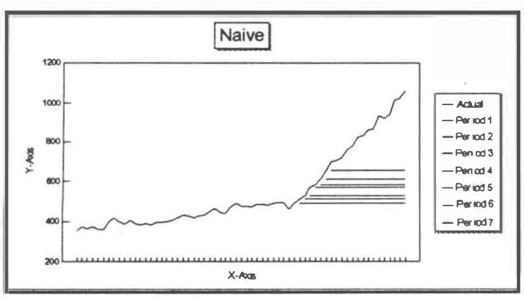

Figure 160

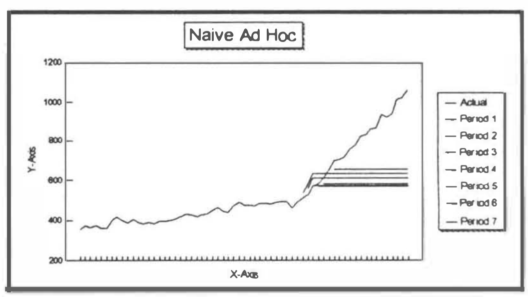

Figure 161

Following are the observed results:

- No technique is particularly good across all horizons and all reported statistics.

- The proposed method is particularly ineffective for horizon 1; however, it ranks about midway for the other horizons for most statistics.

- Other techniques that take anticipated level shifts into account are, in general, more effective than the techniques that do not across all horizons and most statistics.

- The naive technique is particularly ineffective, ranking 11 for all horizons except horizon 1.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid except horizon 1 for symmetry adjusted mean absolute percent error and range of percent error, and horizon 5 for range of percent error.

- The Rank ANOVA test is significant at the $\alpha = 0.05$ level of significance except as follows: at horizon 1, average rank of absolute error, symmetry adjusted mean absolute percent error, and median absolute percent error, mean absolute percent error; and all horizons for range of percent error.

- All models test significantly different from each other, sometimes excepting their nearest alternative by rank or rarely their two nearest alternatives by rank, for all horizons and all statistics where the Kruskal-Wallis statistic tested significant.

Scenario 12, Negative Trend Shift

Scenario 12 is the negative equivalent to Scenario 5. At the time of an anticipated negative level shift, a negative trend shift is simulated.

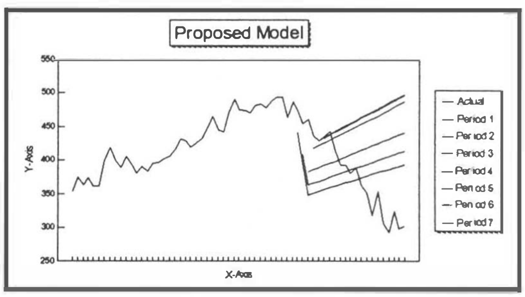

Figure 162

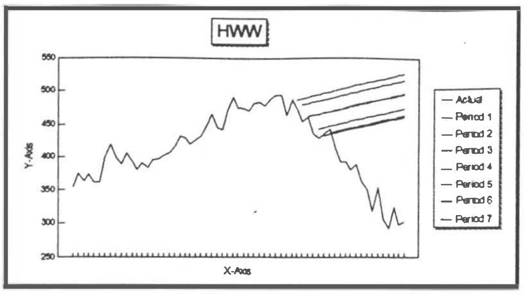

Figure 163

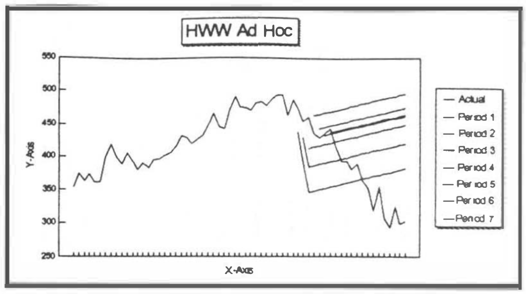

Figure 164

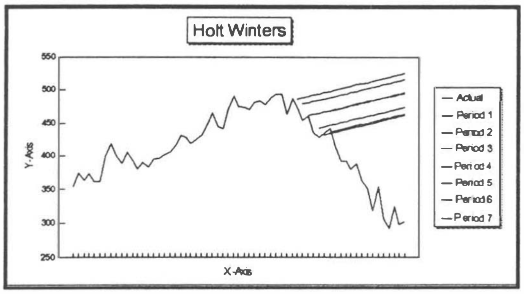

Figure 165

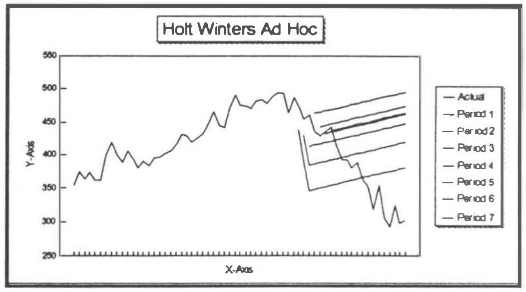

Figure 166

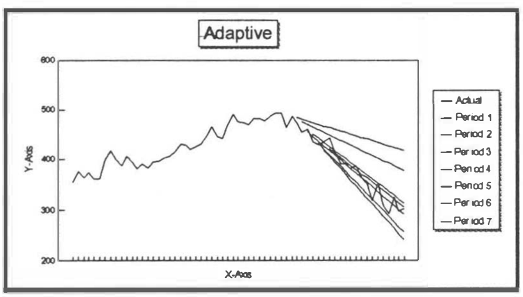

Figure 167

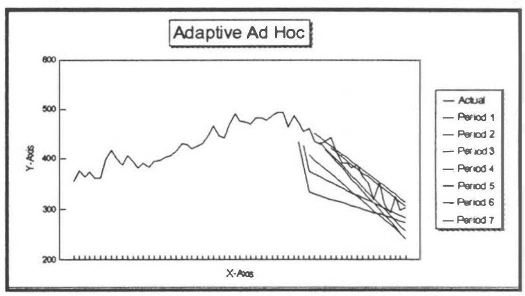

Figure 168

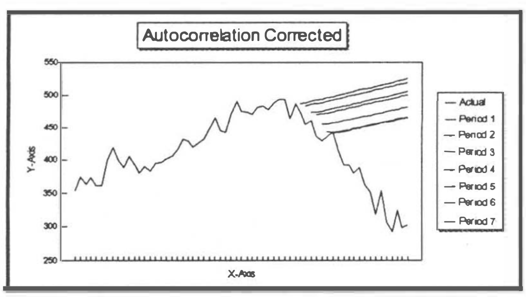

Figure 169

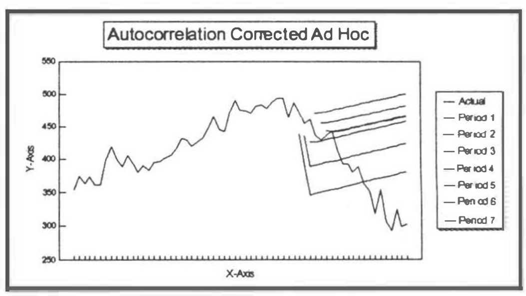

Figure 170

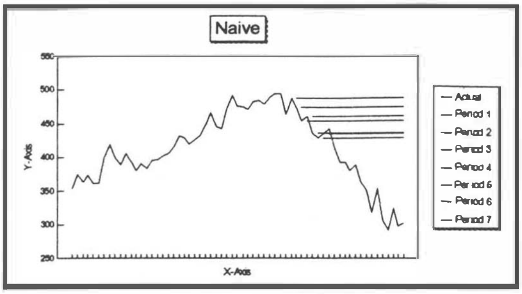

Figure 171

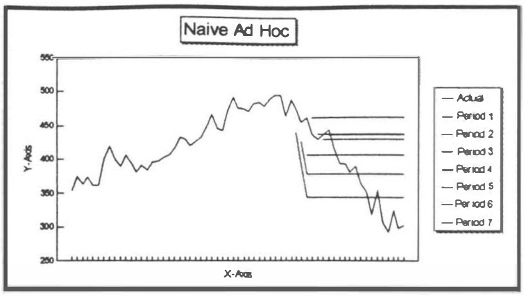

Figure 172

Following are the observed results:

- The proposed technique and the *ad hoc* Holt-Winters-Williams technique perform the poorest for the shorter horizons (periods 1 and 5).

- For those same horizons, in general, the unadjusted models perform better than the *ad hoc* models.

- For longer horizons results are very mixed, although the *ad hoc* adaptive technique does frequently appear to be superior.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid except horizon 15 for the Median Absolute Percent Error.

- The Rank ANOVA test is significant at the α = 0.05 level of significance for horizon 1. Results are very mixed for other horizons.

- In the multiple series comparison analysis, models test different from 7 to 10 of the alternative models for all comparisons where the Kruskal-Wallis result is significant.

Scenario 6, No Change

In scenario 6 no simulated data is added to the actual observations, in effect nothing special happens at the point where a level shift is anticipated.

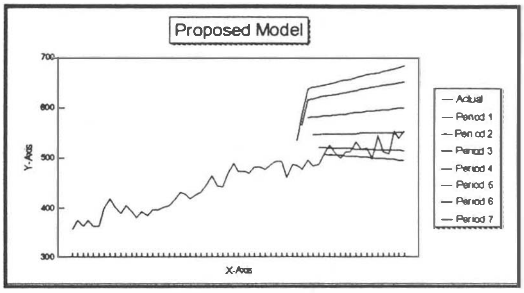

Figure 173

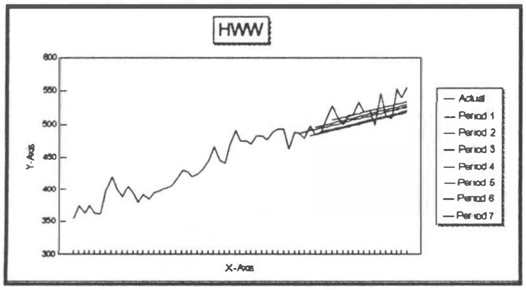

Figure 174

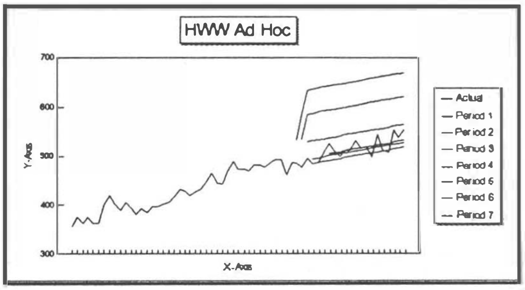

Figure 175

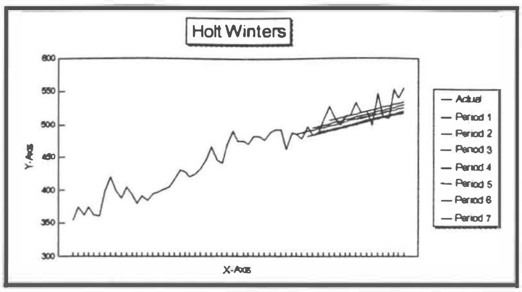

Figure 176

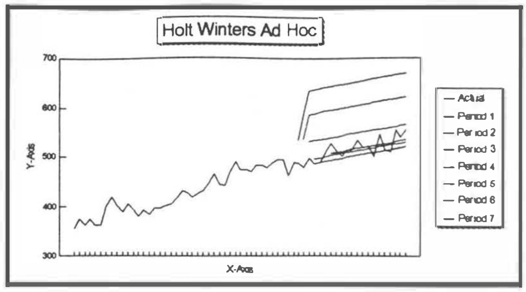

Figure 177

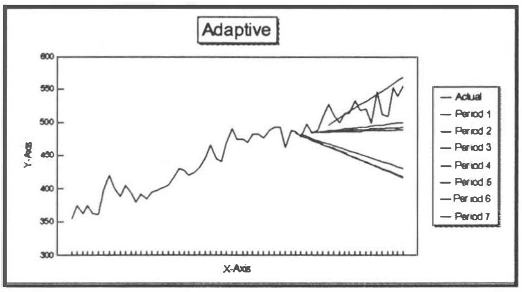

Figure 178

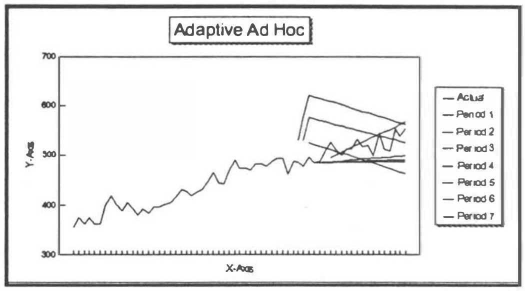

Figure 179

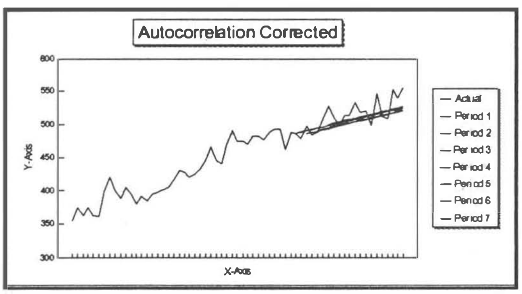

Figure 180

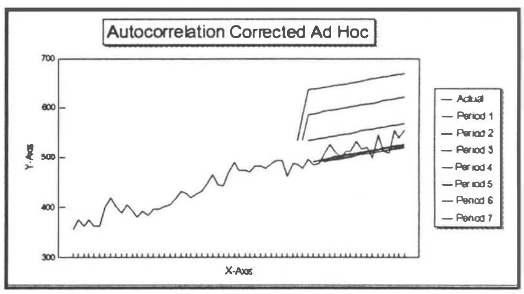

Figure 181

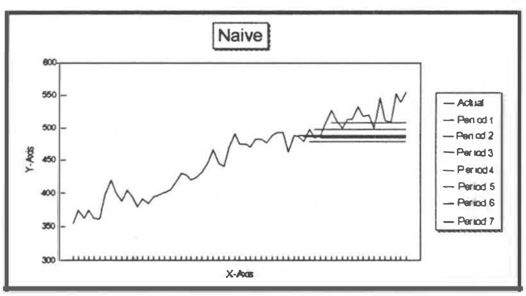

Figure 182

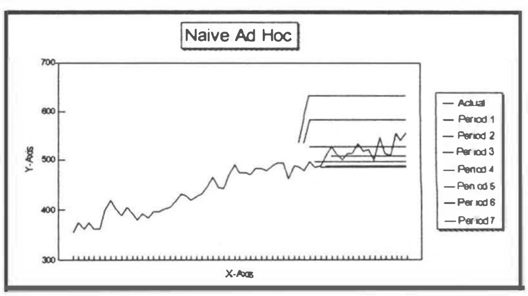

Figure 183

Following is a summary of the observed results:

- The unadjusted models show superior performance across all reported horizons and all reported statistics.

- The rank order of the three most superior models is: (1) the autocorrelation corrected model, (2) the adaptive model, (3) the Holt-Winters models, across all horizons and most statistics.

- The proposed technique ranks tenth out of eleven, just ahead of the *ad hoc* Holt-Winters-Williams model, across all horizons and most statistics.

- Among the *ad hoc* models, the naive technique generally outperforms other naive techniques across all reported horizons and most reported statistics.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid.

- The Rank ANOVA test is significant at the α = 0.05 level of significance for all statistics for which it is valid.

- In the multiple series comparison analysis, models test out as different from 6 to 10 of the alternative models for all comparisons where the Kruskal-Wallis result is significant.

Scenario 7, Variance Shift

In scenario 7 a simulated variance shift that is equal to doubling the amount of variation in the data is added at the point of the anticipated level shift. No simulated level shift is added to the data.

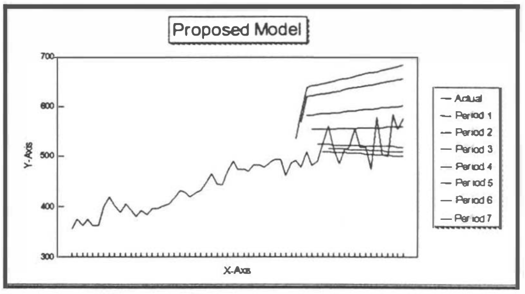

Figure 184

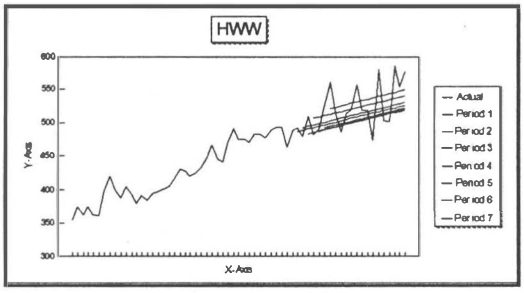

Figure 185

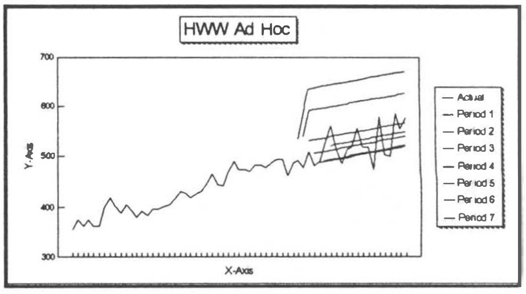

Figure 186

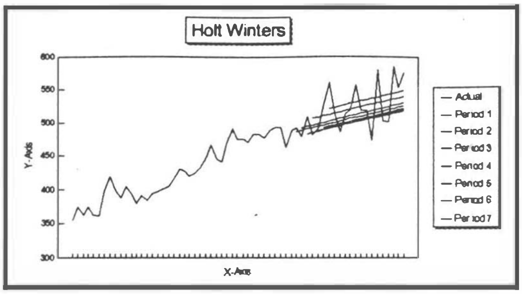

Figure 187

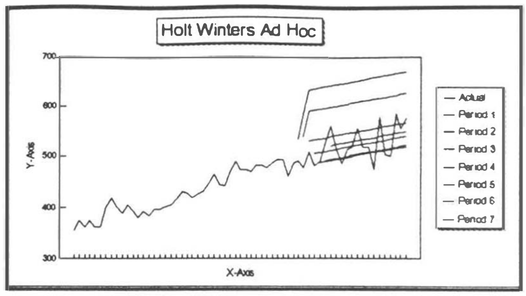

Figure 188

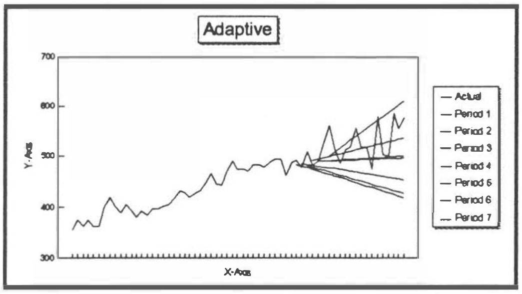

Figure 189

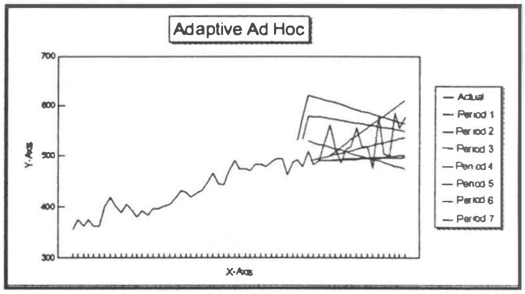

Figure 190

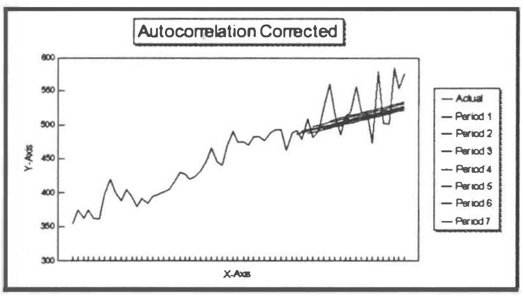

Figure 191

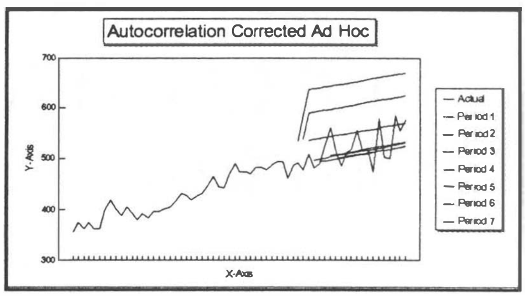

Figure 192

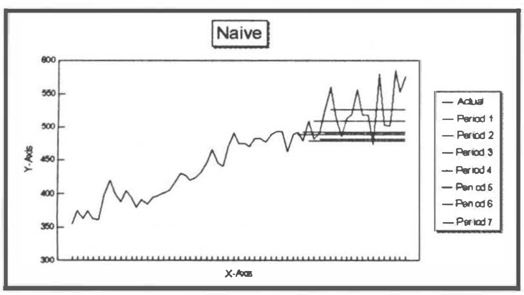

Figure 193

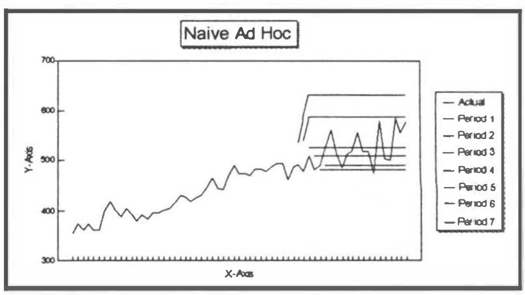

Figure 194

Following are the observed results:

- The unadjusted models are superior to those that include an anticipated level shift across all horizons most statistics.

- The adaptive and the autocorrelation corrected models are most effective.

- The *ad hoc* Holt-Winters-Williams model and the proposed technique are least effective.

- In general the naive model is the least effective of the unadjusted models, while the *ad hoc* naive model is generally more effective than models that take prospective level shifts into account.

- The Kruskal-Wallis test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid.

- The Rank ANOVA test is significant at the $\alpha = 0.05$ level of significance for all statistics for which it is valid for the horizon 1. Results are mixed for other horizons, with the test least frequently significant at horizon 5.

- In the multiple series comparison analysis, models test different from 7 to 10 of the alternative models for all comparisons where the Kruskal-Wallis result is significant except with the median absolute percent error where it ranged to as few as 4.

Scenario 13 Discussed

Scenario 13 address Hypotheses 2a and 2b:

> The alternative techniques and the proposed technique are not equally accurate when used to fit data that has had a level shift in the historical period.
>
> The proposed technique is more accurate than the alternative techniques when used to fit data that has had a level shift in the historical period.

For the last scenario, similar tables are produced and included in Appendix IV. These tables include a comparison of only six models as the retrospective models do not include *ad hoc* models. In this scenario series that have historical level shifts are fit using the proposed technique as a means of explicitly taking historical level shifts into account in exponential smoothing models. These models do not include simulated data. Series are fit through *the*

December 1990 or 1991 as with other scenarios, using the proposed technique during the model fitting stage for the "Adjusted" model. The actual is then updated for six periods.

Because of some of the results, I became concerned that the beta parameter was allowed to be fit for too high a value. I subsequently made a second trial (labeled Scenario 13b in Appendix IV) in which this parameter was restricted to $\beta \le 0.02$; however, the main results were not changed.

The following graphs demonstrate an example of senario 13 (from level shift series 5):

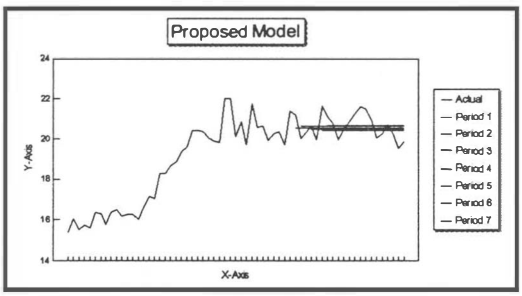

Figure 195

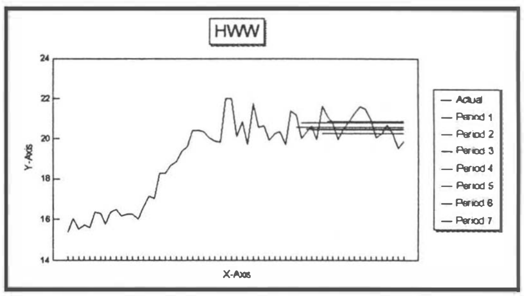

Figure 196

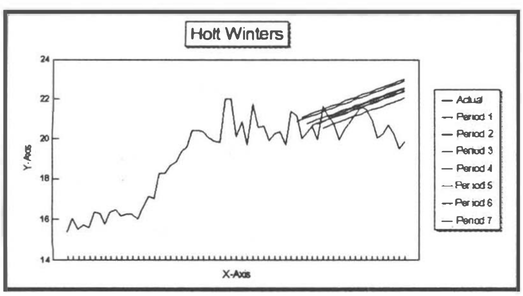

Figure 197

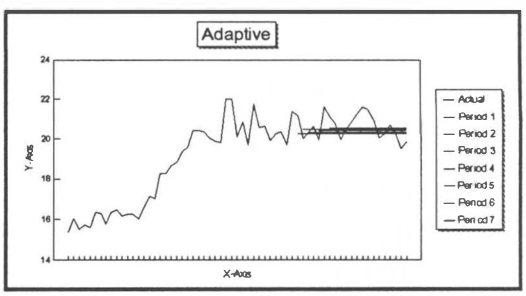

Figure 198

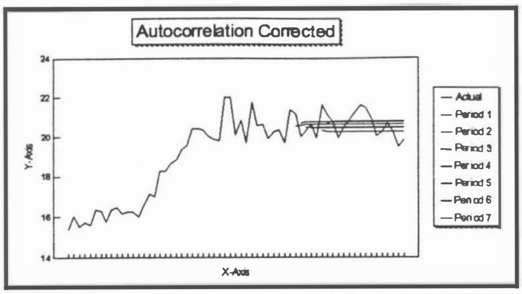

Figure 199

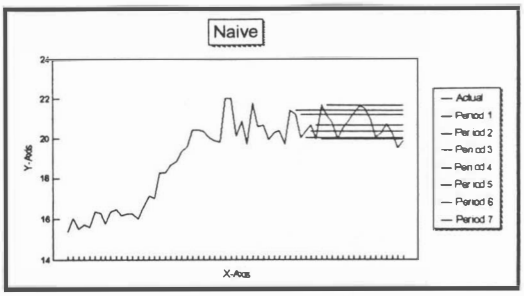

Figure 200

Following are the observed results:

- For horizons 5, 10, and 15, the proposed technique ranked superior for most statistics.

- For horizon 1, the naive technique generally ranked superior for most statistics.

- There was no other discernable pattern of rank order among the techniques.

- The Kruskal-Wallis and Rank ANOVA results are **not statistically significant**.

- These results held for both the initial trial of Scenario 13 and the revised trial with a more restricted β parameter.

Summary

In this chapter I have presented the results from the actual analyses. Detailed tables supporting these results are in Appendix IV.

CHAPTER 9: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

In this chapter I will:

- Provide an overview of the study and its results.
- Provide a discussion of the results presented in last chapter.
- Provide tentative recommendations related to the use of the techniques studied in this dissertation for forecasting when prospective level shifts are anticipated.
- Provide a discussion of the use of the two inferential statistics presented in this dissertation.
- Identify other interesting results of the study.
- Identify areas of needing further study.

An Overview of the Study and its Results

In this study I proposed a technique for incorporating an exogenously estimated level shift into an exponential smoothing model and I conducted two studies to determine whether forecasts made using this technique are more effective than those made with other similar exponential smoothing models. In general, the proposed technique is more effective than other techniques when a simulated actual level shift occurs as expected, even if an unexpected trend shift occurs at the same time. Results are mixed when a level shift occurs when expected, but at a considerably different magnitude. The technique produces a considerably worse result than most other techniques when the level shift fails to occur or manifests itself in an unexpected form such as a trend shift or increase in variability. In a later section of this chapter I argue that this last result can be viewed as a benefit of the technique if one considers it important for the forecaster to be alerted to the fact that the actual events are considerably different from those that are anticipated; however, this benefit would be dependent on further analysis that shows that this result is sufficient to make a difference in the effect of the forecast errors on tracking signals. Results from the second study, which examined whether it is beneficial to use the proposed technique to help fit data series that have historical level shifts, is inconclusive. While the proposed technique results in slightly better forecasts for most statistical measures, actual variation in outcomes is so slight as to cast doubt on whether the proposed technique provided any significant benefits.

Discussion of the Study Results

In the discussion that follows, I will focus on two factors:

- Whether the data analysis warrants some acceptance of the hypothesis with respect to each of the specific scenarios.
- Whether there is a pattern with respect to which scenarios lead to which outcomes such as might lend itself to some guidance for use of these models.

HYPOTHESIS 1a

> The alternative techniques and the proposed technique are not equally accurate in forecasting through periods where policy shifts are anticipated.

The results show that there is a difference in performance between the models for the various scenarios. The Kruskal-Wallis statistics and the Rank ANOVA tests indicate that this difference is significant. For scenarios 1, 2, 8, and 9, i.e., the ones that reflect a simulated actual level shift that compares with the anticipated level shift, the proposed technique outperforms all other techniques. In scenarios 5, 6, 7, and 12, which compare with situations where the level shift does not occur at all or is replaced by some change that is entirely different from the anticipated change, the models that do not include anticipated level shifts perform the best. In general, the proposed method is either the least effective or among the least effective methods under these circumstances. I discuss this further below at the heading "Additional Finding."

The remaining four scenarios include the positive (scenarios 3 and 4) and negative (scenarios 10 and 11) cases where the level shift is considerably less (scenarios 3 and 10) or considerably more (scenarios 4 and 11) than expected. For the scenarios where the level shift was significantly underestimated the models that take an anticipated level shift into account tended to outperform the unadjusted models, although the proposed technique was not the best. However, where the level shift was significantly overestimated the unadjusted models were better.

With respect to the first hypothesis, results were very clear that for each scenario the various models exhibited a pattern of effectiveness, that is the proposed technique was either very effective or very ineffective on all the reported statistics and all the reported horizons. Also, the models that took the prospective level shift into account were all either more effective or less effective than those that did not take prospective level shifts into account.

The general finding is that regardless of the statistic used, the proposed technique differs in effectiveness from the other techniques in a manner that is consistent across various similar scenarios, is consistent with the variation between effectiveness of other methods, i.e., usually follows the same pattern as the difference between other techniques that take anticipated changes into account as compared with those that do no, and is consistent with the common sense expectation of performance, i.e., the proposed technique works when the anticipated change is simulated to actually occur and fails when the anticipated change is simulated to not occur or to vary significantly from the anticipated change.

Nevertheless, with the scenarios that include a change, but not the planned change, some of the results are mixed. Various statistics suggest various techniques to be more effective. Frequently these results are supported by the inferential statistics even when they differ between the various descriptive statistics. This consequence suggests two conclusions: (1) Where significantly conflicting results occur between different descriptive statistics, no particular results should be accepted as superior. (2) It is probably more effective to use multiple descriptive statistics in evaluating various models to avoid being misled by the results of one or two measurements.

HYPOTHESIS 1b

> The proposed technique is more accurate than the alternative techniques when used to forecast through periods where policy shifts are anticipated and such policy changes materialize.

Among the 8 relevant scenarios there are two relevant conditions. Scenarios 1, 2, 8, and 9 simulate accurate or relatively accurate estimates of level shifts. Scenarios 3, 4, 10, and 11 simulate fairly inaccurate estimates of level shifts.

The results are very clear for the accurate estimates of level shifts. In all four of the relevant scenarios, the proposed technique produces superior forecasts by whichever statistic is used to measure accuracy for both near and distant horizons. As examined by scenarios in 2 and 9, this effect is not affected by simulated simultaneous trend shifts as might be expected to arise with actual policy driven level shifts. When this examination is supplemented through the use of non-parametric rank order statistics, extremely high chi squared values suggest that the effect is strong. When there is a reasonable expectation that the externally produced estimate of a level shift is reasonably accurate, the proposed technique can be expected to produce a superior forecast as compared with any other technique examined.

The results are less clear when the estimated level shift is fairly inaccurate. When the simulated actual level shift is twice as large as anticipated it is more effective to use some model in which the level shift is anticipated; however, the proposed method is not necessarily the best model. Other effective models include the *ad hoc* Holt-Winters-Williams model and the *ad hoc* naive model. However, this result does not follow when the level shift is only one fourth as large as expected. In that case, the models in which no level shift is anticipated are the most effective.

HYPOTHESIS 1c

> The proposed technique is more accurate than the subset of the alternative techniques that include use of the *ad hoc* method when used to forecast through periods where policy shifts are anticipated and such policy changes fail to materialize.

The data from scenarios 5, 6, 7, and 12 provide no reasonable evidence that the proposed technique is more accurate or, for the most part, even as accurate as the other *ad hoc* techniques in forecasting through periods of anticipated level shifts when those level shifts *fail to* materialize. In general, where the statistics are in any agreement at all, the proposed technique is among the least accurate techniques under these conditions.

Additional Finding

One of the unanticipated results that has arisen is that when the simulated actual data is considerably different from the anticipated level shift, the proposed technique is generally among the most inaccurate techniques, except where the inaccuracy is in the form of an original underestimation of the actual level shift. On reflection this result is not particularly surprising for two reason:

1. When the level shift is considerably less than the proposed level shift (less than 50% of the proposed level shift), it is natural that the errors from the forecasts that include the level shift would be greater than the errors from the forecasts that do not include the proposed level shift.

2. The parameter setting rules allowed the level parameter for the proposed technique to be set quite low, so when its errors became relatively large it still was not necessarily able to rapidly correct in the direction of the smaller level shift. Meanwhile the *ad hoc*

adjustments were rapidly eliminated from the other models with the first three updates allowing those models to correct to the small level shift much more rapidly.

These findings show that the proposed method is not the most accurate technique when the level shift fails to materialize or materializes in an unexpected way (much smaller, trend shift, or variance shift). **This result is not necessarily undesirable.** When these conditions arise, there is truly an unexpected event underway. In a sense there is something wrong with a forecast that is not adversely affected by data that indicates the materializing future is considerably different from the expected future. This is not because the forecast is wrong, but because it is inexplicably right.

While getting the future right is an objective of a forecast, it should not be its only objective. At the very least, the forecaster should want to be able to replicate the success with additional forecasting. A forecaster should want to know that the data that is being forecast is not behaving as expected. He may be able to find this out without depending on forecast errors (e.g., he may receive a management report that says a policy implementation is delayed). However, it is also possible that he will be dependent on the forecast model to alert him to such unexpected outcomes. In fact, management may look to the forecaster for signals that activities are off track, particularly where there are a large number of activities underway. If a forecast fails because expected events fail to occur as expected, the objective of getting the future right may have failed, but the objective of helping management manage may still be met.

If the forecaster is dependent on the model itself to alert him to the presence of unexpected events, large errors or patterns in errors are desirable when such unexpected events occur. In this case, the proposed method's relatively poor results with descriptive statistics under conditions where the expected level shift does not occur, or is considerably different from the expected level shift, is a benefit rather than a deficit. It suggests that the proposed technique contains, and efficiently summarizes, information that might be developed into a tracking signal that would alert the forecaster to the failure for the expected level shifting event to materialize. Other techniques which fail to detect abnormal conditions, i.e., perform relatively well when, in fact, the expected future fails to materialize, are, on this view, relatively less desirable.

HYPOTHESES 2a and 2b

> The alternative techniques and the proposed technique are not equally accurate when used to fit data that has had a level shift in the historical period.
>
> The proposed technique is more accurate than the alternative techniques when used to fit data that has had a level shift in the historical period.

In scenario 13, 20 data series are fit to each of the six basic models across a period in which there is a historical level shift. Because I suspected possible distortions from the fitting of the β parameter, I fit two versions of scenario 13, in the first, I fit the model with the same grid as with the other scenarios. In the second, I restricted β to not greater than 0.02. Actual results from both are demonstrated, in the Appendix IV; however, results are very similar for the two versions.

The results weakly support the view that the proposed technique can be used to assist in fitting data series that have undergone level shifts, particularly where the forecaster is interested in the longer horizons. The inclusion of a level shift within the model has an effect similar to an intervention variable in a regression or an

ARIMA model. However, the results are not strong enough for statistical significance in either of the non-parametric comparisons. These statistical comparisons suggest that any benefit from using the proposed technique in the model fitting stage is at best relatively weak, particularly since much more significant results were found in the first 12 scenarios. At worst, however, there is little evidence that the proposed technique provides a worse result than do other models.

An Interesting Result

This lack of statistical significance with scenario 13 where the proposed method is used to fit the data series is perhaps the most surprising result of this study. It is particularly surprising because the minimized root mean squared error used in fitting the proposed model is considerably smaller than the minimized root mean squared error for the alternative model for most trials. The selected parameters and model fitting statistics for this study are shown in Appendix V which is summarized in the Table 4 below.

Table 4 Fit All Trials	**Geometric Mean RMSE**	**---Average--- SPE**	**SMPE**	**Geometric Mean SMPE**
Adaptive	50.53	0.13%	5.93%	3.31%
Autocorrelation Corrected	52.67	0.80%	6.49%	3.52%
Holt-Winters	52.23	0.47%	6.53%	3.67%
Proposed Technique	**24.51**	**-0.06%**	**4.85%**	**2.19%**
Holt-Winters-Williams	52.28	0.49%	6.48%	3.64%

However, this result can be seen as consistent with other findings in the literature:

- Everette Gardner and Spyros Makridakis[205] find that success at model fitting is not necessarily a good indicator of forecast accuracy. In this study, the small values of the minimized root mean squared errors clearly shows the proposed technique leads to superior results in model fitting. However, in this scenario, no statistical difference could be found in forecast accuracy with the seven updates.

- Spyros Makridakis and Michele Hibon[206] find no particular advantage in forecast model initialization. The use of the proposed technique in model fitting has a similar effect to model initialization, particularly where the historical data stretches for a large number of periods after the historical level shift. The natural tendency for the model to catch up to the level may dominate when enough periods have passed between the level shift and the end of the historical data.

- Steven Hillmer[207] finds that in exponential smoothing models most of the effect of additive outliers occurs in the next period after the outlier. In effect, the effect exponentiates away for later periods. It seems reasonable, particularly considering the Makridakis and Hibon result, to expect a similar exponential decline in the errors arising from a level shift. Thus, when the level shift is not near the end of the historical period, it has little influence on the forecast.

- Spyros Makridakis, *et. al.*, say: "As a rule of thumb 8 to 3(L) data points are adequate for initial estimation purposes (where L is the length of seasonality)."[208] This result is consistent with Hillmer's result and suggests that the main impact of the level shift is in the first few periods after the shift.

- George C. Canavos and Don M. Miller[209] demonstrate that as α increases, the weight placed on older observations declines dramatically in simple exponential smoothing.

Where α is as great as 0.3, the entire weight placed on all observations exceeding 6 periods (counting the current period) less than 12%. Where α is as great as 0.5, this weight declines to 1.5%. These numbers would have to be adjusted for Holt type models, but they represent the same basic phenomena. Consequently, where the level shift is more than 6 periods old and α is as great as 0.3, the level shift is discounted to a proportionate share of about 12% or less of the overall weighted average projected in the exponential smoothing model. Thus, the use of the proposed technique is unlikely to significantly impact the accuracy of a forecast that has had a level shift 10 or 15 periods or more before the updating period (future period), unless the optimal α would be set particularly low.

When considered from this perspective, the lack of significance in results of this scenario is not unusual. These articles suggest that **results might be significant for forecast made soon after a level shift**. In the following graphs (level shift series 11), the level shift occurs immediately before the updating period. While one series is not sufficient to reach a conclusion, they are very suggestive. It can be seen that the forecast from the proposed model is much more in line with the actual data than the forecasts made with any of the alternative models.* The only other model that comes close is the naive model.

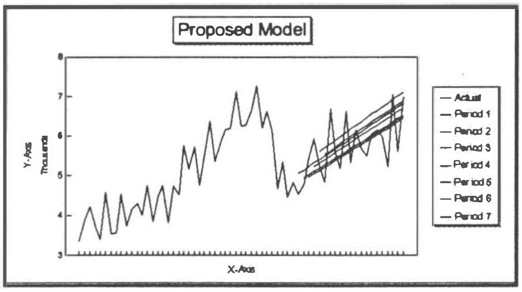

Figure 201

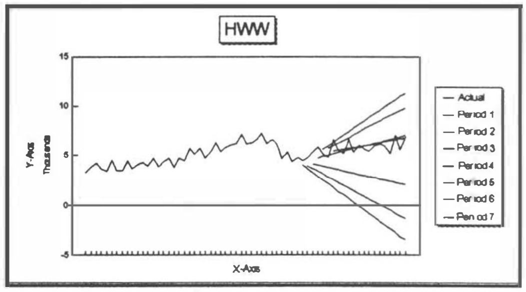

Figure 202

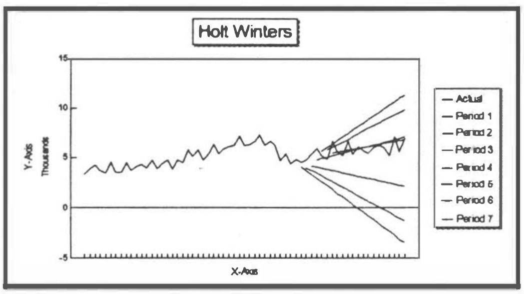

Figure 203

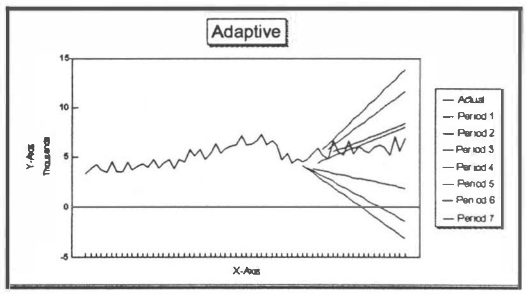

Figure 204

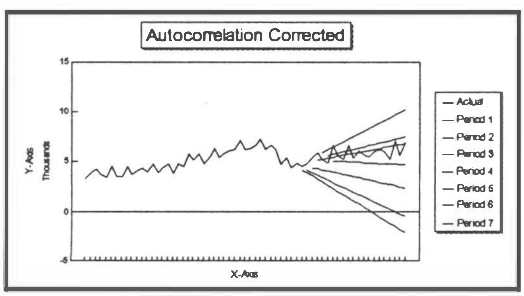

Figure 205

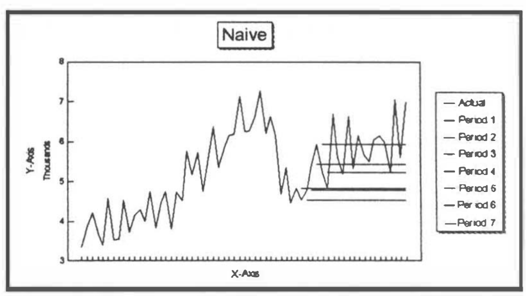

Figure 206

*The forecasts that go below zero most likely would be dampened - see Appendix I - in practical forecasting environments; however, to determine the size of the error generated by the technique, they were allowed to go below zero in this study.

On the other hand, even a dramatic change in level may have little impact after a considerable period has passed as is shown in the following graphs (level shift series 15).

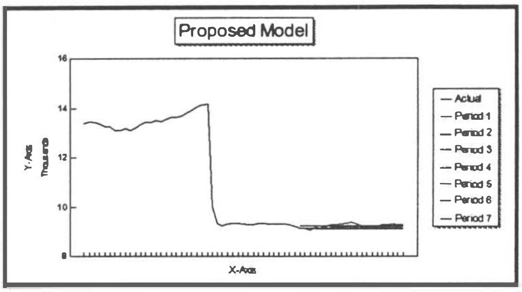

Figure 207

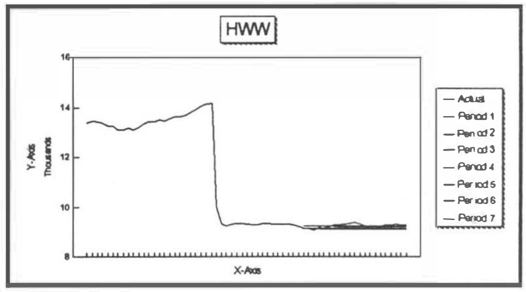

Figure 208

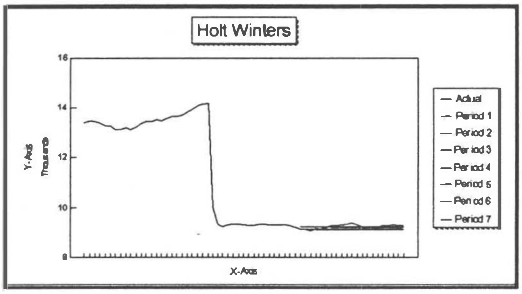

Figure 209

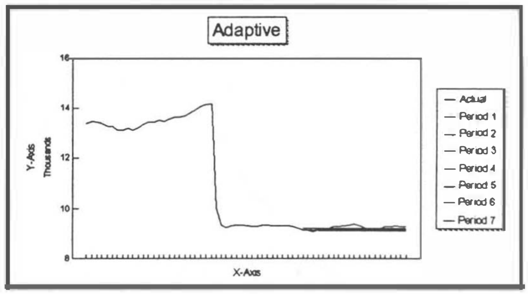

Figure 210

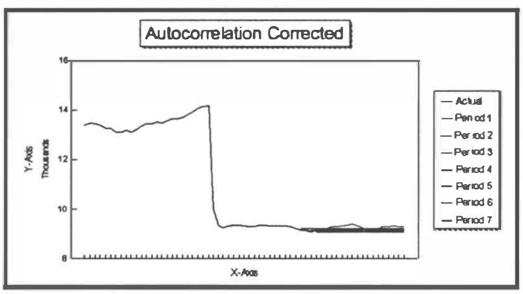

Figure 211

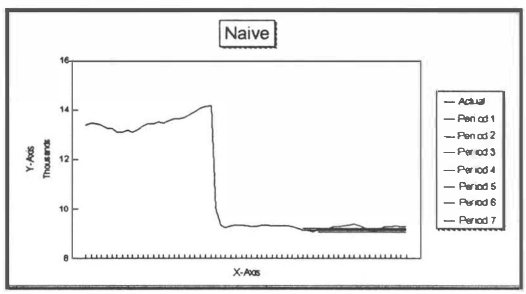

Figure 212

Based on this review, the proposed technique will allow the forecaster to achieve a better model fit; however, that result may not significantly influence the projection of the future. It appears that projections made shortly after the level shift are improved through the use of the proposed technique while projections made well after the level shift are not influenced as much.

Final Remarks about the Proposed Method

The proposed method seems to be a beneficial modification of exponential smoothing forecast models for forecasting when prospective level shifts are expected for the following reasons.

- When the anticipated level shift occurs as expected, the forecast made and updated using the proposed technique is more accurate than forecasts made with other techniques considered in this study.

- When the anticipated level shift occurs, but not as expected, there is little reason to suspect the proposed technique is particularly worse than *ad hoc* techniques considered in this study. The only severe failure of the proposed technique arises when the level shift is considerably smaller than anticipated; however, all forecasts that anticipate level shifts fail with that condition.

- When the anticipated level shift does not occur, forecasts made and updated with the proposed method are likely to have more dramatic errors than forecasts made with other techniques considered in this study. This may serve to alert the forecaster that the future being experienced (at the time of the updates) is not the future that was expected.

- Forecasting with the proposed technique is preferable to the *ad hoc* technique because of an efficiency it produces. With the *ad hoc* technique the forecaster makes a forecast with an exponential smoothing model, in all likelihood using computerized software. He then adds the *ad hoc* adjustment to the result to produce a complete forecast. It is likely that the *ad hoc* adjustment step involves a manual intervention into the forecast, for example, the forecaster may make a forecast using a computerized exponential smoothing model, get a printout of the result, and key both the result and the *ad hoc* adjustment into a spreadsheet to get the final forecast. When the proposed technique is used, the level shift is built into the computerized exponential smoothing model thereby eliminating the manual intervention step.

- The proposed technique also produces an efficiency as compared with other forms of forecast intervention models (e.g., ARIMA models with intervention variables, Duk Bin's model, or the Carreno and Madinaveitia model). This efficiency is found in the permanence of the level shift achieved with the proposed technique. For other intervention models, a level shift is accomplished through the addition of a level shift factor to the base line forecast level, sometimes through the multiplication of a coefficient times a dummy variable. The level shift stays with the forecast only for so long as the addition occurs. To make the level shift permanent, the level shift factor must be individually added to each future observation to the end of the forecast horizon. If more than one level shift occurs, a separate factor must be carried for each, which can lead to the development of fairly complicated models if the data series is subject to considerable policy intervention.

  The proposed technique is much more mathematically efficient. When a level shift occurs, the baseline forecast level is adjusted by the magnitude of the level shift. This new level then **becomes the base line level** of the exponential smoothing model. The level shift stays in the forecast unless it is explicitly taken out. No additional intervention is required to keep the level at the higher (or lower) level for future periods. Additional level shifts can be added to the model in the same manner, adding complication only for those periods during which there are expected level shifts. Thus, the proposed technique provides a comparatively simple intervention model.

- As shown with scenario 13, although there is no requirement to keep the level shift in the model long after the shift has occurred, there is no apparent disadvantage in doing so. Thus, the model built with the prospective level shift that later occurs as expected, will, after enough updates, contain a retrospective level shift that assists with model fitting and may improve forecast accuracy. No additional effort is required except where the level shift does not occur as expected.

Recommendations for Forecasters

Where a forecaster who is using an exponential smoothing model has a reasonably reliable externally supplied estimate of an expected level shift, it would be reasonable for him to modify his exponential smoothing model to apply the method proposed in this dissertation. This application should result in a more accurate forecast as the model is updated while the level shift takes place. Further, the forecast should fail more dramatically should the expected level shift fail to take place. This more dramatic failure might provide the forecaster with a better opportunity to discover that the expected change failed to take place using the forecast errors as a tracking signal.

Where reliability of the externally supplied level shift estimate is unknown or thought to be fairly low, the forecaster may reasonably hesitate to use the proposed technique as other techniques may be more effective. If the forecaster suspects that the externally supplied estimate is smaller than the level shift that may actually materialize, either the proposed technique or the Holt-Winters-Williams technique may be most effective. If he suspects that the externally supplied estimate is significantly larger than the level shift that will actually materialize, it may be best to leave the level shift out of the model unless he wants to use the model for the tracking signal effect briefly mentioned in the last paragraph.

Guidelines

The following guidelines should assist the forecaster:

- **Prospectively**:

1. If you are **reasonably confident in the externally supplied estimate** and that a level shift is the primary consequence of the policy change, use the proposed technique. It will result in a more accurate forecast.

2. If you are **not confident in the externally supplied estimate**, you must choose the technique you use based on goals:

    - If your goal is to have the **most accurate forecast** at all times, you must consider what sort of uncertainty you have:

        - If you suspect that the **policy change will not occur at all** or will materialize in an unexpected manner, leave the forecast alone, possibly after adjusting the level parameter to a high number such as $\alpha = 0.8$.[**]

[**] I did not examine the consequences of simply leaving the model alone; however, it seems likely that raising the level parameter is preferable unless the expected level shift has an insignificant magnitude or has essentially **no** chance of occurring.

- o If you suspect that the **externally supplied estimate is significantly too large**, do not adjust the forecast for the anticipated level change. Leave it alone, possibly after raising the level parameter to a high number such as $\alpha = 0.8$. Alternatively, you may want to correct the externally supplied estimate and select the first option above.

- o If you suspect that the **externally supplied estimate is much too small**, adjust the forecast using an *ad hoc* technique or the proposed technique. Possibly you should also raise the level parameter to a high number such as $\alpha = 0.8$. Alternatively, you may want to correct the externally supplied estimate and select the first option above.

- o If you do not know what to suspect, see the next goal.

o If your goal is to **monitor the data generating function** through the forecast as well as to forecast accurately, use the proposed technique.

If the forecast fails, it may be a signal that the policy change did not take effect as expected.

- **Retrospectively**:

1. If you are **updating** a model that used the proposed technique to include an approximately accurate historical level shift, leave it in. It may improve the forecast if the level shift is recent, or if it does not improve the forecast, there is no evidence that it will make the forecast worse.

2. If you are **fitting** a forecast model to a data series that has a historical level shift:

    - If the **level shift is relatively recent**, include the level shift through the proposed technique.

    - If the **level shift is relatively old**, it is not clear that including the level shift is beneficial for forecasting, although it may improve the model fit. On the other hand, the study does not show that including the level shift will result in a poorer forecast.

- "Relatively recent" and "relatively old" are largely dependent on the parameter selection which is the objective of fitting the forecast. It will be hard to make the judgement suggested above without first fitting the forecast, thus, it may be more effective to always include the level shift in the initial fit of the data.

- The proposed technique affects the fit. **Whatever you do, do not confound the fitting issue by including the level shift through the proposed technique for some parameter combinations and excluding it for others.**

Kruskal-Wallis and Analysis of Variance by Rank

The two non-parametric statistical tests produced overwhelmingly significant results except with scenario 13 where they proved not significant with all statistics and all trials. This last result led me to suspect that there might be something wrong with scenario 13 and after some investigation I came to suspect either (a) the models in scenario 13 were allowed to fit to excessive β parameters, or (b) that the statistics were sensitive to the number of treatments (models) considered. I examined the first suspicion by reducing the possible β parameter range and rerunning this trial. Results have been presented as scenario 13b and are not significantly different from scenario 13. I examined the second suspicion by excluding 5 non-*ad hoc* models from scenario 1 and calculating the Kruskal-Wallis and Rank ANOVA statistics for the rank of absolute error comparison (equivalent to Tables 1-1 and 1-9 in Appendix IV). These results are shown in tables 5 and 6.

Table 5 Inferential Statistics with Fewer Options

Period		Adjust	HWW*	HW*	Adapt*	Auto*	Naive*
1	Average Rank by Series	1.60	3.43	3.10	3.40	4.88	4.60
	Rank of Average Rank	1	4	2	3	6	5
	Kruskal-Wallis Rank Sum	372	1283	1069	1129	1709	1698
	Rank of K-W Rank Sum	1	4	2	3	6	5
	K-W Multi-Comparison Count*	5	5	5	5	4	4
5	Average Rank by Series	1.75	2.83	3.80	3.98	4.80	3.85
	Rank of Average Rank	1	2	3	5	6	4
	Kruskal-Wallis Rank Sum	415.5	903.5	1322.5	1402	1782.5	1434
	Rank of K-W Rank Sum	1	2	3	4	6	5
	K-W Multi-Comparison Count*	5	5	5	4	5	4
10	Average Rank by Series	1.55	2.975	3.825	4.175	4.075	4.4
	Rank of Average Rank	1	2	3	5	4	6
	Kruskal-Wallis Rank Sum	397.5	946.5	1348.5	1494	1531.5	1542
	Rank of K-W Rank Sum	1	2	3	4	5	6
	K-W Multi-Comparison Count*	5	5	5	4	3	4
15	Average Rank by Series	1.58	2.75	3.70	3.98	4.68	4.33
	Rank of Average Rank	1	2	3	4	6	5
	Kruskal-Wallis Rank Sum	383	802	1340.5	1497.5	1712	1525
	Rank of K-W Rank Sum	1	2	3	4	6	5
	K-W Multi-Comparison Count*	5	5	5	4	5	4

Table 6 Rank Anova and Kruskal-Wallis Results

Period		Chi Squared	DF	p value
1	RANK ANOVA	14.52	19	0.7528
	KRUSKAL-WALLIS	50.46	5	0.0000
5	RANK ANOVA	12.15	19	0.8789
	KRUSKAL-WALLIS	47.63	5	0.0000
10	RANK ANOVA	12.41	19	0.8674
	KRUSKAL-WALLIS	43.10	5	0.0000
15	RANK ANOVA	13.96	19	0.7859
	KRUSKAL-WALLIS	53.77	5	0.0000

For the Kruskal-Wallis statistic the values of the statistics change (which should be expected), but the general results do not, that is, the statistics remained significant at the $\alpha = 0.05$ level. However, for the Rank ANOVA test, the statistics are no longer significant. This suggests that the significance of the previous results may be partly attributable to the use of a large number of treatments (forecast models). It is not clear whether this arises because of an increased number of observations or because of some unidentified bias that the tests bring into the analysis.

Following these explorations, I again reviewed the results of scenario 13 and found another reasonable explanation, which is that the actual summarized statistical results in scenario 13 did not vary very much. So, it seems that the lack of significance in scenario 13 as compared with the fairly strong statistical results in the other scenarios could result from the obvious statistical reason, that the different treatments in scenario 13 do not produce particularly different results.

Another problem with these statistics is that in some of the trials rank order was strong, but **inconsistent** between the various descriptive statistics. The non-parametric tests were not sensitive to these inconsistent results. The rank order results were statistically significant with extremely low p values both when the results were consistent between various descriptive statistics and when they were not. This suggests that these rank tests are not sufficient to distinguish superior and inferior forecast models by themselves, but that they may be useful as a supplement to the application of a **battery of descriptive statistics** as presented in this dissertation. If the results are consistent across a battery of descriptive statistics **and** test significant with these tests, the researcher has reason to accept that the treatments are different. Statistical significance is a weaker result while significance without consistency is uninterpretable.

In this dissertation the examination of possible statistical testing of forecast treatments through non-parametric rank order tests was a secondary objective. These results should be considered exploratory. However, it appears that the application of either of these statistical tests in the manner described in this section has some promise when applied across a battery of descriptive statistics. Where results are consistent across the battery of descriptive statistics and the results are significant with one or both of these tests, as occurs with scenarios 1, 2, 8 and 9, it appears that the tests support each other and strengthen the conclusion that the differences in forecast treatments are more than just incidental. Where the results are less consistent across the battery of descriptive statistics, as with scenario 5, or where the statistical tests are not significant, as with scenario 13, results are not firmly supported by the study.

Areas Needing Further Study

I have brought up several topics in this dissertation that need further study. These are:

In chapter 5 I proposed that the method introduced here could be extended to applications with trend shifts using second differences that adjust the trend component of the model. Further examination of this extension seems worthwhile. When a data series undergoes a trend shift, the trend component must respond. However, if the trend parameter is set high, the trend may respond to all the noise in the data series. Since the trend iterates itself for each future period, over-response is likely to lead to extreme variation in forecasts after a few periods. Alternatively, if the parameters are set low, then the models may ignore trend shifts even when they occur. A modification of the proposed technique may allow for a fairly low trend parameter that, nevertheless, does allow for recognition of planned trend changes.

In this study it became apparent that the proposed technique does not necessarily provide a benefit when it is used to fulfill the intervention function for fitting historical level shifts. However, this result may not be universally correct. For example, the technique may be beneficial when the level shift is near to the end of the historical data series, when it is particularly large compared with the prior level, or when it phases in over a large number of periods (thereby emulating a trend shift). Further analysis may provide more insight into why the study of the proposed technique as an intervention variable led to the counter-intuitive results that were achieved.

I have suggested that the forecaster should capitalize on the proposed techniques relatively large errors where the anticipated level shifts do not materialize. Under such circumstances, the forecaster should be able to provide feedback from the forecast to organizational management that the planned event that was expected to produce a level shift is not occurring, or at least, it is not producing the expected level shift. Further analysis is required to determine whether existing tracking signals are adequate for this purpose or whether alternative tracking signals are needed. In particular, the researcher should be alert to the possible problems that will arise when the level shift occurs as expected, but not at precisely the right point in time.

The results concerning simulated actual level shifts that occur when expected but not to the degree expected require further clarification. How accurate should the externally supplied estimate be? How much error is too much? Further study may allow for clarification as to when the technique developed in this study is appropriate and when it is not.

In this study I used two non-parametric statistical tests to evaluate variation in descriptive statistics. Results were promising. I proposed that these statistics might be used in combination with a battery of descriptive statistics as in this dissertation. Further refinement of this approach would be worth pursuing.

Summary

In this chapter I reviewed the results of forecasting level shifting data with various forecast models in 13 scenarios. The models do not produce equally effective forecasts when level shifts are anticipated in the forecast horizon. The proposed technique improves forecast accuracy when the anticipated level shifts actually materializes in the amount expected, even when an unanticipated trend shift also materializes. The proposed technique is among the superior methods when there is a level shift, but it is considerably larger than anticipated. The proposed technique does not produce a particularly effective forecast when the planned level shift does not materialize or materializes in an unexpected manner such as a trend shift, a variance shift, or a considerably smaller than planned level shift. In fact, it produces larger errors than most other techniques in these circumstances. These larger errors might be beneficial if they could be captured for use in a tracking signal. I proposed guidelines for use when prospective level shifts are anticipated. I reviewed the benefits of using non-parametric tests to evaluate forecast models and suggested some further areas of study. **In conclusion, the proposed technique is a beneficial modification of exponential smoothing when used according the proposed guidelines.**

BIBLIOGRAPHY

Aaker, David A., and Robert Jacobson. The sophistication of 'naive' modelling, *International Journal of Forecasting*, 3, no. 3/4 (1987): 449-451.

Ahlburg, Dennis A., and Kenneth C. Land. Population forecasting: Guest editors' introduction, *International Journal of Forecasting*, 8, no. 3 (1992): 289-299.

Archibald, Blyth C. Parameter space of the Holt-Winters' model, *International Journal of Forecasting*, 6, no. 2 (July 1990): 199-210.

Armstrong, J. Scott. "Forecasting with Econometric Methods: Folklore versus Fact," in *The Forecasting Accuracy of Major Time Series Methods*, eds., S. Makridakis, et. *al.*, John Wiley & Sons: Chichester, 1984: 19-34.

Armstrong, J. Scott. *Long-Range Forecasting, From Crystal Ball to Computer*. 2d ed. New York: John Wiley & Sons, 1985.

Armstrong, J. Scott, Roderick J. Brodie, and Shelby H. McIntyre. Forecasting methods for marketing, *International Journal of Forecasting*, 3, no. 3/4 (1987): 355-378.

Armstrong, J. Scott, and Fred Collopy. Error measures for generalizing about forecasting methods: Empirical comparisons, *International Journal of Forecasting*, 8, no. 1 (June 1992): 69-80.

Ashley, Richard. On the Usefulness of Macroeconomic Forecasts as Inputs to Forecasting Models, *Journal of Forecasting*, 2 (1983): 211-223.

Ashley, Richard. On the relative worth of recent macro-economic forecasts, *International Journal of Forecasting*, 4, no. 3 (1988): 363-376.

Barrett, Katherine and Richard Greene. The State of the States, *Financial World* (May 12, 1992): 24 - 43.

Benson, P. George, and Dilek Onkal. The effects of feedback and training on the performance of probability forecasters, *International Journal of Forecasting*, 8, no. 4 (December 1992): 559-573.

Bewley, Ronald, and Denzil G. Fiebig. A flexible logistic growth model with applications in telecommunications, *International Journal of Forecasting*, 4, no. 2 (1988): 177-192.

Bloom, David E., and Sherry Glied. Projecting the number of new AIDS cases in the United States, *International Journal of Forecasting*, 8, no. 3 (1992): 339-365.

Bodo, Giorgio, and Luigi Federico Signorini. Short-term forecasting of the industrial production index, *International Journal of Forecasting*, 3, no. 2 (1987): 245-259.

Bos, Eduard, and Rodolfo A. Bulatao. The demographic impact of AIDS in sub-Saharan Africa, short- and long-term projections, *International Journal of Forecasting*, 8, no. 3 (1992): 367-384.

Bretschneider, Stuart Ira. 1979. Adaptive Time Series Analysis and Forecasting. Ph.D. diss., The Ohio State University.

Bretschneider, Stuart I., Wilpen L. Gorr, Gloria Grizzle, and Earle Klay. Political and organizational influences on the accuracy of forecasting state government revenues, *International Journal of Forecasting*, 5, no. 3 (1989): 307-319.

Buck, Jeffrey A. and John Klemm. Recent Trends in Medicaid Expenditures. *Health Care Financing Review* (1992 Annual Supplement): 271 - 283.

Calmes, Jacqueline. Bricks Without Straw: The Complaints Go On but Congress Keeps Mandating. *Governing* (September 1988): 21-26.

Canavos, George C. *Applied Probability and Statistical Methods*. Boston: Little, Brown and Company, 1984.

Canavos, George C., and Don M. Miller. *An Introduction to Modern Business Statistics*. Belmont, California: Duxbury Press, 1993.

Carbone, Robert, and Spyros Makridakis. Forecasting when pattern changes occur beyond the historical data, *Management Science*, 32, no. 3 (1986): 257-271.

Carreno, Jose Juan, and Jesus Madinaveitia. A modification of time series forecasting methods for handling announced price increases, *International Journal of Forecasting*, 6, no. 4 (1990): 479-484.

Cassidy, Glenn, Mark S. Kamlet, and Daniel S. Nagin. An Empirical examination of bias in revenue forecasts by state governments, *International Journal of Forecasting*, 5, no. 3 (1989): 321-331.

Chambers, John C., Satinder K. Mullick, and Donald D. Smith. How to Choose the Right Forecasting Technique. *Harvard Business Review* (July-August 1971): 45 - 74.

Chatfield, C., The Holt-Winters forecasting procedure. *Applied Statistics*, 27, no. 3 (1978): 264-279.

Cloud, David S. Rhetoric Confronts Reality As the 1991 Budget Battle Gets Under Way. *Governing* (May 1990): 21-25.

Clemen, Robert T. Combining forecasts: a review and annotated bibliography, *International Journal of Forecasting*, 5, no. 4 (1989): 559-583.

Collopy, Fred, and J. Scott Armstrong. Expert opinions about extrapolation and the mystery of the overlooked discontinuities, *International Journal of Forecasting*, 8, no. 4 (December 1992): 575-582.

Congressional Budget Office. *Projections of National Health Expenditures*. Washington, D.C.: Congressional Budget Office, October, 1992.

Coopersmith, Lewis W. Forecasting time series which are inherently discontinuous, *Journal of Forecasting*, 2 (1983): 225-235.

Cox, James E. An assessment of books relevant to forecasting in marketing, *International Journal of Forecasting*, 3, no. 3/4 (1987): 515-527.

Dalrymple, Douglas J. Sales forecasting practices, *International Journal of Forecasting* 3, no. 3/4 (1987): 379-391.

Danaher, Peter J., and Roderick J. Brodie. Predictive accuracy of simple versus complex market share models, *International Journal of Forecasting*, 8, no. 4 (December 1992): 613-626.

Dancer, Robert, and Clifford Gray. An empirical evaluation of constant and adaptive computer forecasting models for inventory control, *Decision Sciences*, 8 (1977): 228-238.

Department of Medical Assistance Services, *The Statistical Record of the Virginia Medicaid Program*. Richmond, Virginia, February, 1993.

Easingwood, Christopher J. An analogical approach to the long term forecasting of major new product sales, *International Journal of Forecasting*, 5, no. 1 (1989): 69-82.

Eilon, Samuel, and Joseph Elmaleh, Adaptive limits in inventory control. *Management Science*, 16, no. 8 (April, 1970): B-533 - B-548.

Ekern, Steinar. Adaptive exponential smoothing revisited, *Journal of the Operational Research Society*, 32 (1981): 775-782.

Feder, Judith, Diane Rowland, John Holahan, Alina Salganicoff, David Heslam. *The Medicaid Cost Explosion, Causes and Consequences*. Baltimore, Maryland: The Kaiser Commission on the Future of Medicaid, February, 1993.

Fildes, R. Quantitative forecasting -- the state of the art: Extrapolative models, *Journal of the Operational Research Society*, 30, no. 8 (1979): 691-710.

Fildes, Robert, and Spyros Makridakis. Forecasting and loss functions, *International Journal of Forecasting*, 4, no. 4 (1988): 545-550.

Fildes, Robert. The evaluation of extrapolative forecasting methods, *International Journal of Forecasting*, 8, no. 1 (April 1992): 81-98.

Flores, Benito E. A pragmatic view of accuracy measurement in forecasting, *Omega*, 14, no. 2 (1986): 93-98.

Flores, Benito E. The utilization of the Wilcoxon test to compare forecasting methods: a note, *International Journal of Forecasting*, 5, no. 4 (1989): 529-535.

Frank, Howard A., and Gerasimos A. Gianakis. Raising the bridge using time series forecasting models, *Public Productivity & Management Review*, XIV, no. 2 (Winter 1990): 171-188.

Fullerton, Thomas M., Jr. A composite approach to forecasting state government revenues: case study of the Idaho sales tax, *International Journal of Forecasting*, 5, no. 3 (1989): 373-380.

Gardner, Everette S., Jr. Automatic monitoring of forecasting errors, *Journal of Forecasting*, 2 (1983): 1-21.

Gardner, Everette S., Jr. Exponential smoothing: The state of the art, *Journal of Forecasting*, 4 (1985): 1-28.

Gardner, Everette S., Jr. and Spyros Makridakis. The future of forecasting, *International Journal of Forecasting*, 4, no. 3 (1988): 325-330.

Gardner, Everette S., Jr. and Ed McKenzie. Seasonal exponential smoothing with dampened trends, *Management Science*, 35, no. 3 (March 1989): 372-376.

Geurts, Michael D., and J. Patrick Kelly. Comments on: "In defense of ARIMA modelling", by D. J. Pack, *International Journal of Forecasting*, 6, no. 4 (1990): 497-499.

Ghali, Moheb A. Seasonality, aggregation and the testing of the production smoothing hypothesis, *The American Economic Review*, 77, no. 3 (June 1987): 464-469.

Gilchrist, Warren. *Statistical Forecasting*. Chichester: John Wiley & Sons, 1977.

Gorr, Wilpen L. Special event data in shared databases, *MIS Quarterly*, 10 (September 1986): 239-250.

Gorr, Wilpen L. Use of special event data in government information systems, *Public Administration Review*, 46, Special Issue (November 1986): 532-539.

Guerrero, Victor M. ARIMA forecasts with restrictions derived from a structural change, *International Journal of Forecasting*, 7, no. 3 (1991): 339-347.

Hanke, John. Forecasting in business schools, a follow-up survey, *International Journal of Forecasting*, 5, no. 2 (1989): 259-262.

Harrison, P. J., and C. F. Stevens. A Bayesian approach to short-term forecasting, *Operational Research Quarterly*, 22, no. 4 (1971): 341-362.

Harrison, P. J., and C. F. Stevens. Bayesian forecasting, *Journal of the Royal Statistical Society* Series B, 34, no. 3 (1976): 205-228.

Harvey, Andrew, and Ralph D. Snyder. Structural time series models in inventory control, *International Journal of Forecasting*, 6, no. 2 (July 1990): 187-198.

Health Care Finance Administration. Appendix 4 HCFA Office of the Actuary forecasting Medicaid expenditures, [photocopy], [*circa* 1992].

HHS News Release. "Administration Acts to Address Medicaid Estimating Problems and Escalating Cost Increases." July 10, 1991.

Hillmer, Steven. Monitoring and Adjusting Forecasts in the Presence of Additive Outliers. *Journal of Forecasting* 3 (1984): 205-215.

Hogwood, Brian W. and Lewis A. Gunn, *Policy Analysis for the Real World*. Oxford: Oxford University Press, 1984.

Holloway, Thomas M. Measuring the cyclical sensitivity of federal receipts and expenditures: simplified estimation procedures, *International Journal of Forecasting*, 5, no. 3 (1989): 347-360.

Howell, Sydney D. Parameter instability in learning curve models, *International Journal of Forecasting*, 6, no. 4 (1990): 541-547.

Hume, David. *An Inquiry Concerning Human Understanding* in *David Hume: On Human Nature and the Understanding*. Antony Flew, ed., Collier Classics in the History of Thought, Crane Brinton and Paul Edwards, eds. New York: The Macmillan Company, 1971.

Jarrett, Jeffrey. *Business Forecasting Methods.* Oxford: Basil Blackwell, 1987.

Jazwinski, A. H. *Stochastic Processes and Filtering Theory* (New York/London; Academic Press, 1970).

Jenkins, Gwilym M. *Practical Experiences with Modelling and Forecasting Time Series* (Jersey, Channel Islands: Gwilym Jenkins & Partners (Overseas) Ltd., 1979).

Jun, Duk Bin and Robert M. Oliver. Bayesian Forecasting Following a Major Level Change in Exponential Smoothing. *Journal of Forecasting* 4 (1985): 293-302.

Jun, Duk B. On Detecting and Estimating a Major Level or Slope Change in General Exponential Smoothing. *Journal of Forecasting* 8 (1989): 55-64.

Kitagawa, Genshiro. Non-Gaussian state-space modelling of nonstationary time series, *Journal of the American Statistical Association*, 82 (December 1987): 1032-1041.

Kolb, R. A., and H. O. Stekler. Are economic forecasts significantly better than naive predictions? An appropriate test, *International Journal of Forecasting*, 9, no. 1 (April 1993): 117-120.

Kirkendall, Nancy J. Monitoring for Outliers and Level Shifts in Kalman Filter Implementations of Exponential Smoothing, *Journal of Forecasting*, 11, no. 6 (1992) 543-560.

LaCivita, Charles J., and Terry G. Seaks. Forecasting accuracy and the choice of first difference or percentage change regression models, *International Journal of Forecasting*, 4, No. 2 (1988): 261-268.

Langbein, Laura Irwin. *Discovering Whether Programs Work: A Guide to Statistical Methods for Program Evaluation.* The Scott, Foresman Public Policy Analysis and Management Science Series, Arnold J. Meltsner and Mark H Moore, eds. Glenview, Illinois: Scott, Foresman and Company, 1980.

Lefrançois, Pierre. Allowing for asymmetry in forecast errors: results from a Monte-Carlo study, *International Journal of Forecasting*, 5. no. 1 (1989): 99-110.

Lemov, Penelope. Climbing out of the Medicaid Trap. *Governing* (October 1991): 49-53.

Lemov, Penelope. States and Medicaid: Ahead of the Feds. *Governing* (July 1993): 27-28.

Lewandowski, Rudolf. Sales forecasting by FORSYS, *Journal of Forecasting*, 1 (1982): 205-214.

Mahajan, Vijay, and Yoram Wind. New product forecasting models: Directions for research and implementation, *International Journal of Forecasting*, 4, no. 3 (1988): 341-358.

Makridakis, S., A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. The accuracy of extrapolation (time series) methods: results of a forecasting competition, *Journal of Forecasting*, 1 (1982): 111-153.

Makridakis, Spyros, Steven C. Wheelwright, and Victor E. McGee. *Forecasting: Methods and Applications*, 3rd ed. New York: John Wiley & Sons, 1983.

Makridakis, Spyros, and Robert L. Winkler. Averages of forecasts, some empirical results, *Management Science*, 29, no. 9 (9/83): 987-996.

Makridakis, S., A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler. *The Forecasting Accuracy of Major Time Series Methods*. Chichester: John Wiley & Sons (1984).

Makridakis, Spyros and Michele Hibon. Accuracy of Forecasting: An Empirical Investigation (with Discussion). Reprinted as chapter 3 in *The Forecasting Accuracy of Major Time Series Methods*, S. Makridakis, et. al., eds., Chichester: John Wiley & Sons, (1984).

Makridakis, Spyros, and Robert Carbone. Forecasting when pattern changes occur beyond the historical data, *Management Science*, 32, no. 3 (1986): 257-271.

Makridakis, Spyros, Michele Hibon, Ed Lusk, and *Moncef* Belhadjali. Confidence intervals: An empirical investigation of the series in the M-competition, *International Journal of Forecasting*, 3, no. 3/4 (1987): 489-508.

Makridakis, Spyros. Metaforecasting: Ways of improving forecasting accuracy and usefulness, *International Journal of Forecasting*, 4, no. 3 (1988): 467-491.

Makridakis, Spyros, and Michele Hibon. Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy, *International Journal of Forecasting*, 7, no. 3 (1991): 317-330.

Makridakis, Spyros, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons. The M2-competition: A real-time judgmentally based forecasting study, *International Journal of Forecasting*, 9, no. 1 (April 1993): 5-22.

Mandy, David M. Forecasting unemployment insurance trust funds: The case of Tennessee *International Journal of Forecasting*, 5 no. 3 (1989): 381-391.

Manton, Kenneth G., Eric Stallard, and Burt Singer. Projecting the future size and health status of the US elderly population, *International Journal of Forecasting*, 8, no. 3 (1992): 433-458.

Martin, Christine A., and Stephen F. Witt. Forecasting tourism demand: A comparison of the accuracy of several quantitative methods, *International Journal of Forecasting*, 5, no. 1 (1989): 7-20.

McClain, John O. Dominant tracking signals, *International Journal of Forecasting*, 4, no. 4 (1988): 563-572.

McCollough, James. State of the Art - Municipal Revenue and Expenditure Forecasting: Current Status and Future Prospects. *Government Finance Review* (October 1990): 38 - 40.

McDonald, James B. Partially adaptive estimation of ARMA time series models, *International Journal of Forecasting*, 5, no. 2 (1989): 217-230.

McDowell, Charles. Not much cuttable budget left after entitlements taken out. *Richmond Times-Dispatch.* (February 9, 1994): A2.

McKenzie, E. Error analysis for Winters' additive seasonal forecasting system, *International Journal of Forecasting*, 2 (1986): 373-382 cited in Paul Newbold and Ted Bos, On exponential smoothing and the assumption of deterministic trend plus white noise data-generating models, *International Journal of Forecasting*, 5, no. 4 (1989): 523-527.

McKenzie, Ed. General exponential smoothing and the equivalent ARMA process, *Journal of Forecasting,* 3 (1984): 333-344.

McLaughlin, Robert L. Measuring the accuracy of forecasts, *Business Economics*, 7 (May 1972): 27-35.

McLaughlin, Robert L. A model of an average recession and recovery, *Journal of Forecasting*, 1 (1982): 55-65.

McNees, Stephen K. On the future of macroeconomic forecasting, *International Journal of Forecasting*, 4, no. 3 (1988): 359-362.

Medicine & Health's Health Reform Insight. Forecasting Health Care Costs: First, Cross Your Fingers. (December 3, 1993): 1.

Montgomery, Douglas C., and Lynwood A. Johnson. *Forecasting and Time Series Analysis*. New York: McGraw-Hill Book Company, 1976.

Nachmias, David and Chava Nachmias. *Research Methods in the Social Sciences*, 2nd ed. New York: St. Martin's Press, 1981.

Naim, Mohamed M., and Denis R. Towill. An engineering approach to LSE modelling of experience curves in the electricity supply industry, *International Journal of Forecasting*, 6, no. 4 (1990): 549-556.

Nerlove, M., and S. Wage. On the optimality of adaptive forecasting, *Management Science*, 10, no. 2 (January, 1964): 207-223.

Newbold, Paul, and Ted Bos. On exponential smoothing and the assumption of deterministic trend plus white noise data-generating models, *International Journal of Forecasting*, 5, no. 4 (1989): 523-527.

Noether, Gottfried E. *Introduction to Statistics.* New York: Houghton Mifflin Company, 1971.

Office of Management and Budget and Department of Health & Human Services, "Better Management for Better Medicaid Estimates," July 10, 1991.

Ord, J. Keith. Future developments in forecasting: The time series connexion, *International Journal of Forecasting*, 4, no. 3 (1988): 389-401.

Pack, David J. In defense of ARIMA modelling, *International Journal of Forecasting*, 6, no. 2 (July 1990): 211-218.

Pankratz, Alan. Time Series Forecasts and Extra-model Information. *Journal of Forecasting*, 8 (1989): 75-83.

Pant, P. Narayan, and William Starbuck. Innocents in the forest: Forecasting and research methods, *Journal of Management*, 16, no. 2 (1990): 433-460.

Poulos, Laurette, Alan Kvanli and Robert Pavur. A comparison of the accuracy of the Box-Jenkins method with that of automated forecasting methods, *International Journal of Forecasting*, 3, no. 2 (1987): 261-267.

Priestly, M. B. Discussion on the paper by Professor Harrison and Mr. Stevens, *Journal of the Royal Statistical Society*, Series B, 34, no. 3 (1976): 228-230.

Report Prepared by the Congressional Research Services for the use of the Subcommittee on Health and the Environment of the Committee on Energy and Commerce, U.S. House of Representatives, Joseph E. Ross, Director, Henry A. Waxman, Chairman. *Medicaid Source Book: Background Data and Analysis (A 1993 Update)*. Washington, D.C.: U.S. Government Printing Office, January, 1993.

Revankar, Nagesh S. On the Problem of Forecasting Prior to 'Price' Control or Decontrol, *Journal of Forecasting*, 11, no. 1 (1992) 1 - 15.

Rosenberg, Alexander. *Philosophy of Social Science.* Dimensions of Philosophy Series, Norman Daniels and Keith Lehrer, eds. (Boulder: Westview Press, 1988).

Rosenfield, Donald. A model for predicting frequencies of random events, *Management Science*, 33, no. 8 (August 1987): 947-954.

SAS Institute Inc., *SAS® User's Guide: Statistics - Version 5 Edition*. Cary, North Carolina: SAS Institute Inc., 1985.

Schnaars, Steven P., and Martin T. Topol. The use of multiple scenarios in sales forecasting, *International Journal of Forecasting*, 3, no. 3/4 (1987): 405-419.

Schroeder, Larry, and Michael Wasylenko. Public sector forecasting in the Third World, *International Journal of Forecasting*, 5, no. 3 (1989): 333-345.

Scott, Claudia Devita. *Forecasting Local Government Spending*. Washington, D.C.: The Urban Institute, 1972.

Shanahan, Eileen. Cracks in the Crystal Ball. *Governing* (December 1991): 29 - 32.

Sharp, John A., and David H. R. Price. Experience curve models in the electronic supply industry, *International Journal of Forecasting*, 6, no. 4 (1990): 531-540.

Shkurti, William J., and Darrell Winefordner. The politics of state revenue forecasting in Ohio, 1984-1987: A case study and research implications, *International Journal of Forecasting*, 5, no. 3 (1989): 361-371.

Smith, Stanley K., and Terry Sincich. Forecasting state and household populations: Evaluating the forecast accuracy and bias of alternative populations projections for states, *International Journal of Forecasting*, 8, no. 3 (1992): 495-508.

Snyder, R. D. Progressive tuning of simple exponential smoothing forecasts, *Journal of the Operational Research Society*, 39, no. 4 (1988): 393-399.

Statistical Abstract of the United States 1992. Washington, D.C.: U.S. Government Printing Office, 1992.

Stekler, H. O. Macroeconomic forecast evaluation techniques, *International Journal of Forecasting*, 7, no. 3 (1991): 375-384.

Steward, Rodney D. *Cost Estimating*. 2nd ed. New York, John Wiley & Sons, Inc., 1990.

Theil, H., and S. Wage. Some observations on adaptive forecasting, *Management Science*, 10, no. 2 (January, 1964): 198-206.

Thompson, Patrick A. An MSE statistic for comparing forecast accuracy across series, *International Journal of Forecasting*, 6, no 2. (July 1990): 219-227.

Thompson, Patrick A. Evaluation of the m-competition forecasts via log mean squared error ratio, *International Journal of Forecasting*, 7, no. 3 (November 1991): 331-334.

Time, The forecasters flunk, (August 27, 1984): 42.

Towill, Denis R.. Forecasting learning curves, *International Journal of Forecasting*, 6, no. 1 (1990): 25-38.

Trigg, D. W., and A. G. Leach. Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, 18, no. 1 (1967): 53-59.

Tsay, Ruey S. Comment: Detecting and modelling changes in time series, *Journal of the American Statistical Association*, 82 (1987): 1056-1059.

Tsay, Ruey S., Outliers, level shifts, and variance changes in time series, *Journal of Forecasting*, 7 (1988): 1-20.

Vandaele, Walter. *Applied Time Series and Box-Jenkins Models*. New York: Harcourt Brace Jovanovich, Publishers, 1983.

Vollmann, Thomas E., William L. Berry, and D. Clay Whybark, *Integrated Production and Inventory Management: Revitalizing the Manufacturing Enterprise*. The Business One Irwin/APICS Library of Integrated Resource Management. Homewood, Illinois 60430: Business One Irwin, 1993.

White, Edna M., and Ronald Dattero. Combining vector forecasts to predict thoroughbred horse race outcomes, *International Journal of Forecasting*, 8, no. 4 (1992): 595-611.

White, Michael J., Ross Clayton, Robert Myrtle, Gilbert Siegel and Aaron Rose, *Managing Public Systems*. Lanham, MD: University Press of America, 1985.

Wilder, L. Douglas, Governor. "Excerpts form the Remarks Made to the Finance and Appropriations Committees of the Virginia General Assembly." Unpublished News Release, Richmond, Virginia, August 23, 1993, 5.

Williams, T. M. Adaptive Holt-Winters forecasting, *Journal of the Operational Research Society*, 38, no. 6 (June 1987): 553-560.

Winters, Peter R. Forecasting sales by exponentially weighted moving averages, *Management Science*, 7 (April 1960): 324-342.

Wooldridge, Blue. Revenue planning in the public sector, [unpublished paper used as required reading in the course "Public Financial Management," Virginia Commonwealth University, Fall 1992 - photocopy].

Woolsey, Gene. The fifth column: The case of the unused forecast or when trying to forecast what's to be, first try y = mx + b, *Interfaces*, 16, no. 6 (November-December 1986): 58-60.

Wun, Lap-Ming, and Wen Lea Pearn. Assessing the statistical characteristics of the mean absolute error [of] forecasting, *International Journal of Forecasting*, 7, no. 3 (November 1991): 335-337.

Yar, Mohammed, and Chris Chatfield. Prediction intervals for the Holt-Winters forecasting procedure, *International Journal of Forecasting*, 6, no. 1 (1990): 127-137.

Young, Peg, and J. Keith Ord. Model selection and estimation for technological growth curves, *International Journal of Forecasting*, 5, no. 4 (1989): 501-513.

ENDNOTES

Endnotes Chapter 1

1. Wilpen L. Gorr, "Special Event Data in Shared Databases," *MIS Quarterly*, 10 (September 1986), 239-250; Wilpen L. Gorr, "Use of Special Event Data in Government Information Systems," *Public Administration Review*, 46, Special Issue (November 1986), 532-539.

Endnotes Chapter 2

2. Stuart I. Bretschneider, Wilpen L. Gorr, Gloria Grizzle, and Earle Klay, "Political and Organizational Influences on the Accuracy of Forecasting State Government Revenues," *International Journal of Forecasting*, 5, no. 3 (1989), 307-319; Glenn Cassidy, Mark S. Kamlet, and Daniel S. Nagin, "An Empirical Examination of Bias in Revenue Forecasts by State Governments," *International Journal of Forecasting*, 5, no. 3 (1989), 321-331; Larry Schroeder and Michael Wasylenko, "Public Sector Forecasting in the Third World," *International Journal of Forecasting*, 5, no. 3 (1989), 333-345; Thomas M. Holloway, "Measuring the Cyclical Sensitivity of Federal Receipts and Expenditures: Simplified Estimation Procedures," *International Journal of Forecasting*, 5, no. 3 (1989), 347-360; William J. Shkurti, and Darrell Winefordner, "The Politics of State Revenue Forecasting in Ohio, 1984 - 1987: A Case Study and Research Implications," *International Journal of Forecasting*, 5, no. 3 (1989), 361-371; Thomas M. Fullerton, Jr. "A Composite Approach to Forecasting State Government Revenues: Case Study of the Idaho Sales Tax," *International Journal of Forecasting*, 5, no. 3 (1989), 373-380; David M. Mandy, "Forecasting Unemployment Insurance Trust Funds: The Case of Tennessee" *International Journal of Forecasting*, 5 no. 3 (1989), 381-391; Howard A. Frank and Gerasimos A. Gianakis, "Raising the Bridge Using Time Series Forecasting Models," *Public Productivity & Management Review*, XIV, no. 2 (Winter 1990), 171-188; Wilpen L. Gorr, "Use of Special Event Data in Government Information Systems," *Public Administration Review*, 46, Special Issue (November 1986), 532-539.

3. Forecaster at Department of Corrections, personal communication, March 5, 1993.

4. Larry Schroeder and Michael Wasylenko, "Public Sector Forecasting in the Third World," *International Journal of Forecasting*, 5, no. 3 (1989), 333-345; Health Care Finance Administration, "Appendix 4 HCFA Office of the Actuary Forecasting Medicaid Expenditures," [photocopy], [*circa* 1992].

5. David E. Bloom and Sherry Glied, "Projecting the Number of New AIDS Cases in the United States," *International Journal of Forecasting*, 8, no. 3 (1992), 339-365; Eduard Bos and Rodolfo A. Bulatao, "The Demographic Impact of AIDS in Sub-Saharan Africa, Short- and Long-Term Projections," *International Journal of Forecasting*, 8, no. 3 (1992), 367-384; Kenneth G. Manton, Eric Stallard, and Burt Singer,

"Projecting the Future Size and Health Status of the US Elderly Population," *International Journal of Forecasting*, 8, no. 3 (1992), 433-458.

6. Brian W. Hogwood and Lewis A. Gunn, *Policy Analysis for the Real World* (Oxford: Oxford University Press, 1984); Michael J. White, Ross Clayton, Robert Myrtle, Gilbert Siegel and Aaron Rose, *Managing Public Systems* (Lanham, MD: University Press of America, 1985); Claudia Devita Scott, *Forecasting Local Government Spending* (Washington, D.C.: The Urban Institute, 1972); Howard A. Frank and Gerasimos A. Gianakis, "Raising the Bridge Using Time Series Forecasting Models," *Public Productivity & Management Review*, XIV, no. 2 (Winter 1990), 171-188; Wilpen L. Gorr, "Use of Special Event Data in Government Information Systems," *Public Administration Review*, 46, Special Issue (November 1986), 532-539; Eileen Shanahan, "Cracks in the Crystal Ball" *Governing* (December 1991) 29 - 32; James McCollough, "State of the Art - Municipal Revenue and Expenditure Forecasting: Current Status and Future Prospects" *Government Finance Review* (October 1990) 38 - 40.

7. Howard A. Frank and Gerasimos A. Gianakis, "Raising the Bridge Using Time Series Forecasting Models," *Public Productivity & Management Review*, XIV, no. 2 (Winter 1990), 171-188.

8. James McCollough, "State of the Art - Municipal Revenue and Expenditure Forecasting: Current Status and Future Prospects," *Government Finance Review* (October 1990) 38 - 40.

9. Stuart Ira Bretschneider, "Adaptive Time Series Analysis and Forecasting," (Ph.D. diss., The Ohio State University, 1979) 38 ff.

10. Eileen Shanahan, "Cracks in the Crystal Ball," *Governing* (December 1991) 29 - 32.

11. Katherine Barrett and Richard Greene, "The State of the States," *Financial World* (May 12, 1992) 24 - 43.

12. Katherine Barrett and Richard Greene, "The State of the States," *Financial World* (May 12, 1992) 28.

13. Department of Medical Assistance Services, *The Statistical Record of the Virginia Medicaid Program* (Richmond, Virginia, February 1993) Table 7-4.

14. Governor L. Douglas Wilder, "Excerpts form the Remarks Made to the Finance and Appropriations Committees of the Virginia General Assembly," Unpublished News Release, (Richmond, Virginia, August 23, 1993), 5.

15. Congressional Budget Office, *Projections of National Health Expenditures* (Washington, D.C.: Congressional Budget Office, October, 1992), 9; Judith Feder, Diane Rowland, John Holahan, Alina Salganicoff, David Heslam, *The Medicaid Cost Explosion, Causes and Consequences* (Baltimore, Maryland: The Kaiser Commission on the Future of Medicaid, February, 1993), xi.

16. Penelope Lemov, "States and Medicaid: Ahead of the Feds" *Governing* (July 1993), 27-28.

17. Report Prepared by the Congressional Research Services for the use of the Subcommittee on Health and the Environment of the Committee on Energy and Commerce, U.S. House of Representatives, Joseph E. Ross, Director, Henry A. Waxman, Chairman, *Medicaid Source Book: Background Data and Analysis (A 1993 Update)* (Washington, D.C.: U.S. Government Printing Office, January, 1993), 475; Charles McDowell, *Richmond Times-Dispatch,* "Not much cuttable budget left after entitlements taken out" (February 9, 1994), A2.

18. Judith Feder, Diane Rowland, John Holahan, Alina Salganicoff, David Heslam, *The Medicaid Cost Explosion, Causes and Consequences* (Baltimore, Maryland: The Kaiser Commission on the Future of Medicaid, February, 1993), 34.

19. *The Medicaid Cost Explosion, Causes and Consequences* (Baltimore, Maryland: The Kaiser Commission on the Future of Medicaid, February, 1993), 32; Report Prepared by the Congressional Research Services for the use of the Subcommittee on Health and the Environment of the Committee on Energy and Commerce, U.S. House of Representatives, Joseph E. Ross, Director, Henry A. Waxman, Chairman, *Medicaid Source Book: Background Data and Analysis (A 1993 Update)* (Washington, D.C.: U.S. Government Printing Office, January, 1993), 493.

20. *Statistical Abstract of the United States 1992* (Washington, D.C.: U.S. Government Printing Office, 1992), 282.

21. Congressional Budget Office, *Projections of National Health Expenditures* (Washington, D.C.; Congressional *Budget* Office, October, 1992), 39; Judith Feder, Diane Rowland,

John Holahan, Alina Salganicoff, David Heslam, *The Medicaid Cost Explosion, Causes and Consequences* (Baltimore, Maryland: The Kaiser Commission on the Future of Medicaid, February, 1993), 8; Charles McDowell, *Richmond Times-Dispatch*, "Not much cuttable budget left after entitlements taken out" (February 9, 1994), A2.

22. HHS News Release, "Administration Acts to Address Medicaid Estimating Problems and Escalating Cost Increases," July 10, 1991.

23. Office of Management and Budget and Department of Health & Human Services, "Better Management for Better Medicaid Estimates," July 10, 1991.

24. Office of Management and Budget and Department of Health & Human Services, "Better Management for Better Medicaid Estimates" (July 10, 1991) 2 - 3.

25. Jeffrey A. Buck and John Klemm, "Recent Trends in Medicaid Expenditures," *Health Care Financing Review* (1992 Annual Supplement) 271 - 283.

26. Jacqueline Calmes, "Bricks Without Straw: The Complaints Go On but Congress Keeps Mandating," *Governing* (September 1988) 21-26; David S. Cloud, "Rhetoric Confronts Reality As the 1991 Budget Battle Gets Under Way," *Governing* (May 1990) 21-25.

27. *Medicine & Health's Health Reform Insight*, "Forecasting Health Care Costs: First, Cross Your Fingers," (December 3, 1993), 1.

28. Governor L. Douglas Wilder, "Excerpts form the Remarks Made to the Finance and Appropriations Committees of the Virginia General Assembly," Unpublished News Release, (Richmond, Virginia, August 23, 1993), 5.

29. Wilpen L. Gorr, "Special Event Data in Shared Databases," *MIS Quarterly*, 10 (September 1986), 239-250; Wilpen L. Gorr, "Use of Special Event Data in Government Information Systems," *Public Administration Review*, 46, Special Issue (November 1986), 532-539.

30. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 905.

31. J. Scott Armstrong, *Long-Range Forecasting* (New York: John Wiley and Son's, 1985), 179.

32. Blue Wooldridge, "Revenue Planning in the Public Sector," [unpublished paper used as required reading in the course "Public Financial Management," Virginia Commonwealth University, Fall 1992-photocopied], 4.

33. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 62 - 183.

34. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 184 - 543.

35. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications* (New York: John Wiley & Sons, 1983), 352-537; J. Keith Ord, "Future Developments in Forecasting: The Time Series Connexion," *International Journal of Forecasting*, 4, no. 3 (1988), 393; Mohammed Yar and Chris Chatfield, "Prediction Intervals for the Holt-Winters Forecasting Procedure," *International Journal of Forecasting*, 6, no. 1 (1990), 127-137; Ed McKenzie, "General Exponential Smoothing and the Equivalent ARMA Process," *Journal of Forecasting,* 3 (1984), 333-344.

36. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 184-346.

37. Douglas J. Dalrymple, "Sales Forecasting Practices," *International Journal of Forecasting*, 3, no. 3/4 (1987), 379-391; John Hanke, "Forecasting in Business Schools, A Follow-Up Survey," *International Journal of Forecasting*, 5, no. 2 (1989), 259-262.

38. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983); Warren Gilchrist, *Statistical Forecasting* (Chichester: John Wiley & Sons, 1977); Douglas C. Montgomery, and Lynwood A. Johnson, *Forecasting and Time Series Analysis* (New York: McGraw-Hill Book Company, 1976); Walter Vandaele, *Applied Time Series and Box-Jenkins Models* (New York: Harcourt Brace Jovanovich, Publishers, 1983); Jeffrey Jarrett, *Business Forecasting Methods* Oxford: Basil Blackwell, 1987).

39. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 84; George C. Canavos and Don M. Miller, *An Introduction to Modern Business Statistics*, (Belmont, California: Duxbury Press: 1993), 677.

40. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 87.

41. Spyros Makridakis and Michele Hibon, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *International Journal of Forecasting*, 7, no. 3 (1991), 317-330.

42. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 97 ff.

43. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 97; T. M. Williams, "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 6 (June 1987), 556.

44. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 103 ff.

45. Peter R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 7 (April 1960), 324-342; Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 105-106; T. M. Williams, "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 6 (June 1978), 556.

46. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 149.

47. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*. 2d ed. (New York: John Wiley & Sons, 1985), 58.

48. George C. Canavos and Don M. Miller, *An Introduction to Modern Business Statistics*, (Belmont, California: Duxbury Press: 1993), 677.

49. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*. 2d ed. (New York: John Wiley & Sons, 1985), 165.

50. S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1 (1982), 111-153; Robert Carbone and Spyros Makridakis, "Forecasting When Pattern Changes Occur Beyond the Historical Data," *Management Science*, 32, no. 3 (1986), 257; R. Fildes, "Quantitative Forecasting -- The State of the Art: Extrapolative Models," *Journal of the Operational Research Society*, 30, no. 8 (1979), 691-710; T. M. Williams, "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 6 (June 1987), 553.

51. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*, 2d ed. (New York: John Wiley & Sons, 1985), 178; J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 366; Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22; P. Narayan Pant and William Starbuck, "Innocents in the Forest: Forecasting and Research Methods," *Journal of Management* 16, no. 2 (1990), 433-460.

52. J. Scott Armstrong, "Forecasting with Econometric Methods: Folklore versus Fact," in *The Forecasting Accuracy of Major Time Series Methods*, eds., S. Makridakis, *et. al.* (John Wiley & Sons: Chichester, 1984), 19 - 34.

53. Everette S. Gardner, Jr. and Spyros Makridakis, "The Future of Forecasting," *International Journal of Forecasting*, 4, no. 3 (1988), 326.

54. Spyros Makridakis, "Ways of Improving Forecasting Accuracy and Usefulness," *International Journal of Forecasting*, 4, no. 3 (1988), 479; *Time*, "The Forecasters Flunk," August 27, 1984; J. Scott Armstrong, *Long Range Forecasting* (New York: John Wiley and Sons, 1985), 179-183.

55. P. Narayan Pant and William H. Starbuck, "Innocents in the Forest: Forecasting and Research Methods," *Journal of Management*, 16, no. 2 (1990), 442; C. Chatfield, "The Holt-Winters Forecasting Procedure," *Applied Statistics*, 27

(1978), 264-279; Robert Carbone and Spyros Makridakis, "Forecasting When Pattern Changes Occur Beyond the Historical Data," *Management Science*, 32, no. 3 (1986), 252-260; David A. Aaker and Robert Jacobson, "The Sophistication of 'Naive' Modeling," *International Journal of Forecasting*, 3, no. 3/4 (1987), 449-451.

56. J. Scott Armstrong, *Long Range Forecasting* (New York: John Wiley and Sons, 1985), 487.

57. Thomas E. Vollmann, William L. Berry, and D. Clay Whybark, *Integrated Production and Inventory Management: Revitalizing the Manufacturing Enterprise*, The Business One Irwin/APICS Library of Integrated Resource Management (Homewood, Illinois 60430: Business One Irwin, 1993), 72, 78-81; J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 383; Douglas J. Dalrymple, "Sales Forecasting Practices: Results from a United States Survey," *International Journal of Forecasting*, 3, no. 3/4 (1987), 383.

58. Everette S. Gardner, Jr., "Exponential Smoothing: The State of the Art," *Journal of Forecasting*, 4 (1985), 1.

59. Gene Woolsey, "The fifth column: The case of the unused forecast or when trying to forecast what's to be, first try y = mx + b," *Interfaces*, 16, no. 6 (November-December 1986), 58-60.

60. R. Fildes "Quantitative Forecasting -- The State of the Art: Extrapolative Models," *Journal of the Operational Research Society*, 30, no. 8 (1979), 700; Giorgio Bodo and Luigi Federico Signorini, "Short-Term Forecasting of the Industrial Production Index," *International Journal of Forecasting*, 3, no. 2 (1987), 245.

61. Giorgio Bodo and Luigi Federico Signorini, "Short-Term Forecasting of the Industrial Production Index," *International Journal of Forecasting*, 3, no. 2 (1987), 245.

62. Thomas E. Vollmann, William L. Berry, and D. Clay Whybark, *Integrated Production and Inventory Management: Revitalizing the Manufacturing Enterprise*, The Business One Irwin/APICS Library of Integrated Resource Management (Homewood, Illinois 60430: Business One Irwin, 1993), 71.

63. Giorgio Bodo and Luigi Federico Signorini, "Short-Term Forecasting of the Industrial Production Index," *International Journal of Forecasting*, 3, no. 2 (1987), 245.

64. Richard Ashley, "On the relative worth of recent macro-economic forecasts," *International Journal of Forecasting*, 4, no. 3 (1988), 363 - 376; Richard Ashley, "On the Usefulness of Macroeconomic Forecasts as Inputs to Forecasting Models," *Journal of Forecasting*, 2 (1983), 211-223.

65. Laura Irwin Langbein, *Discovering Whether Programs Work: A Guide to Statistical Methods for Program Evaluation*, The Scott, Foresman Public Policy Analysis and Management Science Series, Arnold J. Meltsner and Mark H Moore, Eds. (Glenview, Illinois: Scott, Foresman and Company, 1980), 31 - 39; David Nachmias and Chava Nachmias, *Research Methods in the Social Sciences*, 2nd ed. (New York: St. Martin's Press, 1981), 82 ff.

66. Alexander Rosenberg, *Philosophy of Social Science*, Dimensions of Philosophy Series, Norman Daniels and Keith Lehrer, eds. (Boulder: Westview Press, 1988), 9; David Hume, *An Inquiry Concerning Human Understanding* in *David Hume: On Human Nature and the Understanding*, Antony Flew, ed., Collier Classics in the History of Thought, Crane Brinton and Paul Edwards, eds. (New York: The Macmillan Company, 1971), 76 ff.

67. Laura Irwin Langbein, *Discovering Whether Programs Work: A Guide to Statistical Methods for Program Evaluation*, The Scott, Foresman Public Policy Analysis and Management Science Series, Arnold J. Meltsner and Mark H Moore, Eds. (Glenview, Illinois: Scott, Foresman and Company, 1980), 31 - 39; David Nachmias and Chava Nachmias, *Research Methods in the Social Sciences*, 2nd ed. (New York: St. Martin's Press, 1981), 82 ff.

68. J. Scott Armstrong, *Long-Range Forecasting,* 2nd ed. (New York: John Wiley & Sons: 1985) 204; SAS Institute Inc., *SAS® User's Guide: Statistics - Version 5 Edition* (Cary, North Carolina: SAS Institute Inc.: 1985), 763 - 774.

69. Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting,* 8, no. 4 (December 1992), 578; Lewis W. Coopersmith, "Forecasting Time Series Which are Inherently Discontinuous," *Journal of Forecasting,* 2 (1983), 225-235.

70. Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting*, 8, no. 4 (December 1992), 578.

71. Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting*, 8, no. 4 (December 1992), 575-582; Spyros Makridakis, Michele Hibon, Ed Lusk, and Moncef Belhadjali, "Confidence Intervals: An Empirical Investigation of the Series in the M-Competition," *International Journal of Forecasting*, 3, no. 3/4 (1987), 499.

72. Nagesh S. Revankar, On the Problem of Forecasting Prior to 'Price' Control or Decontrol, *Journal of Forecasting*, 11, no. 1 (1992), 1 - 15.

73. D. W. Trigg, and A. G. Leach. Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, 18, no. 1 (1967), 53.

74. D. W. Trigg, and A. G. Leach. Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, 18, no. 1 (1967), 53.

75. D. W. Trigg, and A. G. Leach. Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, 18, no. 1 (1967), 53.

76. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*. 2d ed. (New York: John Wiley & Sons, 1985), 165.

77. J. Scott Armstrong, *Long-Range Forecasting* (New York, John Wiley & Sons, 1985), 171.

78. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*. 2d ed. (New York: John Wiley &

Sons, 1985), 165.

79. D. W. Trigg, and A. G. Leach. Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, 18, no. 1 (1967), 53-59.

80. Forecaster at Department of Corrections, personal communication, March 5, 1993.

81. Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting*, 8, no. 4 (1992), 575-582.

Endnotes Chapter 4

82. Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting*, 8, no. 4 (1992), 578.

83. Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting*, 8, no. 4 (December 1992), 577.

84. Everette S. Gardner, Jr., "Automatic Monitoring of Forecasting Errors," *Journal of Forecasting*, 2 (1983), 1-21.

85. John O. McClain, "Dominant Tracking Signals," *International Journal of Forecasting*, 4, no. 4 (1988), 563-572.

86. L. W. Coopersmith, "Forecasting Time Series Which are Inherently Discontinuous," *Journal of Forecasting*, 2 (1983), 225-235.

87. Thomas E. Vollmann, William L. Berry, and D. Clay Whybark, *Integrated Production and Inventory Management: Revitalizing the Manufacturing Enterprise*, The Business One Irwin/APICS Library of Integrated Resource Management (Homewood, Illinois 60430: Business One Irwin, 1993), 84-85.

88. D. W. Trigg and A. G. Leach, "Exponential Smoothing with an Adaptive Response Rate," *Operational Research Quarterly*, 18, no. 1 (1967), 53-59.

89. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 114.

90. Robert Dancer and Clifford Gray, "An Empirical Evaluation of constant and Adaptive Computer Forecasting Models for Inventory Control," *Decision Sciences*, 8 (1977), 228-238.

91. T. M. Williams, "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 6 (June 1987), 553-560.

92. T. M. Williams, "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 6 (June

1987), 555.

93. T. M. Williams. "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no.6, (June 1987), 556.

94. Blyth C. Archibald, "Parameter Space of the Holt-Winters' model," *International Journal of Forecasting*, 6, no. 2 (July 1990), 200

95. Samuel Eilon and Joseph Elmaleh, "Adaptive Limits in Inventory Control," *Management Science*, 16, no. 8 (April, 1970), B-533 - B-548.

96. Steinar Ekern, "Adaptive Exponential Smoothing Revisited," *Journal of the Operational Research Society*, 32 (1981), 775-782.

97. H. Theil, and S. Wage, "Some Observations on Adaptive Forecasting," *Management Science*, 10, no. 2 (January, 1964), 198-206.

98. M. Nerlove and S. Wage, "On the Optimality of Adaptive Forecasting," *Management Science*, 10, no. 2 (1964), 207-223.

99. T. M. Williams, "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 5 (1987), 560.

100. James B. McDonald, "Partially Adaptive Estimation of ARMA Time Series Models," *International Journal of Forecasting*, 5, no. 2 (1989), 217-230.

101. J. Scott Armstrong, *Long-Range Forecasting* (New York, John Wiley & Sons, 1985), 171.

102. C. Chatfield, "The Holt-Winters Forecasting Procedure," *Applied Statistics*, 27, no. 3 (1978), 264-279.

103. D. W. Trigg and A. G. Leach, "Exponential Smoothing with an Adaptive Response Rate," *Operational Research Quarterly*, 18, no. 1 (1967), 53.

104. Robert Dancer and Clifford Gray, "An Empirical Evaluation of constant and Adaptive Computer Forecasting Models for Inventory Control," *Decision Sciences*, 8 (1977), 231.

105. C. Chatfield, "The Holt-Winters Forecasting Procedure," *Applied Statistics*, 27, no. 3 (1978), 264-279.

106. P. J. Harrison and C. F. Stevens, "Bayesian Forecasting," *Journal of the Royal Statistical Society* Series B, 34, no. 3 (1976), 205-230; Nancy J. Kirkendall, "Monitoring for Outliers and Level Shifts in Kalman Filter Implementation of Exponential Smoothing," *Journal of Forecasting* 11, no. 6 (1992) 543 - 560, Andrew Harvey and Ralph D. Snyder, "Structural Time Series Models in Inventory Control," *International Journal of Forecasting*, 6, no. 2 (July 1990), 187-198.

107. P. J. Harrison and C. F. Stevens, "A Bayesian Approach to Short-term Forecasting," *Operational Research Quarterly*, 22, no. 4 (1971), 241 - 362.

108. Duk Bin Jun and Robert M. Oliver, "Bayesian Forecasting Following a Major Level Change in Exponential Smoothing," *Journal of Forecasting* 4 (1985), 293-302.

109. Duk B. Jun, "On Detecting and Estimating a Major Level or Slope Change in General Exponential Smoothing," *Journal of Forecasting* 8 (1989), 55-64.

110. Genshiro Kitagawa, "Non-Gaussian State-Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82 (December 1987), 1032-1044.

111. Gwilym M. Jenkins, *Practical Experiences with Modelling and Forecasting Time Series* (Jersey, Channel Islands: Gwilym Jenkins & Partners (Overseas) Ltd., 1979), 38-87; Walter Vandaele, *Applied Time Series and Box-Jenkins Models* (New York: Harcourt Brace Jovanovich, Publishers, 1983), 257-300.

112. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 906.

113. Ruey S. Tsay, "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7 (1988), 1-20; Ruey S. Tsay, "Comment: Detecting and Modeling Changes in Time Series," *Journal of the American Statistical Association*, 82 (1987), 1056-1059.

114. David J. Pack, "In Defense of ARIMA Modeling," *International Journal of Forecasting*, 6, no. 2 (July 1990), 211-218.

115. Steven Hillmer, "Monitoring and Adjusting Forecasts in the Presence of Additive Outliers," *Journal of Forecasting* 3 (1984), 205-215.

116. Giorgio Bodo and Luigi Federico Signorini, "Short-Term Forecasting of the Industrial Production Index," *International Journal of Forecasting*, 3, no. 2 (1987), 245.

117. Victor M. Guerrero, "ARIMA Forecasts with Restrictions Derived from a Structural Change," *International Journal of Forecasting*, 7, no. 3 (1991), 339-347.

118. Victor M. Guerrero, "ARIMA Forecasts with Restrictions Derived from a Structural Change," *International Journal of Forecasting*, 7, no. 3 (1991), 339.

119. Victor M. Guerrero, "ARIMA Forecasts with Restrictions Derived from a Structural Change," *International Journal of Forecasting*, 7, no. 3 (1991), 340.

120. Spyros Makridakis and Robert Carbone, "Forecasting When Pattern Changes Occur Beyond the Historical Data," *Management Science*, 32, no. 3 (1986), 257-271.

121. T. M. Williams "Adaptive Holt-Winters Forecasting," *Journal of the Operational Research Society*, 38, no. 6 (June 1987), 557; Everette S. Gardner, Jr., "Exponential Smoothing: The State of the Art," *Journal of Forecasting*, 4 (1985), 11; Steinar Ekern, "Adaptive Exponential Smoothing Revisited," *Journal of the Operational Research Society*, 32, no. 9 (1981), 778; E. McKenzie, "Error Analysis for Winters' Additive Seasonal Forecasting System," *International Journal of Forecasting*, 2 (1986), 373-382 as cited in Paul Newbold and Ted Bos, "On Exponential Smoothing and the Assumption of Deterministic Trend Plus White Noise Data-Generating Models," *International Journal of Forecasting*, 5, no. 4 (1989), 525; D. W. Trigg, and A. G. Leach. Exponential smoothing with an adaptive response rate, *Operational Research Quarterly*, 18, no. 1 (1967), 53.

122. C. Chatfield, "The Holt-Winters Forecasting Procedure," *Applied Statistics*, 27, no. 3 (1978), 264-279.

123. Robert Fildes and Spyros Makridakis, "Forecasting and Loss Functions," *International Journal of Forecasting*, 4, no. 4 (1988), 546; Everette S. Gardner, Jr. and Spyros Makridakis, "The Future of Forecasting," *International Journal of Forecasting*, 4, no. 3 (1988), 326.

124. Wilpen L. Gorr, "Special Event Data in Shared Databases," *MIS Quarterly*, 10 (September 1986), 239-250; Wilpen L. Gorr, "Use of Special Event Data in Government Information Systems," *Public Administration Review*, 46, Special Issue (November 1986), 532-539.

125. Rudolf Lewandowski, "Sales Forecasting by FORSYS," *Journal of Forecasting*, 1 (1982), 205-214.

126. S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1 (1982), 127.

127. Rudolf Lewandowski, "Sales Forecasting by FORSYS," *Journal of Forecasting*, 1 (1982), 209.

128. Rudolf Lewandowski, "Sales Forecasting by FORSYS," *Journal of Forecasting*, 1 (1982), 209-210.

129. Rudolf Lewandowski, "Sales Forecasting by FORSYS," *Journal of Forecasting*, 1 (1982), 209-210.

130. Wilpen L. Gorr, "Special Event Data in Shared Databases," *MIS Quarterly*, 10 (September 1986), 240-241; Wilpen L. Gorr, "Use of Special Event Data in Government Information Systems," *Public Administration Review*, 46, Special Issue (November 1986), 533.

131. Robert L. McLaughlin, "A Model of an Average Recession and Recovery," *Journal of Forecasting*, 1 (1982), 55-65.

132. Jose Juan Carreno and Jesus Madinaveitia, "A Modification of Time Series Forecasting Methods for Handling Announced Price Increases," *International Journal of Forecasting*, 6, no. 4 (1990), 479-484.

133. Jose Juan Carreno and Jesus Madinaveitia, "A Modification of Time Series Forecasting Methods for Handling Announced Price Increases," *International Journal of Forecasting*, 6, no. 4 (1990), 483.

134. J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 365; Vijay Mahajan and Yoram Wind, "New Product Forecasting Models: Directions for Research and Implementation," *International Journal of Forecasting*, 4, no. 3 (1988), 341-358; James E.

Cox, "An Assessment of Books Relevant to Forecasting in Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 519-526; Christopher J. Easingwood, "An Analogical Approach to the Long Term Forecasting of Major New Product Sales," *International Journal of Forecasting*, 5, no. 1 (1989), 69-82; John C. Chambers, Satinder K. Mullick, and Donald D. Smith, "How to Choose the Right Forecasting Technique," *Harvard Business Review* (July-August 1971), 45 - 74.

135. John A. Sharp and David H. R. Price, "Experience Curve Models in the Electronic Supply Industry," *International Journal of Forecasting*, 6, no. 4 (1990), 531-540; Denis R. Towill, "Forecasting Learning Curves," *International Journal of Forecasting*, 6, no. 1 (1990), 25-38; Sydney D. Howell, "Parameter Instability in Learning Curve Models," *International Journal of Forecasting*, 6, no. 4 (199), 541-547; Mohamed M. Naim and Denis R. Towill, "An Engineering Approach to LSE Modelling of Experience Curves in the Electricity Supply Industry," *International Journal of Forecasting*, 6, no. 4 (1990), 549-556; Peg Young and J. Keith Ord, "Model Selection and Estimation for Technological Growth Curves," *International Journal of Forecasting*, 5, no. 4 (1989), 501-513; Ronald Bewley and Denzil G. Fiebig, "A Flexible Logistic Growth Model with Applications in Telecommunications," *International Journal of Forecasting*, 4, no. 2 (1988), 177-192.

136. J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 367.

137. Christopher J. Easingwood "An Analogical Approach to the Long Term Forecasting of Major New Product Sales," *International Journal of Forecasting*, 5, no. 1 (1989), 76.

138. J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 365.

139. Rodney D. Steward, *Cost Estimating*, 2nd ed. (New York, John Wiley & Sons, Inc., 1990).

140. Gene Woolsey, "The Fifth Column: The Case of the Unused Forecast or When Trying to Forecast What's To Be, First Try Y = MX + B," *Interfaces,* 16, no. 6 (November-December 1986), 58-60.

Endnotes Chapter 5

141. Dennis A. Ahlburg and Kenneth C. Land, "Population Forecasting: Guest Editors' Introduction," *International Journal of Forecasting*, 8, no. 3 (1992), 289-299; Stanley K. Smith and Terry Sincich "Forecasting State and Household Populations: Evaluating the Forecast Accuracy and Bias of Alternative Populations Projections for States," *International Journal of Forecasting*, 8, no. 3 (1992), 495-508.

142. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer.* 2d ed. (New York: John Wiley & Sons, 1985), 57 - 64.

143. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer.* 2d ed. (New York: John Wiley & Sons, 1985), 58.

144. J. Scott Armstrong, *Long-Range Forecasting: From Crystal Ball to Computer*, 2d. ed. (New York, John Wiley & Sons, 1985), 81-82.

145. P. J. Harrison and C. F. Stevens, "Bayesian Forecasting," *Journal of the Royal Statistical Society*, Series B, 34, no. 3 (1976), 205-228; M. B. Priestly, "Discussion on the Paper by Professor Harrison and Mr. Stevens,: *Journal of the Royal Statistical Society*, Series B, 34, no. 3 (1976), 229; Thomas E. Vollmann, William L. Berry, and D. Clay Whybark, *Integrated Production and Inventory Management: Revitalizing the Manufacturing Enterprise*, The Business One Irwin/APICS Library of Integrated Resource Management (Homewood, Illinois 60430: Business One Irwin, 1993), 69-90.

146. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed. (New York: John Wiley & Sons, 1983), 149-151.

147. Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22.

148. Thomas E. Vollmann, William L. Berry, and D. Clay Whybark, *Integrated Production and Inventory Management: Revitalizing the Manufacturing Enterprise*, The Business One Irwin/APICS Library of Integrated Resource Management (Homewood, Illinois 60430: Business One Irwin, 1993), 81-82.

149. Personal communication with Don Miller on September 3, 1993.

150. J. Scott Armstrong, *Long-Range Forecasting: From Crystal Ball to Computer*, 2d ed. (New York: John Wiley & Sons, 1985), 61.

151. Alan Pankratz, "Time Series Forecasts and Extra-model Information," *Journal of Forecasting* 8 (1989), 75-83.

Endnotes Chapter 6

152. Stephen K. McNees, "On the Future of Macroeconomic Forecasting," *International Journal of Forecasting*, 4, no. 3 (1988), 359-362.

153. Stephen K. McNees, "On the Future of Macroeconomic Forecasting," *International Journal of Forecasting*, 4, no. 3 (1988), 359-362; J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 365; Michael D. Geurts and J. Patrick Kelly, "Comments on: 'In defense of ARIMA modeling', by D. J. Pack," *International Journal of Forecasting*, 6, no. 4 (1990), 497-499; Spyros Makridakis and Michele Hibon, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *International Journal of Forecasting*, 7, no. 3 (1991), 317-330.

154. Michael D. Geurts and J. Patrick Kelly, "Comments on: 'In defense of ARIMA modeling', by D. J. Pack," *International Journal of Forecasting*, 6, no. 4 (1990), 497-499.

155. Michael D. Geurts and J. Patrick Kelly, "Comments on: 'In defense of ARIMA modeling', by D. J. Pack," *International Journal of Forecasting*, 6, no. 4 (1990), 497-499.

156. Michael D. Geurts and J. Patrick Kelly, "Comments on: 'In defense of ARIMA modeling', by D. J. Pack," *International Journal of Forecasting*, 6, no. 4 (1990), 497-499.

157. J. Scott Armstrong, Roderick J. Brodie, and Shelby H. McIntyre, "Forecasting Methods for Marketing," *International Journal of Forecasting*, 3, no. 3/4 (1987), 365.

158. H. O. Stekler, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, no. 3 (1991), 375-384.

159. H. O. Stekler, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, no. 3 (1991), 375-384.

160. H. O. Stekler, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, no. 3

(1991), 375-384.

161. Patrick A. Thompson, "An MSE Statistic for Comparing Forecast Accuracy Across Series," *International Journal of Forecasting*, 6, no 2. (July 1990), 219-227; Benito E. Flores, "The Utilization of the Wilcoxon Test to Compare Forecasting Methods: A Note," *International Journal of Forecasting*, 5, no. 4 (1989), 529-535; R. A. Kolb and H. O. Stekler, "Are Economic Forecasts Significantly Better than Naive Predictions? An Appropriate Test," *International Journal of Forecasting*, 9, no. 1 (April 1993), 117-120; Peter J. Danaher and Roderick J. Brodie, "Predictive Accuracy of Simple Versus Complex Market Share Models," *International Journal of Forecasting*, 8, no. 4 (December 1992), 613-626; Robert Fildes and Spyros Makridakis, "Forecasting and Loss Functions," *International Journal of Forecasting*, 4, no. 4 (1988), 545-550; Pierre Lefrançois, "Allowing for Asymmetry in Forecast Errors: Results from a Monte-Carlo Study," *International Journal of Forecasting*, 5. no. 1 (1989), 99-110; Patrick A. Thompson, "Evaluation of the M-competition Forecasts via Log Mean Squared Error Ratio,' *International Journal of Forecasting*, 7, no. 3 (November 1991), 331-334; Lap-Ming Wun and Wen Lea Pearn, "Assessing the Statistical Characteristics of the Mean Absolute Error [of] Forecasting," *International Journal of Forecasting*, 7, no. 3 (November 1991), 335-337; J. Scott Armstrong and Fred Collopy, "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons," *International Journal of Forecasting*, 8, no. 1 (June 1992), 69-80; Robert Fildes, "The Evaluation of Extrapolative Forecasting Methods," *International Journal of Forecasting*, 8, no. 1 (April 1992), 81-98; Benito E. Flores, "A Pragmatic View of Accuracy Measurement in Forecasting," *Omega*, 14, no. 2 (1986), 93-98; Robert L. McLaughlin, "Measuring the Accuracy of Forecasts," *Business Economics*, 7 (May 1972), 27-35.

162. Spyros Makridakis, Michele Hibon, Ed Lusk, and Moncef Belhadjali, "Confidence Intervals: An Empirical Investigation of the Series in the M-Competition," *International Journal of Forecasting*, 3, no. 3/4 (1987), 489-508.

163. Stanley K. Smith and Terry Sincich, "Forecasting State and Household Populations: Evaluating the Forecast Accuracy and Bias of Alternative Populations Projections for States," *International Journal of Forecasting*, 8, no. 3 (1992), 495-508; Spyros Makridakis, "Metaforecasting: Ways of Improving Forecasting Accuracy and Usefulness," *International Journal

of Forecasting, 4, no. 3 (1988), 476; Robert Fildes and Spyros Makridakis, "Forecasting and Loss Functions," *International Journal of Forecasting*, 4, no. 4 (1988), 546; Everette S. Gardner, Jr. and Spyros Makridakis, "The Future of Forecasting," *International Journal of Forecasting*, 4, no. 3 (1988), 326.

164. Spyros Makridakis, Steven C. Wheelwright, and Victor E. McGee, *Forecasting: Methods and Applications*, 3rd ed., New York: John Wiley & Sons, 1983, 891.

165. H. O. Stekler, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, no. 3 (1991), 377.

166. Rudolf Lewandowski, "Sales Forecasting by FORSYS," *Journal of Forecasting*, 1 (1982), 205-214; M. Nerlove and S. Wage, "On the Optimality of Adaptive Forecasting," *Management Science*, 10, no. 2 (January, 1964), 207-223.

167. R. D. Snyder, "Progressive Tuning of Simple Exponential Smoothing Forecasts," *Journal of the Operational Research Society*, 39, no. 4 (1988), 393-399; P. J. Harrison and C. F. Stevens, "A Bayesian Approach to Short-term Forecasting," *Operational Research Quarterly* 22, no. 4 (1971), 341-362.

168. Christine A. Martin and Stephen F. Witt, "Forecasting Tourism Demand: A Comparison of the Accuracy of Several Quantitative Methods," *International Journal of Forecasting*, 5, no. 1 (1989), 7-20.

169. S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J Newton, E. Parzen and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1 (1982), 111-153; Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22.

170. Jose Juan Carreno and Jesus Madinaveitia, "A Modification of Time Series Forecasting Methods for Handling Announced Price Increases," *International Journal of Forecasting*, 6, no. 4 (1990), 483; Charles J. LaCivita and Terry G. Seaks, "Forecasting Accuracy and the Choice of First Difference or Percentage Change Regression Models," *International Journal of Forecasting*, 4, No. 2 (1988), 261-

268; Wilpen L. Gorr, "Special Event Data in Shared Databases," *MIS Quarterly*, 10 (September 1986), 239-250; Moheb A. Ghali, "Seasonality, Aggregation and the Testing of the Production Smoothing Hypothesis," *The American Economic Review*, 77, no. 3 (June 1987), 464-469; Donald Rosenfield, "A Model for Predicting Frequencies of Random Events," *Management Science*, 33, no. 8 (August 1987), 947-954; Peter R. Winters, "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, 7 (April 1960), 324-342.

171. Richard Ashley, "On the Relative Worth of Recent Macro-Economic Forecasts," *International Journal of Forecasting*, 4, no. 3 (1988), 363-376.

172. Spyros Makridakis and Robert Carbone, "Forecasting When Pattern Changes Occur Beyond the Historical Data," *Management Science*, 32, no. 3 (1986), 257-271; R. Fildes, "Quantitative Forecasting - The State of the Art, *Journal of the Operational Research Society*, 30, no. 8 (1979), 691-710; P. Narayan Pant and William Starbuck, "Innocents in the Forest: Forecasting and Research Methods," *Journal of Management*, 16, no. 2 (1990), 433-460; Everette S. Gardner, Jr. and Ed McKenzie, "Seasonal Exponential Smoothing with Dampened Trends," *Management Science*, 35, no. 3 (March 1989), 372-376; Laurette Poulos, Alan Kvanli and Robert Pavur, "A Comparison of the Accuracy of the Box-Jenkins Method with that of Automated Forecasting Methods," *International Journal of Forecasting*, 3, no. 2 (1987), 261-267.

173. Douglas J. Dalrymple, "Sales Forecasting Practices," *International Journal of Forecasting* 3, no. 3/4 (1987), 379-391; John Hanke, "Forecasting in Business Schools, A follow-up survey," *International Journal of Forecasting*, 5, no. 2 (1989), 259-262; Fred Collopy and J. Scott Armstrong, "Expert Opinions About Extrapolation and the Mystery of the Overlooked Discontinuities," *International Journal of Forecasting*, 8, no. 4 (December 1992), 575-582.

174. P. Narayan Pant and William Starbuck, "Innocents in the Forest: Forecasting and Research Methods," *Journal of Management*, 16, no. 2 (1990), 433-460; Lewis W. Coopersmith, "Forecasting Times Series Which are Inherently Discontinuous," *Journal of Forecasting*, 2 (1983), 225-235; Ruey S. Tsay, "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7 (1988), 1-20.

175. P. George Benson and Dilek Onkal, "The Effects of Feedback and Training on the Performance of Probability Forecasters," *International Journal of Forecasting*, 8, no. 4 (December, 1992), 559-573; Steven P. Schnaars and Martin T. Topol, "The Use of Multiple Scenarios in Sales Forecasting, *International Journal of Forecasting*, 3, no. 3/4 (1987), 405-419.

176. Stuart Ira Bretschneider, "Adaptive Time Series Analysis and Forecasting," (Ph.D. diss., The Ohio State University, 1979), 38 ff.

177. Stuart Ira Bretschneider, "Adaptive Time Series Analysis and Forecasting," (Ph.D. diss., The Ohio State University, 1979), 37.

178. Stuart Ira Bretschneider, "Adaptive Time Series Analysis and Forecasting," (Ph.D. diss., The Ohio State University, 1979), 41.

179. S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1 (1982), 111-153; Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22.

180. Stuart Ira Bretschneider, "Adaptive Time Series Analysis and Forecasting," (Ph.D. diss., The Ohio State University, 1979) 38 ff.; S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1 (1982), 111-153; Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22.

181. S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting*, 1 (1982), 111-153.

182. Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22.

183. Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons, "The M2-Competition: A Real-Time Judgmentally Based Forecasting Study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 18.

184. Spyros Makridakis and Michele Hibon, "Accuracy of Forecasting: An Empirical Investigation (with Discussion)" reprinted as chapter 3 in *The Forecasting Accuracy of Major Time Series Methods*, S. Makridakis, et. al., eds., (Chichester: John Wiley & Sons, 1984).

185. R. A. Kolb and H. O. Stekler, "Are Economic Forecasts Significantly Better than Naive Predictions? An Appropriate Test," *International Journal of Forecasting*, 9, no. 1 (April 1993), 117-120.

186. Robert L. McLaughlin, "Measuring the Accuracy of Forecasts," *Business Economics*, 7 (May 1972), 27-35.

187. Benito E. Flores, "A Pragmatic View of Accuracy Measurement in Forecasting," *Omega*, 14, no. 2 (1986), 93-98.

188. Benito E. Flores, "A Pragmatic View of Accuracy Measurement in Forecasting," *Omega*, 14, no. 2 (1986), 96.

189. J. Scott Armstrong and Fred Collopy, "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons," *International Journal of Forecasting*, 8, no. 1 (June 1992), 69-80.

190. Patrick A. Thompson, "An MSE Statistic for Comparing Forecast Accuracy Across Series," *International Journal of Forecasting*, 6, no 2. (July 1990), 219-227; Patrick A. Thompson, "Evaluation of the M-competition Forecasts via Log Mean Squared Error Ratio,' *International Journal of Forecasting*, 7, no. 3 (November 1991), 331-334.

191. Patrick A. Thompson, "Evaluation of the M-competition Forecasts via Log Mean Squared Error Ratio,' *International Journal of Forecasting*, 7, no. 3 (November 1991), 331-332.

192. Robert Fildes, "The Evaluation of Extrapolative Forecasting Methods," *International Journal of Forecasting*, 8, no. 1 (April 1992), 81-98.

193. Benito E. Flores, "The Utilization of the Wilcoxon Test to Compare Forecasting Methods: A Note," *International Journal of Forecasting*, 5, no. 4 (1989), 529-535.

194. Benito E. Flores, "The Utilization of the Wilcoxon Test to Compare Forecasting Methods: A Note," *International Journal of Forecasting*, 5, no. 4 (1989), 535.

195. R. A. Kolb and H. O. Stekler, "Are Economic Forecasts Significantly Better than Naive Predictions? An Appropriate Test," *International Journal of Forecasting*, 9, no. 1 (April 1993), 118.

196. H. O. Stekler, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, no. 3 (1991), 375-384.

197. Spyros Makridakis and Michele Hibon, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *International Journal of Forecasting*, 7, no. 3 (1991), 317-330.

198. Spyros Makridakis, Chris Chatfield, Michele Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons "The M2-competition: A real-time judgmentally based forecasting study," *International Journal of Forecasting*, 9, no. 1 (April 1993), 5-22.

199. Gottfried E. Noether, *Introduction to Statistics* (New York: Houghton Mifflin Company, 1971), 147.

Endnotes Chapter 7

200. Gwilym M. Jenkins, *Practical Experiences with Modelling and Forecasting Time Series* (Jersey, Channel Islands: Gwilym Jenkins & Partners (Overseas) Ltd., 1979), 17.

201. Spyros Makridakis, and Michele Hibon, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *International Journal of Forecasting*, 7, no. 3 (1991), 317-330.

202. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*. 2d ed. (New York: John Wiley & Sons, 1985), 162-163.

203. J. Scott Armstrong, *Long-Range Forecasting, From Crystal Ball to Computer*, 2d ed. (New York: John Wiley & Sons, 1985), 161-163.

204. J. Scott Armstrong, *Long-Range Forecasting* (New York, John Wiley & Sons, 1985), 171.

Endnotes Chapter 9

205.Everette S. Gardner, Jr. and Spyros Makridakis, "The Future of Forecasting," *International Journal of Forecasting*, 4, no. 3 (1988), 326.

206. Spyros Makridakis and Michele Hibon, "Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy," *International Journal of Forecasting*, 7, no. 3 (1991) 317-330.

207.Steven Hillmer, "Monitoring and Adjusting Forecasts in the Presence of Additive Outliers," *Journal of Forecasting* 3 (1984): 205-215.

208.Spyros Makridakis, Steven C. Wheelright and Victor E. McGee, *Forecasting Methods and Applications*, 2nd ed. (New York: John Wiley & sons: 1983) 122.

209.George C. Canavos and Don M. Miller *An Introduction to Modern Business Statistics*, (Belmont, California: Duxbury Press: 1993) 679.

APPENDIX I TERMS

Abnormal Errors - Errors that form an identifiable non-random pattern.

Ad Hoc Method/Model - A forecast method that forecasts a data series with a statistical forecast model then adds on a lump sum amount to adjust the forecast for an anticipated policy change.

Ad Hoc Technique - Same as the *ad hoc* method.

Adaptive Forecasting - Forecast models that use some information about forecast error to adjust the value of a forecast parameter.

Adaptive Holt-Winters-Williams - An exponential smoothing forecast technique defined in Appendix II.

Adaptive Techniques - Same as adaptive forecasting.

Additive Seasonality - Seasonality factors that increase or decrease the level by additive factors.

Additive Trend - A trend calculated by adding an incremental value back to a prior level.

Adjusted Holt-Winters-Williams - The forecast technique proposed in this dissertation.

Autocorrelation Corrected Holt-Winters-Williams - A forecast technique that uses the Holt-Winters-Williams exponential smoothing model and also corrects for autocorrelation following the method mentioned by Chatfield.

Analogy - Reasoning from one case to another and borrowing information from the source case, may be either subjective or based on mathematical techniques that rest, in part, on analogy between a new case and old cases.

Analysis of Variance by Rank - See Rank ANOVA.

ARIMA - Autoregressive Integrated Moving Average, a sophisticated use of differences and moving averages in order to forecast data.

Autocorrelation corrected Holt-Winters-Williams - A version of exponential smoothing that uses autocorrelation as a factor similar to a tracking signal.

Autocorrelation - Interdependence (correlation) between observations in the same data series.

Average Percent Error - A relative statistic of bias.

Backward Cusum - A tracking signal.

Cartesian XY Graph - A graph that locates data defined by two variables on a vertical axis and a horizontal axis.

Central Tendency - The average or another statistic that estimates the center of a data series. For forecasts, commonly the level.

Classic Decomposition Models - Forecast models that break down data series by major sources of variation, rather than forecasting aggregated data.

Correlation Based Techniques - Forecasting techniques that use the correlation between two or more data series to forecast one of these series, commonly regression.

Current Period - The period (location on the X axis) of an observation.

Curve - The line formed by connection the individual observations of a series, may be applied to either raw data or to a summarization, such as a forecast. In this usage curves are not necessarily smooth.

Data Series - A set of observations that are organized in order (e.g., over time).

Dampen - This is a forecast technique that makes the trend parameter exponentiate to zero in future periods to reflect an assumption that growth is not permanent.

Decompose - Break down a data series into components, may be additive or multiplicative.

Deseasonalizing - Decomposing a data series by removing the seasonal aspect of the series, may be additive or multiplicative.

Differences - For serial data, the data series made of subtracting the data series X_{t-n} X_{t+1-n}, X_{t+2-n}, ... from the series X_t, X_{t+1}, X_{t+2}, ..., such that $D_t = X_t - X_{t-n}$, etc. When the differences are calculated from raw (undifferenced) numbers, the differences are called first differences. When the series is calculated from a differenced series of order m, it is called the m+1 differences, e.g., second differences are the differences of first differences, etc. When n = 1, the series is the differences of the **first period**, etc. The first difference of the first period is normally called the **first difference**.

Discontinuities - A general term referring to phenomena like level shifts.

Dummy Variable - In correlation based techniques. a variable that has either the value of 1 or 0, generally used to account for a condition that is only sometimes present with the series.

Econometric Techniques - Correlation based techniques, generally referring to the more sophisticated of these techniques.

Endogenous/Exogenous Variables - These terms denote the dichotomy of variables that are both input and output in a forecast model (endogenous variables) and variables that are only input in a forecast model (exogenous variables). Forecast models can generate their own future values of endogenous variables, but are dependent on external sources for future values of exogenous variables.

Engineering Estimates - Cost estimate based on costing out actual component cost generating activities and building up the overall cost from these components.

Error - Actual Observation minus Predicted Observation.

Exponential Smoothing - A forecast technique that extrapolates a series of data through a weighted averaging technique that places more weight on recent data and less weight on older data. Several specific versions of this technique are defined mathematically in Appendix II.

Exponentiated - Generally, this term refers to a number that is multiplied by itself over and over again. When a number is less than 1 and is exponentiated, it soon becomes insignificantly small. In forecasting serial data, a number might be exponentiated as it moves away from the last actual observation in a data series.

Extrapolation - Forecasting by identifying factors like level and trend and projecting them into the future, generally refers to the use of time series techniques.

First Difference - See Difference.

Fit - Calibrate the parameters that best apply a forecast model to a data series.

Forecast - This term may mean forecast model (see forecast model), a projection (see projection) or the practice of applying a model to make a projection.

Forecast Competition - A fairly common form of forecast comparison in which a number of forecast techniques are used to forecast the same data series to determine whether any particular technique is more effective.

Forecast Horizon - The number of periods beyond the current period for which a forecast is desired.

Forecast Model - A set of mathematical formulae that are used to project future values of a data series.

Forecasting - Using any technique to project serial numeric data into the future.

Geometric Average - An average calculated by multiplying observations and finding the root equal to the number of multiplicands of the product.

Geometric Root Mean Squared Error - An error statistic defined in Appendix VII.

Historical Period - For serial data, the period for which there is data.

Hold Out Data - Serial data near the end of the historical period which are not included when fitting forecast model to a data series so that it can be used to evaluate the effectiveness of the model that is fit.

Holt exponential smoothing - Trended exponential smoothing. See formulas in Appendix II.

Holt-Winters exponential smoothing - Trended and seasonalized exponential smoothing. See formulas in Appendix II.

Holt-Winters-Williams exponential smoothing - A modified version of Holt-Winters as defined by T. M. Williams and discussed in Appendix II.

Horizon - the period or periods ahead for which a projection is desired from a forecast model.

Intervention Model - An ARIMA model that uses dummy variables.

Judgmental Adjustments - Generally, correcting a the results of a forecast model for information not included in the data or model fitting.

Kalman Filter - A complex technique similar to adaptive forecasting.

Kruskal-Wallis test - An inferential statistic used in this study and defined in Appendix VII.

Lag - For serial data that is associated by correlation, this term indicates an association that is not concurrent, for example where one variable is located at time t and the other is located at time t-1.

Level - The current period central tendency of a data series. With many forecasting techniques the level is time indexed. (Especially used in discussion of exponential smoothing or moving averages).

Level Shift - Within a data series, one or more observations that change in level by more than the slope (trend) and seasonality of the series.

Log Mean Squared Error Ratio - A statistic used for comparing forecasts. See formulas in Appendix VII.

Loss Function - A statistic, or set of statistics, that is (are) optimized in order to fit a model. A loss function is a statistic that represents the cost of error in the forecast.

Lump Sum Changes - A number that is added in lump to judgmentally adjust a data series.

Mean Absolute Deviation - An error statistic.

Mean Absolute Percent Error - An error statistic. See formulas in Appendix VII.

Mean Deviation - Mean Error.

Mean Error - A statistic that may be a bias measure.

Mean Percent Error - An error statistic. See formulas in Appendix VII.

Mean Squared Error - An error statistic. See formulas in Appendix VII.

Median Absolute Percent Error - An error statistic. See formulas in Appendix VII.

Method - See model.

Model - In this dissertation model, technique and method are used interchangeably. All are used to refer to a means of making a forecast or a set of formulas used to make a forecast.

Moving Average - With N observations of serial data, one can compute m + 1 averages each containing n = N - m serial observations where, N > m. When m > 1 and the m + 1 averages are arranged in serial order, they are referred to as a moving average.

Multiplicative Seasonality - A seasonality factor that adjusts level by multiplicative factors.

Multiplicative Trend - A trend calculated by multiplying a ratio times a prior level.

N-Period Ahead Forecast - The forecast at the observation at t_j+n, where t is the index of the last actual observation and updates by an increment of 1 with each addition of 1 observation to the history of the data, j is the index of the updates, and n is the number of periods from t to the observation measured. The n-period ahead point of a repeated forecast updates moves to a later point in time by the number of additional actual observations added to the history with each update. There is one point observation from each jth update and it is located one period later in time.

Naive 1 - Same as the naive method, a forecast of no change.

Naive 2 - A seasonally adjusted version of Naive 1.

Naive Method - A forecast of no change.

Noise - (white noise) error that is not associated with known causes and does not exhibit any observable pattern.

Non-Stationary - The property of a data series of failing to have either a constant mean or a constant variance, for the purposes of this study, has experienced a level shift.

Optimize - Bring a loss function as close as possible to a desired value. In exponential smoothing, loss functions are usually optimized by bringing them to a minimum value.

Optimizing Technique - A method for bringing a loss function as close as possible to a desired value.

Out of Control - For purposes of this study that would mean it had undergone a level shift.

Outliers - Extraordinary observations, sometimes defined as those observations that exceed 3 standard deviations from some measure of central tendency such as the mean or a moving average.

Parameter - In exponential smoothing, a statistic that is adjusted to optimize the model.

Percent Error - An error statistic.

Planned Policy Change - A level shift that can be anticipated in advance, generally because it results from intentional intervention into the data generating function by a decision maker, also, see policy shift.

Policy Change - See planned policy change and policy shift.

Policy Shift - A portion of a data series that has a steeper or less steep slope than the period before or afterwards and which can be directly associated with an external event, typically a policy decision, see discussion in Chapter 2 regarding other labels that might also be used.

Preprocessing - Adjusting data before forecasting

Prospective shift - See policy shift.

Projection - The values predicted by a forecast model, often this is simply called a forecast.

Ramp - A series of 2 or more data observations that have a different slope from those of the surrounding observations.

Random Noise - See noise.

Random Walk - (Naive Method, Naive 1) Forecasting on the assumption that the next observation (or all future observations) will be the same as the last observation.

Rank ANOVA - Analysis of Variance by Rank (the Friedman Test). An inferential statistic used in this dissertation and defined in Appendix VII.

Regression Models - Forecasts made by generating a regression of the historical (sample) data and extending it into the future by extending the input data into the future (generally by using a forecast of the input data, or sometimes by using lagged values of the input data).

Relative Absolute Error - An error statistic.

Relative Geometric Root Mean Squared Error ratio across time periods - An error statistic.

Root Mean Squared Error - An error statistic.

Sample Data - Data used in fitting a forecast model (the data from the historical period).

Seasonal Factors - Numbers that are used to adjust a forecast for regular variation from the central tendency. For example the observations arising in each march might tend to be 20% greater than the average over the year, or might have 20 extra units compared with the average over the year. Factors can be calculated to adjust a forecast to include these two sorts of expectations. The first would lead to a multiplicative seasonal factor (a number multiplied by the deseasonalized data to get a large enough forecast), the second would lead to an additive seasonal factor (a number added to the data to get a large enough forecast).

Serial Data - See data series.

Serially Correlated Errors - Errors (see errors) that have a pattern. See autocorrelation.

Simple Cusum - A tracking signal.

Simulated Policy Changes - Artificial data used to generate level shifts that are similar to those described in Chapter 3.

Single Exponential Smoothing - (SES) A weighted moving average that places more weight on the more recent observations, see exponential smoothing.

Slope - The difference between two successive observations or projections.

Smoothed Error Tracking Signal - A time indexed error statistic.

Smoothing Constant - A parameter for an exponential smoothing model.

Special Events - See level shift or policy shift, as used here, generally temporary in nature.

Standardized Realization Percent (SR) - An error statistic.

Stationary - The property of a data series of having has a constant mean and a constant variance.

Statistical Error - Error arising in a statistical forecasting model, used in calculating future predicted values.

Statistical Forecast Model - A forecast model that uses information about errors (see error) in calculating future predicted values.

Step - A ramp that includes precisely two observations.

Subjective Estimates - Generally refer to the use of expert or management guesses.

Symmetrically Adjusted MAPE (SMAPE) - An error statistic.

Technique - See model.

Three parameter exponential smoothing - Holt-Winters exponential smoothing.

Theil's U-Coefficients - A statistic used to compare a forecast with the forecast that would have been made using the naive method.

Trace - The forecast for periods t+m through t+m+l where t is a time index, m is the number of time periods before the period of interest, and l is the number of time periods in the period of interest. The trace, therefore, is a vector of forecasts F_{t+m}, F_{t+m+1}, F_{t+m+2}, . . . , F_{t+m+l}.

Tracking Signal - A statistic that is sensitive to the errors near the end of the historical observations of a time series which can be used to indicate that the series is not well fit at that point in the series.

Trading Days - The number of days during which a forecasted data series had an opportunity to occur, frequently the number of business days in a week or the number of days in a month.

Transfer function model - an ARIMA model that uses correlation between data series in making a forecast.

Trend - The slope of a data series. With many forecasting techniques the trend is time indexed.

Two Parameter Exponential Smoothing - Holt Exponential Smoothing.

Unadjusted Model - This term is used to refer to those models included in this study which do not account for prospective level shifts.

Updating - Adding a new period's observation to the historical (observed) data and making a new forecast projection.

Variability - The tendency for data to vary.

Variance - A measure of variation using squared errors.

Windsorize - A technique for adjusting away extraordinary errors.

Winters - A seasonality technique for exponential smoothing.

X11 - A moving average technique developed by the Census Bureau.

Statistical Error - Error arising in a statistical forecasting model, used in calculating future predicted values.

Statistical Forecast Model - A forecast model that uses information about errors (see error) in calculating future predicted values.

Step - A ramp that includes precisely two observations.

Subjective Estimates - Generally refer to the use of expert or management guesses.

Symmetrically Adjusted MAPE (SMAPE) - An error statistic.

Technique - See model.

Three parameter exponential smoothing - Holt-Winters exponential smoothing.

Theil's U-Coefficients - A statistic used to compare a forecast with the forecast that would have been made using the naive method.

Trace - The forecast for periods t+m through t+m+l where t is a time index, m is the number of time periods before the period of interest, and l is the number of time periods in the period of interest. The trace, therefore, is a vector of forecasts F_{t+m}, F_{t+m+1}, F_{t+m+2}, . . . , F_{t+m+l}.

Tracking Signal - A statistic that is sensitive to the errors near the end of the historical observations of a time series which can be used to indicate that the series is not well fit at that point in the series.

Trading Days - The number of days during which a forecasted data series had an opportunity to occur, frequently the number of business days in a week or the number of days in a month.

Transfer function model - an ARIMA model that uses correlation between data series in making a forecast.

Trend - The slope of a data series. With many forecasting techniques, the trend is time indexed.

Two Parameter Exponential Smoothing - Holt Exponential Smoothing.

Unadjusted Model - This term is used to refer to those models included in this study which do not account for prospective level shifts.

Updating - Adding a new period's observation to the historical (observed) data and making a new forecast projection.

Variability - The tendency for data to vary.

Variance - A measure of variation using squared errors.

Windsorize - A technique for adjusting away extraordinary errors.

Winters - A seasonality technique for exponential smoothing.

X11 - A moving average technique developed by the Census Bureau.

APPENDIX II FORMULAS

Holt-Williams Exponential Smoothing

1.	F_{t+m}	= Forecast at time t+m	=	$S_t + (B_t * m)$
2.	S_t	= Level at time t	=	$F_t + \alpha e_t$
3.	B_t	= Trend at time t	=	$B_{t-1} + \beta e_t$
4.	e_t	= Error at time t	=	$X_t - F_t$

Where,

X_t = Observation at time t

α = a level parameter subject to $0 \leq \alpha \leq 1$

β = a trend parameter subject to $0 \leq \beta \leq 1$

t = a time index.

m = the number of periods between the current observation period and a forecast period (horizon).

S_0 = 0, or alternatively an initialized value derived from other techniques.

B_0 = 0, or alternatively an initialized value derived from other techniques.

Holt-Winters-Williams Exponential Smoothing

1. F_{t+m} = Forecast at time t+m = $[S_t+(B_t *m)]*I_{t+m-L}$

2. S_t = Level at time t = $(F_t + \alpha e_t)/I_{t-L}$

3. B_t = Trend at time t = $B_{t-1} + \beta e_t/I_{t-L}$

4. e_t = Error at time t = $X_t - F_t$

5. I_t = Seasonality factor at t = $I_{t-L}+\gamma e_t/(S_{t-1}+B_{t-1})$

Where,

X_t = Observation at time t

α = a level parameter subject to $0 \leq \alpha \leq 1$

β = a trend parameter subject to $0 \leq \beta \leq 1$

t = a time index.

m = the number of periods between the current observation period and a forecast period (horizon).

S_0 = 0, or alternatively an initialized value derived from other techniques.

B_0 = 0, or alternatively an initialized value derived from other techniques.

γ = a seasonality parameter subject to $0 \leq \gamma \leq 1$,

L = the length of the seasonality cycle, and

I_0 = $I_1 = . . . = I_{L-1} = 1$, unless initialized by some other technique, and for future periods $I_t = I_{t-1}$.

Holt-Winters-Williams Adaptive

Holt-Winters-Williams Adaptive is identical with Holt-Winters-Williams with the exception that α is subscripted for time and selected using the following algorithm.

Smoothed Error = $E_t = \phi e_t + (1-\phi)E_{t-1}$

Smoothed Absolute Error = $M_t = \phi \mid e_t \mid + (1-\phi)M_{t-1}$

α_t = Absolute Tracking Signal = $T_t = \mid E_t/M_t \mid$

Holt-Winters-Williams Autocorrelation Corrected

1.	F'	= Autocorr. Corrected F	$= F+(e_{t-1} * \varrho_{e,et-1})$
1.	F_{t+m}	= Forecast at time t+m	$= (S_t+ B_t * m) * I_{t+m-L}$
2.	S_t	= Level at time t	$= (F_t' + \alpha e_t)/I_{t-L}$
3.	B_t	= Trend at time t	$= B_{t-1} + \beta e_t/I_{t-L}$
4.	e_t	= Error at time t	$= X_t - F_t'$
5.	I_t	= Seasonality factor at t	$= I_{t-L}+ e_t/(S_{t-1} + B_{t-1})$

Holt-Williams Exponential Smoothing

1.	F_{t+m}	= Forecast at time t+m	=	$S_t + (B_t * m)$
2.	S_t	= Level at time t	=	$F_t + \alpha e_t$
3.	B_t	= Trend at time t	=	$B_{t-1} + \beta e_t$
4.	e_t	= Error at time t	=	$X_t - F_t$

Where,

X_t = Observation at time t

α = a level parameter subject to $0 \leq \alpha \leq 1$

β = a trend parameter subject to $0 \leq \beta \leq 1$

t = a time index.

m = the number of periods between the current observation period and a forecast period (horizon).

S_0 = 0, or alternatively an initialized value derived from other techniques.

B_0 = 0, or alternatively an initialized value derived from other techniques.

Proposed Technique

(Adjusting the Holt-Winters-Williams Model)

1.a F_t' = Adjusted Forecast at time $t=F_t + P_t$

1. F_t = Initial Forecast at time t $=(S_{t-1} + B_{t-1})*I_{t-L}$

2. S_t = Level at time t $=(F_t' + \alpha e_t)/I_{t-L}$

3. B_t = Trend at time t $= B_{t-1} + \beta e_t/I_{t-L}$

4. e_t = Error at time t $= X_t - F_t'$

5. I_t = Seasonality factor at t $= I_{t-L}+\gamma e_t/S_{t-1}+ B_{t-1})$

6. A_t = Adjustment factor at time $t= P_t - P_{t-1}$

7. P = A periodic estimate of a policy in a vector:

$(\ldots,0,0,0,a,b,c,\ldots,n,n,n,\ldots)$ where,

a, b, c, ..., n all have the same sign, and

$|a| < |b| < |c| < \ldots |n|$.

Other constraints are as with Holt-Winters-Williams.

Holt-Winters-Williams Adaptive

Holt-Winters-Williams Adaptive is identical with Holt-Winters-Williams with the exception that α is subscripted for time and selected using the following algorithm.

$$\text{Smoothed Error} = E_t = \phi e_t + (1-\phi)E_{t-1}$$

$$\text{Smoothed Absolute Error} = M_t = \phi \mid e_t \mid + (1-\phi)M_{t-1}$$

$$\alpha_t = \text{Absolute Tracking Signal} = T_t = \mid E_t/M_t \mid$$

Holt-Winters-Williams Autocorrelation Corrected

1.	F'	= Autocorr. Corrected F	$= F+(e_{t-1} * \varrho_{e,et-1})$
1.	F_{t+m}	= Forecast at time t+m	$= (S_t+ B_t * m) *I_{t+m-L}$
2.	S_t	= Level at time t	$= (F_t'+ \alpha e_t)/I_{t-L}$
3.	B_t	= Trend at time t	$= B_{t-1} + \beta e_t/I_{t-L}$
4.	e_t	= Error at time t	$= X_t - F_t'$
5.	I_t	= Seasonality factor at t	$= I_{t-L}+ e_t/(S_{t-1} + B_{t-1})$

Appendix III Correlation and Squared Correlation Matrices

Correlation for Level Shifting Data After all Preprocessing

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1718
1	1.0000																
2	-0.0808	1.0000															
3	0.1234	-0.0725	1.0000														
4	0.4188	0.0170	0.0765	1.0000													
5	-0.2019	-0.0006	-0.4351	-0.1634	1.0000												
6	-0.0595	0.0290	0.0125	-0.0906	-0.0248	1.0000											
7	0.0893	0.0909	-0.0984	0.3142	-0.1039	-0.0746	1.0000										
8	0.3494	0.0667	-0.1685	0.1299	0.0125	-0.0380	0.0219	1.0000									
9	0.0593	-0.1965	0.0682	0.1132	-0.2241	0.0330	-0.1160	-0.0839	1.0000								
10	-0.1060	0.1017	-0.0147	-0.0895	-0.3643	0.1765	0.3162	-0.1991	-0.0109	1.0000							
11	-0.0443	0.1669	-0.2588	0.1379	0.0977	0.0033	0.0563	0.0096	-0.0038	0.1409	1.0000						
12	-0.0486	-0.0138	-0.0340	-0.0476	0.0784	0.1075	-0.0719	0.0548	-0.0332	0.1899	0.2748	1.0000					
13	0.0706	0.0716	0.0165	0.0800	0.0677	0.1048	-0.1725	-0.0131	0.0024	-0.3128	-0.0210	-0.0846	1.0000				
14	-0.1019	-0.0469	-0.0076	0.1836	-0.1594	-0.0886	-0.0848	0.0099	0.3167	0.0521	-0.0129	0.0351	-0.0778	1.0000			
15	-0.0911	-0.0023	-0.0953	-0.1378	0.1054	0.1322	-0.1068	-0.0019	-0.0357	-0.0224	0.1391	0.2639	0.0776	-0.3969	1.0000		
16	0.1518	-0.3076	-0.0843	0.1151	-0.0655	0.0206	0.0760	0.3096	0.0231	0.0288	0.1192	0.3149	0.0711	0.3322	0.1594	1.0000	
17	-0.1081	-0.0789	0.0600	0.0979	-0.0608	-0.0360	-0.0884	-0.0658	0.1159	0.1211	-0.0843	-0.0379	-0.0254	0.5323	-0.0404	0.4177	1.0000
18	0.0233	-0.0101	-0.0693	0.1999	0.0206	0.1041	-0.0763	0.1569	0.1710	0.0346	0.2320	0.2609	0.0978	-0.2805	0.5021	0.0938	-0.1894
19	0.1518	-0.1186	0.0929	0.0398	0.0627	-0.0889	-0.1483	0.0195	0.1153	-0.0465	-0.0601	-0.1156	-0.0606	-0.0269	-0.0647	-0.1886	-0.1004
20	0.0866	0.0018	0.1720	-0.0106	-0.2069	-0.1830	-0.0573	0.0223	0.0812	-0.1053	-0.3575	-0.2543	-0.0567	-0.0521	-0.0663	-0.3205	-0.1323
Avg r	0.0841	0.0309	0.0142	0.1192	-0.0283	0.0520	0.0383	0.0796	0.0698	0.0445	0.0767	0.0919	0.0418	0.0563	0.0659	0.1133	0.0648

	18	19	20
18	1.0000		
19	0.0051	1.0000	
20	-0.0467	0.0771	1.0000
Avg r	0.1115	0.0272	-0.0204

Note: Although the average is shown at the bottom of the column, it is the average for all correlations for the series

Proposed Technique

(Adjusting the Holt-Winters-Williams Model)

1.a F_t' = Adjusted Forecast at time t=$F_t + P_t$

1. F_t = Initial Forecast at time t =$(S_{t-1} + B_{t-1})*I_{t-L}$

2. S_t = Level at time t =$(F_t' + \alpha e_t)/I_{t-L}$

3. B_t = Trend at time t = $B_{t-1} + \beta e_t/I_{t-L}$

4. e_t = Error at time t = $X_t - F_t'$

5. I_t = Seasonality factor at t = $I_{t-L}+\gamma e_t/S_{t-1}+ B_{t-1})$

6. A_t = Adjustment factor at time t= $P_t - P_{t-1}$

7. P = A periodic estimate of a policy in a vector:

   (...,0,0,0,a,b,c,...,n,n,n,...) where,

   a, b, c, ..., n all have the same sign, and

   $|a| < |b| < |c| < \ldots |n|$.

Other constraints are as with Holt-Winters-Williams.

Correlation Data for Trials 1 Through 12 after All Preprocessing

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1.0000																	
2	-0.1504	1.0000																
3	-0.0169	-0.2776	1.0000															
4	-0.0684	-0.0262	-0.0237	1.0000														
5	-0.2082	-0.0180	-0.1413	0.2947	1.0000													
6	-0.1401	0.1859	-0.0798	0.1115	0.0978	1.0000												
7	-0.0823	0.1992	-0.0673	0.0497	0.1931	0.1174	1.0000											
8	0.1986	-0.1010	-0.0063	-0.2656	-0.2053	0.0153	0.0693	1.0000										
9	0.1560	0.1029	0.0833	-0.2322	-0.0046	-0.1311	-0.2175	0.0507	1.0000									
10	-0.2309	0.0988	-0.0197	-0.0423	0.0936	0.2923	0.2921	-0.1221	0.0909	1.0000								
11	0.1222	0.1924	0.1252	-0.0640	0.0562	-0.0641	0.2752	0.1843	-0.0370	-0.0154	1.0000							
12	-0.0846	0.2036	0.0051	-0.0016	0.0041	0.1490	0.0351	-0.0895	-0.1277	0.0887	-0.1871	1.0000						
13	-0.0199	-0.0200	0.0918	-0.1741	0.0075	-0.1087	-0.2889	0.1004	0.2357	-0.0923	0.1159	-0.0395	1.0000					
14	-0.0337	-0.0863	-0.0065	-0.0470	0.0918	0.1113	0.0220	0.0349	0.1749	-0.0494	0.1175	0.0254	-0.1310	1.0000				
15	0.0029	-0.0276	-0.1050	0.1118	-0.0139	-0.0220	0.1740	-0.1295	-0.0680	-0.0200	-0.1863	0.0092	-0.0998	0.1362	1.0000			
16	0.2035	-0.1381	0.2123	0.1419	-0.0041	-0.0771	-0.0303	-0.1513	0.0608	-0.2321	0.0741	-0.1502	0.0875	-0.1521	0.0893	1.0000		
17	0.0684	-0.0687	0.1060	-0.0019	-0.0231	-0.0940	-0.0106	-0.0273	0.0187	-0.0428	0.0545	-0.0266	0.0397	-0.0114	0.1809	0.1968	1.0000	
18	0.1352	-0.0996	0.1454	-0.1614	-0.0322	-0.0634	-0.0590	-0.0366	0.0824	-0.0727	-0.0320	-0.1161	-0.0221	-0.0502	-0.0424	0.0735	0.2024	
19	-0.0661	0.0486	-0.0276	-0.1488	-0.0530	0.0238	-0.0640	0.0664	0.1031	0.1522	-0.1450	0.0788	0.2352	0.2034	0.1646	-0.1168	0.0156	
20	-0.0014	-0.1585	0.1102	-0.0426	0.0687	-0.0987	-0.1268	0.2092	0.0497	-0.0802	-0.1024	-0.0124	0.0819	0.1372	0.1627	0.0306	0.1297	
Avg r	0.0392	0.0430	0.0554	0.0205	0.0602	0.0613	0.0740	0.0397	0.0695	0.0544	0.0742	0.0382	0.0500	0.0744	0.0659	0.0559	0.0853	

	18	19	20
18	1.0000		
19	0.1003	1.0000	
20	0.0671	0.2496	1.0000
Avg r	0.0509	0.0910	0.0837

Note: Although the average is shown at the bottom of the column, it is the average for all correlations for the series

Appendix III Correlation Matrices

Correlation Data for Trials 1 Through 12 after All Preprocessing

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.0000																
2	-0.1504	1.0000															
3	-0.0169	-0.2776	1.0000														
4	-0.0684	-0.0262	-0.0237	1.0000													
5	-0.2082	-0.0180	-0.1413	0.2947	1.0000												
6	-0.1401	0.1859	-0.0798	0.1115	0.0978	1.0000											
7	-0.0823	0.1992	-0.0673	0.0497	0.1931	0.1174	1.0000										
8	0.1986	-0.1010	-0.0063	-0.2656	-0.2053	0.0153	0.0693	1.0000									
9	0.1560	0.1029	0.0833	-0.2322	-0.0046	-0.1311	-0.2175	0.0507	1.0000								
10	-0.2309	0.0988	-0.0197	-0.0423	0.0936	0.2923	0.2921	-0.1221	0.0909	1.0000							
11	0.1222	0.1924	0.1252	-0.0640	0.0562	-0.0641	0.2752	0.1843	-0.0370	-0.0154	1.0000						
12	-0.0846	0.2036	0.0051	-0.0016	0.0041	0.1490	0.0351	-0.0895	-0.1277	0.0887	-0.1871	1.0000					
13	-0.0199	-0.0200	0.0918	-0.1741	0.0075	-0.1087	-0.2889	0.1004	0.2357	-0.0923	0.1159	-0.0395	1.0000				
14	-0.0337	-0.0863	-0.0065	-0.0470	0.0918	0.1113	0.0220	0.0349	0.1749	-0.0494	0.1175	0.0254	-0.1310	1.0000			
15	0.0029	-0.0276	-0.1050	0.1118	-0.0139	-0.0220	0.1740	-0.1295	-0.0680	-0.0200	-0.1863	0.0092	-0.0998	0.1362	1.0000		
16	0.2035	-0.1381	0.2123	0.1419	-0.0041	-0.0771	-0.0303	-0.1513	0.0608	-0.2321	0.0741	-0.1502	0.0875	-0.1521	0.0893	1.0000	
17	0.0684	-0.0687	0.1060	-0.0019	-0.0231	-0.0940	-0.0106	-0.0273	0.0187	-0.0428	0.0545	-0.0266	0.0397	-0.0114	0.1809	0.1968	1.0000
18	0.1352	-0.0996	0.1454	-0.1614	-0.0322	-0.0634	-0.0590	-0.0366	0.0824	-0.0727	-0.0320	-0.1161	-0.0221	-0.0502	-0.0424	0.0735	0.2024
19	-0.0661	0.0486	-0.0276	-0.1488	-0.0530	0.0238	-0.0640	0.0664	0.1031	0.1522	-0.1450	0.0788	0.2352	0.2034	0.1646	-0.1168	0.0156
20	-0.0014	-0.1585	0.1102	-0.0426	0.0687	-0.0987	-0.1268	0.2092	0.0497	-0.0802	-0.1024	-0.0124	0.0819	0.1372	0.1627	0.0306	0.1297
Avg r	0.0392	0.0430	0.0554	0.0205	0.0602	0.0613	0.0740	0.0397	0.0695	0.0544	0.0742	0.0382	0.0500	0.0744	0.0659	0.0559	0.0853

	18	19	20
18	1.0000		
19	0.1003	1.0000	
20	0.0671	0.2496	1.0000
Avg r	0.0509	0.0910	0.0837

Note: Although the average is shown at the bottom of the column, it is the average for all correlations for the series

Appendix IV Tables from Scenarios 1 Through 13b

Level Shift as Planned

Scenario 1

Table: 1 - 1

Average Rank of Absolute Error

Period		Adjust	HWW	HW	Adapt	Auto	Naive	HWW*	HW*	Adapt*	Auto*	Naive*
1	Average	3.54	6.75	6.98	7.87	7.21	6.38	5.27	5.04	5.04	6.18	5.94
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Geometric Mean	3.33	6.68	6.93	7.79	7.03	6.33	5.22	4.93	4.93	6.12	5.82
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Average Rank by Series	1.65	7.63	8.43	9.50	8.03	7.00	4.15	3.80	3.80	6.55	5.80
	Rank of Average Rank	1	8	10	11	9	7	4	2	2	6	5
	Kruskal-Wallis Rank Sum	478.0	2,880.5	3,131.5	3,691.5	3,084.0	2,576.0	1,405.0	1,324.5	1,324.5	2,345.0	2,126.0
	Rank of K-W Rank Sum	1	8	10	11	9	7	4	2	2	6	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	9	8	8	10	10
5	Average	3.10	6.85	7.15	7.85	6.95	6.69	4.51	5.28	5.58	6.44	5.59
	Rank	1	8	10	11	9	7	2	3	4	6	5
	Geometric Mean	2.90	6.77	7.10	7.78	6.88	6.57	4.38	5.04	5.47	6.32	5.50
	Rank	1	8	10	11	9	7	2	3	4	6	5
	Average Rank by Series	1.78	7.95	8.50	9.03	7.53	7.13	3.35	4.70	4.60	6.53	4.93
	Rank of Average Rank	1	9	10	11	8	7	2	4	3	6	5
	Kruskal-Wallis Rank Sum	431.0	2,864.5	3,118.5	3,574.5	2,937.0	2,706.5	1,025.0	1,652.0	1,762.0	2,478.0	1,761.0
	Rank of K-W Rank Sum	1	8	10	11	9	7	2	3	5	6	4
	K-W Multi-Comparison Count*	10	9	10	10	9	10	10	10	9	10	9
10	Average	3.06	6.81	7.12	7.52	6.97	6.66	4.35	5.84	5.84	6.22	6.04
	Rank	1	8	10	11	9	7	2	3	3	6	5
	Geometric Mean	2.85	6.69	7.04	7.40	6.74	6.50	4.12	5.67	5.67	6.00	5.83
	Rank	1	8	10	11	9	7	2	3	3	6	5
	Average Rank by Series	1.85	7.25	7.9	8.2	7.7	6.95	3.575	5.625	5.625	5.925	6.025
	Rank of Average Rank	1	8	10	11	9	7	2	3	3	5	6
	Kruskal-Wallis Rank Sum	512.0	2,704.0	2,940.5	3,224.5	2,852.5	2,622.5	1,120.0	2,031.5	2,031.5	2,316.0	2,189.5
	Rank of K-W Rank Sum	1	8	10	11	9	7	2	3	3	6	5
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	10	10	10	10
15	Average	2.80	6.78	7.10	7.85	7.12	6.61	3.96	5.31	5.76	6.67	6.05
	Rank	1	8	9	11	10	6	2	3	4	7	5
	Geometric Mean	2.69	6.63	7.00	7.71	6.97	6.35	3.71	4.95	5.52	6.50	5.85
	Rank	1	8	10	11	9	6	2	3	4	7	5
	Average Rank by Series	1.68	7.30	7.65	8.43	7.93	6.65	3.28	4.93	5.33	6.73	6.13
	Rank of Average Rank	1	8	9	11	10	6	2	3	4	7	5
	Kruskal-Wallis Rank Sum	398.0	2,677.5	2,886.0	3,302.0	2,920.0	2,510.0	988.5	1,787.0	2,041.5	2,601.5	2,198.0
	Rank of K-W Rank Sum	1	8	9	11	10	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	9	9	10	9	10	10	10	10	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Appendix IV Tables from Scenarios 1 Through 13b

Level Shift as Planned

Scenario 1

Table: 1 - 1

Average Rank of Absolute Error

Period		Adjust	HWW	HW	Adapt	Auto	Naive	HWW*	HW*	Adapt*	Auto*	Naive*
1	Average	3.54	6.75	6.98	7.87	7.21	6.38	5.27	5.04	5.04	6.18	5.94
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Geometric Mean	3.33	6.68	6.93	7.79	7.03	6.33	5.22	4.93	4.93	6.12	5.82
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Average Rank by Series	1.65	7.63	8.43	9.50	8.03	7.00	4.15	3.80	3.80	6.55	5.80
	Rank of Average Rank	1	8	10	11	9	7	4	2	2	6	5
	Kruskal-Wallis Rank Sum	478.0	2,880.5	3,131.5	3,691.5	3,084.0	2,576.0	1,405.0	1,324.5	1,324.5	2,345.0	2,126.0
	Rank of K-W Rank Sum	1	8	10	11	9	7	4	2	2	6	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	9	8	8	10	10
5	Average	3.10	6.85	7.15	7.85	6.95	6.69	4.51	5.28	5.58	6.44	5.59
	Rank	1	8	10	11	9	7	2	3	4	6	5
	Geometric Mean	2.90	6.77	7.10	7.78	6.88	6.57	4.38	5.04	5.47	6.32	5.50
	Rank	1	8	10	11	9	7	2	3	4	6	5
	Average Rank by Series	1.78	7.95	8.50	9.03	7.53	7.13	3.35	4.70	4.60	6.53	4.93
	Rank of Average Rank	1	9	10	11	8	7	2	4	3	6	5
	Kruskal-Wallis Rank Sum	431.0	2,864.5	3,118.5	3,574.5	2,937.0	2,706.5	1,025.0	1,652.0	1,762.0	2,478.0	1,761.0
	Rank of K-W Rank Sum	1	8	10	11	9	7	2	3	5	6	4
	K-W Multi-Comparison Count*	10	9	10	10	9	10	10	10	9	10	9
10	Average	3.06	6.81	7.12	7.52	6.97	6.66	4.35	5.84	5.84	6.22	6.04
	Rank	1	8	10	11	9	7	2	3	3	6	5
	Geometric Mean	2.85	6.69	7.04	7.40	6.74	6.50	4.12	5.67	5.67	6.00	5.83
	Rank	1	8	10	11	9	7	2	3	3	6	5
	Average Rank by Series	1.85	7.25	7.9	8.2	7.7	6.95	3.575	5.625	5.625	5.925	6.025
	Rank of Average Rank	1	8	10	11	9	7	2	3	3	5	6
	Kruskal-Wallis Rank Sum	512.0	2,704.0	2,940.5	3,224.5	2,852.5	2,622.5	1,120.0	2,031.5	2,031.5	2,316.0	2,189.5
	Rank of K-W Rank Sum	1	8	10	11	9	7	2	3	3	6	5
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	10	10	10	10
15	Average	2.80	6.78	7.10	7.85	7.12	6.61	3.96	5.31	5.76	6.67	6.05
	Rank	1	8	9	11	10	6	2	3	4	7	5
	Geometric Mean	2.69	6.63	7.00	7.71	6.97	6.35	3.71	4.95	5.52	6.50	5.85
	Rank	1	8	10	11	9	6	2	3	4	7	5
	Average Rank by Series	1.68	7.30	7.65	8.43	7.93	6.65	3.28	4.93	5.33	6.73	6.13
	Rank of Average Rank	1	8	9	11	10	6	2	3	4	7	5
	Kruskal-Wallis Rank Sum	398.0	2,677.5	2,886.0	3,302.0	2,920.0	2,510.0	988.5	1,787.0	2,041.5	2,601.5	2,198.0
	Rank of K-W Rank Sum	1	8	9	11	10	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	9	9	10	9	10	10	10	10	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level Shift as Planned

Period: Scenario 1 Range of Percent Error

Table: 1-2

Period:		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	8.74%	14.77%	14.53%	15.90%	16.09%	16.30%	12.80%	10.61%	10.67%	13.60%	12.46%
	Rank	1	8	7	9	10	11	5	2	3	6	4
	Geometric Mean	4.02%	12.83%	12.50%	13.20%	14.32%	14.00%	11.19%	6.85%	7.08%	9.95%	9.50%
	Rank	1	8	7	9	11	10	6	2	3	5	4
	Average Rank by Series	1.95	7.50	7.00	7.80	8.68	8.75	5.75	3.70	3.40	6.20	5.28
	Rank of Average Rank	1	8	7	9	10	11	5	3	2	6	4
	Kruskal-Wallis Rank Sum	1,425.0	2,595.0	2,520.0	2,622.0	2,799.5	2,730.0	2,284.0	1,602.0	1,590.0	2,133.0	2,009.5
	Rank of K-W Rank Sum	1	8	7	9	11	10	6	3	2	5	4
	K-W Multi-Comparison Count*	10	8	9	9	9	9	10	9	9	10	10
5	Average	8.03%	26.32%	26.00%	22.51%	22.66%	28.28%	19.84%	15.00%	16.05%	18.20%	17.54%
	Rank	1	10	9	7	8	11	6	2	3	5	4
	Geometric Mean	5.07%	25.38%	25.07%	21.56%	21.52%	27.46%	18.48%	11.11%	13.16%	15.70%	15.49%
	Rank	1	10	9	8	7	11	6	2	3	5	4
	Average Rank by Series	1.55	8.13	7.88	6.15	6.40	9.65	6.10	4.13	4.33	5.90	5.80
	Rank of Average Rank	1	10	9	7	8	11	6	2	3	5	4
	Kruskal-Wallis Rank Sum	742.0	3,050.5	3,020.5	2,522.0	2,489.0	3,305.0	2,122.0	1,543.5	1,636.5	1,963.0	1,916.0
	Rank of K-W Rank Sum	1	10	9	8	7	11	6	2	3	5	4
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	10	10	9	9
10	Average	10.08%	31.37%	31.30%	33.23%	29.59%	27.69%	20.58%	22.04%	24.40%	30.83%	26.31%
	Rank	1	10	9	11	7	6	2	3	4	8	5
	Geometric Mean	6.71%	29.26%	29.53%	30.52%	27.46%	26.81%	19.84%	14.32%	17.85%	24.93%	20.84%
	Rank	1	9	10	11	8	7	4	2	3	6	5
	Average Rank by Series	1.775	7.65	7.7	7.225	6.6	7.85	4.375	4.95	5.2	6.675	6
	Rank of Average Rank	1	9	10	8	6	11	2	3	4	7	5
	Kruskal-Wallis Rank Sum	732.5	2,754.0	2,802.0	2,827.5	2,559.0	2,536.0	1,669.5	1,767.0	2,044.0	2,485.5	2,133.0
	Rank of K-W Rank Sum	1	9	10	11	8	7	2	3	4	6	5
	K-W Multi-Comparison Count*	10	8	8	8	8	8	10	10	10	8	10
15	Average	10.01%	36.16%	37.48%	40.05%	34.72%	26.46%	19.78%	26.84%	30.15%	39.74%	32.35%
	Rank	1	8	9	11	7	3	2	4	5	10	6
	Geometric Mean	6.46%	31.93%	33.74%	34.32%	30.01%	25.65%	19.06%	14.73%	18.85%	28.05%	23.89%
	Rank	1	9	10	11	8	6	4	2	3	7	5
	Average Rank by Series	1.88	7.95	7.88	6.85	6.58	7.30	5.03	4.75	4.98	6.65	6.18
	Rank of Average Rank	1	11	10	8	6	9	4	2	3	7	5
	Kruskal-Wallis Rank Sum	779.5	2,702.0	2,814.5	2,806.0	2,593.5	2,382.0	1,677.5	1,777.0	2,074.5	2,513.0	2,190.5
	Rank of K-W Rank Sum	1	9	11	10	8	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	10	9	9	9	10	10	10	10	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is signficant.

Level Shift as Planned — Table: 1-3

Period: Scenario 1 Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	3.63%	7.45%	7.49%	8.73%	7.99%	7.15%	6.16%	4.86%	4.93%	6.22%	5.71%
	Rank	1	8	9	11	10	7	5	2	3	6	4
	Geometric Mean	1.93%	6.78%	6.83%	7.91%	7.17%	6.35%	5.49%	3.49%	3.67%	5.03%	4.63%
	Rank	1	8	9	11	10	7	6	2	3	5	4
	Average Rank by Series	1.55	8.18	8.53	9.60	8.45	7.55	5.50	3.28	3.28	5.35	4.75
	Rank of Average Rank	1	8	10	11	9	7	6	2	2	5	4
	Kruskal-Wallis Rank Sum	1,260.0	2,677.5	2,694.5	3,007.0	2,810.0	2,468.0	2,201.0	1,573.5	1,589.5	2,087.0	1,942.0
	Rank of K-W Rank Sum	1	8	9	11	10	7	6	2	3	5	4
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	9	9	10	10
5	Average	4.39%	13.13%	13.62%	15.20%	13.59%	11.96%	8.06%	8.25%	9.03%	10.79%	9.29%
	Rank	1	8	10	11	9	7	2	3	4	6	5
	Geometric Mean	2.88%	12.37%	13.01%	14.56%	13.09%	11.28%	7.79%	6.20%	7.48%	9.31%	7.97%
	Rank	1	8	9	11	10	7	4	2	3	6	5
	Average Rank by Series	1.50	8.38	8.98	9.20	8.35	7.10	4.00	4.18	4.33	5.75	4.25
	Rank of Average Rank	1	9	10	11	8	7	2	3	5	6	4
	Kruskal-Wallis Rank Sum	747.0	2,801.5	2,961.5	3,239.0	2,987.0	2,482.0	1,424.0	1,680.5	1,854.5	2,225.0	1,908.0
	Rank of K-W Rank Sum	1	8	9	11	10	7	2	3	4	6	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	10	9	10	9
10	Average	5.63%	17.05%	18.26%	21.11%	18.06%	13.79%	9.26%	12.81%	14.52%	17.61%	14.80%
	Rank	1	7	10	11	9	4	2	3	5	8	6
	Geometric Mean	3.88%	14.93%	16.40%	19.16%	16.20%	12.85%	8.64%	8.34%	10.50%	13.78%	11.65%
	Rank	1	8	10	11	9	6	3	2	4	7	5
	Average Rank by Series	1.65	7.875	7.975	8.65	7.85	6.85	3.85	4.875	4.825	6.2	5.4
	Rank of Average Rank	1	9	10	11	8	7	2	4	3	6	5
	Kruskal-Wallis Rank Sum	829.0	2,562.5	2,754.5	3,067.0	2,734.0	2,262.0	1,454.0	1,830.5	2,120.5	2,545.0	2,151.0
	Rank of K-W Rank Sum	1	8	10	11	9	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	9	9	10	9	10	10	10	9	9	9
15	Average	5.79%	20.29%	22.30%	27.35%	21.92%	15.24%	9.17%	16.03%	18.65%	24.11%	18.75%
	Rank	1	7	9	11	8	3	2	4	5	10	6
	Geometric Mean	4.03%	16.60%	18.87%	23.33%	18.55%	14.21%	8.53%	8.99%	11.74%	17.16%	13.34%
	Rank	1	7	10	11	9	6	2	3	4	8	5
	Average Rank by Series	1.65	7.78	7.93	8.90	7.80	6.75	3.90	4.58	4.88	6.55	5.30
	Rank of Average Rank	1	8	10	11	9	7	2	3	4	6	5
	Kruskal-Wallis Rank Sum	807.0	2,499.5	2,722.5	3,067.0	2,713.0	2,270.0	1,405.0	1,850.5	2,174.5	2,619.0	2,182.0
	Rank of K-W Rank Sum	1	7	10	11	9	6	2	3	4	8	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	10	9	10	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level Shift as Planned

Period: Scenario 1 Root Mean Squared Error Table: 1-4

Period:		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	48.94	164.28	162.50	187.79	177.28	161.67	138.16	90.05	95.14	129.96	120.46
	Rank	1	9	8	11	10	7	6	2	3	5	4
	Average Rank by Series	1.30	8.18	8.03	9.55	8.55	8.20	5.80	3.03	3.13	5.45	4.80
	Rank of Average Rank	1	8	7	11	10	9	6	2	3	5	4
5	Geometric Mean	73.65	335.36	341.67	366.72	340.29	321.00	222.38	158.27	190.39	236.21	211.53
	Rank	1	8	10	11	9	7	5	2	3	6	4
	Average Rank by Series	1.25	8.33	8.83	8.95	8.30	7.40	4.00	4.23	4.23	5.50	5.00
	Rank of Average Rank	1	9	10	11	8	7	2	3	3	6	5
10	Geometric Mean	103.23	412.12	440.36	500.55	424.32	360.08	251.84	219.43	275.72	377.54	310.90
	Rank	1	8	10	11	9	6	3	2	4	7	5
	Average Rank by Series	1.65	7.775	8.125	8.5	7.25	7	3.8	4.775	4.875	6.5	5.75
	Rank of Average Rank	1	9	10	11	8	7	2	3	4	6	5
15	Geometric Mean	109.66	464.33	516.09	611.23	493.13	390.42	251.29	242.17	314.75	457.78	361.97
	Rank	1	8	10	11	9	6	3	2	4	7	5
	Average Rank by Series	1.50	7.68	7.88	8.70	7.40	6.75	3.95	4.78	5.08	6.80	5.50
	Rank of Average Rank	1	9	10	11	8	6	2	3	4	7	5

Level Shift as Planned

Period: Scenario 1 Geometric Root Mean Squared Error Table: 1-5

		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	26.32	86.26	94.50	124.35	98.39	73.32	59.06	44.53	47.32	68.54	64.51
	Rank	1	8	9	11	10	7	4	2	3	6	5
	Average Rank by Series	2.05	7.68	8.28	9.60	8.40	6.10	5.00	3.58	3.98	5.80	5.55
	Rank of Average Rank	1	8	9	11	10	7	4	2	3	6	5
5	Geometric Mean	47.09	164.81	196.95	234.14	193.97	140.94	87.98	95.67	115.16	141.12	109.71
	Rank	1	8	10	11	9	6	2	3	5	7	4
	Average Rank by Series	2.25	8.23	8.78	9.15	8.50	6.60	4.00	4.23	4.38	5.30	4.60
	Rank of Average Rank	1	8	10	11	9	7	2	3	4	6	5
10	Geometric Mean	55.46	213.34	246.74	299.24	257.65	192.15	97.81	137.02	168.85	200.22	175.52
	Rank	1	8	9	11	10	6	2	3	4	7	5
	Average Rank by Series	2	7.725	8.025	8.5	8.35	6.55	4	4.425	5.025	5.95	5.45
	Rank of Average Rank	1	8	9	11	10	7	2	3	4	6	5
15	Geometric Mean	67.91	246.61	295.91	405.33	326.87	237.79	112.09	147.96	194.63	276.15	216.18
	Rank	1	7	9	11	10	6	2	3	4	8	5
	Average Rank by Series	1.80	7.63	8.03	9.05	8.50	6.90	3.25	4.23	4.68	6.40	5.55
	Rank of Average Rank	1	8	9	11	10	7	2	3	4	6	5

Level Shift as Planned		Average Rank of Absolute Error		Table 1 - 6
Period	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	62.68	19	0.0000
	KRUSKAL-WALLIS	120.15	10	0.0000
5	RANK ANOVA	55.38	19	0.0000
	KRUSKAL-WALLIS	114.09	10	0.0000
10	RANK ANOVA	40.12	19	0.0032
	KRUSKAL-WALLIS	82.38	10	0.0000
15	RANK ANOVA	44.95	19	0.0007
	KRUSKAL-WALLIS	93.77	10	0.0000

Level Shift as Planned		Symmetry Adjusted MAPE		Table 1 - 7
	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	76.71	19	0.0000
	KRUSKAL-WALLIS	42.55	10	0.0000
5	RANK ANOVA	72.92	19	0.0000
	KRUSKAL-WALLIS	84.67	10	0.0000
10	RANK ANOVA	56.54	19	0.0000
	KRUSKAL-WALLIS	57.83	10	0.0000
15	RANK ANOVA	58.39	19	0.0000
	KRUSKAL-WALLIS	57.56	10	0.0000

Level Shift as Planned		Range of Percent Error		Table 1 - 8
	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	52.83	19	0.0000
	KRUSKAL-WALLIS	30.28	10	0.0008
5	RANK ANOVA	50.16	19	0.0001
	KRUSKAL-WALLIS			
	RANK ANOVA	35.47	19	0.0122
10	KRUSKAL-WALLIS	49.81	10	0.0000
	RANK ANOVA	32.88	19	0.0248
15	KRUSKAL-WALLIS	46.49	10	0.0000

Level Shift as Planned		Geometric Root Mean Square Error		Table 1 - 9
	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	56.20	19	0.0000
5	RANK ANOVA	58.23	19	0.0000
10	RANK ANOVA	45.14	19	0.0007
15	RANK ANOVA	56.24	19	0.0000

	Level Shift as Planned	Log Mean Squared Error Ratio		Table 1 - 10
Period	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	56.20	19	0.0000
	KRUSKAL-WALLIS	82.89	10	0.0000
5	RANK ANOVA	58.23	19	0.0000
	KRUSKAL-WALLIS	70.33	10	0.0000
10	RANK ANOVA	45.14	19	0.0007
	KRUSKAL-WALLIS	55.80	10	0.0000
15	RANK ANOVA	56.24	19	0.0000
	KRUSKAL-WALLIS	60.82	10	0.0000

Level Shift as Planned		Mean Absolute Percent Error		Table 1 - 11
	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	72.64	19	0.0000
	KRUSKAL-WALLIS	40.66	10	0.0000
5	RANK ANOVA	68.09	19	0.0000
	KRUSKAL-WALLIS	72.90	10	0.0000
10	RANK ANOVA	47.51	19	0.0003
	KRUSKAL-WALLIS	51.57	10	0.0000
15	RANK ANOVA	48.88	19	0.0002
	KRUSKAL-WALLIS	52.48	10	0.0000

Level Shift as Planned		Median Absolute Percent Error		Table 1 - 12
	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	54.55	19	0.0000
	KRUSKAL-WALLIS	34.52	10	0.0002
5	RANK ANOVA	62.78	19	0.0000
	KRUSKAL-WALLIS	59.36	10	0.0000
10	RANK ANOVA	34.47	19	0.0162
	KRUSKAL-WALLIS	43.64	10	0.0000
15	RANK ANOVA	43.85	19	0.0010
	KRUSKAL-WALLIS	46.21	10	0.0000

Level Shift as Planned		Root Mean Square Error		Table 1 - 13
	Scenario 1	Chi Square	DF	p Value
1	RANK ANOVA	77.22	19	0.0000
5	RANK ANOVA	66.50	19	0.0000
10	RANK ANOVA	45.59	19	0.0006
15	RANK ANOVA	46.12	19	0.0005

Level and Trend Shift

Table: 2 - 1

Period: Scenario 2

Average Rank of Absolute Error

Period		Adjust	HWW	HW	Adapt	Auto	Naive	HWW*	HW*	Adapt*	Auto*	Naive*
1	Average	3.55	6.72	6.98	7.66	7.24	6.40	5.21	5.03	5.03	6.16	5.98
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Geometric Mean	3.34	6.65	6.94	7.79	7.06	6.35	5.16	4.94	4.94	6.10	5.88
	Rank	1	8	9	11	10	7	4	3	2	6	5
	Average Rank by Series	1.70	7.60	8.40	9.45	8.15	7.00	4.15	3.83	3.83	6.43	5.78
	Rank of Average Rank	1	8	10	11	9	7	4	3	2	6	5
	Kruskal-Wallis Rank Sum	489.0	2855.5	3136.0	3685.0	3108.0	2581.5	1393.0	1299.0	1299.0	2333.5	2175.0
	Rank of K-W Rank Sum	1	8	10	11	9	7	4	2	2	6	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	9	9	10	10
5	Average	3.13	6.92	7.15	7.89	6.86	6.73	4.48	5.33	5.61	6.42	5.49
	Rank	1	8	10	11	8	7	2	3	5	6	4
	Geometric Mean	2.95	6.83	7.10	7.83	6.78	6.62	4.34	5.10	5.50	6.31	5.37
	Rank	1	8	10	11	8	7	2	3	5	6	4
	Average Rank by Series	1.58	8.18	8.60	9.00	7.50	7.18	3.20	5.03	4.83	6.33	4.60
	Rank of Average Rank	1	9	10	11	8	7	2	5	3	6	3
	Kruskal-Wallis Rank Sum	414.0	2916.5	3135.5	3595.0	2858.5	2752.5	998.0	1686.5	1791.5	2464.0	1698.0
	Rank of K-W Rank Sum	1	8	10	11	8	7	2	3	5	6	4
	K-W Multi-Comparison Count*	10	9	10	10	9	10	10	9	10	10	9
10	Average	3.03	6.79	7.12	7.56	6.96	6.79	4.32	5.84	5.84	6.24	5.99
	Rank	1	8	10	11	9	7	2	3	3	6	5
	Geometric Mean	2.84	6.66	7.05	7.44	6.73	6.64	4.09	5.67	5.67	6.02	5.78
	Rank	1	8	10	11	9	7	2	4	3	6	5
	Average Rank by Series	1.73	7.28	7.85	8.28	7.73	7.15	3.60	5.45	5.45	6.05	5.90
	Rank of Average Rank	1	8	10	11	9	7	2	3	3	6	5
	Kruskal-Wallis Rank Sum	489.5	2697.5	2944.5	3244.5	2851.0	2695.5	1104.0	2009.5	2009.5	2331.0	2152.0
	Rank of K-W Rank Sum	1	8	10	11	9	7	2	3	3	6	5
	K-W Multi-Comparison Count*	10	9	10	10	10	9	10	10	10	10	10
15	Average	2.87	6.72	7.12	7.87	7.10	6.65	3.99	5.28	5.75	6.68	5.98
	Rank	1	8	9	11	9	6	2	3	4	7	5
	Geometric Mean	2.76	6.57	7.02	7.72	6.95	6.40	3.74	4.90	5.51	6.51	5.78
	Rank	1	8	10	11	9	6	2	3	4	7	5
	Average Rank by Series	1.73	7.10	7.90	8.53	7.78	6.73	3.33	4.80	5.35	6.83	5.95
	Rank of Average Rank	1	8	9	11	9	6	2	3	4	7	5
	Kruskal-Wallis Rank Sum	414.0	2631.5	2901.0	3324.0	2913.5	2541.5	995.5	1770.5	2041.0	2608.0	2169.5
	Rank of K-W Rank Sum	1	8	9	11	10	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	9	9	10	9	9	10	10	10	8	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level and Trend Shift

Period: Scenario 2 Range of Percent Error Table: 2-2

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	8.77%	14.79%	14.56%	15.96%	16.17%	16.29%	12.69%	10.59%	10.64%	13.61%	12.50%
	Rank	1	8	7	9	10	11	5	2	3	6	4
	Geometric Mean	4.07%	12.89%	12.57%	13.29%	14.41%	14.02%	11.09%	6.87%	7.07%	9.94%	9.50%
	Rank	1	8	7	9	11	10	6	2	3	5	4
	Average Rank by Series	1.95	7.60	6.90	7.80	8.68	8.75	5.75	3.70	3.40	6.20	5.28
	Rank of Average Rank	1	8	7	9	10	11	5	3	2	6	4
	Kruskal-Wallis Rank Sum	1428.0	2603.0	2519.0	2628.0	2803.5	2733.0	2267.0	1603.0	1590.0	2127.0	2008.5
	Rank of K-W Rank Sum	1	8	7	9	11	10	6	3	2	5	4
	K-W Multi-Comparison Count*	10	9	10	9	9	9	10	9	9	10	10
5	Average	8.12%	26.23%	25.89%	22.64%	22.78%	28.24%	19.52%	14.77%	15.80%	18.10%	17.43%
	Rank	1	10	9	7	8	11	6	2	3	5	4
	Geometric Mean	5.15%	25.37%	25.04%	21.74%	21.73%	27.45%	18.10%	11.12%	13.06%	15.62%	15.40%
	Rank	1	10	9	8	7	11	6	2	3	5	4
	Average Rank by Series	1.55	8.18	7.88	6.18	6.45	9.65	6.15	4.08	4.23	5.83	5.85
	Rank of Average Rank	1	10	9	7	8	11	6	2	3	5	5
	Kruskal-Wallis Rank Sum	751.0	3062.5	3021.5	2551.5	2510.0	3318.0	2090.0	1514.5	1604.5	1967.5	1919.0
	Rank of K-W Rank Sum	1	10	9	8	7	11	6	2	3	5	4
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	10	10	9	9
10	Average	9.91%	31.20%	31.01%	33.02%	29.28%	27.69%	20.19%	21.83%	24.12%	30.54%	25.96%
	Rank	1	9	9	11	7	6	2	3	4	8	5
	Geometric Mean	6.55%	29.22%	29.32%	30.38%	27.22%	26.79%	19.40%	14.10%	17.77%	24.76%	20.63%
	Rank	1	9	10	11	8	7	4	2	3	6	5
	Average Rank by Series	1.78	7.85	7.65	7.18	6.60	7.55	4.33	5.15	5.25	6.78	5.90
	Rank of Average Rank	1	10	9	8	6	9	2	3	4	7	5
	Kruskal-Wallis Rank Sum	721.5	2763.0	2796.0	2831.5	2578.0	2562.0	1649.5	1773.0	2036.0	2483.5	2116.0
	Rank of K-W Rank Sum	1	9	10	11	8	7	2	3	4	6	5
	K-W Multi-Comparison Count*	10	8	8	8	9	8	10	10	9	9	9
15	Average	9.92%	36.13%	37.22%	39.81%	34.56%	26.34%	19.38%	26.57%	29.80%	39.40%	31.88%
	Rank	1	8	9	11	7	3	2	4	5	10	6
	Geometric Mean	6.52%	32.17%	33.68%	34.28%	30.14%	25.54%	18.66%	14.95%	19.15%	27.96%	23.66%
	Rank	1	9	10	11	8	6	3	2	4	7	5
	Average Rank by Series	1.65	8.10	7.85	6.85	6.83	7.20	5.05	4.75	5.05	6.70	5.98
	Rank of Average Rank	1	11	10	8	7	9	4	2	3	6	5
	Kruskal-Wallis Rank Sum	767.0	2756.0	2845.0	2811.0	2607.5	2384.0	1626.0	1766.0	2076.0	2509.0	2162.5
	Rank of K-W Rank Sum	1	9	11	10	8	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	9	9	8	10	10	10	10	10	10	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level and Trend Shift | Table: 2-3

Period: | Scenario 2 | Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	3.63%	7.46%	7.49%	8.73%	7.99%	7.18%	6.13%	4.86%	4.92%	6.21%	5.70%
	Rank	1	8	9	11	10	7	5	2	3	6	4
	Geometric Mean	1.93%	6.80%	6.84%	7.91%	7.18%	6.38%	5.46%	3.50%	3.66%	5.02%	4.61%
	Rank	1	8	9	11	10	7	6	2	3	5	4
	Average Rank by Series	1.45	8.13	8.53	9.60	8.45	7.60	5.55	3.33	3.33	5.30	4.75
	Rank of Average Rank	1	8	10	11	9	7	6	3	2	5	4
	Kruskal-Wallis Rank Sum	1252.0	2677.5	2696.5	3005.0	2814.0	2483.0	2192.0	1574.5	1592.5	2089.0	1934.0
	Rank of K-W Rank Sum	1	8	9	11	10	7	6	2	3	5	4
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	9	9	10	10
5	Average	4.41%	13.15%	13.63%	15.20%	13.59%	12.08%	7.98%	8.26%	9.00%	10.73%	9.23%
	Rank	1	8	10	11	9	7	2	3	4	6	5
	Geometric Mean	2.95%	12.41%	13.03%	14.56%	13.09%	11.41%	7.72%	6.28%	7.49%	9.30%	7.94%
	Rank	1	8	9	11	10	7	4	2	3	6	5
	Average Rank by Series	1.50	8.38	8.98	9.25	8.35	7.15	4.00	4.23	4.38	5.55	4.25
	Rank of Average Rank	1	9	10	11	8	7	2	3	5	6	4
	Kruskal-Wallis Rank Sum	747.0	2808.5	2969.5	3253.0	2981.0	2523.0	1405.0	1680.5	1851.5	2198.0	1893.0
	Rank of K-W Rank Sum	1	8	9	11	10	7	2	3	4	6	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	10	9	10	9
10	Average	5.69%	17.04%	18.22%	21.03%	17.99%	14.09%	9.18%	12.76%	14.42%	17.48%	14.66%
	Rank	1	7	10	11	9	4	2	3	5	8	6
	Geometric Mean	3.95%	14.97%	16.42%	19.11%	16.16%	13.14%	8.55%	8.41%	10.52%	13.74%	11.59%
	Rank	1	8	10	11	9	6	3	2	4	7	5
	Average Rank by Series	1.65	7.88	8.03	8.90	7.95	6.95	3.80	4.88	4.68	6.00	5.30
	Rank of Average Rank	1	9	10	11	9	7	2	4	3	6	5
	Kruskal-Wallis Rank Sum	823.0	2571.5	2760.5	3063.0	2725.0	2311.0	1437.0	1835.5	2113.5	2531.0	2139.0
	Rank of K-W Rank Sum	1	8	10	11	9	6	2	3	4	7	5
	K-W Multi-Comparison Count*	10	9	9	10	9	10	10	10	9	9	9
15	Average	5.82%	20.21%	22.18%	27.18%	21.73%	15.67%	9.07%	15.92%	18.45%	23.88%	18.48%
	Rank	1	7	9	11	8	3	2	4	5	10	6
	Geometric Mean	4.02%	16.65%	18.89%	23.28%	18.46%	14.61%	8.42%	9.06%	11.81%	17.17%	13.18%
	Rank	1	7	10	11	9	6	2	3	4	8	5
	Average Rank by Series	1.70	7.83	7.[illegible]	8.90	7.90	6.75	3.90	4.53	4.88	6.60	5.10
	Rank of Average Rank	1	9	10	11	9	7	2	3	4	6	5
	Kruskal-Wallis Rank Sum	788.0	2500.5	2727.5	3071.0	2702.0	2320.0	1383.0	1852.5	2179.5	2615.0	2171.0
	Rank of K-W Rank Sum	1	7	10	11	9	6	2	3	4	8	4
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	10	9	10	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level and Trend Shift

Period:		Scenario 2 Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
			Root Mean Square Error					Table:	2-4			
1	Geometric Mean	49.36	164.82	162.99	188.33	177.83	162.41	137.49	90.46	95.00	129.90	120.45
	Rank	1	9	8	11	10	7	6	2	3	5	4
	Average Rank by Series	1.30	8.18	8.08	9.65	8.55	8.20	5.55	3.08	3.18	5.45	4.80
	Rank of Average Rank	1	8	7	11	10	9	6	2	3	5	4
5	Geometric Mean	75.32	336.92	342.76	367.95	341.54	324.11	220.39	160.48	190.89	236.84	211.37
	Rank	1	8	10	11	9	7	5	2	3	6	4
	Average Rank by Series	1.25	8.38	8.88	8.95	8.45	7.45	4.05	4.13	4.18	5.40	4.90
	Rank of Average Rank	1	9	10	11	9	7	2	3	4	6	5
10	Geometric Mean	105.02	415.04	442.00	501.58	424.95	366.82	249.00	222.33	277.65	378.57	310.74
	Rank	1	8	10	11	9	6	3	2	4	7	5
	Average Rank by Series	1.50	7.93	8.23	8.50	7.35	7.10	3.75	4.78	4.83	6.45	5.60
	Rank of Average Rank	1	9	10	11	8	7	2	3	4	6	5
15	Geometric Mean	110.40	468.99	518.61	612.92	493.97	400.37	248.39	246.13	319.55	460.77	360.74
	Rank	1	8	10	11	9	6	3	2	4	7	5
	Average Rank by Series	1.40	7.73	7.88	8.70	7.45	6.75	3.95	4.78	5.13	6.80	5.45
	Rank of Average Rank	1	9	10	11	8	6	2	3	4	7	5

Level and Trend Shift

Period:		Scenario 2 Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
			Geometric Root Mean Square Error					Table:	2-5			
1	Geometric Mean	26.70	86.96	95.85	124.48	97.15	74.67	59.42	43.69	47.76	68.56	62.84
	Rank	1	8	9	11	10	7	4	2	3	6	5
	Average Rank by Series	2.05	7.63	8.33	9.55	8.35	6.30	5.00	3.58	3.93	5.80	5.50
	Rank of Average Rank	1	8	9	11	10	7	4	2	3	6	5
5	Geometric Mean	49.47	167.80	199.07	237.87	193.68	0.00	89.28	98.63	115.19	139.46	109.97
	Rank	1	8	10	11	9	1	3	4	6	7	5
	Average Rank by Series	2.15	8.33	8.88	9.30	8.30	6.35	3.90	4.48	4.43	5.30	4.60
	Rank of Average Rank	1	8	10	11	8	7	2	4	3	6	5
10	Geometric Mean	62.27	214.08	253.44	292.72	258.72	203.82	106.73	138.11	172.19	193.15	175.53
	Rank	1	8	9	11	10	7	2	3	4	6	5
	Average Rank by Series	2.1	7.775	8.325	8.35	8.3	7.1	3.9	4.275	5.075	5.45	5.35
	Rank of Average Rank	1	8	9	10	9	7	2	3	4	5	5
15	Geometric Mean	66.81	232.26	291.35	415.49	323.95	254.39	111.97	142.19	186.30	271.24	210.09
	Rank	1	6	9	11	10	7	2	3	4	8	5
	Average Rank by Series	2.00	7.18	8.23	9.15	8.35	7.00	3.75	3.98	4.73	6.15	5.50
	Rank of Average Rank	1	8	9	11	10	7	2	3	4	6	5

Level and Trend Shift Average Rank of Absolute Error Table 2 - 6

Period Scenario 2		Chi Square	DF	p Value	Period
1	RANK ANOVA	61.72	19	0.0000	
	KRUSKAL-WALLIS	120.73	10	0.0000	
5	RANK ANOVA	58.70	19	0.0000	
	KRUSKAL-WALLIS	116.73	10	0.0000	
10	RANK ANOVA	42.00	19	0.0018	
	KRUSKAL-WALLIS	85.29	10	0.0000	
15	RANK ANOVA	45.10	19	0.0007	
	KRUSKAL-WALLIS	93.58	10	0.0000	

Level and Trend Shift Log Mean Square Error Ratio Table 2 - 10

Scenario 2		Chi Square	DF	p Value
1	RANK ANOVA	56.00	19	0.0000
	KRUSKAL-WALLIS	83.23	10	0.0000
5	RANK ANOVA	58.36	19	0.0000
	KRUSKAL-WALLIS	66.70	10	0.0000
10	RANK ANOVA	47.08	19	0.0003
	KRUSKAL-WALLIS	58.02	10	0.0000
15	RANK ANOVA	52.36	19	0.0001
	KRUSKAL-WALLIS	60.86	10	0.0000

Level and Trend Shift Symmetry Adjusted MAPE Table 2 -7

	Scenario 2	Chi Square	DF	p Value
1	RANK ANOVA	77.66	19	0.0000
	KRUSKAL-WALLIS	42.79	10	0.0000
5	RANK ANOVA	73.65	19	0.0000
	KRUSKAL-WALLIS	86.30	10	0.0000
10	RANK ANOVA	58.26	19	0.0000
	KRUSKAL-WALLIS	58.33	10	0.0000
15	RANK ANOVA	62.37	19	0.0000
	KRUSKAL-WALLIS	58.16	10	0.0000

Level and Trend Shift Mean Absolute Percent Error Table 2 - 11

	Scenario 2	Chi Square	DF	p Value
1	RANK ANOVA	72.98	19	0.0000
	KRUSKAL-WALLIS	40.98	10	0.0000
5	RANK ANOVA	68.32	19	0.0000
	KRUSKAL-WALLIS	74.21	10	0.0000
10	RANK ANOVA	50.45	19	0.0001
	KRUSKAL-WALLIS	52.02	10	0.0000
15	RANK ANOVA	49.54	19	0.0002
	KRUSKAL-WALLIS	53.66	10	0.0000

Level and Trend Shift Range of Percent Error Table 2 -8

	Scenario 2	Chi Square	DF	p Value
1	RANK ANOVA	52.95	19	0.0000
	KRUSKAL-WALLIS	30.43	10	0.0007
5	RANK ANOVA	51.01	19	0.0001
	KRUSKAL-WALLIS	73.51	10	0.0000
	RANK ANOVA	34.70	19	0.0152
10	KRUSKAL-WALLIS	50.96	10	0.0000
	RANK ANOVA	35.34	19	0.0127
15	KRUSKAL-WALLIS	49.13	10	0.0000

Level and Trend Shift Median Absolute Percent Error Table 2 - 12

	Scenario 2	Chi Square	DF	p Value
1	RANK ANOVA	53.99	19	0.0000
	KRUSKAL-WALLIS	34.45	10	0.0002
5	RANK ANOVA	64.07	19	0.0000
	KRUSKAL-WALLIS	60.34	10	0.0000
10	RANK ANOVA	36.00	19	0.0106
	KRUSKAL-WALLIS	43.72	10	0.0000
15	RANK ANOVA	44.70	19	0.0008
	KRUSKAL-WALLIS	46.76	10	0.0000

Level and Trend Shift Geometric Root Mean Square Error Table 2 - 9

	Scenario 2	Chi Square	DF	p Value
1	RANK ANOVA	56.00	19	0.0000
5	RANK ANOVA	59.24	19	0.0000
10	RANK ANOVA	47.08	19	0.0003
15	RANK ANOVA	52.36	19	0.0001

Level and Trend Shift Root Mean Square Error Table 2 - 13

	Scenario 2	Chi Square	DF	p Value
1	RANK ANOVA	77.74	19	0.0000
5	RANK ANOVA	68.63	19	0.0000
10	RANK ANOVA	48.89	19	0.0002
15	RANK ANOVA	47.35	19	0.0003

Period: Scenario 3 25% Level Shift

Average Rank of Absolute Error

Table: 3-1

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	7.78	5.04	4.95	4.96	4.83	5.40	9.01	6.00	6.00	6.17	5.89
	Rank	10	4	2	3	1	5	11	7	7	9	6
	Geometric Mean	7.61	4.90	4.83	4.81	4.67	5.21	8.91	5.95	5.95	6.07	5.79
	Rank	10	4	3	2	1	5	11	7	7	9	6
	Average Rank by Series	8.50	4.33	3.93	3.60	3.73	5.35	10.90	6.63	6.63	6.28	5.85
	Rank of Average Rank	10	4	3	1	2	5	11	8	8	7	6
	Kruskal-Wallis Rank Sum	3,334.5	1,513.0	1,417.5	1,392.5	1,321.0	1,825.0	3,976.0	2,266.0	2,266.0	2,391.5	2,245.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	7	7	9	6
	K-W Multi-Comparison Count*	10	10	9	8	9	10	10	8	8	10	8
5	Average	7.18	5.18	5.23	5.55	4.90	5.66	8.24	5.73	5.88	6.39	6.09
	Rank	10	2	3	4	1	5	11	6	7	9	8
	Geometric Mean	6.90	4.90	5.07	5.34	4.69	5.42	8.05	5.52	5.75	6.22	5.94
	Rank	10	2	3	4	1	5	11	6	7	9	8
	Average Rank by Series	7.20	4.95	4.80	5.13	3.70	5.90	9.48	5.90	5.85	6.88	6.23
	Rank of Average Rank	10	3	2	4	1	6	11	6	5	9	8
	Kruskal-Wallis Rank Sum	2,887.5	1,746.0	1,705.0	1,867.0	1,426.5	2,056.5	3,510.0	2,115.5	2,194.5	2,562.5	2,340.5
	Rank of K-W Rank Sum	10	3	2	4	1	5	11	6	7	9	8
	K-W Multi-Comparison Count*	10	9	9	10	10	9	10	8	9	10	10
10	Average	6.85	5.37	5.48	5.29	5.17	5.92	7.82	5.96	5.96	6.15	6.25
	Rank	10	3	4	2	1	5	11	6	6	8	9
	Geometric Mean	6.42	5.04	5.23	4.89	4.98	5.57	7.49	5.83	5.83	6.02	6.14
	Rank	10	3	4	1	2	5	11	6	6	8	9
	Average Rank by Series	6.7	5.425	5.5	5	4.45	5.925	8.45	6.1	6.1	6.05	6.675
	Rank of Average Rank	10	3	4	2	1	5	11	7	7	6	9
	Kruskal-Wallis Rank Sum	2,656.5	1,859.0	1,885.0	1,739.5	1,700.0	2,207.5	3,183.0	2,271.0	2,271.0	2,349.0	2,436.5
	Rank of K-W Rank Sum	10	3	4	2	1	5	11	6	6	8	9
	K-W Multi-Comparison Count*	10	9	9	9	9	8	10	8	8	9	10
15	Average	6.58	5.41	5.45	5.59	5.21	6.29	7.39	5.78	5.93	6.29	6.09
	Rank	10	2	3	4	1	9	11	5	6	8	7
	Geometric Mean	5.95	5.00	5.09	5.16	5.02	5.79	6.79	5.55	5.81	6.09	5.87
	Rank	9	1	3	4	2	6	11	5	7	10	8
	Average Rank by Series	5.975	5.5	5.525	5.425	4.8	6.375	7.575	6.075	6.175	6.375	6.2
	Rank of Average Rank	5	3	4	2	1	9	11	6	7	9	8
	Kruskal-Wallis Rank Sum	2,438.0	1,926.0	1,916.0	1,944.0	1,755.0	2,412.0	2,905.5	2,168.5	2,284.0	2,434.5	2,316.5
	Rank of K-W Rank Sum	10	3	2	4	1	8	11	5	6	9	7
	K-W Multi-Comparison Count*	8	8	8	8	10	8	10	10	9	8	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Period: Scenario 3 25% Level Shift Table: 3-2

Range of Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	16.25%	12.51%	12.43%	10.78%	11.36%	13.68%	22.50%	14.27%	14.31%	12.39%	13.44%
	Rank	10	5	4	1	2	7	11	8	9	3	6
	Geometric Mean	13.19%	7.27%	7.05%	7.10%	7.33%	8.08%	20.77%	11.13%	11.19%	10.28%	10.63%
	Rank	10	3	1	2	4	5	11	8	9	6	7
	Average Rank by Series	7.80	4.83	4.43	3.28	4.45	5.40	10.20	6.78	6.63	5.73	6.50
	Rank of Average Rank	10	4	2	1	3	5	11	9	8	6	7
	Kruskal-Wallis Rank Sum	2,161.0	1,419.0	1,416.5	1,372.0	1,437.0	1,525.0	3,150.0	1,790.0	1,802.5	1,714.5	2,335.0
	Rank of K-W Rank Sum	9	3	2	1	4	5	11	7	8	6	10
	K-W Multi-Comparison Count*	10	7	7	7	7	10	10	8	9	9	10
5	Average	18.62%	15.06%	14.77%	12.84%	12.72%	17.27%	38.83%	18.89%	18.68%	17.38%	18.28%
	Rank	8	4	3	2	1	5	11	10	9	6	7
	Geometric Mean	17.45%	11.95%	11.65%	10.36%	10.66%	13.75%	37.28%	17.19%	16.95%	15.83%	17.30%
	Rank	10	4	3	1	2	5	11	8	7	6	9
	Average Rank by Series	7.45	4.98	4.33	3.45	3.85	6.50	10.55	6.33	6.08	5.90	6.60
	Rank of Average Rank	10	4	3	1	2	8	11	7	6	5	9
	Kruskal-Wallis Rank Sum	2,397.0	1,625.0	1,565.5	1,381.0	1,463.0	1,879.0	3,986.0	2,368.0	2,313.5	2,139.0	2,512.0
	Rank of K-W Rank Sum	9	4	3	1	2	5	11	8	7	6	10
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	9	9	10	10
10	Average	24.41%	17.59%	17.28%	16.60%	15.47%	17.83%	37.12%	19.30%	18.46%	19.64%	19.26%
	Rank	10	4	3	2	1	5	11	8	6	9	7
	Geometric Mean	21.56%	15.02%	15.07%	14.19%	13.59%	14.82%	35.64%	17.52%	16.61%	18.05%	17.51%
	Rank	10	4	5	2	1	3	11	8	6	9	7
	Average Rank by Series	7.175	5.225	5.2	4.35	4	5.55	9.875	6.275	5.95	6.45	5.95
	Rank of Average Rank	10	4	3	2	1	5	11	8	6	9	6
	Kruskal-Wallis Rank Sum	2,744.5	1,870.5	1,891.5	1,775.0	1,655.5	1,820.0	3,812.5	2,136.0	2,052.0	2,250.0	2,190.0
	Rank of K-W Rank Sum	10	4	5	2	1	3	11	7	6	9	8
	K-W Multi-Comparison Count*	10	8	8	9	10	7	10	9	10	9	8
15	Average	28.68%	18.82%	18.58%	19.74%	17.62%	16.05%	41.47%	18.05%	17.28%	19.84%	17.78%
	Rank	10	7	6	8	3	1	11	5	2	9	4
	Geometric Mean	23.12%	15.39%	15.89%	16.41%	14.97%	13.60%	39.35%	16.41%	15.35%	18.20%	15.51%
	Rank	10	4	6	8	2	1	11	7	3	9	5
	Average Rank by Series	6.925	5.575	5.3	5.55	5.35	5.2	9.725	5.775	5	6.65	4.95
	Rank of Average Rank	10	7	4	6	5	3	11	8	2	9	1
	Kruskal-Wallis Rank Sum	2,805.5	1,975.5	2,062.5	2,233.0	1,979.0	1,705.0	3,907.5	2,132.0	2,017.0	2,376.0	1,940.0
	Rank of K-W Rank Sum	10	3	6	8	4	1	11	7	5	9	2
	K-W Multi-Comparison Count*	10	7	8	10	7	10	10	9	6	10	7

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Period:	Scenario 3 25% Level Shift							Table: 3-3				
	Mean Absolute Percent Error											
		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	9.68%	5.59%	5.56%	5.09%	5.09%	6.04%	13.57%	6.74%	6.76%	6.49%	6.20%
	Rank	10	4	3	2	1	5	11	8	9	7	6
	Geometric Mean	8.27%	3.91%	3.82%	3.78%	3.79%	4.24%	12.58%	5.59%	5.63%	5.53%	5.38%
	Rank	10	4	3	1	2	5	11	8	9	7	6
	Average Rank by Series	9.00	4.03	3.78	2.90	3.15	5.45	11.00	7.13	6.98	6.45	6.15
	Rank of Average Rank	10	4	3	1	2	5	11	9	8	7	6
	Kruskal-Wallis Rank Sum	1,795.0	921.0	924.5	736.0	783.0	1,118.0	2,840.0	1,336.0	1,327.5	1,191.0	2,218.0
	Rank of K-W Rank Sum	9	3	4	1	2	5	11	8	7	6	10
	K-W Multi-Comparison Count*	10	9	9	9	9	9	10	9	9	9	10
5	Average	12.37%	8.29%	8.35%	7.88%	7.52%	8.95%	18.74%	8.93%	9.24%	9.55%	9.35%
	Rank	10	3	4	2	1	6	11	5	7	9	8
	Geometric Mean	11.55%	6.87%	7.01%	6.93%	6.63%	7.62%	18.07%	8.30%	8.72%	9.07%	8.96%
	Rank	10	2	4	3	1	5	11	6	7	9	8
	Average Rank by Series	7.45	4.58	4.63	4.35	3.80	5.70	10.30	6.08	5.93	6.70	6.50
	Rank of Average Rank	10	3	4	2	1	5	11	7	6	9	8
	Kruskal-Wallis Rank Sum	2,197.0	1,276.0	1,285.5	1,174.0	1,099.0	1,447.0	3,424.0	1,521.0	1,602.5	1,724.0	2,317.0
	Rank of K-W Rank Sum	9	3	4	2	1	5	11	6	7	8	10
	K-W Multi-Comparison Count*	10	9	9	9	9	9	10	9	10	10	10
10	Average	15.50%	10.65%	10.92%	10.41%	9.91%	11.15%	21.28%	11.14%	11.85%	12.20%	12.06%
	Rank	10	3	4	2	1	6	11	5	7	9	8
	Geometric Mean	14.05%	8.56%	9.17%	9.05%	8.62%	9.49%	20.39%	9.62%	10.61%	11.36%	10.92%
	Rank	10	1	4	3	2	5	11	6	7	9	8
	Average Rank by Series	7	5.225	5.125	4.75	4.35	6.1	9.6	5.625	5.675	6.15	6.4
	Rank of Average Rank	10	4	3	2	1	7	11	5	6	8	9
	Kruskal-Wallis Rank Sum	2,517.0	1,605.0	1,674.5	1,606.0	1,492.0	1,702.0	3,387.0	1,834.0	1,985.5	2,055.0	2,253.0
	Rank of K-W Rank Sum	10	2	4	3	1	5	11	6	7	8	9
	K-W Multi-Comparison Count*	10	8	7	8	10	9	10	10	9	9	10
15	Average	17.73%	11.86%	12.34%	12.55%	10.97%	12.68%	22.35%	12.49%	13.40%	14.40%	13.15%
	Rank	10	2	3	5	1	6	11	4	8	9	7
	Geometric Mean	14.68%	9.12%	10.10%	10.28%	9.21%	10.78%	19.67%	10.71%	11.91%	12.67%	11.38%
	Rank	10	1	3	4	2	6	11	5	8	9	7
	Average Rank by Series	6.3	5.125	5.025	5.25	4.45	6.1	8.95	5.975	6.025	6.95	5.85
	Rank of Average Rank	9	3	2	4	1	8	11	6	7	10	5
	Kruskal-Wallis Rank Sum	2,461.0	1,791.0	1,897.5	1,891.0	1,648.0	1,850.0	3,088.0	1,946.0	2,140.5	2,334.0	2,166.0
	Rank of K-W Rank Sum	10	2	5	4	1	3	11	6	7	9	8
	K-W Multi-Comparison Count*	10	9	7	8	10	7	10	9	9	10	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Period: Scenario 3 25% Level Shift Table: 3-4

Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	159.60	81.91	79.70	80.89	81.95	89.19	248.13	118.45	119.48	116.70	117.17
	Rank	10	3	1	2	4	5	11	8	9	6	7
	Average Rank by Series	8.10	4.28	4.03	3.20	3.60	5.55	11.00	6.83	6.83	5.75	6.85
	Rank of Average Rank	10	4	3	1	2	5	11	7	7	6	9
5	Geometric Mean	235.90111	146.76204	147.168	144.86492	139.800891	164.888	395.38813	186.88931	191.135	191.74174	197.194
	Rank	10	3	4	2	1	5	11	6	7	8	9
	Average Rank by Series	7.40	4.78	4.53	4.35	3.90	5.75	10.20	6.33	5.88	6.40	6.50
	Rank of Average Rank	10	4	3	2	1	5	11	7	6	8	9
10	Geometric Mean	302.23	193.97	203.68	202.62	191.94	210.73	453.54	225.60	238.57	252.71	242.43
	Rank	10	2	4	3	1	5	11	6	7	9	8
	Average Rank by Series	6.65	4.825	4.725	4.95	4.4	5.85	10	5.975	5.925	6.35	6.35
	Rank of Average Rank	10	3	2	4	1	5	11	7	6	8	8
15	Geometric Mean	330.67	209.60	228.35	233.92	208.84	236.99	468.35	245.51	264.30	287.28	257.27
	Rank	10	2	3	4	1	5	11	6	8	9	7
	Average Rank by Series	6.45	5.125	5.225	5.05	4.35	6.15	9.3	5.775	5.825	6.85	5.9
	Rank of Average Rank	9	3	4	2	1	8	11	5	6	10	7

Period: Scenario 3 25% Level Shift Table: 3-5

Geometric Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	109.06	47.38	46.65	48.10	46.85	52.69	155.40	63.59	63.73	65.39	58.91
	Rank	10	3	1	4	2	5	11	7	8	9	6
	Average Rank by Series	8.75	3.68	3.58	3.95	4.30	5.15	11.00	6.63	6.63	6.50	5.85
	Rank of Average Rank	10	2	1	3	4	5	11	8	8	7	6
5	Geometric Mean	165.68	93.86	99.21	100.10	93.66	99.60	226.64	95.05	105.94	128.91	117.02
	Rank	10	2	4	6	1	5	11	3	7	9	8
	Average Rank by Series	7.05	4.83	5.23	5.10	4.20	5.20	9.50	5.38	5.73	7.45	6.35
	Rank of Average Rank	9	2	5	3	1	4	11	6	7	10	8
10	Geometric Mean	196.91	113.30	125.64	131.42	122.59	138.89	268.45	115.07	141.51	160.63	155.73
	Rank	10	1	4	5	3	6	11	2	7	9	8
	Average Rank by Series	6.75	5.275	5.275	5.15	5	6.55	8.75	4.925	5.525	6.45	6.35
	Rank of Average Rank	10	4	4	3	2	9	11	1	6	8	7
15	Geometric Mean	190.47	128.66	150.94	155.40	132.96	164.27	230.73	142.99	175.80	182.90	161.28
	Rank	10	1	4	5	2	7	11	3	8	9	6
	Average Rank by Series	6.45	5.225	5.225	5.4	4.5	6.4	7.65	6.175	6.325	6.55	6.1
	Rank of Average Rank	9	2	2	4	1	8	11	6	7	10	5

25% Level Shift	Average Rank of Absolute Error			Table 3- 6
Period	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	52.60	19	0.0001
	KRUSKAL-WALLIS	89.34	10	0.0000
5	RANK ANOVA	24.45	19	0.1796
	KRUSKAL-WALLIS	43.31	10	0.0000
10	RANK ANOVA	12.03	19	0.8842
	KRUSKAL-WALLIS	23.76	10	0.0083
15	RANK ANOVA	5.86	19	0.9982
	KRUSKAL-WALLIS	13.31	10	0.2068

25% Level Shift		Symmetry Adjusted MAPE		Table 3-7
	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	61.41	19	0.0000
	KRUSKAL-WALLIS	68.62	10	0.0000
5	RANK ANOVA	27.41	19	0.0955
	KRUSKAL-WALLIS	54.66	10	0.0000
10	RANK ANOVA	16.68	19	0.6117
	KRUSKAL-WALLIS	31.88	10	0.0004
15	RANK ANOVA	12.37	19	0.8693
	KRUSKAL-WALLIS	15.71	10	0.1082

25% Level Shift		Range of Percent Error		Table 3-8
	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	38.24	19	0.0055
	KRUSKAL-WALLIS	43.53	10	0.0000
5	RANK ANOVA	40.60	19	0.0027
	KRUSKAL-WALLIS	68.45	10	0.0000
	RANK ANOVA	26.39	19	0.1196
10	KRUSKAL-WALLIS	46.50	10	0.0000
	RANK ANOVA	20.57	19	0.3610
15	KRUSKAL-WALLIS	44.21	10	0.0000

25% Level Shift	Geometric Root Mean Square Error			Table 3-9
	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	55.35	19	0.0000
5	RANK ANOVA	24.18	19	0.1894
10	RANK ANOVA	13.99	19	0.7843
15	RANK ANOVA	8.21	19	0.9844

25% Level Shift	Log Mean Square Error Ratio			Table 3-10
Period	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	55.35	19	0.0000
	KRUSKAL-WALLIS	41.94	10	0.0000
5	RANK ANOVA	24.18	19	0.1894
	KRUSKAL-WALLIS	20.97	10	0.0213
10	RANK ANOVA	13.99	19	0.7843
	KRUSKAL-WALLIS	13.88	10	0.1786
15	RANK ANOVA	8.21	19	0.9844
	KRUSKAL-WALLIS	9.05	10	0.5270

25% Level Shift	Mean Absolute Percent Error			Table 3 -11
	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	78.12	19	0.0000
	KRUSKAL-WALLIS	109.94	10	0.0000
5	RANK ANOVA	58.34	19	0.0000
	KRUSKAL-WALLIS	69.55	10	0.0000
10	RANK ANOVA	47.62	19	0.0003
	KRUSKAL-WALLIS	42.77	10	0.0000
15	RANK ANOVA	42.33	19	0.0016
	KRUSKAL-WALLIS	35.88	10	0.0001

25% Level Shift	Median Absolute Percent Error			Table 3 -12
	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	50.69	19	0.0001
	KRUSKAL-WALLIS	59.30	10	0.0000
5	RANK ANOVA	10.95	19	0.9255
	KRUSKAL-WALLIS	31.36	10	0.0005
10	RANK ANOVA	14.65	19	0.7455
	KRUSKAL-WALLIS	21.36	10	0.0187
15	RANK ANOVA	5.52	19	0.9988
	KRUSKAL-WALLIS	7.17	10	0.7096

25% Level Shift	Root Mean Square Error			Table 3 -13
	Scenario 3	Chi Square	DF	p Value
1	RANK ANOVA	54.85	19	0.0000
5	RANK ANOVA	32.79	19	0.0254
10	RANK ANOVA	24.85	19	0.1654
15	RANK ANOVA	18.16	19	0.5119

200% Level Shift

Table: 4-1

Period:		Scenario 4	Average Rank of Absolute Error									
		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	5.56	6.78	7.05	7.99	6.82	5.81	4.47	5.04	5.04	6.21	5.47
	Rank	5	8	10	11	9	6	1	2	2	7	4
	Geometric Mean	5.35	6.66	6.98	7.91	6.71	5.71	4.18	4.94	4.94	6.13	5.35
	Rank	5	8	10	11	9	6	1	2	2	7	4
	Average Rank by Series	5.23	7.33	8.43	9.50	7.93	5.60	3.25	4.13	4.13	6.28	4.83
	Rank of Average Rank	5	8	10	11	9	6	1	2	2	7	4
	Kruskal-Wallis Rank Sum	1813.0	2831.5	3128.0	3708.5	2877.5	2005.0	1153.0	1399.5	1399.5	2396.0	1769.5
	Rank of K-W Rank Sum	5	8	10	11	9	6	1	2	2	7	4
	K-W Multi-Comparison Count*	9	9	10	10	9	10	9	10	10	10	9
5	Average	4.95	6.70	7.02	7.68	7.00	5.88	4.19	5.10	5.52	6.31	5.66
	Rank	2	8	10	11	9	6	1	3	4	7	5
	Geometric Mean	4.66	6.61	6.98	7.63	6.97	5.78	3.97	4.91	5.41	6.21	5.58
	Rank	2	8	10	11	9	6	1	3	4	7	5
	Average Rank by Series	4.48	7.53	8.33	9.13	8.18	5.88	2.75	4.13	4.63	6.13	4.88
	Rank of Average Rank	3	8	10	11	9	6	1	2	4	7	5
	Kruskal-Wallis Rank Sum	1460.0	2800.0	3096.5	3579.0	3081.5	2030.0	855.5	1472.0	1738.0	2385.0	1812.5
	Rank of K-W Rank Sum	2	8	10	11	9	6	1	3	4	7	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	10	9	10	9
10	Average	4.75	6.55	6.97	7.53	6.93	5.85	4.31	5.64	5.64	6.41	5.81
	Rank	2	8	10	11	9	6	1	3	3	7	5
	Geometric Mean	4.41	6.42	6.89	7.44	6.82	5.64	4.11	5.42	5.42	6.27	5.64
	Rank	2	8	10	11	9	5	1	3	3	7	6
	Average Rank by Series	4.23	7.08	7.78	8.70	8.33	5.93	3.13	4.73	4.73	6.55	5.38
	Rank of Average Rank	2	8	9	11	10	6	1	3	3	7	5
	Kruskal-Wallis Rank Sum	1388.0	2591.0	2936.0	3317.5	2962.5	2059.0	1012.5	1881.0	1881.0	2472.0	2026.5
	Rank of K-W Rank Sum	2	8	9	11	10	6	1	3	3	7	5
	K-W Multi-Comparison Count*	10	10	9	10	9	9	10	10	10	10	9
15	Average	4.50	6.67	7.04	7.64	6.79	5.91	3.92	5.32	5.81	6.57	5.83
	Rank	2	8	10	11	9	6	1	3	4	7	5
	Geometric Mean	4.21	6.53	6.96	7.52	6.66	5.60	3.72	5.03	5.61	6.39	5.57
	Rank	2	8	10	11	9	5	1	3	6	7	4
	Average Rank by Series	3.95	7.23	7.50	8.70	7.65	6.20	2.75	4.73	5.15	6.53	5.63
	Rank of Average Rank	2	8	9	11	10	6	1	3	4	7	5
	Kruskal-Wallis Rank Sum	1226.0	2694.0	2944.5	3285.0	2782.5	2080.5	839.5	1755.5	2047.5	2584.5	2070.5
	Rank of K-W Rank Sum	2	8	10	11	9	6	1	3	4	7	5
	K-W Multi-Comparison Count*	10	10	10	10	10	8	10	10	8	10	8

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

200% Level Shift

Period: Scenario 4 Range of Percent Error Table: 4-2

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	12.74%	18.67%	18.36%	21.33%	22.25%	19.88%	12.19%	12.90%	12.60%	16.60%	16.77%
	Rank	3	8	7	10	11	9	1	4	2	5	6
	Geometric Mean	11.43%	18.06%	17.72%	19.77%	21.40%	19.23%	9.48%	11.30%	10.86%	13.91%	14.68%
	Rank	4	8	7	10	11	9	1	3	2	5	6
	Average Rank by Series	3.18	7.50	7.10	8.88	9.65	8.00	2.73	3.60	3.10	5.98	6.30
	Rank of Average Rank	3	8	7	10	11	9	1	4	2	5	6
	Kruskal-Wallis Rank Sum	1510.5	2650.0	2568.0	2861.5	3091.0	2821.0	1467.5	1570.0	1514.0	2088.5	2168.0
	Rank of K-W Rank Sum	2	8	7	10	11	9	1	4	3	5	6
	K-W Multi-Comparison Count*	8	10	10	9	10	9	8	9	7	9	9
5	Average	18.44%	37.98%	38.11%	33.39%	33.59%	39.90%	16.06%	25.92%	25.78%	27.43%	27.13%
	Rank	2	9	10	7	8	11	1	4	3	6	5
	Geometric Mean	16.31%	37.33%	37.47%	32.39%	32.49%	39.65%	14.53%	24.62%	24.55%	26.61%	26.21%
	Rank	2	9	10	7	8	11	1	4	3	6	5
	Average Rank by Series	2.53	8.63	8.63	6.78	7.05	10.00	1.93	5.08	4.93	5.08	5.40
	Rank of Average Rank	2	9	9	7	8	11	1	4	3	4	6
	Kruskal-Wallis Rank Sum	919.5	3257.5	3275.5	2652.5	2701.0	3570.0	648.5	1730.5	1736.5	1935.5	1883.0
	Rank of K-W Rank Sum	2	9	10	7	8	11	1	3	4	6	5
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	9	9	9	9
10	Average	24.27%	47.74%	47.32%	48.44%	43.76%	39.39%	23.70%	33.63%	36.04%	41.31%	36.00%
	Rank	2	10	9	11	8	6	1	3	5	7	4
	Geometric Mean	20.64%	44.94%	45.01%	44.83%	40.55%	38.95%	19.06%	28.31%	30.86%	34.03%	30.35%
	Rank	2	10	11	9	8	7	1	3	5	6	4
	Average Rank by Series	2.38	8.55	8.58	7.90	7.68	7.25	2.43	4.75	5.28	5.65	5.58
	Rank of Average Rank	1	10	11	9	8	7	2	3	4	6	5
	Kruskal-Wallis Rank Sum	1224.5	2888.0	2924.5	2884.0	2642.5	2570.0	1217.5	1708.0	1951.5	2297.0	2002.5
	Rank of K-W Rank Sum	2	10	11	9	8	7	1	3	4	6	5
	K-W Multi-Comparison Count*	9	8	8	8	9	9	9	10	9	10	9
15	Average	30.77%	54.81%	57.91%	62.76%	52.69%	38.66%	30.06%	41.81%	45.04%	55.97%	44.79%
	Rank	2	8	10	11	7	3	1	4	6	9	5
	Geometric Mean	23.78%	48.52%	51.94%	54.59%	46.27%	38.32%	21.24%	30.70%	34.47%	42.59%	35.41%
	Rank	2	9	10	11	8	6	1	3	4	7	5
	Average Rank by Series	2.88	8.25	8.70	8.18	7.53	7.05	2.28	4.90	5.15	5.73	5.38
	Rank of Average Rank	2	10	11	9	8	7	1	3	4	6	5
	Kruskal-Wallis Rank Sum	1503.5	2724.0	2878.0	2884.5	2589.5	2371.0	1396.5	1685.0	1941.0	2341.5	1995.5
	Rank of K-W Rank Sum	2	9	10	11	8	7	1	3	4	6	5
	K-W Multi-Comparison Count*	10	10	9	9	10	9	10	10	9	9	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

200% Level Shift — Table: 4-3

Period: Scenario 4 Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	8.68%	10.04%	10.14%	12.10%	10.93%	9.10%	6.37%	7.02%	7.11%	9.17%	8.36%
	Rank	5	8	9	11	10	6	1	2	3	7	4
	Geometric Mean	7.73%	9.69%	9.82%	11.46%	10.22%	8.63%	4.92%	6.57%	6.70%	8.35%	7.59%
	Rank	5	8	9	11	10	7	1	2	3	6	4
	Average Rank by Series	4.75	7.93	8.38	9.95	8.80	6.35	2.65	3.13	3.58	5.85	4.65
	Rank of Average Rank	5	8	9	11	10	7	1	2	3	6	4
	Kruskal-Wallis Rank Sum	1996.0	2726.5	2784.5	3145.0	2834.0	2303.0	1466.0	1447.5	1500.5	2204.0	1903.0
	Rank of K-W Rank Sum	5	8	9	11	10	7	2	1	3	6	4
	K-W Multi-Comparison Count*	10	9	8	10	9	10	8	8	8	10	10
5	Average	13.28%	17.79%	18.60%	21.54%	19.63%	14.80%	9.65%	13.44%	14.44%	17.46%	15.68%
	Rank	2	8	9	11	10	5	1	3	4	7	6
	Geometric Mean	12.81%	17.06%	17.96%	20.88%	19.05%	14.43%	8.61%	12.14%	13.27%	16.33%	14.79%
	Rank	3	8	9	11	10	5	1	2	4	7	6
	Average Rank by Series	4.55	7.78	8.38	9.50	9.20	5.60	2.10	3.63	3.98	5.85	5.45
	Rank of Average Rank	4	8	9	11	10	6	1	2	3	7	5
	Kruskal-Wallis Rank Sum	1641.0	2592.5	2809.5	3322.0	3029.0	1950.0	876.0	1606.5	1854.5	2517.0	2112.0
	Rank of K-W Rank Sum	3	8	9	11	10	5	1	2	4	7	6
	K-W Multi-Comparison Count*	10	9	10	10	10	10	10	10	10	9	10
10	Average	17.47%	23.26%	25.38%	30.01%	26.05%	16.38%	14.53%	19.48%	21.84%	26.76%	22.72%
	Rank	3	7	8	11	9	2	1	4	5	10	6
	Geometric Mean	15.83%	20.39%	22.51%	27.27%	23.39%	15.85%	11.79%	15.08%	17.38%	22.54%	18.81%
	Rank	3	7	8	11	10	4	1	2	5	9	6
	Average Rank by Series	3.95	7.28	7.93	9.55	8.45	5.45	2.05	4.23	4.63	6.75	5.75
	Rank of Average Rank	2	8	9	11	10	5	1	3	4	7	6
	Kruskal-Wallis Rank Sum	1877.0	2354.5	2596.5	3034.0	2700.0	1816.0	1399.0	1743.5	2038.5	2588.0	2163.0
	Rank of K-W Rank Sum	4	7	9	11	10	3	1	2	5	8	6
	K-W Multi-Comparison Count*	9	10	9	10	10	8	10	9	10	9	10
15	Average	20.90%	28.42%	31.74%	38.61%	32.20%	17.60%	18.10%	24.74%	28.54%	35.63%	29.30%
	Rank	3	5	8	11	9	1	2	4	6	10	7
	Geometric Mean	17.57%	22.80%	25.94%	32.58%	26.85%	17.00%	12.85%	17.06%	20.37%	27.38%	22.10%
	Rank	4	7	8	11	9	2	1	3	5	10	6
	Average Rank by Series	3.80	7.38	7.73	9.00	8.40	5.50	2.05	4.48	4.93	6.65	6.10
	Rank of Average Rank	2	8	9	11	10	5	1	3	4	7	6
	Kruskal-Wallis Rank Sum	1869.0	2302.5	2530.5	2954.0	2681.0	1887.0	1474.0	1772.5	2078.5	2576.0	2205.0
	Rank of K-W Rank Sum	3	7	8	11	10	4	1	2	5	9	6
	K-W Multi-Comparison Count*	9	10	9	10	10	9	10	10	10	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

200% Level Shift

Period: Scenario 4 Root Mean Square Error Table: 4-4

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	215.26	270.27	268.08	315.75	296.97	254.93	154.08	186.48	185.80	238.12	219.91
	Rank	4	9	8	11	10	7	1	3	2	6	5
	Average Rank by Series	4.70	7.83	7.88	9.80	9.05	6.40	2.75	3.13	3.18	5.90	5.40
	Rank of Average Rank	4	8	9	11	10	7	1	2	3	6	5
5	Geometric Mean	387.58	583.25	595.88	653.85	617.50	531.90	267.63	394.46	420.14	503.19	465.85
	Rank	2	8	9	11	10	7	1	3	4	6	5
	Average Rank by Series	3.00	8.13	8.73	9.65	9.00	6.50	1.50	3.83	4.23	6.00	5.45
	Rank of Average Rank	2	8	9	11	10	7	1	3	4	6	5
10	Geometric Mean	480.68	699.82	750.57	864.39	753.33	568.37	371.89	496.17	562.60	707.09	593.71
	Rank	2	7	9	11	10	5	1	3	4	8	6
	Average Rank by Series	3.20	7.68	8.38	9.30	8.45	6.05	1.85	4.33	4.88	6.50	5.40
	Rank of Average Rank	2	8	9	11	10	6	1	3	4	7	5
15	Geometric Mean	552.65	790.00	880.72	1052.50	885.19	601.54	410.44	562.38	665.09	871.21	706.77
	Rank	2	7	9	11	10	4	1	3	5	8	6
	Average Rank by Series	3.10	7.68	8.33	9.00	8.00	5.80	2.15	4.48	5.13	6.80	5.55
	Rank of Average Rank	2	8	10	11	9	6	1	3	4	7	5

200% Level Shift

Period: Scenario 4 Geometric Root Mean Square Error Table: 4-5

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	139.88	141.63	162.67	203.00	146.68	102.84	79.64	105.19	119.30	147.20	120.91
	Rank	6	7	10	11	8	2	1	3	4	9	5
	Average Rank by Series	5.70	6.88	7.88	9.45	7.20	4.80	3.65	3.93	4.68	6.85	5.00
	Rank of Average Rank	6	8	10	11	9	4	1	2	3	7	5
5	Geometric Mean	245.22	254.10	285.25	362.76	313.31	177.24	153.20	209.44	237.52	306.70	270.47
	Rank	5	6	8	11	10	2	1	3	4	9	7
	Average Rank by Series	5.55	7.13	7.23	8.50	8.30	4.45	3.20	4.38	4.78	5.90	6.60
	Rank of Average Rank	5	8	9	11	10	3	1	2	4	6	7
10	Geometric Mean	332.73	296.64	355.46	468.78	416.76	239.59	223.57	264.35	318.49	428.11	377.80
	Rank	6	4	7	11	9	2	1	3	5	10	8
	Average Rank by Series	5.45	5.825	6.775	8.2	7.55	4.9	3.45	4.625	5.125	7.5	6.6
	Rank of Average Rank	5	6	8	11	10	3	1	2	4	9	7
15	Geometric Mean	371.49	392.38	446.65	611.29	503.39	292.41	254.49	335.25	404.69	538.85	456.34
	Rank	4	5	7	11	9	2	1	3	6	10	8
	Average Rank by Series	4.80	6.83	7.13	8.55	7.60	5.45	2.50	4.58	5.13	6.95	6.50
	Rank of Average Rank	3	7	9	11	10	5	1	2	4	8	6

200% Level Shift Average Rank of Absolute Error				Table 4- 6
Period	Scenario 4	Chi Square	DF	p Value
1	RANK ANOVA	45.27	19	0.0006
	KRUSKAL-WALLIS	87.44	10	0.0000
5	RANK ANOVA	44.03	19	0.0009
	KRUSKAL-WALLIS	88.28	10	0.0000
10	RANK ANOVA	35.89	19	0.0109
	KRUSKAL-WALLIS	63.01	10	0.0000
15	RANK ANOVA	33.13	19	0.0232
	KRUSKAL-WALLIS	68.04	10	0.0000

200% Level Shift		Symmetry Adjusted MAPE		Table 4- 7
	Scenario 4	Chi Square	DF	p Value
1	RANK ANOVA	66.49	19	0.0000
	KRUSKAL-WALLIS	45.48	10	0.0000
5	RANK ANOVA	72.73	19	0.0000
	KRUSKAL-WALLIS	84.85	10	0.0000
10	RANK ANOVA	60.09	19	0.0000
	KRUSKAL-WALLIS	38.94	10	0.0000
15	RANK ANOVA	57.02	19	0.0000
	KRUSKAL-WALLIS	30.83	10	0.0006

200% Level Shift		Range of Percent Error		Table 4- 8
	Scenario 4	Chi Square	DF	p Value
1	RANK ANOVA	65.01	19	0.0000
	KRUSKAL-WALLIS	47.47	10	0.0000
5	RANK ANOVA	66.44	19	0.0000
	KRUSKAL-WALLIS	114.26	10	0.0000
	RANK ANOVA	51.93	19	0.0001
10	KRUSKAL-WALLIS	50.19	10	0.0000
	RANK ANOVA	48.94	19	0.0002
15	KRUSKAL-WALLIS	35.88	10	0.0001

200% Level Shift Geometric Root Mean Square Error				Table 4- 9
	Scenario 4	Chi Square	DF	p Value
1	RANK ANOVA	34.30	19	0.0169
5	RANK ANOVA	30.98	19	0.0406
10	RANK ANOVA	22.56	19	0.2574
15	RANK ANOVA	30.66	19	0.0440

200% Level Shift	Log Mean Square Error Ratio			Table	4-10
Period	Scenario 4	Chi Square	DF	p Value	
1	RANK ANOVA	34.30	19	0.0169	
	KRUSKAL-WALLIS	31.89	10	0.0004	
5	RANK ANOVA	30.98	19	0.0406	
	KRUSKAL-WALLIS	25.43	10	0.0046	
10	RANK ANOVA	22.56	19	0.2574	
	KRUSKAL-WALLIS	14.61	10	0.1471	
15	RANK ANOVA	30.66	19	0.0440	
	KRUSKAL-WALLIS	17.95	10	0.0559	

200% Level Shift	Mean Absolute Percent Error			Table	4-11
	Scenario 4	Chi Square	DF	p Value	
1	RANK ANOVA	64.67	19	0.0000	
	KRUSKAL-WALLIS	45.01	10	0.0000	
5	RANK ANOVA	61.69	19	0.0000	
	KRUSKAL-WALLIS	63.90	10	0.0000	
10	RANK ANOVA	52.23	19	0.0001	
	KRUSKAL-WALLIS	29.68	10	0.0010	
15	RANK ANOVA	46.53	19	0.0004	
	KRUSKAL-WALLIS	24.35	10	0.0067	

200% Level Shift	Median Absolute Percent Error			Table	4-12
	Scenario 4	Chi Square	DF	p Value	
1	RANK ANOVA	25.99	19	0.1306	
	KRUSKAL-WALLIS	22.52	10	0.0127	
5	RANK ANOVA	24.91	19	0.1635	
	KRUSKAL-WALLIS	34.83	10	0.0001	
10	RANK ANOVA	19.38	19	0.4325	
	KRUSKAL-WALLIS	27.05	10	0.0026	
15	RANK ANOVA	27.05	19	0.1035	
	KRUSKAL-WALLIS	24.69	10	0.0060	

200% Level Shift	Root Mean Square Error			Table	4-13
	Scenario 4	Chi Square	DF	p Value	
1	RANK ANOVA	62.52	19	0.0000	
5	RANK ANOVA	75.33	19	0.0000	
10	RANK ANOVA	57.81	19	0.0000	
15	RANK ANOVA	50.88	19	0.0001	

Trend Shift — Table: 5-1

Period:		Scenario 5 Adjusted	Average Rank of Absolute Error HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	6.45	5.62	5.42	5.98	6.27	6.08	8.24	5.18	5.18	5.85	5.74
	Rank	10	4	3	7	9	8	11	1	1	6	5
	Geometric Mean	6.29	5.46	5.23	5.55	5.88	5.66	8.06	5.13	5.13	5.67	5.55
	Rank	10	4	3	5	9	7	11	1	1	8	6
	Average Rank by Series	6.30	5.13	5.33	6.58	7.10	5.93	9.58	4.15	4.15	5.93	5.75
	Rank of Average Rank	8	3	4	9	10	6	11	1	1	6	5
	Kruskal-Wallis Rank Sum	2591.5	1931.0	1857.0	2383.5	2418.0	2226.5	3631.0	1510.0	1510.0	2151.0	2079.0
	Rank of K-W Rank Sum	10	4	3	8	9	7	11	1	1	6	5
	K-W Multi-Comparison Count*	10	9	9	9	9	9	10	9	9	8	9
5	Average	5.09	6.93	7.00	7.07	6.35	8.15	5.96	4.81	4.89	5.14	4.61
	Rank	4	8	9	10	7	11	6	2	3	5	1
	Geometric Mean	4.59	6.84	6.89	6.75	6.15	7.86	5.58	4.68	4.70	4.67	4.28
	Rank	2	9	10	8	7	11	6	4	5	3	1
	Average Rank by Series	4.18	7.68	8.08	7.83	6.68	8.90	5.75	3.90	4.15	4.88	4.00
	Rank of Average Rank	4	8	10	9	7	11	6	1	3	5	2
	Kruskal-Wallis Rank Sum	1659.5	2872.0	2938.0	2825.5	2448.0	3365.5	2139.0	1400.0	1510.0	1769.0	1383.5
	Rank of K-W Rank Sum	4	9	10	8	7	11	6	2	3	5	1
	K-W Multi-Comparison Count*	10	8	9	9	10	10	10	9	10	10	9
10	Average	5.25	7.10	7.28	6.89	6.16	8.59	4.44	5.49	5.49	5.08	4.39
	Rank	4	9	10	8	7	11	2	5	5	3	1
	Geometric Mean	4.63	6.98	7.09	6.55	5.93	8.26	3.89	5.21	5.21	4.62	4.09
	Rank	4	9	10	8	7	11	1	5	5	3	2
	Average Rank by Series	5.13	7.83	7.90	6.93	6.38	9.15	3.80	5.48	5.48	4.58	3.70
	Rank of Average Rank	4	9	10	8	7	11	2	5	5	3	1
	Kruskal-Wallis Rank Sum	1792.0	2886.0	2995.5	2669.5	2295.0	3489.0	1409.5	1955.0	1955.0	1704.0	1277.0
	Rank of K-W Rank Sum	4	9	10	8	7	11	2	5	5	3	1
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	10	10	10	10
15	Average	5.50	7.16	7.30	6.54	5.94	8.76	4.62	5.49	5.58	4.89	4.23
	Rank	5	9	10	8	7	11	2	4	6	3	1
	Geometric Mean	4.87	7.02	7.07	6.13	5.66	8.45	4.02	5.29	5.28	4.41	3.89
	Rank	4	9	10	8	7	11	2	6	5	3	1
	Average Rank by Series	5.50	7.83	7.93	6.58	6.28	9.25	4.00	5.10	5.58	4.35	3.63
	Rank of Average Rank	5	9	10	8	7	11	2	4	6	3	1
	Kruskal-Wallis Rank Sum	1908.5	2884.0	2974.0	2468.0	2177.5	3548.0	1512.0	1950.5	2013.0	1622.0	1252.5
	Rank of K-W Rank Sum	4	9	10	8	7	11	2	5	6	3	1
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	9	9	10	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Trend Shift

Period: Scenario 5 Range of Percent Error Table: 5-2

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	11.24%	12.21%	12.44%	14.14%	12.91%	12.29%	16.74%	11.80%	11.90%	13.50%	13.73%
	Rank	1	4	6	10	7	5	11	2	3	8	9
	Geometric Mean	9.22%	7.93%	8.98%	10.38%	10.05%	7.02%	15.51%	8.39%	8.72%	9.93%	10.86%
	Rank	6	2	5	9	8	1	11	3	4	7	10
	Average Rank by Series	5.90	5.10	4.98	7.23	6.38	4.65	8.15	4.65	4.83	6.88	7.28
	Rank of Average Rank	6	5	4	9	7	1	11	1	3	8	10
	Kruskal-Wallis Rank Sum	2148.0	2035.0	2052.5	2291.5	2249.5	1954.0	2983.0	1975.0	1980.5	2266.5	2374.5
	Rank of K-W Rank Sum	6	4	5	9	7	1	11	2	3	8	10
	K-W Multi-Comparison Count*	10	7	7	8	8	8	10	6	6	8	10
5	Average	19.81%	16.70%	17.55%	16.11%	17.26%	14.46%	22.32%	17.51%	16.75%	16.55%	15.61%
	Rank	10	5	9	3	7	1	11	8	6	4	2
	Geometric Mean	19.33%	13.40%	14.81%	14.50%	15.89%	11.01%	21.72%	15.22%	14.22%	13.29%	13.11%
	Rank	10	4	7	6	9	1	11	8	5	3	2
	Average Rank by Series	7.00	5.28	5.68	5.40	5.70	5.05	8.30	6.28	6.38	5.70	5.25
	Rank of Average Rank	10	3	5	4	6	1	11	8	9	6	2
	Kruskal-Wallis Rank Sum	2639.0	2072.5	2194.5	2003.0	2202.0	1742.0	3038.0	2231.5	2147.5	2104.0	1936.0
	Rank of K-W Rank Sum	10	4	7	3	8	1	11	9	6	5	2
	K-W Multi-Comparison Count*	10	7	7	8	7	10	10	8	6	8	9
10	Average	24.34%	19.67%	21.15%	20.27%	19.73%	15.56%	32.77%	19.35%	18.82%	19.47%	18.98%
	Rank	10	6	9	8	7	1	11	4	2	5	3
	Geometric Mean	22.79%	16.29%	18.05%	17.82%	18.08%	12.43%	32.15%	17.81%	17.20%	17.39%	17.44%
	Rank	10	2	8	7	9	1	11	6	3	4	5
	Average Rank by Series	6.78	5.15	5.10	4.85	5.05	4.60	10.08	6.15	5.95	6.35	5.95
	Rank of Average Rank	10	5	4	2	3	1	11	8	6	9	6
	Kruskal-Wallis Rank Sum	2685.5	2038.0	2245.0	2126.0	2090.0	1421.0	3643.5	2024.0	1960.0	2053.0	2024.0
	Rank of K-W Rank Sum	10	5	9	8	7	1	11	3	2	6	3
	K-W Multi-Comparison Count*	10	5	10	8	5	10	10	5	7	5	5
15	Average	22.69%	19.44%	21.56%	22.48%	20.21%	13.34%	35.98%	16.73%	16.83%	19.82%	17.47%
	Rank	10	5	8	9	7	1	11	2	3	6	4
	Geometric Mean	19.36%	15.36%	17.32%	17.98%	16.94%	11.33%	33.93%	15.98%	16.25%	18.61%	16.65%
	Rank	10	2	7	8	6	1	11	3	4	9	5
	Average Rank by Series	6.05	5.68	5.68	5.80	5.75	4.65	10.30	5.23	5.18	6.40	5.30
	Rank of Average Rank	9	5	5	8	7	1	11	3	2	10	4
	Kruskal-Wallis Rank Sum	2369.0	1965.5	2239.5	2241.0	2094.0	1294.0	3631.0	1989.5	2027.5	2361.0	2098.0
	Rank of K-W Rank Sum	10	2	7	8	5	1	11	3	4	9	6
	K-W Multi-Comparison Count*	9	8	9	9	8	10	10	8	6	9	8

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Trend Shift Table: 5-3

Period: Scenario 5 Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	7.15%	6.40%	6.20%	7.18%	7.09%	7.04%	10.78%	6.09%	5.97%	6.88%	6.48%
	Rank	9	4	3	10	8	7	11	2	1	6	5
	Geometric Mean	6.43%	5.53%	5.22%	5.70%	5.72%	6.36%	10.25%	5.17%	4.99%	5.44%	5.36%
	Rank	10	6	3	7	8	9	11	2	1	5	4
	Average Rank by Series	6.95	5.53	5.18	6.10	6.00	6.55	10.30	4.58	4.28	5.45	5.10
	Rank of Average Rank	10	6	4	8	7	9	11	2	1	5	3
	Kruskal-Wallis Rank Sum	2333.0	2041.5	1970.5	2190.0	2187.0	2301.0	3347.0	1933.5	1897.5	2095.0	2014.0
	Rank of K-W Rank Sum	10	5	3	8	7	9	11	2	1	6	4
	K-W Multi-Comparison Count*	9	7	6	9	9	9	10	7	8	9	7
5	Average	12.86%	17.38%	16.61%	17.00%	16.30%	20.38%	15.15%	11.87%	11.10%	11.77%	11.42%
	Rank	5	10	8	9	7	11	6	4	1	3	2
	Geometric Mean	12.50%	16.26%	15.32%	15.15%	14.79%	20.24%	14.85%	9.96%	9.03%	8.55%	8.85%
	Rank	5	10	9	8	6	11	7	4	3	1	2
	Average Rank by Series	4.00	8.33	8.48	7.95	7.40	9.40	5.85	3.63	3.53	4.05	3.40
	Rank of Average Rank	4	9	10	8	7	11	6	3	2	5	1
	Kruskal-Wallis Rank Sum	1684.0	2769.5	2601.5	2620.0	2459.0	3483.0	2221.0	1635.5	1529.5	1705.0	1602.0
	Rank of K-W Rank Sum	4	10	8	9	7	11	6	3	1	5	2
	K-W Multi-Comparison Count*	9	10	9	9	10	10	10	8	9	8	8
10	Average	25.09%	27.92%	26.60%	25.73%	25.76%	33.07%	21.46%	22.82%	21.49%	20.54%	20.68%
	Rank	6	10	9	7	8	11	3	5	4	1	2
	Geometric Mean	24.71%	26.07%	24.50%	22.92%	23.65%	32.91%	21.06%	20.35%	18.71%	16.72%	17.96%
	Rank	9	10	8	6	7	11	5	4	3	1	2
	Average Rank by Series	5.00	8.38	8.33	7.50	7.25	9.45	3.75	4.43	4.38	4.00	3.55
	Rank of Average Rank	6	10	9	8	7	11	2	5	4	3	1
	Kruskal-Wallis Rank Sum	2107.0	2728.5	2559.5	2410.0	2369.0	3505.0	1567.0	1898.5	1774.5	1728.0	1663.0
	Rank of K-W Rank Sum	6	10	9	8	7	11	1	5	4	3	2
	K-W Multi-Comparison Count*	10	10	10	9	9	10	10	10	9	8	9
15	Average	34.69%	35.34%	33.62%	31.84%	32.20%	42.34%	30.29%	31.00%	29.28%	27.44%	27.86%
	Rank	9	10	8	6	7	11	4	5	3	1	2
	Geometric Mean	34.38%	33.01%	30.98%	28.53%	29.74%	42.24%	29.95%	28.21%	26.12%	23.39%	24.98%
	Rank	10	9	8	5	6	11	7	4	3	1	2
	Average Rank by Series	5.25	8.33	8.33	7.00	6.95	9.50	3.85	4.68	4.73	3.80	3.60
	Rank of Average Rank	6	9	9	8	7	11	3	4	5	2	1
	Kruskal-Wallis Rank Sum	2200.0	2661.5	2509.5	2262.0	2208.0	3563.0	1616.0	1997.5	1888.5	1747.0	1657.0
	Rank of K-W Rank Sum	6	10	9	8	7	11	1	5	4	3	2
	K-W Multi-Comparison Count*	9	10	10	8	9	10	9	10	10	10	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Trend Shift
Period: Scenario 5 Root Mean Square Error Table: 5-4

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	146.05	129.06	122.48	134.37	134.91	143.47	218.14	121.78	117.44	129.89	131.70
	Rank	10	4	3	7	8	9	11	2	1	5	6
	Average Rank by Series	6.40	5.48	5.23	6.85	6.15	6.45	9.40	3.93	3.93	6.30	5.90
	Rank of Average Rank	8	4	3	10	6	9	11	1	1	7	5
5	Geometric Mean	344.33	404.44	384.03	384.61	374.20	482.56	383.47	269.02	243.94	230.41	236.14
	Rank	5	10	8	9	6	11	7	4	3	1	2
	Average Rank by Series	4.35	8.03	8.13	8.30	7.15	9.30	5.80	3.68	3.53	4.45	3.30
	Rank of Average Rank	4	8	9	10	7	11	6	3	2	5	1
10	Geometric Mean	747.02	762.28	722.29	693.10	697.52	922.69	693.39	605.91	554.73	505.38	536.15
	Rank	9	10	8	5	7	11	6	4	3	1	2
	Average Rank by Series	5.30	8.13	8.28	7.55	7.10	9.30	4.00	4.48	4.43	4.10	3.35
	Rank of Average Rank	6	9	10	8	7	11	2	5	4	3	1
15	Geometric Mean	1148.17	1107.97	1048.86	996.13	1004.80	1353.92	1061.12	946.72	877.47	809.85	837.76
	Rank	10	9	7	5	6	11	8	4	3	1	2
	Average Rank by Series	5.40	8.13	8.23	7.20	6.95	9.25	4.15	4.68	4.58	4.05	3.40
	Rank of Average Rank	6	9	10	8	7	11	3	5	4	2	1

Trend Shift
Period: Scenario 5 Geometric Root Mean Square Error Table: 5-5

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	100.95	74.28	67.89	80.02	83.31	98.19	156.55	71.08	66.55	73.91	69.70
	Rank	10	6	2	7	8	9	11	4	1	5	3
	Average Rank by Series	7.30	5.43	5.08	6.00	6.45	6.55	10.35	4.78	4.48	5.10	4.50
	Rank of Average Rank	10	6	4	7	8	9	11	3	1	5	2
5	Geometric Mean	215.86	310.03	287.08	273.09	277.49	441.20	271.21	173.12	152.82	142.70	161.99
	Rank	5	10	9	7	8	11	6	4	2	1	3
	Average Rank by Series	4.30	8.28	8.13	7.85	7.15	9.35	5.80	3.73	3.43	3.80	4.20
	Rank of Average Rank	5	10	9	8	7	11	6	2	1	3	4
10	Geometric Mean	614.33	610.07	582.98	512.32	555.54	876.79	446.56	484.47	466.38	397.47	425.50
	Rank	10	9	8	6	7	11	3	5	4	1	2
	Average Rank by Series	5.15	8.025	8.325	7.25	7.15	9.25	3.7	4.475	4.875	4.1	3.7
	Rank of Average Rank	6	9	10	8	7	11	1	4	5	3	1
15	Geometric Mean	1027.32	929.30	855.55	691.12	836.89	1337.53	817.37	815.27	757.70	596.00	723.89
	Rank	10	9	8	2	7	11	6	5	4	1	3
	Average Rank by Series	5.20	8.23	7.93	6.75	6.95	9.60	3.65	4.93	4.93	3.90	3.95
	Rank of Average Rank	6	10	9	7	8	11	1	4	4	2	3

Trend Shift Average Rank of Absolute Error — Table 5-6

Period	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	23.63	19	0.2106
	KRUSKAL-WALLIS	42.10	10	0.0000
5	RANK ANOVA	37.77	19	0.0063
	KRUSKAL-WALLIS	62.58	10	0.0000
10	RANK ANOVA	33.60	19	0.0205
	KRUSKAL-WALLIS	62.62	10	0.0000
15	RANK ANOVA	33.42	19	0.0215
	KRUSKAL-WALLIS	59.76	10	0.0000

Trend Shift Symmetry Adjusted MAPE — Table 5- 7

	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	27.46	19	0.0944
	KRUSKAL-WALLIS	15.90	10	0.1026
5	RANK ANOVA	57.47	19	0.0000
	KRUSKAL-WALLIS	48.05	10	0.0000
10	RANK ANOVA	50.34	19	0.0001
	KRUSKAL-WALLIS	40.50	10	0.0000
15	RANK ANOVA	44.59	19	0.0008
	KRUSKAL-WALLIS	36.66	10	0.0001

Trend Shift Range of Percent Error — Table 5-8

	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	16.74	19	0.6078
	KRUSKAL-WALLIS	10.72	10	0.3797
5	RANK ANOVA	10.08	19	0.9510
	KRUSKAL-WALLIS	15.32	10	0.1208
	RANK ANOVA	24.55	19	0.1758
10	KRUSKAL-WALLIS	38.41	10	0.0000
	RANK ANOVA	24.02	19	0.1955
15	KRUSKAL-WALLIS	37.96	10	0.0000

Trend Shift Geometric Root Mean Square Error — Table 5- 9

	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	30.91	19	0.0413
5	RANK ANOVA	50.94	19	0.0001
10	RANK ANOVA	43.65	19	0.0011
15	RANK ANOVA	42.31	19	0.0016

Trend Shift Log Mean Square Error Ratio — Table 5-10

Period	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	30.91	19	0.0413
	KRUSKAL-WALLIS	39.45	10	0.0000
5	RANK ANOVA	50.94	19	0.0001
	KRUSKAL-WALLIS	62.00	10	0.0000
10	RANK ANOVA	43.65	19	0.0011
	KRUSKAL-WALLIS	46.44	10	0.0000
15	RANK ANOVA	42.31	19	0.0016
	KRUSKAL-WALLIS	48.50	10	0.0000

Trend Shift Mean Absolute Percent Error — Table 5-11

	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	28.35	19	0.0769
	KRUSKAL-WALLIS	20.10	10	0.0283
5	RANK ANOVA	57.93	19	0.0000
	KRUSKAL-WALLIS	49.51	10	0.0000
10	RANK ANOVA	50.39	19	0.0001
	KRUSKAL-WALLIS	41.66	10	0.0000
15	RANK ANOVA	46.43	19	0.0004
	KRUSKAL-WALLIS	38.85	10	0.0000

Trend Shift Median Absolute Percent Error — Table 5- 12

	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	24.33	19	0.1838
	KRUSKAL-WALLIS	22.74	10	0.0118
5	RANK ANOVA	42.59	19	0.0015
	KRUSKAL-WALLIS	31.74	10	0.0004
10	RANK ANOVA	39.66	19	0.0036
	KRUSKAL-WALLIS	32.04	10	0.0004
15	RANK ANOVA	29.80	19	0.0544
	KRUSKAL-WALLIS	30.89	10	0.0006

Trend Shift Root Mean Square Error — Table 5- 13

	Scenario 5	Chi Square	DF	p Value
1	RANK ANOVA	23.70	19	0.2079
5	RANK ANOVA	52.64	19	0.0001
10	RANK ANOVA	46.42	19	0.0004
15	RANK ANOVA	42.66	19	0.0014

No Change | Table: 6-1

Period:		Scenario 6	Average Rank of Absolute Error									
		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	8.08	4.89	4.61	4.49	4.24	5.66	9.17	6.07	6.07	6.50	5.99
	Rank	10	4	3	2	1	5	11	7	7	9	6
	Geometric Mean	7.94	4.79	4.47	4.42	4.19	5.54	9.06	6.05	6.05	6.45	5.95
	Rank	10	4	3	2	1	5	11	7	7	9	6
	Average Rank by Series	9.08	4.15	3.35	2.98	2.28	5.65	10.98	6.65	6.65	7.55	5.98
	Rank of Average Rank	10	4	3	2	1	5	11	7	7	9	6
	Kruskal-Wallis Rank Sum	3629.5	1327.0	1143.0	973.0	774.5	2043.0	4074.5	2424.0	2424.0	2878.0	2353.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	7	7	9	6
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	9	9	10	9
5	Average	7.95	5.33	4.88	4.59	4.27	6.09	8.86	6.17	5.96	6.08	5.82
	Rank	10	4	3	2	1	8	11	9	6	7	5
	Geometric Mean	7.72	5.20	4.69	4.38	4.09	5.86	8.71	6.11	5.93	5.91	5.73
	Rank	10	4	3	2	1	6	11	9	8	7	5
	Average Rank by Series	8.33	4.73	3.80	3.88	2.90	6.63	10.20	6.75	6.30	6.33	6.18
	Rank of Average Rank	10	4	2	3	1	8	11	9	6	7	5
	Kruskal-Wallis Rank Sum	3374.5	1698.5	1373.0	1222.0	990.0	2422.0	3843.5	2493.5	2324.0	2374.0	2195.0
	Rank of K-W Rank Sum	10	4	3	2	1	8	11	9	6	7	5
	K-W Multi-Comparison Count*	10	10	10	10	10	8	10	9	9	8	10
10	Average	7.68	5.24	5.14	4.83	4.35	6.24	8.54	6.02	6.02	6.34	5.72
	Rank	10	4	3	2	1	8	11	6	6	9	5
	Geometric Mean	7.43	5.05	4.92	4.56	4.07	5.91	8.32	5.95	5.95	6.21	5.64
	Rank	10	4	3	2	1	6	11	7	7	9	5
	Average Rank by Series	7.78	4.78	4.68	4.25	3.30	6.45	9.53	6.53	6.53	6.70	5.80
	Rank of Average Rank	10	4	3	2	1	6	11	7	7	9	5
	Kruskal-Wallis Rank Sum	3193.5	1686.0	1613.0	1427.5	1104.5	2512.5	3622.0	2352.5	2352.5	2493.5	2056.5
	Rank of K-W Rank Sum	10	4	3	2	1	9	11	6	6	8	5
	K-W Multi-Comparison Count*	10	9	9	10	10	9	10	10	10	9	10
15	Average	7.39	5.53	5.15	5.05	4.22	6.51	8.03	6.17	6.03	6.25	5.67
	Rank	10	4	3	2	1	9	11	7	6	8	5
	Geometric Mean	6.94	5.35	4.82	4.69	3.94	6.14	7.59	6.13	5.96	6.11	5.53
	Rank	10	4	3	2	1	9	11	8	6	7	5
	Average Rank by Series	7.10	5.30	4.45	4.35	3.28	6.98	8.60	6.85	6.78	6.60	5.73
	Rank of Average Rank	10	4	3	2	1	9	11	8	7	6	5
	Kruskal-Wallis Rank Sum	2896.0	1906.0	1679.5	1592.5	1070.0	2602.5	3274.0	2468.5	2373.5	2439.5	2008.0
	Rank of K-W Rank Sum	10	4	3	2	1	9	11	8	6	7	5
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	9	9	8	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

No Change
Period: Scenario 6 Range of Percent Error Table: 6-2

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	20.54%	11.94%	11.92%	10.39%	10.88%	13.09%	27.16%	17.37%	17.42%	14.11%	15.59%
	Rank	10	4	3	1	2	5	11	8	9	6	7
	Geometric Mean	17.20%	4.77%	4.67%	4.67%	4.87%	5.44%	25.18%	14.51%	14.57%	12.50%	13.40%
	Rank	10	3	1	2	4	5	11	8	9	6	7
	Average Rank by Series	8.35	4.15	3.90	3.38	3.50	4.90	10.35	7.65	7.40	5.68	6.75
	Rank of Average Rank	10	4	3	1	2	5	11	9	8	6	7
	Kruskal-Wallis Rank Sum	2844.0	1591.0	1588.0	1648.5	1663.0	1691.0	3469.0	2544.0	2552.0	2294.5	2425.0
	Rank of K-W Rank Sum	10	2	1	3	4	5	11	8	9	6	7
	K-W Multi-Comparison Count*	10	7	7	6	6	8	10	9	9	10	10
5	Average	25.39%	13.12%	13.04%	11.23%	11.13%	15.39%	47.87%	25.32%	25.33%	25.80%	25.93%
	Rank	8	4	3	2	1	5	11	6	7	9	10
	Geometric Mean	23.72%	8.05%	7.69%	7.03%	6.80%	9.05%	46.29%	23.60%	23.30%	24.02%	23.97%
	Rank	8	4	3	2	1	5	11	7	6	10	9
	Average Rank by Series	7.45	4.13	3.58	3.30	3.10	5.10	10.65	6.93	7.13	7.35	7.30
	Rank of Average Rank	10	4	3	2	1	5	11	6	7	9	8
	Kruskal-Wallis Rank Sum	2681.0	1393.5	1368.5	1257.0	1244.0	1558.0	3954.0	2677.5	2684.5	2732.0	2760.0
	Rank of K-W Rank Sum	7	4	3	2	1	5	11	6	8	9	10
	K-W Multi-Comparison Count*	7	9	9	9	9	10	10	8	6	6	7
10	Average	34.10%	14.95%	14.60%	12.71%	12.25%	15.39%	45.78%	24.57%	24.48%	24.42%	26.47%
	Rank	10	4	3	2	1	5	11	8	7	6	9
	Geometric Mean	29.93%	10.28%	10.13%	8.80%	8.29%	9.11%	43.12%	23.62%	23.53%	23.18%	25.52%
	Rank	10	5	4	2	1	3	11	8	7	6	9
	Average Rank by Series	8.50	4.90	4.43	3.08	3.08	4.80	10.20	6.75	6.53	6.78	6.98
	Rank of Average Rank	10	5	3	1	1	4	11	7	6	8	9
	Kruskal-Wallis Rank Sum	3125.0	1541.0	1509.5	1280.5	1254.5	1487.0	3856.0	2508.0	2501.5	2531.5	2715.5
	Rank of K-W Rank Sum	10	5	4	2	1	3	11	7	6	8	9
	K-W Multi-Comparison Count*	10	8	8	9	9	8	10	8	8	8	10
15	Average	40.72%	14.16%	13.92%	14.52%	11.96%	12.89%	54.03%	22.14%	22.31%	24.71%	24.48%
	Rank	10	4	3	5	1	2	11	6	7	9	8
	Geometric Mean	32.22%	8.96%	9.30%	9.31%	7.84%	8.23%	49.80%	21.57%	21.76%	23.12%	23.40%
	Rank	10	3	4	5	1	2	11	6	7	8	9
	Average Rank by Series	8.08	4.70	4.28	4.63	3.68	4.10	10.13	6.20	6.23	6.93	7.08
	Rank of Average Rank	10	5	3	4	1	2	11	6	7	8	9
	Kruskal-Wallis Rank Sum	3098.5	1521.0	1482.5	1576.5	1290.5	1369.0	3903.5	2398.0	2421.5	2594.5	2654.5
	Rank of K-W Rank Sum	10	4	3	5	1	2	11	6	7	8	9
	K-W Multi-Comparison Count*	10	8	9	9	9	9	10	9	9	9	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

No Change Table: 6-3

Period: Scenario 6 Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	12.89%	5.41%	5.39%	4.63%	4.61%	5.93%	17.02%	7.93%	7.93%	7.83%	7.63%
	Rank	10	4	3	2	1	5	11	9	8	7	6
	Geometric Mean	11.20%	2.61%	2.51%	2.39%	2.38%	3.26%	15.74%	6.84%	6.83%	6.83%	6.70%
	Rank	10	4	3	2	1	5	11	9	8	7	6
	Average Rank by Series	9.85	3.68	2.98	2.60	2.35	5.40	11.00	7.28	6.98	6.95	6.95
	Rank of Average Rank	10	4	3	2	1	5	11	9	8	6	6
	Kruskal-Wallis Rank Sum	3208.0	1577.5	1562.5	1477.0	1463.0	1716.0	3711.0	2411.5	2405.5	2404.0	2374.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	9	8	7	6
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	7	7	7	7
5	Average	17.18%	6.62%	6.54%	5.79%	5.71%	7.75%	24.10%	10.35%	10.40%	10.43%	10.51%
	Rank	10	4	3	2	1	5	11	6	7	8	9
	Geometric Mean	16.17%	4.28%	4.04%	3.90%	3.67%	5.91%	23.27%	9.84%	9.92%	10.01%	10.08%
	Rank	10	4	3	2	1	5	11	6	7	8	9
	Average Rank by Series	9.20	4.08	3.13	3.35	2.95	5.05	10.85	6.78	6.58	6.95	7.10
	Rank of Average Rank	10	4	2	3	1	5	11	7	6	8	9
	Kruskal-Wallis Rank Sum	3469.0	1449.5	1415.5	1285.0	1291.0	1694.0	4038.0	2373.5	2397.5	2441.0	2456.0
	Rank of K-W Rank Sum	10	4	3	1	2	5	11	6	7	8	9
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	8	7	7	8
10	Average	21.35%	8.65%	8.56%	7.80%	7.31%	10.24%	27.72%	11.80%	11.96%	12.05%	11.87%
	Rank	10	4	3	2	1	5	11	6	8	9	7
	Geometric Mean	19.44%	5.76%	5.72%	5.44%	4.99%	8.08%	26.48%	10.76%	11.03%	11.32%	11.02%
	Rank	10	4	3	2	1	5	11	6	8	9	7
	Average Rank by Series	8.50	4.63	3.98	3.55	3.10	5.95	10.70	6.63	6.38	6.50	6.10
	Rank of Average Rank	10	4	3	2	1	5	11	9	7	8	6
	Kruskal-Wallis Rank Sum	3403.0	1596.5	1565.5	1473.0	1411.0	1869.0	3946.0	2173.5	2253.5	2364.0	2255.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	6	7	9	8
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	9	8	10	9
15	Average	24.62%	9.33%	9.17%	8.82%	7.43%	11.86%	29.68%	12.25%	12.35%	12.64%	11.88%
	Rank	10	4	3	2	1	5	11	7	8	9	6
	Geometric Mean	19.82%	6.28%	6.28%	5.87%	5.09%	9.31%	25.47%	11.30%	11.47%	11.76%	10.92%
	Rank	10	3	4	2	1	5	11	7	8	9	6
	Average Rank by Series	8.00	4.48	4.03	4.05	2.95	5.90	9.85	6.93	6.63	6.80	6.40
	Rank of Average Rank	10	4	2	3	1	5	11	9	7	8	6
	Kruskal-Wallis Rank Sum	3178.0	1634.5	1620.5	1563.0	1350.0	2065.0	3668.0	2303.5	2341.5	2362.0	2224.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	7	8	9	6
	K-W Multi-Comparison Count*	10	8	8	8	10	10	10	7	8	8	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

No Change

Period: Scenario 6 Root Mean Square Error — Table: 6-4

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	201.81	52.16	50.14	48.36	48.54	63.34	290.18	141.00	141.39	137.86	138.55
	Rank	10	4	3	1	2	5	11	8	9	6	7
	Average Rank by Series	9.45	3.73	3.23	2.55	2.50	5.55	11.00	7.33	7.18	6.60	6.90
	Rank of Average Rank	10	4	3	2	1	5	11	9	8	6	7
5	Geometric Mean	308.06	86.73	82.17	78.90	73.97	118.36	470.79	216.90	218.13	218.76	224.07
	Rank	10	4	3	2	1	5	11	6	7	8	9
	Average Rank by Series	9.15	4.18	3.38	3.60	2.60	5.50	10.80	6.53	6.38	6.80	7.10
	Rank of Average Rank	10	4	2	3	1	5	11	7	6	8	9
10	Geometric Mean	390.87	122.13	121.06	113.93	103.57	161.95	546.11	248.24	251.49	254.44	253.69
	Rank	10	4	3	2	1	5	11	6	7	9	8
	Average Rank by Series	8.25	4.68	4.23	3.55	3.10	5.75	10.50	6.48	6.43	6.60	6.45
	Rank of Average Rank	10	4	3	2	1	5	11	8	6	9	7
15	Geometric Mean	421.31	132.29	133.01	127.57	108.60	188.49	567.46	254.52	258.69	268.08	252.26
	Rank	10	3	4	2	1	5	11	7	8	9	6
	Average Rank by Series	7.85	4.58	4.08	3.90	2.85	5.60	10.05	6.88	6.68	7.05	6.50
	Rank of Average Rank	10	4	3	2	1	5	11	8	7	9	6

No Change

Period: Scenario 6 Geometric Root Mean Square Error — Table: 6-5

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	138.93	29.28	27.81	26.81	26.63	40.23	183.36	58.87	55.38	64.69	60.19
	Rank	10	4	3	2	1	5	11	7	6	9	8
	Average Rank by Series	9.30	3.78	3.63	2.70	2.50	5.50	10.95	6.93	6.28	7.70	6.75
	Rank of Average Rank	10	4	3	2	1	5	11	8	6	9	7
5	Geometric Mean	211.18	51.19	47.98	48.21	45.72	71.91	268.47	91.35	90.56	94.83	90.25
	Rank	10	4	2	3	1	5	11	8	7	9	6
	Average Rank by Series	8.50	4.18	3.43	3.60	3.30	5.30	10.60	6.83	6.63	7.20	6.45
	Rank of Average Rank	10	4	2	3	1	5	11	8	7	9	6
10	Geometric Mean	246.00	67.75	71.80	70.83	63.62	115.33	327.98	101.41	114.36	121.52	104.19
	Rank	10	2	4	3	1	8	11	5	7	9	6
	Average Rank by Series	7.85	4.225	4.575	4.35	3.55	6.35	10.2	5.725	6.125	7.05	6
	Rank of Average Rank	10	2	4	3	1	8	11	5	7	9	6
15	Geometric Mean	242.87	89.08	88.09	79.12	71.57	145.28	284.51	130.59	132.29	129.05	123.54
	Rank	10	4	3	2	1	9	11	7	8	6	5
	Average Rank by Series	7.35	4.83	4.58	4.15	3.20	7.05	8.60	6.78	6.78	6.45	6.25
	Rank of Average Rank	10	4	3	2	1	9	11	7	7	6	5

No Change	Average Rank of Absolute Error			Table	6- 6
Perlod	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	75.92	19	0.0000	
	KRUSKAL-WALLIS	145.27	10	0.0000	
5	RANK ANOVA	47.14	19	0.0003	
	KRUSKAL-WALLIS	93.99	10	0.0000	
10	RANK ANOVA	32.00	19	0.0312	
	KRUSKAL-WALLIS	69.65	10	0.0000	
15	RANK ANOVA	25.22	19	0.1533	
	KRUSKAL-WALLIS	49.34	10	0.0000	

No Change		Symmetry Adjusted MAPE		Table	6- 7
	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	86.31	19	0.0000	
	KRUSKAL-WALLIS	63.24	10	0.0000	
5	RANK ANOVA	63.39	19	0.0000	
	KRUSKAL-WALLIS	91.68	10	0.0000	
10	RANK ANOVA	45.14	19	0.0007	
	KRUSKAL-WALLIS	69.50	10	0.0000	
15	RANK ANOVA	35.63	19	0.0117	
	KRUSKAL-WALLIS	51.42	10	0.0000	

No Change		Range of Percent Error		Table	6- 8
	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	54.58	19	0.0000	
	KRUSKAL-WALLIS	48.41	10	0.0000	
5	RANK ANOVA	58.00	19	0.0000	
	KRUSKAL-WALLIS	97.78	10	0.0000	
	RANK ANOVA	50.99	19	0.0001	
10	KRUSKAL-WALLIS	90.30	10	0.0000	
	RANK ANOVA	41.09	19	0.0023	
15	KRUSKAL-WALLIS	86.89	10	0.0000	

No Change	Geometric Root Mean Square Error			Table	6- 9
	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	77.21	19	0.0000	
5	RANK ANOVA	56.29	19	0.0000	
10	RANK ANOVA	38.28	19	0.0055	
15	RANK ANOVA	27.41	19	0.0955	

	No Change	Log Mean Square Error Ratio			Table 6- 10
Perlod	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	77.21	19	0.0000	
	KRUSKAL-WALLIS	125.57	10	0.0000	
5	RANK ANOVA	56.29	19	0.0000	
	KRUSKAL-WALLIS	72.35	10	0.0000	
10	RANK ANOVA	38.28	19	0.0055	
	KRUSKAL-WALLIS	50.92	10	0.0000	
15	RANK ANOVA	27.41	19	0.0955	
	KRUSKAL-WALLIS	34.59	10	0.0001	

	No Change	Mean Absolute Percent Error			Table 6- 11
	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	87.85	19	0.0000	
	KRUSKAL-WALLIS	68.50	10	0.0000	
5	RANK ANOVA	69.16	19	0.0000	
	KRUSKAL-WALLIS	102.16	10	0.0000	
10	RANK ANOVA	52.05	19	0.0001	
	KRUSKAL-WALLIS	80.90	10	0.0000	
15	RANK ANOVA	42.36	19	0.0016	
	KRUSKAL-WALLIS	61.33	10	0.0000	

	No Change	Median Absolute Percent Error			Table 6- 12
	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	79.00	19	0.0000	
	KRUSKAL-WALLIS	54.98	10	0.0000	
5	RANK ANOVA	43.56	19	0.0011	
	KRUSKAL-WALLIS	60.15	10	0.0000	
10	RANK ANOVA	34.11	19	0.0178	
	KRUSKAL-WALLIS	54.41	10	0.0000	
15	RANK ANOVA	18.58	19	0.4841	
	KRUSKAL-WALLIS	35.55	10	0.0001	

	No Change	Root Mean Square Error			Table 6- 13
	Scenario 6	Chi Square	DF	p Value	
1	RANK ANOVA	82.07	19	0.0000	
5	RANK ANOVA	66.05	19	0.0000	
10	RANK ANOVA	48.02	19	0.0003	
15	RANK ANOVA	44.87	19	0.0007	

Variance Shift Table: 7-1

Period: Scenario 7 Average Rank of Absolute Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	7.56	5.39	5.19	4.64	4.69	5.74	8.72	6.12	6.12	5.87	5.85
	Rank	10	4	3	1	2	5	11	8	8	7	6
	Geometric Mean	7.36	5.27	5.08	4.53	4.61	5.61	8.57	6.09	6.09	5.76	5.75
	Rank	10	4	3	1	2	5	11	8	8	7	6
	Average Rank by Series	8.00	4.85	4.23	2.98	3.28	5.88	10.50	6.95	6.95	5.85	6.15
	Rank of Average Rank	10	4	3	1	2	6	11	8	8	5	7
	Kruskal-Wallis Rank Sum	3273.0	1738.0	1521.5	1031.0	1035.5	2157.0	3955.5	2540.0	2540.0	2177.0	2239.5
	Rank of K-W Rank Sum	10	4	3	1	2	5	11	8	8	6	7
	K-W Multi-Comparison Count*	10	10	10	9	9	9	10	10	10	8	9
5	Average	7.24	5.36	5.12	5.39	4.72	5.98	8.12	5.85	5.81	6.45	5.96
	Rank	10	3	2	4	1	8	11	6	5	9	7
	Geometric Mean	7.04	5.21	4.97	5.19	4.52	5.68	7.96	5.75	5.74	6.34	5.88
	Rank	10	4	2	3	1	5	11	7	6	9	8
	Average Rank by Series	7.68	5.00	4.43	4.95	3.50	6.88	9.60	5.98	5.38	7.03	5.60
	Rank of Average Rank	10	4	2	3	1	8	11	7	5	9	6
	Kruskal-Wallis Rank Sum	3136.5	1701.5	1504.5	1707.5	1191.5	2367.5	3613.5	2151.0	2103.5	2652.5	2180.5
	Rank of K-W Rank Sum	10	3	2	4	1	8	11	6	5	9	7
	K-W Multi-Comparison Count*	10	9	10	9	10	10	10	8	8	10	8
10	Average	7.35	5.42	5.22	5.06	4.55	5.89	8.17	6.05	6.05	6.30	5.87
	Rank	10	4	3	2	1	6	11	7	7	9	5
	Geometric Mean	7.05	5.25	5.04	4.80	4.34	5.53	7.91	6.00	6.00	6.12	5.77
	Rank	10	4	3	2	1	5	11	7	7	9	6
	Average Rank by Series	7.65	5.20	4.88	4.18	3.25	6.85	9.13	6.33	6.33	6.15	5.65
	Rank of Average Rank	10	4	3	2	1	9	11	7	7	6	5
	Kruskal-Wallis Rank Sum	3029.5	1801.0	1653.0	1483.5	1161.0	2321.5	3451.0	2385.5	2385.5	2410.5	2170.0
	Rank of K-W Rank Sum	10	4	3	2	1	6	11	7	7	9	5
	K-W Multi-Comparison Count*	10	10	10	10	10	9	10	7	7	8	10
15	Average	7.33	5.47	5.20	5.25	4.54	6.18	8.08	6.10	5.96	6.18	5.72
	Rank	10	4	2	3	1	9	11	7	6	8	5
	Geometric Mean	7.01	5.25	4.91	4.88	4.36	5.80	7.79	6.00	5.89	5.98	5.63
	Rank	10	4	3	2	1	6	11	9	7	8	5
	Average Rank by Series	7.50	5.20	4.75	4.60	3.45	6.90	8.90	6.45	6.30	6.38	5.58
	Rank of Average Rank	10	4	3	2	1	9	11	8	6	7	5
	Kruskal-Wallis Rank Sum	2967.0	1852.0	1697.0	1701.0	1145.0	2486.0	3371.5	2388.0	2287.5	2336.5	2078.5
	Rank of K-W Rank Sum	10	4	2	3	1	9	11	8	6	7	5
	K-W Multi-Comparison Count*	10	10	9	9	10	10	10	9	9	8	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Variance Shift

Period: Scenario 7 Range of Percent Error Table: 7-2

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	34.61%	29.91%	29.81%	24.16%	25.10%	33.08%	40.29%	34.95%	35.02%	28.22%	29.58%
	Rank	8	6	5	1	2	7	11	9	10	3	4
	Geometric Mean	22.20%	10.59%	10.19%	10.89%	10.40%	12.02%	31.36%	20.23%	20.42%	18.65%	19.29%
	Rank	10	3	1	4	2	5	11	8	9	6	7
	Average Rank by Series	7.85	4.18	3.63	4.25	3.70	5.90	9.80	6.98	7.13	6.25	6.35
	Rank of Average Rank	10	3	1	4	2	5	11	8	9	6	7
	Kruskal-Wallis Rank Sum	2554.0	1750.5	1722.5	1865.0	1828.0	1868.0	3060.0	2436.5	2455.5	2361.0	2409.0
	Rank of K-W Rank Sum	10	2	1	4	3	5	11	8	9	6	7
	K-W Multi-Comparison Count*	10	8	9	8	7	8	10	7	8	8	7
5	Average	41.92%	30.84%	31.19%	28.73%	24.32%	34.72%	60.46%	40.74%	40.77%	41.11%	37.02%
	Rank	10	3	4	2	1	5	11	7	8	9	6
	Geometric Mean	31.90%	14.64%	15.61%	14.55%	12.85%	16.25%	52.70%	31.06%	31.09%	29.44%	28.59%
	Rank	10	3	4	2	1	5	11	8	9	7	6
	Average Rank by Series	7.58	4.18	4.50	4.23	2.75	5.30	10.48	6.83	6.85	7.18	6.15
	Rank of Average Rank	10	2	4	3	1	5	11	7	8	9	6
	Kruskal-Wallis Rank Sum	2605.5	1666.5	1713.0	1617.5	1503.0	1773.0	3427.5	2546.5	2547.0	2471.5	2439.0
	Rank of K-W Rank Sum	10	3	4	2	1	5	11	8	9	7	6
	K-W Multi-Comparison Count*	9	8	8	9	10	9	10	8	7	7	9
10	Average	39.60%	28.55%	28.52%	25.27%	22.62%	30.76%	51.47%	36.66%	36.31%	32.25%	31.61%
	Rank	10	4	3	2	1	5	11	9	8	7	6
	Geometric Mean	35.05%	16.40%	17.16%	17.16%	15.21%	16.10%	48.48%	30.17%	29.95%	28.97%	28.73%
	Rank	10	3	5	4	1	2	11	9	8	7	6
	Average Rank by Series	7.35	4.35	4.10	4.13	3.13	4.60	9.70	7.65	7.40	6.73	6.88
	Rank of Average Rank	8	4	2	3	1	5	11	10	9	6	7
	Kruskal-Wallis Rank Sum	2819.0	1678.0	1682.0	1752.5	1515.5	1691.0	3497.0	2462.0	2435.0	2395.5	2382.5
	Rank of K-W Rank Sum	10	2	3	5	1	4	11	9	8	7	6
	K-W Multi-Comparison Count*	10	7	7	7	10	7	10	7	7	7	7
15	Average	46.92%	23.36%	23.87%	24.38%	21.43%	24.81%	59.28%	32.29%	32.40%	36.44%	34.76%
	Rank	10	2	3	4	1	5	11	6	7	9	8
	Geometric Mean	39.30%	15.62%	17.06%	16.23%	14.78%	14.67%	54.88%	28.44%	28.72%	31.73%	31.29%
	Rank	10	3	5	4	2	1	11	6	7	9	8
	Average Rank by Series	7.85	4.48	4.48	4.15	3.68	5.00	9.45	6.38	6.38	7.30	6.88
	Rank of Average Rank	10	3	3	2	1	5	11	6	6	9	8
	Kruskal-Wallis Rank Sum	2961.0	1668.5	1701.5	1727.0	1595.5	1638.0	3619.0	2239.5	2266.5	2448.0	2445.5
	Rank of K-W Rank Sum	10	3	4	5	1	2	11	6	7	9	8
	K-W Multi-Comparison Count*	10	6	7	8	8	7	10	9	9	9	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Variance Shift Table: 7-3

Period: Scenario 7 Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	16.81%	12.42%	12.40%	10.52%	10.58%	13.15%	20.65%	14.22%	14.22%	12.69%	12.50%
	Rank	10	4	3	1	2	7	11	9	8	6	5
	Geometric Mean	12.93%	5.69%	5.62%	5.14%	5.13%	6.23%	17.67%	9.66%	9.64%	9.06%	9.11%
	Rank	10	4	3	2	1	5	11	9	8	6	7
	Average Rank by Series	8.20	4.28	4.18	2.70	2.70	6.00	10.55	7.68	7.28	6.00	6.45
	Rank of Average Rank	10	4	3	1	1	5	11	9	8	5	7
	Kruskal-Wallis Rank Sum	2754.0	1875.5	1867.5	1681.0	1739.0	1997.0	3164.0	2353.5	2342.5	2262.0	2274.0
	Rank of K-W Rank Sum	10	4	3	1	2	5	11	9	8	6	7
	K-W Multi-Comparison Count*	10	9	9	9	9	10	10	8	7	8	7
5	Average	20.08%	13.08%	12.97%	12.91%	11.42%	13.83%	26.48%	16.00%	16.07%	16.38%	15.14%
	Rank	10	4	3	2	1	5	11	7	8	9	6
	Geometric Mean	17.72%	7.74%	7.68%	7.78%	7.07%	8.94%	24.55%	13.00%	13.19%	13.75%	13.04%
	Rank	10	3	2	4	1	5	11	6	8	9	7
	Average Rank by Series	7.85	4.48	4.18	4.25	3.40	6.15	9.90	6.48	6.23	6.75	6.35
	Rank of Average Rank	10	4	2	3	1	5	11	8	6	9	7
	Kruskal-Wallis Rank Sum	2849.0	1774.5	1747.5	1747.0	1660.0	1816.0	3444.0	2278.5	2312.5	2377.0	2304.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	6	8	9	7
	K-W Multi-Comparison Count*	10	7	7	7	10	7	10	8	7	8	7
10	Average	24.14%	14.52%	14.48%	14.10%	12.44%	15.16%	29.91%	17.25%	17.46%	17.65%	16.63%
	Rank	10	4	3	2	1	5	11	7	8	9	6
	Geometric Mean	22.16%	9.03%	9.54%	9.39%	8.50%	10.44%	28.53%	14.16%	14.73%	15.08%	14.45%
	Rank	10	2	4	3	1	5	11	6	8	9	7
	Average Rank by Series	8.05	4.43	4.23	4.10	3.30	6.00	9.80	6.73	6.38	6.55	6.45
	Rank of Average Rank	10	4	3	2	1	5	11	9	6	8	7
	Kruskal-Wallis Rank Sum	3053.0	1738.5	1742.5	1762.0	1581.0	1826.0	3521.0	2223.5	2267.5	2339.0	2256.0
	Rank of K-W Rank Sum	10	2	3	4	1	5	11	6	8	9	7
	K-W Multi-Comparison Count*	10	8	8	7	10	9	10	8	7	9	8
15	Average	26.69%	13.47%	13.34%	13.83%	11.58%	14.47%	32.09%	16.78%	16.88%	17.70%	16.15%
	Rank	10	3	2	4	1	5	11	7	8	9	6
	Geometric Mean	23.45%	8.61%	9.17%	8.85%	7.68%	11.04%	29.67%	14.29%	14.72%	15.04%	14.26%
	Rank	10	2	4	3	1	5	11	7	8	9	6
	Average Rank by Series	8.15	4.58	4.08	3.85	3.05	5.60	9.50	7.13	6.73	6.90	6.45
	Rank of Average Rank	10	4	3	2	1	5	11	9	7	8	6
	Kruskal-Wallis Rank Sum	3092.0	1754.5	1751.5	1746.0	1526.0	1924.0	3545.0	2209.5	2258.5	2307.0	2196.0
	Rank of K-W Rank Sum	10	4	3	2	1	5	11	7	8	9	6
	K-W Multi-Comparison Count*	10	8	8	8	10	10	10	8	7	9	8

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Variance Shift

Period: Scenario 7 Root Mean Square Error Table: 7-4

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	235.41	113.44	111.04	107.70	105.41	122.60	324.18	194.58	195.10	184.78	188.09
	Rank	10	4	3	2	1	5	11	8	9	6	7
	Average Rank by Series	7.55	4.48	4.18	2.85	3.25	6.45	9.60	7.63	7.08	6.05	6.90
	Rank of Average Rank	9	4	3	1	2	6	11	10	8	5	7
5	Geometric Mean	329.07	156.98	157.13	156.25	142.00	181.61	480.20	278.10	280.43	280.09	269.93
	Rank	10	3	4	2	1	5	11	7	9	8	6
	Average Rank by Series	7.45	4.83	4.48	4.55	3.50	6.20	9.50	6.63	6.23	6.50	6.15
	Rank of Average Rank	10	4	2	3	1	6	11	9	7	8	5
10	Geometric Mean	437.18	198.09	208.45	203.94	184.31	223.56	573.69	321.19	328.10	331.71	316.66
	Rank	10	2	4	3	1	5	11	7	8	9	6
	Average Rank by Series	7.20	4.93	4.68	4.15	3.30	6.35	9.25	6.88	6.38	6.60	6.30
	Rank of Average Rank	10	4	3	2	1	6	11	9	7	8	5
15	Geometric Mean	482.66	192.48	206.01	199.05	172.02	241.17	614.30	321.19	328.77	344.31	323.00
	Rank	10	2	4	3	1	5	11	6	8	9	7
	Average Rank by Series	7.25	4.63	4.28	4.10	3.15	6.05	9.35	7.08	6.58	7.00	6.55
	Rank of Average Rank	10	4	3	2	1	5	11	9	7	8	6

Variance Shift

Period: Scenario 7 Geometric Root Mean Square Error Table: 7-5

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	153.97	69.06	70.72	54.32	56.25	75.38	205.86	96.24	97.74	84.76	83.41
	Rank	10	3	4	1	2	5	11	8	9	7	6
	Average Rank by Series	7.60	5.13	5.33	2.90	2.80	5.75	9.80	7.58	7.48	5.75	5.90
	Rank of Average Rank	10	3	4	2	1	5	11	9	8	5	7
5	Geometric Mean	214.00	91.44	87.59	97.89	88.09	103.96	268.02	122.86	123.53	141.90	127.71
	Rank	10	3	1	4	2	5	11	6	7	9	8
	Average Rank by Series	7.90	4.68	4.23	4.55	4.40	5.55	9.10	6.08	5.93	7.00	6.60
	Rank of Average Rank	10	4	1	3	2	5	11	7	6	9	8
10	Geometric Mean	301.37	112.67	112.60	117.50	103.22	140.63	365.53	158.97	160.27	178.32	161.00
	Rank	10	3	2	4	1	5	11	6	7	9	8
	Average Rank by Series	8	4.975	4.475	4.3	3.4	6.55	9.25	6.325	6.175	6.7	5.85
	Rank of Average Rank	10	4	3	2	1	8	11	7	6	9	5
15	Geometric Mean	314.65	119.15	120.93	113.28	98.54	149.23	385.52	172.51	175.22	166.48	152.31
	Rank	10	3	4	2	1	5	11	8	9	7	6
	Average Rank by Series	7.75	5.13	4.68	4.50	3.15	5.75	9.30	6.98	6.98	6.25	5.55
	Rank of Average Rank	10	4	3	2	1	6	11	8	8	7	5

Variance Shift	Average Rank of Absolute Error			Table 7- 6
Period	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	50.54	19	0.0001
	KRUSKAL-WALLIS	98.03	10	0.0000
5	RANK ANOVA	30.67	19	0.0438
	KRUSKAL-WALLIS	63.07	10	0.0000
10	RANK ANOVA	28.45	19	0.0752
	KRUSKAL-WALLIS	55.00	10	0.0000
15	RANK ANOVA	24.21	19	0.1881
	KRUSKAL-WALLIS	47.56	10	0.0000

Variance Shift		Symmetry Adjusted MAPE		Table 7- 7
	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	53.34	19	0.0000
	KRUSKAL-WALLIS	20.51	10	0.0248
5	RANK ANOVA	24.59	19	0.1744
	KRUSKAL-WALLIS	31.94	10	0.0004
10	RANK ANOVA	31.29	19	0.0375
	KRUSKAL-WALLIS	38.95	10	0.0000
15	RANK ANOVA	31.98	19	0.0314
	KRUSKAL-WALLIS	38.17	10	0.0000

Variance Shift		Range of Percent Error		Table 7- 8
	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	39.64	19	0.0036
	KRUSKAL-WALLIS	22.77	10	0.0116
5	RANK ANOVA	47.43	19	0.0003
	KRUSKAL-WALLIS	44.06	10	0.0000
	RANK ANOVA	43.76	19	0.0010
10	KRUSKAL-WALLIS	46.01	10	0.0000
	RANK ANOVA	34.39	19	0.0165
15	KRUSKAL-WALLIS	51.28	10	0.0000

Variance Shift	Geometric Root Mean Square Error			Table 7- 9
	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	45.11	19	0.0007
5	RANK ANOVA	25.89	19	0.1333
10	RANK ANOVA	30.23	19	0.0489
15	RANK ANOVA	30.79	19	0.0426

Variance Shift	Log Mean Square Error Ratio			Table 7-10
Period	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	45.11	19	0.0007
	KRUSKAL-WALLIS	58.51	10	0.0000
5	RANK ANOVA	25.89	19	0.1333
	KRUSKAL-WALLIS	43.41	10	0.0000
10	RANK ANOVA	30.23	19	0.0489
	KRUSKAL-WALLIS	40.62	10	0.0000
15	RANK ANOVA	30.79	19	0.0426
	KRUSKAL-WALLIS	34.53	10	0.0002

Variance Shift	Mean Absolute Percent Error			Table 7-11
	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	61.11	19	0.0000
	KRUSKAL-WALLIS	25.02	10	0.0053
5	RANK ANOVA	37.04	19	0.0078
	KRUSKAL-WALLIS	37.75	10	0.0000
10	RANK ANOVA	38.34	19	0.0054
	KRUSKAL-WALLIS	44.87	10	0.0000
15	RANK ANOVA	40.99	19	0.0024
	KRUSKAL-WALLIS	46.34	10	0.0000

Variance Shift	Median Absolute Percent Error			Table 7-12
	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	42.77	19	0.0014
	KRUSKAL-WALLIS	21.52	10	0.0177
5	RANK ANOVA	10.61	19	0.9362
	KRUSKAL-WALLIS	19.94	10	0.0299
10	RANK ANOVA	25.72	19	0.1381
	KRUSKAL-WALLIS	34.67	10	0.0001
15	RANK ANOVA	24.36	19	0.1827
	KRUSKAL-WALLIS	29.16	10	0.0012

Variance Shift	Root Mean Square Error			Table 7-13
	Scenario 7	Chi Square	DF	p Value
1	RANK ANOVA	45.58	19	0.0006
5	RANK ANOVA	28.80	19	0.0692
10	RANK ANOVA	28.74	19	0.0702
15	RANK ANOVA	34.01	19	0.0183

Period: Level Shift as Planned (N) Scenario 8 Table: 8 - 1

Average Rank of Absolute Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	3.50	6.74	6.87	7.86	7.16	6.05	5.51	5.25	5.25	6.04	5.86
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Geometric Mean	3.29	6.68	6.82	7.80	7.06	5.99	5.34	5.18	5.18	5.97	5.80
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Average Rank by Series	1.48	7.45	7.50	9.80	8.85	6.15	4.35	4.35	4.35	5.90	5.68
	Rank of Average Rank	1	8	9	11	10	7	2	2	2	6	5
	Kruskal-Wallis Rank Sum	437.5	2,883.5	3,037.5	3,772.0	3,175.5	2,212.0	1,700.0	1,437.5	1,437.5	2,220.0	1,952.5
	Rank of K-W Rank Sum	1	8	9	11	10	6	4	2	2	7	5
	K-W Multi-Comparison Count*	10	10	10	10	10	9	10	9	9	9	10
5	Average	3.91	6.38	6.38	7.26	7.09	5.56	6.04	5.67	5.67	6.03	6.01
	Rank	1	9	8	11	10	2	7	4	3	6	5
	Geometric Mean	3.51	6.33	6.29	7.18	6.92	5.37	5.81	5.36	5.42	5.79	5.90
	Rank	1	9	8	11	10	3	6	2	4	5	7
	Average Rank by Series	2.73	6.88	6.73	8.55	7.40	5.38	5.88	5.85	5.33	5.78	5.53
	Rank of Average Rank	1	9	8	11	10	3	7	6	2	5	4
	Kruskal-Wallis Rank Sum	889.0	2,477.0	2,460.5	3,210.5	2,900.0	1,853.0	2,110.0	2,008.5	1,998.5	2,256.0	2,147.0
	Rank of K-W Rank Sum	1	9	8	11	10	2	5	4	3	7	6
	K-W Multi-Comparison Count*	10	9	9	10	10	10	9	9	9	10	9
10	Average	3.76	6.50	6.63	7.14	6.71	5.71	5.71	6.06	6.06	5.96	5.82
	Rank	1	8	9	11	10	3	2	6	6	5	4
	Geometric Mean	3.40	6.41	6.55	7.04	6.55	5.44	5.51	5.82	5.82	5.72	5.71
	Rank	1	8	10	11	9	2	3	6	6	5	4
	Average Rank by Series	2.5	7	6.825	7.825	7.175	5.725	5.575	6.25	6.25	5.3	5.475
	Rank of Average Rank	1	9	8	11	10	5	4	6	6	2	3
	Kruskal-Wallis Rank Sum	809.0	2,565.5	2,652.0	3,076.0	2,704.5	2,014.5	1,920.0	2,274.5	2,274.5	2,075.5	1,967.0
	Rank of K-W Rank Sum	1	8	9	11	10	4	2	6	6	5	3
	K-W Multi-Comparison Count*	10	10	9	10	9	8	9	9	9	9	8
15	Average	3.99	6.50	6.73	6.90	6.49	6.02	5.68	6.03	6.21	6.00	5.45
	Rank	1	9	10	11	8	5	3	6	7	4	2
	Geometric Mean	3.68	6.40	6.68	6.68	6.24	5.62	5.44	5.60	5.92	5.68	5.14
	Rank	1	9	10	11	8	5	3	4	7	6	2
	Average Rank by Series	2.975	7.075	7.15	7.325	6.275	6.4	6.05	6.15	6.225	5.475	4.9
	Rank of Average Rank	1	9	10	11	7	8	4	5	6	3	2
	Kruskal-Wallis Rank Sum	917.0	2,518.5	2,688.5	2,832.0	2,478.0	2,270.0	1,946.5	2,305.5	2,387.0	2,157.5	1,809.5
	Rank of K-W Rank Sum	1	9	10	11	8	5	3	6	7	4	2
	K-W Multi-Comparison Count*	10	9	10	10	9	9	10	9	10	10	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level Shift as Planned (N) Scenario 8 Table: 8-2

Period: Range of Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	15.14%	24.73%	24.30%	24.71%	27.52%	24.50%	20.72%	18.76%	19.22%	20.24%	22.26%
	Rank	1	10	7	9	11	8	5	2	3	4	6
	Geometric Mean	6.72%	19.10%	18.46%	19.78%	22.12%	18.79%	17.82%	11.28%	12.47%	15.58%	16.88%
	Rank	1	9	7	10	11	8	6	2	3	4	5
	Average Rank by Series	2.53	7.58	6.98	7.23	8.48	7.55	6.68	4.03	4.03	5.18	5.78
	Rank of Average Rank	1	10	7	8	11	9	6	2	2	4	5
	Kruskal-Wallis Rank Sum	1,470.5	2,511.5	2,444.5	2,580.5	2,733.5	2,458.0	2,421.5	1,634.5	1,677.5	2,125.5	2,252.5
	Rank of K-W Rank Sum	1	9	7	10	11	8	6	2	3	4	5
	K-W Multi-Comparison Count*	10	7	7	9	10	7	8	9	9	10	10
5	Average	14.29%	36.27%	36.98%	35.63%	37.72%	36.12%	39.24%	25.49%	27.88%	30.86%	32.45%
	Rank	1	8	9	6	10	7	11	2	3	4	5
	Geometric Mean	8.88%	33.40%	33.78%	33.69%	35.56%	31.64%	38.75%	18.21%	21.91%	26.84%	26.86%
	Rank	1	7	9	8	10	6	11	2	3	4	5
	Average Rank by Series	2.13	6.83	6.73	6.53	7.58	6.70	8.33	4.78	4.78	5.48	6.18
	Rank of Average Rank	1	9	8	6	10	7	11	2	2	4	5
	Kruskal-Wallis Rank Sum	861.5	2,476.5	2,477.5	2,601.5	2,700.5	2,361.0	3,132.5	1,576.5	1,707.5	2,140.5	2,274.5
	Rank of K-W Rank Sum	1	7	8	9	10	6	11	2	3	4	5
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	10	10	10	10
10	Average	16.17%	40.75%	41.93%	50.22%	45.54%	31.54%	38.94%	32.70%	37.52%	44.55%	37.31%
	Rank	1	7	8	11	10	2	6	3	5	9	4
	Geometric Mean	11.20%	36.75%	37.48%	46.09%	42.86%	28.31%	37.63%	22.12%	27.40%	35.28%	30.25%
	Rank	1	7	8	11	10	4	9	2	3	6	5
	Average Rank by Series	2.45	6.65	6.3	7.775	7.3	5.25	6.8	5.35	5.45	6.775	5.9
	Rank of Average Rank	1	7	6	11	10	2	9	3	4	8	5
	Kruskal-Wallis Rank Sum	866.0	2,438.0	2,454.0	2,932.5	2,838.0	1,804.0	2,495.0	1,839.0	2,087.0	2,408.5	2,148.0
	Rank of K-W Rank Sum	1	7	8	11	10	2	9	3	4	6	5
	K-W Multi-Comparison Count*	10	7	7	10	10	9	8	9	9	8	9
15	Average	15.98%	44.71%	50.02%	64.09%	51.75%	27.28%	37.48%	38.61%	45.42%	57.68%	45.26%
	Rank	1	5	8	11	9	2	3	4	7	10	6
	Geometric Mean	10.67%	38.60%	42.36%	53.84%	46.65%	25.91%	36.39%	23.30%	28.63%	40.23%	34.54%
	Rank	1	7	9	11	10	3	6	2	4	8	5
	Average Rank by Series	2.5	6.8	6.6	8.125	7.575	4.7	6.75	5.4	5.45	6.325	5.775
	Rank of Average Rank	1	9	7	11	10	2	8	3	4	6	5
	Kruskal-Wallis Rank Sum	908.0	2,414.0	2,590.0	3,035.5	2,811.5	1,593.0	2,438.0	1,824.0	2,107.0	2,448.5	2,140.5
	Rank of K-W Rank Sum	1	6	9	11	10	2	7	3	4	8	5
	K-W Multi-Comparison Count*	10	8	10	10	10	10	8	10	9	8	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level Shift as Planned (N) Scenario 8 Table: 8-3

Period: Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	6.13%	11.70%	11.81%	14.56%	13.60%	10.96%	9.51%	8.45%	8.61%	10.88%	10.11%
	Rank	1	8	9	11	10	7	4	2	3	6	5
	Geometric Mean	3.19%	10.26%	10.44%	12.39%	11.35%	9.04%	8.52%	5.98%	6.40%	8.50%	8.01%
	Rank	1	8	9	11	10	7	6	2	3	5	4
	Average Rank by Series	1.40	7.88	7.98	9.50	9.30	6.40	5.45	3.63	3.78	5.50	5.20
	Rank of Average Rank	1	8	9	11	10	7	5	2	3	6	4
	Kruskal-Wallis Rank Sum	1,291.0	2,613.5	2,661.5	2,892.0	2,761.0	2,324.0	2,215.0	1,629.5	1,655.5	2,188.0	2,079.0
	Rank of K-W Rank Sum	1	8	9	11	10	7	6	2	3	5	4
	K-W Multi-Comparison Count*	10	9	9	10	10	10	9	9	9	9	10
5	Average	7.76%	17.28%	18.11%	22.08%	20.59%	14.15%	16.89%	13.01%	14.32%	16.72%	14.94%
	Rank	1	8	9	11	10	3	7	2	4	6	5
	Geometric Mean	5.06%	16.18%	17.01%	20.68%	19.38%	12.99%	16.09%	9.51%	11.45%	14.13%	13.08%
	Rank	1	8	9	11	10	4	7	2	3	6	5
	Average Rank by Series	2.35	7.13	7.28	9.10	8.25	5.05	6.95	4.53	4.58	5.70	5.10
	Rank of Average Rank	1	8	9	11	10	4	7	2	3	6	5
	Kruskal-Wallis Rank Sum	942.0	2,419.5	2,567.5	3,113.0	2,922.0	1,806.0	2,421.0	1,783.5	1,980.5	2,265.0	2,090.0
	Rank of K-W Rank Sum	1	7	9	11	10	3	8	2	4	6	5
	K-W Multi-Comparison Count*	10	9	10	10	10	9	9	9	10	10	10
10	Average	9.42%	21.79%	24.04%	28.84%	24.87%	15.43%	17.71%	18.86%	21.44%	25.09%	20.58%
	Rank	1	7	8	11	9	2	3	4	6	10	5
	Geometric Mean	6.61%	19.10%	21.18%	26.01%	22.86%	13.99%	16.56%	13.10%	15.90%	19.24%	16.87%
	Rank	1	7	9	11	10	3	5	2	4	8	6
	Average Rank by Series	2.5	6.775	6.875	8.5	7.6	5.15	6.35	5.675	5.775	5.5	5.3
	Rank of Average Rank	1	8	9	11	10	2	7	5	6	4	3
	Kruskal-Wallis Rank Sum	973.0	2,308.5	2,545.5	3,047.0	2,782.0	1,656.0	2,077.0	1,959.5	2,253.5	2,493.0	2,215.0
	Rank of K-W Rank Sum	1	7	9	11	10	2	4	3	6	8	5
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	10	8	9	9
15	Average	9.45%	25.59%	28.82%	35.15%	28.10%	16.45%	17.36%	22.93%	26.68%	32.37%	24.05%
	Rank	1	6	9	11	8	2	3	4	7	10	5
	Geometric Mean	6.78%	21.22%	23.83%	28.67%	23.89%	15.17%	16.19%	14.30%	17.29%	21.77%	16.93%
	Rank	1	7	9	11	10	3	4	2	6	8	5
	Average Rank by Series	2.4	7.125	7.075	7.8	7.1	6.15	6.4	5.575	5.875	5.75	4.75
	Rank of Average Rank	1	10	8	11	9	6	7	3	5	4	2
	Kruskal-Wallis Rank Sum	971.0	2,431.5	2,643.5	2,946.0	2,683.0	1,830.0	2,015.0	1,982.5	2,293.5	2,469.0	2,045.0
	Rank of K-W Rank Sum	1	7	9	11	10	2	4	3	6	8	5
	K-W Multi-Comparison Count*	10	9	9	10	9	10	8	8	10	9	8

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level Shift as Planned (N) Scenario 8 Table: 8-4

Period: Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	48.94	151.69	151.51	171.88	165.98	139.76	140.50	87.84	94.87	118.59	115.65
	Rank	1	9	8	11	10	6	7	2	3	5	4
	Average Rank by Series	1.55	8.08	7.98	9.25	8.85	6.65	6.50	3.88	3.83	4.80	4.65
	Rank of Average Rank	1	9	8	11	10	7	6	3	2	5	4
5	Geometric Mean	73.65	242.82	250.02	289.98	275.29	202.89	268.54	140.80	169.40	204.31	189.24
	Rank	1	7	8	11	10	5	9	2	3	6	4
	Average Rank by Series	2.35	6.875	6.975	8.65	8.25	4.75	7.6	4.825	4.775	5.6	5.35
	Rank of Average Rank	1	7	8	11	10	2	9	4	3	6	5
10	Geometric Mean	103.23	305.77	331.42	394.84	347.21	225.76	281.24	202.74	246.74	296.84	256.80
	Rank	1	8	9	11	10	3	6	2	4	7	5
	Average Rank by Series	2.6	6.925	6.875	7.85	7.15	5.2	6.5	5.875	5.975	5.8	5.25
	Rank of Average Rank	1	9	8	11	10	2	7	5	6	4	3
15	Geometric Mean	109.66	340.92	380.72	457.01	383.62	238.26	287.06	229.77	278.08	349.06	279.78
	Rank	1	7	9	11	10	3	6	2	4	8	5
	Average Rank by Series	2.45	7.075	6.975	7.75	6.8	5.45	6.9	5.575	5.925	5.9	5.2
	Rank of Average Rank	1	10	9	11	7	3	8	4	6	5	2

Level Shift as Planned (N) Scenario 8 Table: 8-5

Period: Geometric Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	26.32	87.33	92.79	121.68	94.17	72.33	57.88	48.64	50.67	73.43	64.81
	Rank	1	8	9	11	10	6	4	2	3	7	5
	Average Rank by Series	1.65	7.075	7.275	9.7	8.25	6.65	4.85	4.275	4.525	6.3	5.45
	Rank of Average Rank	1	8	9	11	10	7	4	2	3	6	5
5	Geometric Mean	47.09	128.19	136.27	181.21	169.58	95.17	104.13	79.94	95.60	116.44	108.18
	Rank	1	8	9	11	10	3	5	2	4	7	6
	Average Rank by Series	2.95	7.175	6.925	9.05	8.1	4.75	6.45	4.925	4.925	5.7	5.05
	Rank of Average Rank	1	9	8	11	10	2	7	3	3	6	5
10	Geometric Mean	55.46	167.70	195.37	223.20	205.65	123.47	110.85	124.13	146.71	161.54	154.71
	Rank	1	8	9	11	10	3	2	4	5	7	6
	Average Rank by Series	2.65	6.525	7.425	7.9	7.6	5.7	5.6	5.825	6.125	5.2	5.45
	Rank of Average Rank	1	8	9	11	10	5	4	6	7	2	3
15	Geometric Mean	67.91	186.96	210.97	255.38	218.53	150.33	126.55	134.56	157.61	199.60	156.23
	Rank	1	7	9	11	10	4	2	3	6	8	5
	Average Rank by Series	2.55	6.475	6.925	8.2	7.3	6.55	6	5.175	5.975	5.75	5.1
	Rank of Average Rank	1	7	9	11	10	8	6	3	5	4	2

Level Shift as Planned (N)		Average Rank of Absolute Error		
	Scenario 8	Table:	8-6	
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	57.95	19	0.0000
	KRUSKAL-WALLIS	112.37	10	0.0000
5	RANK ANOVA	23.06	19	0.2346
	KRUSKAL-WALLIS	44.24	10	0.0000
10	RANK ANOVA	21.20	19	0.3258
	KRUSKAL-WALLIS	43.00	10	0.0000
15	RANK ANOVA	16.35	19	0.6337
	KRUSKAL-WALLIS	33.71	10	0.0002

Level Shift as Planned (N)		Log Mean Squared Error Ratio		
	Scenario 8	Table:	8	-10
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	50.18	19	0.0001
	KRUSKAL-WALLIS	83.95	10	0.0000
5	RANK ANOVA	32.08	19	0.0306
	KRUSKAL-WALLIS	42.85	10	0.0000
10	RANK ANOVA	22.33	19	0.2681
	KRUSKAL-WALLIS	32.12	10	0.0004
15	RANK ANOVA	22.78	19	0.2471
	KRUSKAL-WALLIS	25.90	10	0.0039

Level Shift as Planned (N)		Range of Percent Error		
	Scenario 8	Table:	8-7	
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	36.45	19	0.0093
	KRUSKAL-WALLIS	22.63	10	0.0122
5	RANK ANOVA	29.86	19	0.0536
	KRUSKAL-WALLIS	48.02	10	0.0000
10	RANK ANOVA	21.87	19	0.2907
	KRUSKAL-WALLIS	40.43	10	0.0000
15	RANK ANOVA	24.80	19	0.1671
	KRUSKAL-WALLIS	44.16	10	0.0000

Level Shift as Planned (N)		Median Absolute Percent Error		
	Scenario 8	Table:	8	-11
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	64.07	19	0.0000
	KRUSKAL-WALLIS	35.67	10	0.0001
5	RANK ANOVA	23.98	19	0.1968
	KRUSKAL-WALLIS	35.24	10	0.0001
10	RANK ANOVA	19.78	19	0.4079
	KRUSKAL-WALLIS	31.90	10	0.0004
15	RANK ANOVA	20.15	19	0.3854
	KRUSKAL-WALLIS	28.59	10	0.0014

Level Shift as Planned (N)		Symmetry Adjusted MAPE		
	Scenario 8	Table:	8-8	
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	66.71	19	0.0000
	KRUSKAL-WALLIS	30.18	10	0.0008
5	RANK ANOVA	31.29	19	0.0376
	KRUSKAL-WALLIS	38.70	10	0.0000
10	RANK ANOVA	19.30	19	0.4375
	KRUSKAL-WALLIS	33.74	10	0.0002
15	RANK ANOVA	21.82	19	0.2933
	KRUSKAL-WALLIS	29.74	10	0.0009

Level Shift as Planned (N)		Mean Absolute Percent Error		
	Scenario 8	Table:	8	-12
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	66.72	19	0.0000
	KRUSKAL-WALLIS	32.77	10	0.0003
5	RANK ANOVA	39.85	19	0.0034
	KRUSKAL-WALLIS	43.95	10	0.0000
10	RANK ANOVA	25.69	19	0.1391
	KRUSKAL-WALLIS	38.87	10	0.0000
15	RANK ANOVA	23.26	19	0.2259
	KRUSKAL-WALLIS	35.45	10	0.0001

Level Shift as Planned (N)		Geometric Root Mean Squared Error		
	Scenario 8	Table:	8-9	
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	50.18	19	0.0001
5	RANK ANOVA	32.08	19	0.0306
10	RANK ANOVA	22.33	19	0.2681
15	RANK ANOVA	22.78	19	0.2471

Level Shift as Planned (N)			Root Mean Squared Error		
	Scenario 8		Table:	8	-13
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	62.79	19	0.0000
	5	RANK ANOVA	36.64	19	0.0088
	10	RANK ANOVA	20.77	19	0.3496
	15	RANK ANOVA	21.73	19	0.2978

Period: Level and Trend Shift (N) Scenario 9 Table: 9-1

Average Rank of Absolute Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	3.61	6.70	6.85	7.87	7.07	6.04	5.62	5.23	5.23	6.04	5.79
	Rank	1	8	9	11	10	6	4	2	2	7	5
	Geometric Mean	3.38	6.64	6.80	7.82	6.97	5.97	5.46	5.16	5.16	5.97	5.72
	Rank	1	8	9	11	10	7	4	2	2	6	5
	Average Rank by Series	1.48	7.40	7.48	9.88	8.65	6.38	4.75	4.30	4.30	5.88	5.35
	Rank of Average Rank	1	8	9	11	10	7	4	2	2	6	5
	Kruskal-Wallis Rank Sum	460.0	2,855.5	3,036.5	3,784.0	3,106.5	2,221.0	1,813.0	1,429.5	1,429.5	2,228.0	1,881.5
	Rank of K-W Rank Sum	1	8	9	11	10	6	4	2	2	7	5
	K-W Multi-Comparison Count*	10	10	9	10	9	9	9	9	9	9	9
5	Average	3.89	6.41	6.39	7.23	7.05	5.59	6.03	5.73	5.70	5.99	5.99
	Rank	1	9	8	11	10	2	7	4	3	6	5
	Geometric Mean	3.50	6.36	6.30	7.14	6.88	5.43	5.82	5.44	5.46	5.76	5.88
	Rank	1	9	8	11	10	2	6	3	4	5	7
	Average Rank by Series	2.68	7.05	6.88	8.55	7.13	5.25	6.00	5.80	5.30	5.78	5.60
	Rank of Average Rank	1	9	8	11	10	2	7	6	3	5	4
	Kruskal-Wallis Rank Sum	869.0	2,512.5	2,496.0	3,187.0	2,861.5	1,853.5	2,113.0	2,026.0	2,021.0	2,242.0	2,128.5
	Rank of K-W Rank Sum	1	9	8	11	10	2	5	4	3	7	6
	K-W Multi-Comparison Count*	10	9	9	10	10	10	9	9	9	10	9
10	Average	3.86	6.44	6.57	7.10	6.69	5.71	5.75	6.07	6.07	5.97	5.84
	Rank	1	8	9	11	10	2	3	6	6	5	4
	Geometric Mean	3.50	6.34	6.48	6.98	6.54	5.45	5.53	5.82	5.82	5.73	5.72
	Rank	1	8	9	11	10	2	3	6	6	5	4
	Average Rank by Series	2.875	6.875	6.8	7.75	7.1	5.575	5.775	6.125	6.125	5.325	5.45
	Rank of Average Rank	1	9	8	11	10	4	5	6	6	2	3
	Kruskal-Wallis Rank Sum	869.0	2,507.0	2,595.0	3,060.0	2,706.0	2,005.0	1,953.0	2,279.5	2,279.5	2,100.0	1,978.5
	Rank of K-W Rank Sum	1	8	9	11	10	4	2	6	6	5	3
	K-W Multi-Comparison Count*	10	10	10	10	10	8	8	9	9	10	8
15	Average	4.01	6.49	6.71	6.82	6.52	6.01	5.76	6.04	6.22	5.92	5.50
	Rank	1	8	10	11	9	5	3	6	7	4	2
	Geometric Mean	3.68	6.39	6.66	6.57	6.27	5.58	5.49	5.63	5.92	5.58	5.19
	Rank	1	9	11	10	8	5	3	6	7	4	2
	Average Rank by Series	3.025	7.075	7.025	7	6.25	6.475	6.175	6.2	6.325	5.525	4.925
	Rank of Average Rank	1	11	10	9	6	8	4	5	7	3	2
	Kruskal-Wallis Rank Sum	947.5	2,510.0	2,661.0	2,776.5	2,489.5	2,262.0	2,009.0	2,301.5	2,401.5	2,122.5	1,829.0
	Rank of K-W Rank Sum	1	9	10	11	8	5	3	6	7	4	2
	K-W Multi-Comparison Count*	10	9	10	10	9	9	10	9	10	10	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level and Trend Shift (N) Scenario 9 Table: 9-2

Period: Range of Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	15.16%	24.60%	24.18%	24.57%	27.41%	24.30%	20.76%	18.65%	19.13%	20.13%	22.17%
	Rank	1	10	7	9	11	8	5	2	3	4	6
	Geometric Mean	6.77%	19.00%	18.35%	19.62%	22.01%	18.62%	17.90%	11.17%	12.46%	15.54%	16.87%
	Rank	1	9	7	10	11	8	6	2	3	4	5
	Average Rank by Series	2.48	7.63	7.03	7.13	8.48	7.35	6.83	4.03	4.03	5.18	5.88
	Rank of Average Rank	1	10	7	8	11	9	6	2	2	4	5
	Kruskal-Wallis Rank Sum	1,470.5	2,506.5	2,446.5	2,567.5	2,733.5	2,442.0	2,431.5	1,638.5	1,685.5	2,122.5	2,265.5
	Rank of K-W Rank Sum	1	9	8	10	11	7	6	2	3	4	5
	K-W Multi-Comparison Count*	10	6	7	9	10	7	7	9	9	10	10
5	Average	14.12%	35.68%	36.43%	35.20%	37.46%	35.64%	39.14%	25.12%	27.53%	30.56%	32.18%
	Rank	1	8	9	6	10	7	11	2	3	4	5
	Geometric Mean	8.94%	33.00%	33.41%	33.39%	35.37%	31.47%	38.65%	18.13%	21.82%	26.68%	26.74%
	Rank	1	7	9	8	10	6	11	2	3	4	5
	Average Rank by Series	2.13	6.78	6.68	6.65	7.48	6.60	8.48	4.73	4.83	5.55	6.13
	Rank of Average Rank	1	9	8	7	10	6	11	2	3	4	5
	Kruskal-Wallis Rank Sum	842.5	2,477.5	2,482.5	2,616.0	2,698.5	2,357.0	3,139.5	1,567.5	1,706.5	2,157.0	2,265.5
	Rank of K-W Rank Sum	1	7	8	9	10	6	11	2	3	4	5
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	10	10	10	10
10	Average	15.82%	40.36%	41.85%	50.22%	45.39%	31.41%	39.12%	32.58%	37.38%	44.72%	37.24%
	Rank	1	7	8	11	10	2	6	3	5	9	4
	Geometric Mean	10.93%	36.35%	37.36%	46.15%	42.68%	28.20%	37.82%	22.13%	27.45%	35.52%	30.39%
	Rank	1	7	8	11	10	4	9	2	3	6	5
	Average Rank by Series	2.2	6.725	6.55	7.875	7.15	5.2	6.75	5.375	5.4	6.875	5.9
	Rank of Average Rank	1	7	6	11	10	2	8	3	4	9	5
	Kruskal-Wallis Rank Sum	836.0	2,420.5	2,458.0	2,965.5	2,819.0	1,777.0	2,511.0	1,831.5	2,084.0	2,464.5	2,143.0
	Rank of K-W Rank Sum	1	6	7	11	10	2	9	3	4	8	5
	K-W Multi-Comparison Count*	10	8	7	10	10	9	8	9	9	7	9
15	Average	15.88%	44.12%	49.80%	63.77%	51.66%	26.76%	37.80%	38.19%	45.17%	57.61%	45.20%
	Rank	1	5	8	11	9	2	3	4	6	10	7
	Geometric Mean	10.94%	37.96%	42.02%	53.87%	46.76%	25.16%	36.83%	23.53%	28.97%	39.95%	34.67%
	Rank	1	7	9	11	10	3	6	2	4	8	5
	Average Rank by Series	2.6	6.975	6.575	8.025	7.725	4.5	6.85	5.325	5.275	6.375	5.775
	Rank of Average Rank	1	9	7	11	10	2	8	4	3	6	5
	Kruskal-Wallis Rank Sum	894.0	2,395.5	2,580.5	3,032.5	2,841.5	1,539.0	2,473.0	1,825.5	2,117.5	2,459.5	2,151.5
	Rank of K-W Rank Sum	1	6	9	11	10	2	8	3	4	7	5
	K-W Multi-Comparison Count*	10	8	10	10	10	10	8	10	9	8	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level and Trend Shift (N) Scenario 9 Table: 9-3

Period:	Mean Absolute Percent Error	Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	6.12%	11.65%	11.77%	14.49%	13.54%	10.94%	9.52%	8.44%	8.61%	10.83%	10.06%
	Rank	1	8	9	11	10	7	4	2	3	6	5
	Geometric Mean	3.19%	10.21%	10.39%	12.33%	11.30%	9.01%	8.55%	5.98%	6.41%	8.49%	7.99%
	Rank	1	8	9	11	10	7	6	2	3	5	4
	Average Rank by Series	1.40	7.88	7.98	9.50	9.20	6.45	5.45	3.63	3.78	5.60	5.15
	Rank of Average Rank	1	8	9	11	10	7	5	2	3	6	4
	Kruskal-Wallis Rank Sum	1,287.0	2,617.5	2,660.5	2,885.0	2,751.0	2,323.0	2,223.0	1,630.5	1,656.5	2,195.0	2,081.0
	Rank of K-W Rank Sum	1	8	9	11	10	7	6	2	3	5	4
	K-W Multi-Comparison Count*	10	9	9	10	10	10	9	9	9	9	10
5	Average	7.77%	17.15%	17.97%	21.90%	20.41%	14.10%	16.98%	13.06%	14.35%	16.66%	14.88%
	Rank	1	8	9	11	10	3	7	2	4	6	5
	Geometric Mean	5.17%	16.09%	16.91%	20.55%	19.26%	13.00%	16.19%	9.71%	11.52%	14.14%	13.06%
	Rank	1	7	9	11	10	4	8	2	3	6	5
	Average Rank by Series	2.35	7.03	7.13	9.10	8.30	5.00	7.10	4.58	4.68	5.75	5.00
	Rank of Average Rank	1	7	9	11	10	4	8	2	3	6	4
	Kruskal-Wallis Rank Sum	933.0	2,399.5	2,549.5	3,100.0	2,923.0	1,814.0	2,453.0	1,789.5	1,989.5	2,272.0	2,087.0
	Rank of K-W Rank Sum	1	7	9	11	10	3	8	2	4	6	5
	K-W Multi-Comparison Count*	10	9	10	10	10	9	9	9	10	10	10
10	Average	9.49%	21.73%	23.95%	28.67%	24.74%	15.65%	17.84%	19.00%	21.53%	25.07%	20.59%
	Rank	1	7	8	11	9	2	3	4	6	10	5
	Geometric Mean	6.73%	19.05%	21.11%	25.85%	22.78%	14.23%	16.68%	13.51%	16.09%	19.25%	16.94%
	Rank	1	7	9	11	10	3	5	2	4	8	6
	Average Rank by Series	2.5	6.825	6.875	8.45	7.55	5.15	6.45	5.675	5.825	5.5	5.2
	Rank of Average Rank	1	8	9	11	10	2	7	5	6	4	3
	Kruskal-Wallis Rank Sum	962.0	2,300.5	2,543.5	3,032.0	2,770.0	1,672.0	2,097.0	1,965.5	2,256.5	2,490.0	2,221.0
	Rank of K-W Rank Sum	1	7	9	11	10	2	4	3	6	8	5
	K-W Multi-Comparison Count*	10	8	9	10	10	10	10	10	8	9	8
15	Average	9.45%	25.47%	28.65%	34.97%	27.97%	16.85%	17.52%	23.08%	26.71%	32.29%	24.01%
	Rank	1	6	9	11	8	2	3	4	7	10	5
	Geometric Mean	6.75%	21.13%	23.70%	28.64%	23.90%	15.50%	16.36%	14.92%	17.72%	21.78%	16.97%
	Rank	1	7	9	11	10	3	4	2	6	8	5
	Average Rank by Series	2.3	7.075	7.025	7.85	7.15	6.1	6.35	5.575	5.925	5.9	4.75
	Rank of Average Rank	1	9	8	11	10	6	7	3	5	4	2
	Kruskal-Wallis Rank Sum	955.0	2,399.5	2,611.5	2,944.0	2,680.0	1,880.0	2,027.0	1,998.5	2,301.5	2,484.0	2,029.0
	Rank of K-W Rank Sum	1	7	9	11	10	2	4	3	6	8	5
	K-W Multi-Comparison Count*	10	10	9	10	9	10	8	8	10	10	8

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Level and Trend Shift (N) Scenario 9 Table: 9-4

Period: Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	49.36	151.21	151.07	171.27	165.42	139.29	141.18	87.96	95.23	118.80	115.71
	Rank	1	9	8	11	10	6	7	2	3	5	4
	Average Rank by Series	1.55	8.18	8.08	9.15	8.75	6.75	6.45	3.88	3.83	4.75	4.65
	Rank of Average Rank	1	9	8	11	10	7	6	3	2	5	4
5	Geometric Mean	75.32	241.82	249.37	289.37	274.81	203.05	271.09	143.82	170.94	204.82	189.69
	Rank	1	7	8	11	10	5	9	2	3	6	4
	Average Rank by Series	2.45	6.875	6.975	8.65	8.25	4.75	7.65	4.775	4.725	5.6	5.3
	Rank of Average Rank	1	7	8	11	10	3	9	4	2	6	5
10	Geometric Mean	105.02	305.14	331.77	396.21	348.46	228.72	284.99	209.51	251.09	299.60	259.19
	Rank	1	8	9	11	10	3	6	2	4	7	5
	Average Rank by Series	2.65	7.025	6.825	7.85	7.1	5.15	6.5	5.875	5.925	5.8	5.3
	Rank of Average Rank	1	9	8	11	10	2	7	5	6	4	3
15	Geometric Mean	110.40	341.01	381.81	461.42	386.82	243.19	292.41	240.19	286.55	352.54	283.07
	Rank	1	7	9	11	10	3	6	2	5	8	4
	Average Rank by Series	2.3	7.025	7.025	7.7	6.95	5.45	7	5.525	5.875	6	5.15
	Rank of Average Rank	1	9	9	11	7	3	8	4	5	6	2

Level and Trend Shift (N) Scenario 9 Table: 9-5

Period: Geometric Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	26.70	86.35	91.63	121.48	93.92	73.20	58.51	48.78	50.92	73.43	64.23
	Rank	1	8	9	11	10	6	4	2	3	7	5
	Average Rank by Series	1.75	6.975	7.175	9.8	8.2	6.6	5	4.225	4.475	6.45	5.35
	Rank of Average Rank	1	8	9	11	10	7	4	2	3	6	5
5	Geometric Mean	49.47	131.03	137.69	179.92	167.29	0.00	107.60	85.77	95.81	114.80	108.93
	Rank	2	8	9	11	10	1	5	3	4	7	6
	Average Rank by Series	2.9	7.125	6.875	9	8.05	4.75	6.75	5.075	4.825	5.75	4.9
	Rank of Average Rank	1	9	8	11	10	2	7	5	3	6	4
10	Geometric Mean	62.27	171.44	198.17	215.14	209.14	132.46	116.00	131.16	151.56	151.69	155.03
	Rank	1	8	9	11	10	4	2	3	5	6	7
	Average Rank by Series	2.5	6.725	7.425	7.8	7.65	5.85	5.75	5.775	6.275	4.85	5.4
	Rank of Average Rank	1	8	9	11	10	6	4	5	7	2	3
15	Geometric Mean	66.81	186.74	208.97	266.68	211.67	165.06	128.29	144.54	169.20	198.18	150.34
	Rank	1	7	9	11	10	5	2	3	6	8	4
	Average Rank by Series	2.45	6.525	6.875	8.3	6.95	7	5.8	5.325	6.125	5.7	4.95
	Rank of Average Rank	1	7	8	11	9	10	5	3	6	4	2

Level and Trend Shift (N) — Average Rank of Absolute Error

Scenario 9 — Table: 9 -6

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	56.69	19	0.0000
	KRUSKAL-WALLIS	108.97	10	0.0000
5	RANK ANOVA	23.43	19	0.2189
	KRUSKAL-WALLIS	43.99	10	0.0000
10	RANK ANOVA	17.80	19	0.5361
	KRUSKAL-WALLIS	39.29	10	0.0000
15	RANK ANOVA	15.02	19	0.7212
	KRUSKAL-WALLIS	31.19	10	0.0005

Level and Trend Shift (N) — Log Mean Squared Error Ratio

Scenario 9 — Table: 9 -12

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	49.55	19	0.0002
	KRUSKAL-WALLIS	83.81	10	0.0000
5	RANK ANOVA	32.31	19	0.0289
	KRUSKAL-WALLIS	36.99	10	0.0001
10	RANK ANOVA	24.12	19	0.1914
	KRUSKAL-WALLIS	31.05	10	0.0006
15	RANK ANOVA	24.00	19	0.1963
	KRUSKAL-WALLIS	27.92	10	0.0019

Level and Trend Shift (N) — Range of Percent Error

Scenario 9 — Table: 9 -7

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	36.43	19	0.0093
	KRUSKAL-WALLIS	22.31	10	0.0136
5	RANK ANOVA	30.06	19	0.0510
	KRUSKAL-WALLIS	49.08	10	0.0000
10	RANK ANOVA	24.30	19	0.1848
	KRUSKAL-WALLIS	42.47	10	0.0000
15	RANK ANOVA	25.58	19	0.1424
	KRUSKAL-WALLIS	45.91	10	0.0000

Level and Trend Shift (N) — Median Absolute Percent Error

Scenario 9 — Table: 9 -11

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	63.16	19	0.0000
	KRUSKAL-WALLIS	35.46	10	0.0001
5	RANK ANOVA	23.07	19	0.2341
	KRUSKAL-WALLIS	35.16	10	0.0001
10	RANK ANOVA	18.70	19	0.4760
	KRUSKAL-WALLIS	30.15	10	0.0008
15	RANK ANOVA	18.33	19	0.5008
	KRUSKAL-WALLIS	27.86	10	0.0019

Level and Trend Shift (N) — Symmetry Adjusted MAPE

Scenario 9 — Table: 9 -8

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	66.04	19	0.0000
	KRUSKAL-WALLIS	29.60	10	0.0010
5	RANK ANOVA	30.93	19	0.0411
	KRUSKAL-WALLIS	38.82	10	0.0000
10	RANK ANOVA	19.27	19	0.4395
	KRUSKAL-WALLIS	33.56	10	0.0002
15	RANK ANOVA	21.70	19	0.2994
	KRUSKAL-WALLIS	30.41	10	0.0007

Level and Trend Shift (N) — Mean Absolute Percent Error

Scenario 9 — Table: 9 -12

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	66.08	19	0.0000
	KRUSKAL-WALLIS	32.59	10	0.0003
5	RANK ANOVA	39.64	19	0.0036
	KRUSKAL-WALLIS	43.72	10	0.0000
10	RANK ANOVA	25.57	19	0.1427
	KRUSKAL-WALLIS	38.37	10	0.0000
15	RANK ANOVA	23.99	19	0.1966
	KRUSKAL-WALLIS	34.99	10	0.0001

Level and Trend Shift (N) — Geometric Root Mean Squared Error

Scenario 9 — Table: 9 -9

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	49.55	19	0.0002
5	RANK ANOVA	32.24	19	0.0293
10	RANK ANOVA	24.12	19	0.1914
15	RANK ANOVA	24.00	19	0.1963

Level and Trend Shift (N) — Root Mean Squared Error

Scenario 9 — Table: 9 -13

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	62.63	19	0.0000
5	RANK ANOVA	36.39	19	0.0095
10	RANK ANOVA	20.43	19	0.3690
15	RANK ANOVA	23.28	19	0.2252

Period: 25% Level Shift (N) Scenario 10 Table: 10-1

Average Rank of Absolute Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	8.10	4.67	4.83	5.41	4.99	4.89	9.25	6.09	6.09	5.87	5.90
	Rank	10	1	2	5	4	3	11	8	8	6	7
	Geometric Mean	7.95	4.49	4.74	5.34	4.85	4.58	9.14	6.05	6.05	5.76	5.77
	Rank	10	1	3	5	4	2	11	8	8	6	7
	Average Rank by Series	8.90	3.80	3.93	5.10	4.23	4.63	10.75	6.95	6.95	5.30	5.90
	Rank of Average Rank	10	1	2	5	3	4	11	8	8	6	7
	Kruskal-Wallis Rank Sum	3,576.0	1,261.0	1,255.5	1,779.0	1,483.0	1,561.0	4,036.0	2,524.0	2,524.0	2,211.5	2,242.0
	Rank of K-W Rank Sum	10	2	1	5	3	4	11	8	8	6	7
	K-W Multi-Comparison Count*	10	9	9	10	9	9	10	10	10	9	9
5	Average	8.17	4.48	4.65	4.69	4.66	4.67	9.17	6.65	6.74	6.08	6.04
	Rank	10	1	2	5	3	4	11	8	9	7	6
	Geometric Mean	8.01	4.32	4.59	4.54	4.45	4.40	9.04	6.56	6.68	5.91	5.89
	Rank	10	1	5	4	3	2	11	8	9	7	6
	Average Rank by Series	8.90	3.75	3.48	3.83	3.83	4.03	10.10	7.80	7.70	6.45	6.15
	Rank of Average Rank	10	2	1	3	3	5	11	9	8	7	6
	Kruskal-Wallis Rank Sum	3,540.5	1,207.5	1,242.5	1,257.0	1,319.0	1,426.5	3,904.5	2,843.0	2,922.0	2,324.5	2,323.0
	Rank of K-W Rank Sum	10	1	2	3	4	5	11	8	9	7	6
	K-W Multi-Comparison Count*	10	8	7	7	8	10	10	9	9	9	9
10	Average	8.27	4.69	4.96	4.66	4.37	5.05	9.28	6.78	6.78	5.84	5.49
	Rank	10	3	4	2	1	5	11	8	8	7	6
	Geometric Mean	8.09	4.47	4.86	4.40	4.14	4.68	9.20	6.64	6.64	5.66	5.32
	Rank	10	3	5	2	1	4	11	8	8	7	6
	Average Rank by Series	8.875	4.025	4.45	3.875	3.55	4.775	10.5	7.6	7.6	6.175	4.95
	Rank of Average Rank	10	3	4	2	1	5	11	8	8	7	6
	Kruskal-Wallis Rank Sum	3,523.0	1,461.0	1,578.0	1,330.0	1,169.0	1,709.0	3,969.0	2,813.5	2,813.5	2,155.5	1,917.5
	Rank of K-W Rank Sum	10	3	4	2	1	5	11	8	8	7	6
	K-W Multi-Comparison Count*	10	10	10	10	10	10	10	10	10	10	10
15	Average	8.32	4.50	4.80	4.62	4.40	5.21	9.14	6.62	6.77	5.83	5.80
	Rank	10	2	4	3	1	5	11	8	9	7	6
	Geometric Mean	8.07	4.27	4.66	4.34	4.16	4.75	9.01	6.50	6.67	5.60	5.66
	Rank	10	2	4	3	1	5	11	8	9	6	7
	Average Rank by Series	8.575	3.975	4.3	3.75	3.525	5.175	10.125	7.55	7.575	5.7	5.75
	Rank of Average Rank	10	3	4	2	1	5	11	8	9	6	7
	Kruskal-Wallis Rank Sum	3,455.5	1,281.5	1,482.5	1,321.0	1,193.0	1,805.0	3,862.5	2,767.5	2,823.0	2,149.5	2,169.0
	Rank of K-W Rank Sum	10	2	4	3	1	5	11	8	9	6	7
	K-W Multi-Comparison Count*	10	9	10	9	10	10	10	9	9	9	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

25% Level Shift (N) Scenario 10 Table: 10-2

Period: Range of Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	18.09%	13.71%	13.72%	11.42%	12.84%	14.26%	24.02%	17.58%	17.58%	14.36%	14.72%
	Rank	10	3	4	1	2	5	11	8	9	6	7
	Geometric Mean	14.69%	7.44%	7.50%	7.36%	7.93%	6.82%	22.37%	13.79%	13.79%	12.40%	12.14%
	Rank	10	3	4	2	5	1	11	8	9	7	6
	Average Rank by Series	8.08	4.60	4.30	3.78	4.33	3.90	9.58	7.80	7.60	5.88	6.18
	Rank of Average Rank	10	5	3	1	4	2	11	9	8	6	7
	Kruskal-Wallis Rank Sum	2,647.5	1,732.0	1,739.0	1,763.5	1,841.5	1,659.0	3,328.5	2,484.0	2,487.0	2,320.5	2,307.5
	Rank of K-W Rank Sum	10	2	3	4	5	1	11	8	9	7	6
	K-W Multi-Comparison Count*	10	7	7	7	9	8	10	9	9	9	9
5	Average	24.19%	14.74%	15.00%	12.77%	13.43%	16.12%	49.22%	29.24%	28.93%	25.27%	25.72%
	Rank	6	3	4	1	2	5	11	10	9	7	8
	Geometric Mean	21.69%	10.58%	11.08%	10.06%	10.13%	10.05%	48.43%	27.90%	27.55%	24.41%	25.13%
	Rank	6	4	5	2	3	1	11	10	9	7	8
	Average Rank by Series	6.60	3.68	3.63	3.20	3.60	3.65	10.75	8.43	8.38	6.95	7.15
	Rank of Average Rank	6	5	3	1	2	4	11	10	9	7	8
	Kruskal-Wallis Rank Sum	2,446.0	1,354.5	1,366.5	1,239.0	1,302.0	1,368.0	4,005.0	2,991.5	2,936.5	2,591.0	2,710.0
	Rank of K-W Rank Sum	6	3	4	1	2	5	11	10	9	7	8
	K-W Multi-Comparison Count*	10	7	7	9	6	7	10	9	9	10	10
10	Average	31.57%	16.66%	17.15%	16.52%	15.32%	16.41%	47.36%	27.77%	27.42%	25.45%	24.48%
	Rank	10	4	5	3	1	2	11	9	8	7	6
	Geometric Mean	27.49%	12.92%	13.85%	13.97%	12.88%	11.51%	46.29%	26.60%	26.20%	24.31%	23.38%
	Rank	10	3	4	5	2	1	11	9	8	7	6
	Average Rank by Series	6.775	4.325	4.425	3.525	3.125	4.3	10.575	7.975	8.075	6.625	6.275
	Rank of Average Rank	8	4	5	2	1	3	11	9	10	7	6
	Kruskal-Wallis Rank Sum	2,812.5	1,471.5	1,522.5	1,508.5	1,336.5	1,329.0	3,995.5	2,739.5	2,710.5	2,486.5	2,397.5
	Rank of K-W Rank Sum	10	3	5	4	2	1	11	9	8	7	6
	K-W Multi-Comparison Count*	10	8	8	8	9	9	10	9	9	10	10
15	Average	38.56%	16.54%	17.64%	19.32%	15.92%	13.76%	54.57%	26.75%	26.22%	26.22%	23.51%
	Rank	10	3	4	5	2	1	11	9	8	7	6
	Geometric Mean	30.46%	14.01%	15.45%	15.96%	14.08%	10.78%	51.99%	25.00%	24.26%	23.85%	22.15%
	Rank	10	2	4	5	3	1	11	9	8	7	6
	Average Rank by Series	7.175	3.975	4.325	4.725	4.2	3.15	10.575	7.425	7.225	6.825	6.4
	Rank of Average Rank	8	2	4	5	3	1	11	10	9	7	6
	Kruskal-Wallis Rank Sum	2,888.5	1,521.5	1,630.5	1,789.5	1,454.0	1,184.0	3,990.5	2,618.5	2,525.5	2,447.5	2,260.0
	Rank of K-W Rank Sum	10	3	4	5	2	1	11	9	8	7	6
	K-W Multi-Comparison Count*	10	9	10	10	9	10	10	10	9	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

25% Level Shift (N) Scenario 10 — Table: 10-3

Period: Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	10.81%	6.15%	6.17%	6.10%	5.87%	6.40%	14.96%	7.80%	7.83%	6.91%	6.88%
	Rank	10	3	4	2	1	5	11	8	9	7	6
	Geometric Mean	9.46%	3.93%	3.99%	4.24%	4.09%	3.88%	14.01%	6.57%	6.61%	6.12%	6.06%
	Rank	10	2	3	5	4	1	11	8	9	7	6
	Average Rank by Series	8.45	4.13	4.08	4.80	3.75	4.40	10.70	7.08	7.13	5.85	5.65
	Rank of Average Rank	10	3	2	5	1	4	11	8	9	7	6
	Kruskal-Wallis Rank Sum	3,037.0	1,637.5	1,647.5	1,751.0	1,707.0	1,688.0	3,578.0	2,395.5	2,400.5	2,241.0	2,227.0
	Rank of K-W Rank Sum	10	1	2	5	4	3	11	8	9	7	6
	K-W Multi-Comparison Count*	10	7	7	8	6	6	10	9	9	9	9
5	Average	17.41%	7.16%	7.34%	7.33%	7.16%	7.46%	25.09%	13.27%	13.42%	12.16%	11.71%
	Rank	10	2	4	3	1	5	11	8	9	7	6
	Geometric Mean	16.74%	5.59%	5.94%	6.14%	5.97%	5.43%	24.61%	12.18%	12.39%	11.52%	11.14%
	Rank	10	2	3	5	4	1	11	8	9	7	6
	Average Rank by Series	8.95	3.03	2.98	3.45	3.10	3.60	10.95	8.03	7.93	7.10	6.90
	Rank of Average Rank	10	2	1	4	3	5	11	9	8	7	6
	Kruskal-Wallis Rank Sum	3,322.0	1,303.5	1,326.5	1,361.0	1,349.0	1,380.0	4,039.0	2,633.5	2,658.5	2,513.0	2,424.0
	Rank of K-W Rank Sum	10	1	2	4	3	5	11	8	9	7	6
	K-W Multi-Comparison Count*	10	6	6	6	6	6	10	9	9	10	10
10	Average	22.53%	9.16%	9.53%	9.72%	8.69%	9.73%	29.45%	15.01%	15.28%	14.54%	12.76%
	Rank	10	2	3	4	1	5	11	8	9	7	6
	Geometric Mean	20.45%	7.07%	7.77%	7.96%	7.17%	7.17%	28.34%	13.43%	13.79%	13.23%	11.89%
	Rank	10	1	4	5	2	3	11	8	9	7	6
	Average Rank by Series	8.8	3.625	3.625	3.4	3.2	4.2	10.65	7.575	7.575	6.8	6.55
	Rank of Average Rank	10	3	3	2	1	5	11	8	8	7	6
	Kruskal-Wallis Rank Sum	3,303.0	1,449.5	1,499.5	1,596.0	1,418.0	1,522.0	3,909.0	2,459.5	2,518.5	2,429.0	2,206.0
	Rank of K-W Rank Sum	10	2	3	5	1	4	11	8	9	7	6
	K-W Multi-Comparison Count*	10	7	8	9	9	7	10	8	9	9	10
15	Average	26.39%	10.12%	10.68%	11.55%	9.28%	11.53%	32.55%	15.99%	16.40%	16.02%	13.46%
	Rank	10	2	3	5	1	4	11	7	9	8	6
	Geometric Mean	22.33%	7.99%	8.78%	9.35%	7.97%	8.80%	30.34%	14.32%	14.82%	14.08%	12.37%
	Rank	10	2	3	5	1	4	11	8	9	7	6
	Average Rank by Series	8.5	3.425	3.575	3.65	3.55	4.85	10.55	7.525	7.875	6.5	6
	Rank of Average Rank	10	1	3	4	2	5	11	8	9	7	6
	Kruskal-Wallis Rank Sum	3,212.0	1,484.5	1,595.5	1,704.0	1,365.0	1,699.0	3,816.0	2,427.5	2,506.5	2,382.0	2,118.0
	Rank of K-W Rank Sum	10	2	3	5	1	4	11	8	9	7	6
	K-W Multi-Comparison Count*	10	10	10	9	10	9	10	8	9	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

25% Level Shift (N) Scenario 10 — Table: 10-4

Period: Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	172.70	73.85	74.35	76.75	76.68	72.21	259.06	128.90	129.06	117.70	118.17
	Rank	10	2	3	5	4	1	11	8	9	6	7
	Average Rank by Series	8.60	3.88	3.78	4.00	3.75	4.30	11.00	7.63	7.58	5.60	5.90
	Rank of Average Rank	10	3	2	4	1	5	11	9	8	6	7
5	Geometric Mean	306.93	105.39	111.14	111.13	108.52	102.67	479.73	247.10	247.56	231.47	225.02
	Rank	10	2	5	4	3	1	11	8	9	7	6
	Average Rank by Series	9	3.075	3.125	3.15	2.85	3.8	11	8.175	8.025	6.95	6.85
	Rank of Average Rank	10	2	3	4	1	5	11	9	8	7	6
10	Geometric Mean	385.42	140.95	154.25	157.73	140.56	141.09	543.79	277.13	278.81	268.84	246.17
	Rank	10	2	4	5	1	3	11	8	9	7	6
	Average Rank by Series	8.55	3.575	3.675	3.55	3.15	4.45	10.8	7.825	7.575	6.85	6
	Rank of Average Rank	10	3	4	2	1	5	11	9	8	7	6
15	Geometric Mean	438.18	162.64	178.30	186.58	159.84	167.42	614.12	293.74	298.87	289.74	257.29
	Rank	10	2	4	5	1	3	11	8	9	7	6
	Average Rank by Series	8.5	3.375	3.625	3.6	3.25	4.3	10.65	7.775	7.875	6.75	6.3
	Rank of Average Rank	10	2	4	3	1	5	11	8	9	7	6

25% Level Shift (N) Scenario 10 — Table: 10-5

Period: Geometric Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	105.85	38.96	40.54	47.65	42.82	42.28	147.01	57.60	59.24	61.11	58.54
	Rank	10	1	2	5	4	3	11	6	8	9	7
	Average Rank by Series	8.2	4.175	4.475	5.15	4.2	4.6	10	6.375	6.575	6.25	6
	Rank of Average Rank	10	1	3	5	2	4	11	8	9	7	6
5	Geometric Mean	203.25	62.51	66.81	71.47	67.99	60.11	260.67	116.73	125.60	115.66	108.87
	Rank	10	2	3	5	4	1	11	8	9	7	6
	Average Rank by Series	9	3.275	3.425	3.95	3.65	3.15	10.4	7.425	7.925	7.1	6.7
	Rank of Average Rank	10	2	3	5	4	1	11	8	9	7	6
10	Geometric Mean	257.64	80.86	89.74	94.01	83.99	88.03	332.73	141.09	151.60	140.86	122.80
	Rank	10	1	4	5	2	3	11	8	9	7	6
	Average Rank by Series	8.65	3.875	3.925	4.35	3.75	4.85	9.85	7.075	7.425	6.5	5.75
	Rank of Average Rank	10	2	3	4	1	5	11	8	9	7	6
15	Geometric Mean	289.95	100.78	109.00	119.79	97.38	126.42	365.16	176.48	183.03	173.56	146.44
	Rank	10	2	3	4	1	5	11	8	9	7	6
	Average Rank by Series	8.15	3.625	3.875	4.25	3.35	5.5	9.85	7.425	7.725	6.5	5.75
	Rank of Average Rank	10	2	3	4	1	5	11	8	9	7	6

25% Level Shift (N) — Average Rank of Absolute Error
Scenario 10 — Table: 10 -6

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	50.08	19	0.0001
	KRUSKAL-WALLIS	102.13	10	0.0000
5	RANK ANOVA	59.15	19	0.0000
	KRUSKAL-WALLIS	121.33	10	0.0000
10	RANK ANOVA	54.64	19	0.0000
	KRUSKAL-WALLIS	105.70	10	0.0000
15	RANK ANOVA	49.97	19	0.0001
	KRUSKAL-WALLIS	103.09	10	0.0000

25% Level Shift (N) — Range of Percent Error
Scenario 10 — Table: 10 -7

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	42.08	19	0.0017
	KRUSKAL-WALLIS	33.38	10	0.0002
5	RANK ANOVA	70.02	19	0.0000
	KRUSKAL-WALLIS	107.75	10	0.0000
10	RANK ANOVA	55.41	19	0.0000
	KRUSKAL-WALLIS	89.38	10	0.0000
15	RANK ANOVA	48.97	19	0.0002
	KRUSKAL-WALLIS	81.04	10	0.0000

25% Level Shift (N) — Symmetry Adjusted MAPE
Scenario 10 — Table: 10 -8

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	58.33	19	0.0000
	KRUSKAL-WALLIS	54.74	10	0.0000
5	RANK ANOVA	88.73	19	0.0000
	KRUSKAL-WALLIS	117.13	10	0.0000
10	RANK ANOVA	69.17	19	0.0000
	KRUSKAL-WALLIS	87.36	10	0.0000
15	RANK ANOVA	63.10	19	0.0000
	KRUSKAL-WALLIS	73.94	10	0.0000

25% Level Shift (N) — Geometric Root Mean Squared Error
Scenario 10 — Table: 10 -9

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	34.81	19	0.0147
5	RANK ANOVA	70.92	19	0.0000
10	RANK ANOVA	45.54	19	0.0006
15	RANK ANOVA	47.60	19	0.0003

25% Level Shift (N) — Log Mean Squared Error Ratio
Scenario 10 — Table: 10 -10

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	34.81	19	0.0147
	KRUSKAL-WALLIS	46.08	10	0.0000
5	RANK ANOVA	70.92	19	0.0000
	KRUSKAL-WALLIS	84.99	10	0.0000
10	RANK ANOVA	45.54	19	0.0006
	KRUSKAL-WALLIS	44.20	10	0.0000
15	RANK ANOVA	47.60	19	0.0003
	KRUSKAL-WALLIS	43.49	10	0.0000

25% Level Shift (N) — Median Absolute Percent Error
Scenario 10 — Table: 10 -11

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	45.10	19	0.0007
	KRUSKAL-WALLIS	39.10	10	0.0000
5	RANK ANOVA	57.26	19	0.0000
	KRUSKAL-WALLIS	68.06	10	0.0000
10	RANK ANOVA	52.29	19	0.0001
	KRUSKAL-WALLIS	56.02	10	0.0000
15	RANK ANOVA	41.25	19	0.0022
	KRUSKAL-WALLIS	46.30	10	0.0000

25% Level Shift (N) — Mean Absolute Percent Error
Scenario 10 — Table: 10 -12

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	49.35	19	0.0002
	KRUSKAL-WALLIS	49.46	10	0.0000
5	RANK ANOVA	85.47	19	0.0000
	KRUSKAL-WALLIS	109.25	10	0.0000
10	RANK ANOVA	67.58	19	0.0000
	KRUSKAL-WALLIS	84.50	10	0.0000
15	RANK ANOVA	61.26	19	0.0000
	KRUSKAL-WALLIS	72.71	10	0.0000

25% Level Shift (N) — Root Mean Squared Error
Scenario 10 — Table: 10 -13

Period:		Chi Squared	DF	p Value
1	RANK ANOVA	61.33	19	0.0000
5	RANK ANOVA	88.02	19	0.0000
10	RANK ANOVA	66.96	19	0.0000
15	RANK ANOVA	67.00	19	0.0000

Period: 200% Level Shift (N) Scenario 11 Table: 11-1

Average Rank of Absolute Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	5.43	6.78	7.14	8.02	7.21	5.41	4.17	5.11	5.11	6.29	5.66
	Rank	5	8	9	11	10	4	1	2	2	7	6
	Geometric Mean	5.21	6.63	7.05	7.97	7.11	5.33	3.96	5.03	5.03	6.21	5.56
	Rank	4	8	9	11	10	5	1	2	2	7	6
	Average Rank by Series	4.73	7.30	8.30	9.83	8.65	4.73	2.65	4.25	4.25	6.65	5.30
	Rank of Average Rank	4	8	9	11	10	4	1	2	2	7	6
	Kruskal-Wallis Rank Sum	1,708.0	2,823.0	3,148.0	3,759.5	3,171.0	1,716.0	870.5	1,474.0	1,474.0	2,473.0	1,926.5
	Rank of K-W Rank Sum	4	8	9	11	10	5	1	2	2	7	6
	K-W Multi-Comparison Count*	9	10	9	10	9	9	10	10	10	10	10
5	Average	5.26	6.54	6.87	7.64	7.56	5.28	4.44	4.74	5.23	6.19	6.24
	Rank	4	8	9	11	10	5	1	2	3	6	7
	Geometric Mean	4.95	6.46	6.81	7.60	7.51	5.10	4.27	4.54	5.03	6.09	6.14
	Rank	3	8	9	11	10	5	1	2	4	6	7
	Average Rank by Series	4.75	7.10	7.93	9.40	8.90	5.08	3.03	3.30	4.35	5.95	6.23
	Rank of Average Rank	4	8	9	11	10	5	1	2	3	6	7
	Kruskal-Wallis Rank Sum	1,647.5	2,577.5	2,883.5	3,551.0	3,452.0	1,620.5	1,022.5	1,277.0	1,618.5	2,327.5	2,332.5
	Rank of K-W Rank Sum	5	8	9	11	10	4	1	2	3	6	7
	K-W Multi-Comparison Count*	8	10	10	10	10	8	10	10	8	9	9
10	Average	4.81	6.56	6.96	7.66	7.02	5.44	4.31	5.54	5.54	6.56	5.94
	Rank	2	8	9	11	10	3	1	4	4	7	6
	Geometric Mean	4.50	6.46	6.89	7.58	6.96	5.12	4.12	5.32	5.32	6.45	5.89
	Rank	2	8	9	11	10	3	1	4	4	7	6
	Average Rank by Series	4.075	7.125	7.925	9.2	8.3	5.35	3.275	4.7	4.7	6.45	5.475
	Rank of Average Rank	2	8	9	11	10	5	1	3	3	7	6
	Kruskal-Wallis Rank Sum	1,349.0	2,600.5	2,931.0	3,489.0	3,020.5	1,847.0	982.5	1,813.5	1,813.5	2,604.5	2,042.5
	Rank of K-W Rank Sum	2	7	9	11	10	5	1	3	3	8	6
	K-W Multi-Comparison Count*	10	9	10	10	10	9	10	9	9	9	10
15	Average	4.76	6.58	6.89	7.24	6.89	5.54	4.26	5.46	5.93	6.44	6.01
	Rank	2	8	9	11	10	4	1	3	5	7	6
	Geometric Mean	4.40	6.46	6.83	7.15	6.71	4.98	4.02	5.04	5.67	6.30	5.89
	Rank	2	8	10	11	9	3	1	4	5	7	6
	Average Rank by Series	4.175	7.225	7.4	8.25	7.65	5.925	3.4	4.825	5.5	6.05	5.6
	Rank of Average Rank	2	8	9	11	10	6	1	3	4	7	5
	Kruskal-Wallis Rank Sum	1,370.0	2,563.0	2,797.5	3,100.0	2,824.5	2,021.0	1,051.0	1,898.5	2,122.5	2,422.0	2,140.0
	Rank of K-W Rank Sum	2	8	9	11	10	4	1	3	5	7	6
	K-W Multi-Comparison Count*	10	10	9	10	9	10	10	10	9	10	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

200% Level Shift (N) Scenario 11 Table: 11-2

Period: Range of Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	58.67%	63.13%	61.16%	77.21%	75.98%	64.12%	52.42%	50.58%	48.78%	67.50%	63.53%
	Rank	4	6	5	11	10	8	3	2	1	9	7
	Geometric Mean	41.46%	50.47%	47.84%	52.79%	57.56%	52.62%	36.04%	35.71%	31.32%	47.27%	45.45%
	Rank	4	8	7	10	11	9	3	2	1	6	5
	Average Rank by Series	5.23	7.05	6.38	7.60	8.55	6.85	4.28	4.30	3.28	5.95	6.55
	Rank of Average Rank	4	9	6	10	11	8	2	3	1	5	7
	Kruskal-Wallis Rank Sum	2,207.5	2,388.0	2,302.5	2,452.0	2,631.0	2,397.0	2,001.5	1,810.0	1,668.5	2,224.0	2,228.0
	Rank of K-W Rank Sum	4	8	7	10	11	9	3	2	1	5	6
	K-W Multi-Comparison Count*	8	8	8	8	10	8	10	10	10	7	7
5	Average	70.22%	118.11%	120.27%	114.32%	124.13%	116.96%	64.59%	83.23%	83.61%	97.49%	97.25%
	Rank	2	9	10	7	11	8	1	3	4	6	5
	Geometric Mean	61.49%	110.09%	111.36%	107.61%	115.87%	109.05%	52.41%	72.35%	72.22%	90.61%	89.87%
	Rank	2	9	10	7	11	8	1	4	3	6	5
	Average Rank by Series	3.03	7.60	7.63	8.03	8.18	7.80	2.88	4.00	4.08	6.33	6.48
	Rank of Average Rank	2	7	8	10	11	9	1	3	4	5	6
	Kruskal-Wallis Rank Sum	1,220.5	2,824.0	2,818.5	2,801.5	2,948.5	2,825.0	1,197.5	1,622.0	1,588.5	2,252.5	2,211.5
	Rank of K-W Rank Sum	2	9	8	7	11	10	1	4	3	6	5
	K-W Multi-Comparison Count*	9	7	7	7	10	7	9	9	9	9	9
10	Average	77.66%	131.65%	133.72%	155.47%	138.12%	97.59%	67.51%	90.88%	103.53%	129.44%	109.69%
	Rank	2	8	9	11	10	4	1	3	5	7	6
	Geometric Mean	69.29%	120.78%	121.74%	143.80%	129.26%	89.33%	53.29%	71.78%	81.11%	113.92%	97.11%
	Rank	2	8	9	11	10	5	1	3	4	7	6
	Average Rank by Series	3.3	7.9	7.85	9.55	8.1	4.95	2.55	4.25	4.75	6.8	6
	Rank of Average Rank	2	9	8	11	10	5	1	3	4	7	6
	Kruskal-Wallis Rank Sum	1,492.0	2,668.0	2,682.0	3,122.0	2,930.0	2,012.0	1,218.0	1,580.0	1,836.0	2,543.0	2,227.0
	Rank of K-W Rank Sum	2	8	9	11	10	5	1	3	4	7	6
	K-W Multi-Comparison Count*	10	9	9	10	10	10	10	10	10	10	10
15	Average	89.97%	146.51%	162.66%	194.64%	159.06%	91.61%	79.98%	109.51%	127.00%	164.86%	127.45%
	Rank	2	7	9	11	8	3	1	4	5	10	6
	Geometric Mean	71.70%	119.69%	131.02%	159.88%	137.15%	79.50%	56.44%	75.22%	87.59%	120.56%	102.83%
	Rank	2	7	9	11	10	4	1	3	5	8	6
	Average Rank by Series	3.45	8.025	8.2	8.975	8.325	5.15	2.95	4.175	4.5	6.125	6.125
	Rank of Average Rank	2	8	9	11	10	5	1	3	4	6	6
	Kruskal-Wallis Rank Sum	1,695.0	2,503.5	2,637.0	3,052.5	2,805.5	1,827.0	1,413.0	1,696.5	1,934.0	2,507.5	2,238.5
	Rank of K-W Rank Sum	2	7	9	11	10	4	1	3	5	8	6
	K-W Multi-Comparison Count*	10	9	10	10	10	10	10	10	10	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

200% Level Shift (N) Scenario 11 Table: 11-3

Period: Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	28.75%	28.38%	28.81%	36.51%	33.88%	24.30%	22.77%	21.42%	21.84%	29.53%	27.48%
	Rank	7	6	8	11	10	4	3	1	2	9	5
	Geometric Mean	23.58%	25.82%	26.45%	31.67%	28.51%	21.17%	17.36%	18.76%	19.38%	24.26%	22.28%
	Rank	6	8	9	11	10	4	1	2	3	7	5
	Average Rank by Series	5.70	7.33	7.98	9.50	8.50	4.30	3.25	3.48	4.03	6.35	5.60
	Rank of Average Rank	6	8	9	11	10	4	1	2	3	7	5
	Kruskal-Wallis Rank Sum	2,300.0	2,475.5	2,554.5	2,878.0	2,614.0	1,928.0	1,820.0	1,655.5	1,738.5	2,262.0	2,084.0
	Rank of K-W Rank Sum	7	8	9	11	10	4	3	1	2	6	5
	K-W Multi-Comparison Count*	9	9	8	10	9	10	10	10	10	9	10
5	Average	40.70%	52.64%	55.48%	67.83%	63.46%	40.86%	30.13%	37.40%	41.16%	53.57%	49.42%
	Rank	3	7	9	11	10	4	1	2	5	8	6
	Geometric Mean	37.81%	49.41%	52.17%	63.85%	60.49%	38.41%	25.94%	32.64%	36.06%	48.13%	45.89%
	Rank	4	8	9	11	10	5	1	2	3	7	6
	Average Rank by Series	4.50	7.83	8.23	9.60	9.60	4.90	2.25	3.33	3.73	6.10	5.95
	Rank of Average Rank	4	8	9	10	10	5	1	2	3	7	6
	Kruskal-Wallis Rank Sum	1,802.0	2,503.5	2,688.5	3,251.0	3,146.0	1,728.0	1,161.0	1,502.5	1,754.5	2,442.0	2,331.0
	Rank of K-W Rank Sum	5	8	9	11	10	3	1	2	4	7	6
	K-W Multi-Comparison Count*	8	9	10	10	10	8	10	10	8	9	10
10	Average	46.77%	62.83%	70.09%	85.23%	74.32%	39.36%	38.72%	50.70%	58.55%	74.87%	62.89%
	Rank	3	6	8	11	9	2	1	4	5	10	7
	Geometric Mean	42.46%	54.08%	60.00%	76.42%	68.19%	37.06%	31.39%	37.76%	43.86%	61.38%	53.55%
	Rank	4	7	8	11	10	2	1	3	5	9	6
	Average Rank by Series	4.2	7.125	7.525	9.6	9.05	4.5	2.95	3.975	4.475	6.5	6.1
	Rank of Average Rank	3	8	9	11	10	5	1	2	4	7	6
	Kruskal-Wallis Rank Sum	1,894.0	2,365.5	2,562.5	3,187.0	2,944.0	1,522.0	1,465.0	1,570.5	1,882.5	2,582.0	2,335.0
	Rank of K-W Rank Sum	5	7	8	11	10	2	1	3	4	9	6
	K-W Multi-Comparison Count*	9	9	9	10	10	8	9	9	9	9	9
15	Average	52.28%	73.01%	83.25%	103.17%	84.39%	38.36%	44.72%	62.49%	74.11%	94.77%	75.03%
	Rank	3	5	8	11	9	1	2	4	6	10	7
	Geometric Mean	41.88%	55.62%	62.73%	79.51%	67.85%	36.32%	30.88%	41.19%	48.71%	65.73%	55.26%
	Rank	4	7	8	11	10	2	1	3	5	9	6
	Average Rank by Series	3.85	7.075	7.475	8.95	8.3	5.15	2.6	4.775	5.275	6.7	5.85
	Rank of Average Rank	2	8	9	11	10	4	1	3	5	7	6
	Kruskal-Wallis Rank Sum	1,939.0	2,376.5	2,531.5	2,938.0	2,768.0	1,678.0	1,484.0	1,723.5	1,997.5	2,557.0	2,317.0
	Rank of K-W Rank Sum	4	7	8	11	10	2	1	3	5	9	6
	K-W Multi-Comparison Count*	9	9	9	10	10	9	10	9	9	9	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

200% Level Shift (N) Scenario 11 — Table: 11-4

Period: Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	201.81	253.91	253.11	295.73	282.99	230.75	146.97	175.69	176.32	218.51	205.66
	Rank	4	9	8	11	10	7	1	2	3	6	5
	Average Rank by Series	4.75	8.08	8.08	9.70	9.15	6.30	2.40	3.18	3.43	5.65	5.30
	Rank of Average Rank	4	8	8	11	10	7	1	2	3	6	5
5	Geometric Mean	308.06	463.67	476.27	561.16	534.35	386.47	220.91	292.73	315.42	412.16	388.86
	Rank	3	8	9	11	10	5	1	2	4	7	6
	Average Rank by Series	3.65	7.825	8.325	10	9.4	5.4	2.15	3.275	3.775	6.35	5.85
	Rank of Average Rank	3	8	9	11	10	5	1	2	4	7	6
10	Geometric Mean	390.87	564.87	610.29	739.24	663.52	403.36	294.63	377.11	434.31	581.40	509.26
	Rank	3	7	9	11	10	4	1	2	5	8	6
	Average Rank by Series	3.5	7.625	8.025	9.4	8.7	5.3	2.35	4.175	4.775	6.45	5.7
	Rank of Average Rank	2	8	9	11	10	5	1	3	4	7	6
15	Geometric Mean	421.31	593.22	663.09	820.71	709.98	387.63	322.29	421.04	498.43	661.22	561.83
	Rank	4	7	9	11	10	2	1	3	5	8	6
	Average Rank by Series	3.5	7.325	7.725	8.7	8.25	5.15	2.5	4.875	5.325	6.8	5.85
	Rank of Average Rank	2	8	9	11	10	4	1	3	5	7	6

200% Level Shift (N) Scenario 11 — Table: 11-5

Period: Geometric Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	138.93	142.91	156.69	187.18	157.79	94.76	84.85	105.63	114.59	138.27	124.86
	Rank	7	8	9	11	10	2	1	3	4	6	5
	Average Rank by Series	6	7.275	7.975	8.95	8	4.3	3.35	3.875	4.375	6.4	5.5
	Rank of Average Rank	6	8	9	11	10	3	1	2	4	7	5
5	Geometric Mean	211.18	208.95	235.55	319.21	293.56	144.40	132.58	148.00	176.30	253.81	231.90
	Rank	6	5	8	11	10	2	1	3	4	9	7
	Average Rank by Series	5.7	6.875	6.925	9.35	9.05	4.2	2.75	3.425	4.075	6.95	6.7
	Rank of Average Rank	5	7	8	11	10	4	1	2	3	9	6
10	Geometric Mean	246.00	259.56	306.07	397.11	364.94	181.45	178.79	211.06	245.80	352.88	306.61
	Rank	5	6	7	11	10	2	1	3	4	9	8
	Average Rank by Series	4.6	6.275	7.425	8.6	8.25	5.35	3.4	4.125	4.425	7.35	6.2
	Rank of Average Rank	4	7	9	11	10	5	1	2	3	8	6
15	Geometric Mean	242.87	303.79	334.87	435.75	361.40	207.23	175.57	248.59	297.91	391.14	326.65
	Rank	3	6	8	11	9	2	1	4	5	10	7
	Average Rank by Series	4.6	6.775	6.925	8.3	7.45	6.15	2.95	4.675	5.225	6.95	6
	Rank of Average Rank	2	7	8	11	10	6	1	3	4	9	5

200% Level Shift (N)		Average Rank of Absolute Error		
Scenario 11		Table:	11	-6
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	55.36	19	0.0000
	KRUSKAL-WALLIS	104.91	10	0.0000
5	RANK ANOVA	48.60	19	0.0002
	KRUSKAL-WALLIS	89.50	10	0.0000
10	RANK ANOVA	39.79	19	0.0035
	KRUSKAL-WALLIS	74.32	10	0.0000
15	RANK ANOVA	24.61	19	0.1736
	KRUSKAL-WALLIS	47.66	10	0.0000

	200% Level Shift (N)		Log Mean Squared Error Ratio		
		Scenario 11	Table:	11	-10
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	37.68	19	0.0065
		KRUSKAL-WALLIS	39.30	10	0.0000
	5	RANK ANOVA	50.21	19	0.0001
		KRUSKAL-WALLIS	43.75	10	0.0000
	10	RANK ANOVA	32.71	19	0.0259
		KRUSKAL-WALLIS	17.85	10	0.0576
	15	RANK ANOVA	24.88	19	0.1644
		KRUSKAL-WALLIS	12.71	10	0.2403

200% Level Shift (N)		Range of Percent Error		
Scenario 11		Table:	11	-7
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	26.79	19	0.1097
	KRUSKAL-WALLIS	9.97	10	0.4428
5	RANK ANOVA	46.21	19	0.0005
	KRUSKAL-WALLIS	58.73	10	0.0000
10	RANK ANOVA	52.12	19	0.0001
	KRUSKAL-WALLIS	48.99	10	0.0000
15	RANK ANOVA	47.68	19	0.0003
	KRUSKAL-WALLIS	34.67	10	0.0001

	200% Level Shift (N)		Median Absolute Percent Error		
		Scenario 11	Table:	11	-11
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	35.36	19	0.0126
		KRUSKAL-WALLIS	23.64	10	0.0086
	5	RANK ANOVA	36.85	19	0.0063
		KRUSKAL-WALLIS	33.50	10	0.0002
	10	RANK ANOVA	22.72	19	0.2500
		KRUSKAL-WALLIS	23.08	10	0.0105
	15	RANK ANOVA	11.87	19	0.8909
		KRUSKAL-WALLIS	13.02	10	0.2223

200% Level Shift (N)		Symmetry Adjusted MAPE		
Scenario 11		Table:	11	-8
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	48.60	19	0.0002
	KRUSKAL-WALLIS	21.01	10	0.0210
5	RANK ANOVA	53.33	19	0.0000
	KRUSKAL-WALLIS	42.03	10	0.0000
10	RANK ANOVA	40.31	19	0.0030
	KRUSKAL-WALLIS	30.08	10	0.0008
15	RANK ANOVA	19.67	19	0.4144
	KRUSKAL-WALLIS	17.04	10	0.0735

	200% Level Shift (N)		Mean Absolute Percent Error		
		Scenario 11	Table:	11	-12
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	47.60	19	0.0003
		KRUSKAL-WALLIS	19.58	10	0.0335
	5	RANK ANOVA	67.14	19	0.0000
		KRUSKAL-WALLIS	56.16	10	0.0000
	10	RANK ANOVA	49.76	19	0.0001
		KRUSKAL-WALLIS	42.45	10	0.0000
	15	RANK ANOVA	38.78	19	0.0047
		KRUSKAL-WALLIS	28.01	10	0.0018

200% Level Shift (N)		Geometric Root Mean Squared Error		
Scenario 11		Table:	11	-9
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	37.68	19	0.0065
5	RANK ANOVA	50.21	19	0.0001
10	RANK ANOVA	32.71	19	0.0259
15	RANK ANOVA	24.88	19	0.1644

200% Level Shift (N)			Root Mean Squared Error		
		Scenario 11	Table:	11	-13
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	65.06	19	0.0000
	5	RANK ANOVA	72.81	19	0.0000
	10	RANK ANOVA	53.33	19	0.0000
	15	RANK ANOVA	40.78	19	0.0026

Period: Trend Shift (N) Scenario 12 Table: 12-1

Average Rank of Absolute Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	7.55	5.03	4.78	5.39	5.12	5.15	8.70	6.04	6.04	5.84	6.03
	Rank	10	2	1	5	3	4	11	8	8	6	7
	Geometric Mean	7.36	4.83	4.63	5.13	4.88	4.93	8.54	5.98	5.98	5.77	5.90
	Rank	10	2	1	5	3	4	11	8	8	6	7
	Average Rank by Series	8.08	4.63	3.93	5.28	4.18	4.65	10.35	6.43	6.43	6.03	5.68
	Rank of Average Rank	10	3	1	5	2	4	11	8	8	7	6
	Kruskal-Wallis Rank Sum	3,262.5	1,561.0	1,267.5	1,814.0	1,617.0	1,592.0	3,839.5	2,355.0	2,355.0	2,176.0	2,274.5
	Rank of K-W Rank Sum	10	2	1	5	4	3	11	8	8	6	7
	K-W Multi-Comparison Count*	10	8	10	10	8	8	10	9	9	10	9
5	Average	6.64	5.33	5.22	5.99	5.94	4.95	7.90	5.94	5.96	6.14	6.00
	Rank	10	3	2	7	5	1	11	4	6	9	8
	Geometric Mean	6.32	5.18	4.99	5.62	5.65	4.74	7.72	5.85	5.88	5.99	5.78
	Rank	10	3	2	4	5	1	11	7	8	9	6
	Average Rank by Series	6.65	4.63	4.28	6.18	5.78	4.05	9.08	6.18	6.45	6.68	6.08
	Rank of Average Rank	9	3	2	6	4	1	11	6	8	10	5
	Kruskal-Wallis Rank Sum	2,636.0	1,703.5	1,651.0	2,229.0	2,138.0	1,417.0	3,462.0	2,177.0	2,234.5	2,360.0	2,302.0
	Rank of K-W Rank Sum	10	3	2	6	4	1	11	5	7	9	8
	K-W Multi-Comparison Count*	10	9	9	8	9	10	10	8	7	9	7
10	Average	5.90	6.15	6.13	7.19	7.20	5.78	7.10	4.78	4.78	5.44	5.48
	Rank	6	8	7	10	11	5	9	1	1	3	4
	Geometric Mean	5.14	6.02	5.96	6.80	6.92	5.55	6.77	4.61	4.61	5.16	5.11
	Rank	4	8	7	10	11	6	9	1	1	5	3
	Average Rank by Series	5.65	6.325	6.125	7.625	7.825	5.925	7.575	3.95	3.95	5.425	5.4
	Rank of Average Rank	5	8	7	10	11	6	9	1	1	4	3
	Kruskal-Wallis Rank Sum	2,123.5	2,352.0	2,350.0	2,908.0	2,903.5	2,105.5	2,824.0	1,384.5	1,384.5	1,891.5	2,011.5
	Rank of K-W Rank Sum	6	8	7	11	10	5	9	1	1	3	4
	K-W Multi-Comparison Count*	9	9	9	9	8	9	9	9	9	10	10
15	Average	6.09	6.19	6.21	7.31	7.55	5.74	6.31	4.56	4.56	5.68	5.79
	Rank	6	7	8	10	11	4	9	1	2	3	5
	Geometric Mean	5.10	6.01	5.96	6.74	7.29	5.49	5.53	4.40	4.33	5.11	5.33
	Rank	3	9	8	10	11	6	7	2	1	4	5
	Average Rank by Series	6.375	6.2	6.375	7.4	8.15	5.6	6.55	4	3.925	5.75	5.675
	Rank of Average Rank	7	6	7	10	11	3	9	2	1	5	4
	Kruskal-Wallis Rank Sum	2,292.0	2,340.5	2,364.5	2,854.5	3,072.5	2,058.5	2,421.0	1,334.0	1,355.0	2,088.0	2,129.5
	Rank of K-W Rank Sum	6	7	8	10	11	3	9	1	2	4	5
	K-W Multi-Comparison Count*	8	7	7	10	10	8	8	9	9	8	8

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Trend Shift (N) Scenario 12 Table: 12-2

Period: Range of Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	18.41%	14.73%	14.85%	13.23%	14.52%	15.62%	25.58%	18.39%	18.57%	15.47%	15.94%
	Rank	9	3	4	1	2	6	11	8	10	5	7
	Geometric Mean	13.67%	7.60%	8.10%	8.32%	8.77%	6.39%	23.28%	13.35%	13.70%	12.44%	12.57%
	Rank	9	2	3	4	5	1	11	8	10	6	7
	Average Rank by Series	7.88	3.93	4.33	4.25	5.10	4.10	9.78	7.03	7.23	5.80	6.60
	Rank of Average Rank	10	1	4	3	5	2	11	8	9	6	7
	Kruskal-Wallis Rank Sum	2,553.5	1,714.5	1,739.5	1,831.0	1,929.0	1,657.0	3,336.5	2,405.5	2,452.5	2,321.0	2,370.0
	Rank of K-W Rank Sum	10	2	3	4	5	1	11	8	9	6	7
	K-W Multi-Comparison Count*	10	8	9	10	10	9	10	8	9	9	8
5	Average	30.39%	23.43%	23.83%	23.42%	22.62%	23.98%	47.20%	28.79%	28.39%	26.62%	27.77%
	Rank	10	3	4	2	1	5	11	9	8	6	7
	Geometric Mean	28.48%	15.34%	16.66%	17.78%	17.16%	12.97%	46.70%	25.39%	24.86%	24.02%	24.84%
	Rank	10	2	3	5	4	1	11	9	8	6	7
	Average Rank by Series	7.20	3.75	3.90	5.05	4.23	4.35	9.95	6.95	6.65	6.75	7.23
	Rank of Average Rank	9	1	2	5	3	4	11	8	6	7	10
	Kruskal-Wallis Rank Sum	2,695.0	1,613.0	1,629.0	1,806.0	1,715.5	1,587.0	3,670.0	2,449.0	2,387.0	2,341.0	2,417.5
	Rank of K-W Rank Sum	10	2	3	5	4	1	11	9	7	6	8
	K-W Multi-Comparison Count*	10	8	8	10	10	8	10	8	7	8	7
10	Average	56.68%	33.00%	33.21%	31.00%	32.71%	31.09%	63.48%	38.74%	37.67%	34.76%	35.36%
	Rank	10	4	5	1	3	2	11	9	8	6	7
	Geometric Mean	47.03%	24.55%	25.46%	25.31%	27.19%	19.47%	58.43%	31.01%	29.88%	28.85%	28.41%
	Rank	10	2	4	3	5	1	11	9	8	7	6
	Average Rank by Series	7.125	4.95	4.775	5.225	5.5	4.2	9.025	6.6	6.225	6.375	6
	Rank of Average Rank	10	3	2	4	5	1	11	9	7	8	6
	Kruskal-Wallis Rank Sum	2,928.5	1,857.0	1,861.5	1,865.5	2,002.0	1,580.0	3,441.5	2,300.0	2,204.5	2,103.5	2,166.0
	Rank of K-W Rank Sum	10	2	3	4	5	1	11	9	8	6	7
	K-W Multi-Comparison Count*	10	8	8	8	10	10	10	10	9	9	8
15	Average	89.82%	39.78%	40.87%	41.86%	43.34%	36.13%	86.77%	52.34%	48.91%	47.61%	50.46%
	Rank	11	2	3	4	5	1	10	9	7	6	8
	Geometric Mean	75.37%	31.45%	34.47%	35.39%	37.38%	23.14%	75.25%	44.09%	39.09%	37.08%	41.36%
	Rank	11	2	3	4	6	1	10	9	7	5	8
	Average Rank by Series	7.575	5	4.775	4.85	5.875	3.75	7.675	6.9	6.225	6.55	6.825
	Rank of Average Rank	10	4	2	3	5	1	11	9	6	7	8
	Kruskal-Wallis Rank Sum	3,144.5	1,792.0	1,834.5	1,833.0	2,015.5	1,490.0	3,221.5	2,367.0	2,174.5	2,102.0	2,335.5
	Rank of K-W Rank Sum	10	2	4	3	5	1	11	9	7	6	8
	K-W Multi-Comparison Count*	9	8	8	8	10	10	9	9	9	9	9

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Trend Shift (N) Scenario 12 — Table: 12-3

Period: Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Average	10.58%	6.79%	6.70%	7.04%	6.78%	6.92%	14.72%	8.44%	8.36%	7.86%	7.79%
	Rank	10	3	1	5	2	4	11	9	8	7	6
	Geometric Mean	9.70%	4.38%	4.17%	4.60%	4.53%	4.38%	14.11%	7.04%	6.90%	6.79%	6.72%
	Rank	10	3	1	5	4	2	11	9	8	7	6
	Average Rank by Series	8.00	4.33	3.93	4.65	4.25	4.40	10.50	6.93	6.78	6.05	6.20
	Rank of Average Rank	10	3	1	5	2	4	11	9	8	6	7
	Kruskal-Wallis Rank Sum	2,905.0	1,680.5	1,654.5	1,843.0	1,823.0	1,712.0	3,499.0	2,350.5	2,313.5	2,268.0	2,261.0
	Rank of K-W Rank Sum	10	2	1	5	4	3	11	9	8	7	6
	K-W Multi-Comparison Count*	10	8	8	9	9	8	10	9	7	8	8
5	Average	18.98%	13.26%	12.89%	13.68%	14.16%	13.14%	26.65%	15.32%	15.00%	14.65%	14.57%
	Rank	10	3	1	4	5	2	11	9	8	7	6
	Geometric Mean	17.69%	11.77%	11.30%	11.95%	12.31%	11.51%	26.08%	13.85%	13.39%	13.37%	13.29%
	Rank	10	3	1	4	5	2	11	9	8	7	6
	Average Rank by Series	6.60	4.18	4.13	5.10	5.35	3.90	10.65	6.73	6.58	6.45	6.35
	Rank of Average Rank	9	3	2	4	5	1	11	10	8	7	6
	Kruskal-Wallis Rank Sum	2,828.0	1,781.5	1,727.5	1,927.0	1,952.0	1,771.0	3,800.0	2,163.5	2,108.5	2,133.0	2,118.0
	Rank of K-W Rank Sum	10	3	1	4	5	2	11	9	6	8	7
	K-W Multi-Comparison Count*	10	8	8	9	9	8	10	7	7	7	7
10	Average	32.52%	25.97%	25.15%	26.27%	27.53%	25.46%	37.54%	21.39%	20.50%	20.80%	21.45%
	Rank	10	7	5	8	9	6	11	3	1	2	4
	Geometric Mean	27.21%	24.10%	23.08%	23.48%	25.17%	23.49%	34.34%	18.63%	17.42%	17.90%	18.31%
	Rank	10	8	5	6	9	7	11	4	1	2	3
	Average Rank by Series	4.95	6.925	6.875	7.6	7.8	6.45	7.8	4.275	3.975	4.65	4.7
	Rank of Average Rank	5	8	7	9	10	6	10	2	1	3	4
	Kruskal-Wallis Rank Sum	2,531.0	2,337.5	2,213.5	2,290.0	2,447.0	2,242.0	3,160.0	1,781.5	1,698.5	1,768.0	1,841.0
	Rank of K-W Rank Sum	10	8	5	7	9	6	11	3	1	2	4
	K-W Multi-Comparison Count*	10	9	8	7	10	8	10	8	9	7	8
15	Average	55.25%	41.29%	39.94%	41.91%	45.15%	40.37%	55.67%	32.28%	30.94%	33.13%	35.21%
	Rank	10	7	5	8	9	6	11	2	1	3	4
	Geometric Mean	46.41%	37.16%	35.61%	36.67%	40.96%	36.25%	48.30%	27.96%	25.64%	28.00%	29.93%
	Rank	10	8	5	7	9	6	11	2	1	3	4
	Average Rank by Series	6.3	6.975	6.925	7.5	8.15	6.15	6.45	3.875	3.625	4.9	5.15
	Rank of Average Rank	6	9	8	10	11	5	7	2	1	3	4
	Kruskal-Wallis Rank Sum	2,713.0	2,341.5	2,254.5	2,312.0	2,583.0	2,250.0	2,810.0	1,726.5	1,648.5	1,767.0	1,904.0
	Rank of K-W Rank Sum	10	8	6	7	9	5	11	2	1	3	4
	K-W Multi-Comparison Count*	10	9	8	7	10	8	10	8	9	9	10

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Trend Shift (N) Scenario 12 Table: 12-4

Period: Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	170.36	77.90	75.88	79.58	80.11	76.13	252.70	130.10	129.57	123.24	123.43
	Rank	10	3	1	4	5	2	11	9	8	6	7
	Average Rank by Series	8.55	4.13	3.58	4.05	3.60	4.55	10.95	7.38	7.33	5.85	6.05
	Rank of Average Rank	10	4	1	3	2	5	11	9	8	6	7
5	Geometric Mean	283.27	179.55	174.46	181.34	166.79	170.83	444.21	226.23	220.26	215.54	216.31
	Rank	10	3	2	4	5	1	11	9	8	6	7
	Average Rank by Series	6.75	3.925	3.625	4.85	4.95	3.8	10.65	7.225	6.975	6.65	6.6
	Rank of Average Rank	8	3	1	4	5	2	11	10	9	7	6
10	Geometric Mean	383.43	325.76	316.52	314.77	336.27	306.71	487.93	259.48	246.03	249.16	249.65
	Rank	10	8	7	6	9	5	11	4	1	2	3
	Average Rank by Series	5.25	7.075	6.775	7.5	7.4	6	8.7	4.375	3.925	4.55	4.45
	Rank of Average Rank	5	8	7	10	9	6	11	2	1	4	3
15	Geometric Mean	568.47	440.36	428.93	432.03	478.63	417.31	596.93	344.81	317.82	332.29	354.13
	Rank	10	8	6	7	9	5	11	3	1	2	4
	Average Rank by Series	6.05	7.175	6.925	7.5	8.3	5.95	6.55	4.125	3.775	4.6	5.05
	Rank of Average Rank	6	9	8	10	11	5	7	2	1	3	4

Trend Shift (N) Scenario 12 Table: 12-5

Period: Geometric Root Mean Squared Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive	HWW*	HW*	Adaptive*	Auto*	Naive*
1	Geometric Mean	104.70	44.84	41.88	50.72	47.54	47.17	145.42	66.28	61.19	65.05	65.00
	Rank	10	2	1	5	4	3	11	9	6	8	7
	Average Rank by Series	7.9	4.775	4.125	4.95	4.4	4.7	9.85	6.525	6.075	6.3	6.4
	Rank of Average Rank	10	4	1	5	2	3	11	9	6	7	8
5	Geometric Mean	180.15	126.63	117.78	129.24	128.09	131.85	254.65	142.34	134.14	135.90	127.34
	Rank	10	2	1	5	4	6	11	9	7	8	3
	Average Rank by Series	7	5.425	5.325	5.9	5.8	4.85	9.75	5.725	5.475	6	4.75
	Rank of Average Rank	10	4	3	8	7	2	11	6	5	9	1
10	Geometric Mean	246.01	243.08	225.72	251.57	259.11	243.24	330.04	163.21	151.41	160.84	168.98
	Rank	8	6	5	9	10	7	11	3	1	2	4
	Average Rank by Series	5.8	6.875	6.925	7.9	7.9	5.95	7.95	3.725	3.675	4.5	4.8
	Rank of Average Rank	5	7	8	9	9	6	11	2	1	3	4
15	Geometric Mean	369.81	352.12	330.01	352.28	386.81	366.35	421.56	233.16	220.00	239.84	265.29
	Rank	9	6	5	7	10	8	11	2	1	3	4
	Average Rank by Series	6.3	7.075	7.025	7.6	7.85	6.55	6.75	3.425	3.475	4.7	5.25
	Rank of Average Rank	5	9	8	10	11	6	7	1	2	3	4

	Trend Shift (N)	Average Rank of Absolute Error		
	Scenario 12	Table:	12	-6
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	38.01	19	0.0059
	KRUSKAL-WALLIS	75.35	10	0.0000
5	RANK ANOVA	20.65	19	0.3562
	KRUSKAL-WALLIS	36.84	10	0.0001
10	RANK ANOVA	18.16	19	0.5116
	KRUSKAL-WALLIS	34.48	10	0.0002
15	RANK ANOVA	17.04	19	0.5874
	KRUSKAL-WALLIS	34.48	10	0.0002

	Trend Shift (N)		Log Mean Squared Error Ratio		
		Scenario 12	Table:	12	-10
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	31.05	19	0.0398
		KRUSKAL-WALLIS	41.67	10	0.0000
	5	RANK ANOVA	20.46	19	0.3674
		KRUSKAL-WALLIS	30.04	10	0.0008
	10	RANK ANOVA	28.59	19	0.0728
		KRUSKAL-WALLIS	35.62	10	0.0001
	15	RANK ANOVA	25.93	19	0.1321
		KRUSKAL-WALLIS	32.58	10	0.0003

	Trend Shift (N)	Range of Percent Error		
	Scenario 12	Table:	12	-7
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	37.27	19	0.0073
	KRUSKAL-WALLIS	31.06	10	0.0006
5	RANK ANOVA	38.69	19	0.0048
	KRUSKAL-WALLIS	49.43	10	0.0000
10	RANK ANOVA	18.93	19	0.4615
	KRUSKAL-WALLIS	35.28	10	0.0001
15	RANK ANOVA	17.23	19	0.5740
	KRUSKAL-WALLIS	36.58	10	0.0001

	Trend Shift (N)		Median Absolute Percent Error		
		Scenario 12	Table:	12	-11
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	31.64	19	0.0343
		KRUSKAL-WALLIS	32.17	10	0.0004
	5	RANK ANOVA	25.19	19	0.1543
		KRUSKAL-WALLIS	38.71	10	0.0000
	10	RANK ANOVA	17.79	19	0.5367
		KRUSKAL-WALLIS	17.65	10	0.0612
	15	RANK ANOVA	16.79	19	0.6042
		KRUSKAL-WALLIS	12.24	10	0.2691

	Trend Shift (N)	Symmetry Adjusted MAPE		
	Scenario 12	Table:	12	-8
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	54.23	19	0.0000
	KRUSKAL-WALLIS	48.68	10	0.0000
5	RANK ANOVA	46.90	19	0.0004
	KRUSKAL-WALLIS	63.06	10	0.0000
10	RANK ANOVA	30.54	19	0.0453
	KRUSKAL-WALLIS	33.02	10	0.0003
15	RANK ANOVA	23.28	19	0.2251
	KRUSKAL-WALLIS	23.27	10	0.0098

	Trend Shift (N)		Mean Absolute Percent Error		
		Scenario 12	Table:	12	-12
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	42.45	19	0.0015
		KRUSKAL-WALLIS	40.75	10	0.0000
	5	RANK ANOVA	37.62	19	0.0066
		KRUSKAL-WALLIS	45.57	10	0.0000
	10	RANK ANOVA	24.01	19	0.1957
		KRUSKAL-WALLIS	22.98	10	0.0108
	15	RANK ANOVA	22.48	19	0.2609
		KRUSKAL-WALLIS	20.02	10	0.0291

	Trend Shift (N)	Geometric Root Mean Squared Error		
	Scenario 12	Table:	12	-9
Period:		Chi Squared	DF	p Value
1	RANK ANOVA	31.05	19	0.0398
5	RANK ANOVA	20.46	19	0.3674
10	RANK ANOVA	28.59	19	0.0728
15	RANK ANOVA	25.93	19	0.1321

Trend Shift (N)	Root Mean Squared Error				
		Scenario 12	Table:	12	-13
Period:	Chi Squared		DF	p Value	
	1	RANK ANOVA	58.49	19	0.0000
	5	RANK ANOVA	44.87	19	0.0007
	10	RANK ANOVA	26.83	19	0.1086
	15	RANK ANOVA	22.83	19	0.2451

Historical Level Shift — Table: 13-1

Scenario 13 Average Rank of Absolute Error

Period:		Adjusted	HWW	HW	Adaptive	Auto	Naive
1	Average	3.61	3.49	3.49	3.59	3.26	3.16
	Rank	6	4	3	5	2	1
	Geometric Mean	3.49	3.41	3.42	3.43	3.20	3.02
	Rank	6	3	4	5	2	1
	Average Rank by Series	4.00	3.68	3.45	3.80	3.18	2.90
	Rank of Average Rank	6	4	3	5	2	1
	Kruskal-Wallis Rank Sum	1385.5	1261.5	1237.0	1331.0	1120.5	924.5
	Rank of K-W Rank Sum	6	4	3	5	2	1
	K-W Multi-Comparison Count*	5	4	4	5	5	5
5	Average	3.11	3.56	3.49	3.68	3.37	3.39
	Rank	1	5	4	6	2	3
	Geometric Mean	2.88	3.42	3.39	3.43	3.27	3.22
	Rank	1	5	4	6	3	2
	Average Rank by Series	2.95	3.68	3.63	3.98	3.30	3.48
	Rank of Average Rank	1	5	4	6	2	3
	Kruskal-Wallis Rank Sum	987.5	1266.5	1239.0	1393.0	1184.5	1189.5
	Rank of K-W Rank Sum	1	5	4	6	2	3
	K-W Multi-Comparison Count*	5	4	4	5	4	4
10	Average	3.08	3.64	3.56	3.58	3.23	3.51
	Rank	1	6	4	5	2	3
	Geometric Mean	2.84	3.48	3.43	3.39	3.11	3.25
	Rank	1	6	5	4	2	3
	Average Rank by Series	2.90	3.80	3.68	3.78	3.05	3.80
	Rank of Average Rank	1	5	3	4	2	5
	Kruskal-Wallis Rank Sum	984.0	1337.5	1284.0	1301.0	1099.0	1254.5
	Rank of K-W Rank Sum	1	6	4	5	2	3
	K-W Multi-Comparison Count*	5	4	3	3	5	4
15	Average	3.11	3.64	3.64	3.51	3.19	3.52
	Rank	1	5	5	3	2	4
	Geometric Mean	2.88	3.45	3.45	3.28	3.06	3.16
	Rank	1	5	6	4	2	3
	Average Rank by Series	3.38	3.80	3.65	3.55	2.95	3.68
	Rank of Average Rank	2	6	4	3	1	5
	Kruskal-Wallis Rank Sum	1050.0	1310.0	1297.0	1259.5	1078.5	1265.0
	Rank of K-W Rank Sum	1	6	5	3	2	4
	K-W Multi-Comparison Count*	5	4	2	3	5	3

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Historical Level Shift		Table:	13-2			
Period:	Scenario 13	Range of Percent Error				
	Adjusted	HWW	HW	Adaptive	Auto	Naive
1 Average	8.69%	8.44%	8.39%	8.56%	8.23%	9.46%
Rank	5	3	2	4	1	6
Geometric Mean	4.76%	4.85%	4.84%	4.75%	4.58%	4.84%
Rank	3	6	5	2	1	4
Average Rank by Series	3.45	3.80	3.55	3.15	3.20	3.85
Rank of Average Rank	3	5	4	1	2	6
Kruskal-Wallis Rank Sum	1218.0	1209.0	1203.0	1220.0	1179.0	1231.0
Rank of K-W Rank Sum	4	3	2	5	1	6
K-W Multi-Comparison Count*	1	0	0	0	2	1
5 Average	17.13%	18.02%	17.94%	16.79%	17.64%	21.11%
Rank	2	5	4	1	3	6
Geometric Mean	8.03%	7.62%	7.77%	7.55%	7.83%	7.25%
Rank	6	3	4	2	5	1
Average Rank by Series	3.55	3.40	3.35	3.60	3.45	3.65
Rank of Average Rank	4	2	1	5	3	6
Kruskal-Wallis Rank Sum	1247.0	1194.0	1203.0	1215.0	1223.0	1178.0
Rank of K-W Rank Sum	6	2	3	4	5	1
K-W Multi-Comparison Count*	4	1	1	0	2	2
10 Average	19.66%	21.53%	21.52%	19.36%	20.39%	16.29%
Rank	3	6	5	2	4	1
Geometric Mean	9.03%	8.75%	8.80%	8.50%	8.82%	8.36%
Rank	6	3	4	2	5	1
Average Rank by Series	3.20	3.40	3.05	3.45	3.80	4.10
Rank of Average Rank	2	3	1	4	5	6
Kruskal-Wallis Rank Sum	1241.0	1207.0	1210.0	1172.0	1231.0	1199.0
Rank of K-W Rank Sum	6	3	4	1	5	2
K-W Multi-Comparison Count*	2	0	0	2	2	0
15 Average	15.98%	17.49%	17.08%	15.72%	14.26%	13.21%
Rank	4	6	5	3	2	1
Geometric Mean	9.09%	7.94%	7.89%	7.46%	7.11%	7.08%
Rank	6	5	4	3	2	1
Average Rank by Series	3.20	3.85	3.55	3.70	3.30	3.40
Rank of Average Rank	1	6	4	5	2	3
Kruskal-Wallis Rank Sum	1297.0	1230.0	1232.0	1177.0	1161.0	1163.0
Rank of K-W Rank Sum	6	4	5	3	1	2
K-W Multi-Comparison Count*	5	4	4	3	3	3

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Historical Level Shift Table: 13-3

Period: Scenario 13 Mean Absolute Percent Error

Period		Adjusted	HWW	HW	Adaptive	Auto	Naive
1	Average	5.06%	5.71%	5.71%	5.00%	5.41%	4.59%
	Rank	3	5	6	2	4	1
	Geometric Mean	2.83%	3.00%	3.01%	2.90%	2.87%	2.65%
	Rank	2	5	6	4	3	1
	Average Rank by Series	3.80	3.90	3.80	4.05	2.80	2.65
	Rank of Average Rank	3	5	3	6	2	1
	Kruskal-Wallis Rank Sum	1219.0	1231.0	1230.0	1246.0	1185.0	1149.0
	Rank of K-W Rank Sum	3	5	4	6	2	1
	K-W Multi-Comparison Count*	2	2	2	2	4	4
5	Average	7.56%	11.64%	11.57%	8.12%	11.69%	8.56%
	Rank	1	5	4	2	6	3
	Geometric Mean	4.70%	5.74%	5.69%	4.99%	5.90%	5.18%
	Rank	1	5	4	2	6	3
	Average Rank by Series	3.05	3.90	3.80	4.05	3.15	3.05
	Rank of Average Rank	1	5	4	6	3	1
	Kruskal-Wallis Rank Sum	1166.0	1241.0	1229.0	1172.0	1248.0	1204.0
	Rank of K-W Rank Sum	1	5	4	2	6	3
	K-W Multi-Comparison Count*	3	2	2	3	3	1
10	Average	10.04%	19.15%	19.18%	11.19%	18.84%	11.19%
	Rank	1	5	6	2	4	3
	Geometric Mean	6.10%	8.33%	8.44%	6.50%	7.96%	7.59%
	Rank	1	5	6	2	4	3
	Average Rank by Series	3.00	3.95	3.95	3.60	2.90	3.60
	Rank of Average Rank	2	5	5	3	1	3
	Kruskal-Wallis Rank Sum	1133.0	1263.0	1269.0	1137.0	1215.0	1243.0
	Rank of K-W Rank Sum	1	5	6	2	3	4
	K-W Multi-Comparison Count*	4	3	3	4	4	2
15	Average	10.72%	24.83%	24.95%	12.73%	24.59%	12.11%
	Rank	1	5	6	3	4	2
	Geometric Mean	6.86%	9.49%	9.87%	7.41%	8.67%	8.86%
	Rank	1	5	6	2	3	4
	Average Rank by Series	3.20	3.95	3.75	3.45	2.95	3.70
	Rank of Average Rank	2	6	5	3	1	4
	Kruskal-Wallis Rank Sum	1137.0	1247.0	1274.0	1134.0	1194.0	1274.0
	Rank of K-W Rank Sum	2	4	5	1	3	5
	K-W Multi-Comparison Count*	4	3	3	4	5	3

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Historical Level Shift Table: 13-4

Period: Scenario 13 Root Mean Square Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive
1	Geometric Mean	32.05	34.64	34.72	33.65	32.95	31.75
	Rank	2	5	6	4	3	1
	Average Rank by Series	3.60	3.90	4.00	3.90	2.60	3.00
	Rank of Average Rank	3	4	6	4	1	2
5	Geometric Mean	54.77	65.19	65.48	57.23	66.90	58.68
	Rank	1	4	5	2	6	3
	Average Rank by Series	2.95	3.80	3.65	4.05	3.00	3.55
	Rank of Average Rank	1	5	4	6	2	3
10	Geometric Mean	71.07	94.66	95.84	75.33	92.20	86.73
	Rank	1	5	6	2	4	3
	Average Rank by Series	2.85	3.85	3.85	3.70	3.05	3.70
	Rank of Average Rank	1	5	5	3	2	3
15	Geometric Mean	81.17	110.33	114.70	86.87	100.71	101.90
	Rank	1	5	6	2	3	4
	Average Rank by Series	3.15	4.05	3.80	3.50	2.85	3.65
	Rank of Average Rank	2	6	5	3	1	4

Historical Level Shift Table: 13-5

Period: Scenario 13 Geometric Root Mean Square Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive
1	Geometric Mean	21.64	22.64	22.02	20.94	22.45	19.21
	Rank	3	6	4	2	5	1
	Average Rank by Series	3.75	3.60	3.45	3.70	3.45	3.05
	Rank of Average Rank	6	4	2	5	2	1
5	Geometric Mean	34.24	43.61	39.54	38.87	43.82	0.00
	Rank	2	5	4	3	6	1
	Average Rank by Series	3.05	3.95	3.60	3.95	3.05	3.40
	Rank of Average Rank	1	5	4	5	1	3
10	Geometric Mean	44.41	64.44	64.86	50.74	59.01	0.00
	Rank	2	5	6	3	4	1
	Average Rank by Series	3.15	3.75	3.75	3.8	3.1	3.45
	Rank of Average Rank	2	4	4	6	1	3
15	Geometric Mean	56.89	79.64	82.64	64.49	75.22	80.18
	Rank	1	4	6	2	3	5
	Average Rank by Series	3.30	3.85	3.70	3.40	3.05	3.70
	Rank of Average Rank	2	6	4	3	1	4

Historical Level Shift		Average Rank of Absolute Error		Table 13- 6
Period	Scenario 13	Chi Square	DF	p Value
1	RANK ANOVA	3.00	19	1.0000
	KRUSKAL-WALLIS	5.72	5	0.3348
5	RANK ANOVA	2.57	19	1.0000
	KRUSKAL-WALLIS	3.64	5	0.6022
10	RANK ANOVA	3.02	19	1.0000
	KRUSKAL-WALLIS	3.94	5	0.5578
15	RANK ANOVA	2.28	19	1.0000
	KRUSKAL-WALLIS	2.72	5	0.7423

Historical Level Shift		Symmetry Adjusted MAPE		Table 13- 7
	Scenario 13	Chi Square	DF	p Value
1	RANK ANOVA	4.95	19	0.9995
	KRUSKAL-WALLIS	0.24	5	0.9986
5	RANK ANOVA	3.30	19	1.0000
	KRUSKAL-WALLIS	0.23	5	0.9987
10	RANK ANOVA	3.32	19	1.0000
	KRUSKAL-WALLIS	0.73	5	0.9814
15	RANK ANOVA	2.57	19	1.0000
	KRUSKAL-WALLIS	0.88	5	0.9716

Historical Level Shift		Range of Percent Error		Table 13-8
	Scenario 13	Chi Square	DF	p Value
1	RANK ANOVA	2.22	19	1.0000
	KRUSKAL-WALLIS	0.07	5	0.9999
5	RANK ANOVA	1.53	19	1.0000
	KRUSKAL-WALLIS	0.12	5	0.9997
	RANK ANOVA	2.84	19	1.0000
10	KRUSKAL-WALLIS	0.12	5	0.9997
	RANK ANOVA	1.98	19	1.0000
15	KRUSKAL-WALLIS	0.58	5	0.9887

Historical Level Shift		Geometric Root Mean Square Error		Table13-9
	Scenario 13	Chi Square	DF	p Value
1	RANK ANOVA	2.01	19	1.0000
5	RANK ANOVA	2.98	19	1.0000
10	RANK ANOVA	2.35	19	1.0000
15	RANK ANOVA	2.27	19	1.0000

Historical Level Shift		Log Mean Square Error Ratio		Table 13 - 10
Period	Scenario 13	Chi Square	DF	p Value
1	RANK ANOVA	2.01	19	1.0000
	KRUSKAL-WALLIS	1.90	5	0.8624
5	RANK ANOVA	3.07	19	1.0000
	KRUSKAL-WALLIS	2.17	5	0.8246
10	RANK ANOVA	2.21	19	1.0000
	KRUSKAL-WALLIS	1.52	5	0.9107
15	RANK ANOVA	2.27	19	1.0000
	KRUSKAL-WALLIS	1.18	5	0.9464

Historical Level Shift		Mean Absolute Percent Error		Table 13 - 11
	Scenario 13	Chi Square	DF	p Value
1	RANK ANOVA	4.93	19	0.9995
	KRUSKAL-WALLIS	0.27	5	0.9981
5	RANK ANOVA	3.46	19	1.0000
	KRUSKAL-WALLIS	0.26	5	0.9984
10	RANK ANOVA	3.37	19	1.0000
	KRUSKAL-WALLIS	0.77	5	0.9788
15	RANK ANOVA	2.73	19	1.0000
	KRUSKAL-WALLIS	0.86	5	0.9727

Historical Level Shift		Median Absolute Percent Error		Table 13 - 12
Scenario 13		Chi Square	DF	p Value
1	RANK ANOVA	2.34	19	1.0000
	KRUSKAL-WALLIS	0.47	5	0.9931
5	RANK ANOVA	2.24	19	1.0000
	KRUSKAL-WALLIS	0.17	5	0.9994
10	RANK ANOVA	3.56	19	1.0000
	KRUSKAL-WALLIS	1.12	5	0.9526
15	RANK ANOVA	3.08	19	1.0000
	KRUSKAL-WALLIS	1.16	5	0.9484

Historical Level Shift		Root Mean Square Error		Table 13 - 13
Scenario 13		Chi Square	DF	p Value
1	RANK ANOVA	4.52	19	0.9997
5	RANK ANOVA	3.25	19	1.0000
10	RANK ANOVA	3.21	19	1.0000
15	RANK ANOVA	3.23	19	1.0000

Historical Level Shift (Restricted) Table: 13B-1

Period:		Scenario 13b Adjusted	HWW	Average Rank of Absolute Error HW	Adaptive	Auto	Naive
1	Average	3.62	3.53	3.49	3.59	3.26	3.16
	Rank	6	4	3	5	2	1
	Geometric Mean	3.49	3.44	3.43	3.43	3.20	3.02
	Rank	6	5	4	3	2	1
	Average Rank by Series	3.95	3.73	3.45	3.80	3.18	2.90
	Rank of Average Rank	6	4	3	5	2	1
	Kruskal-Wallis Rank Sum	1372.0	1292.5	1239.0	1324.5	1114.0	927.5
	Rank of K-W Rank Sum	6	4	3	5	2	1
	K-W Multi-Comparison Count*	5	4	5	4	5	5
5	Average	2.91	3.62	3.46	3.73	3.45	3.47
	Rank	1	5	3	6	2	4
	Geometric Mean	2.71	3.49	3.34	3.52	3.35	3.28
	Rank	1	5	3	6	4	2
	Average Rank by Series	2.65	3.80	3.55	4.08	3.40	3.53
	Rank of Average Rank	1	5	4	6	2	3
	Kruskal-Wallis Rank Sum	860.0	1312.5	1224.5	1414.0	1233.5	1231.0
	Rank of K-W Rank Sum	1	5	2	6	4	3
	K-W Multi-Comparison Count*	5	5	3	5	3	3
10	Average	2.85	3.72	3.57	3.63	3.30	3.58
	Rank	1	6	3	5	2	4
	Geometric Mean	2.59	3.57	3.42	3.45	3.17	3.30
	Rank	1	6	4	5	2	3
	Average Rank by Series	2.75	3.88	3.60	3.85	3.10	3.83
	Rank of Average Rank	1	6	3	5	2	4
	Kruskal-Wallis Rank Sum	850.5	1381.0	1289.0	1335.5	1140.0	1281.5
	Rank of K-W Rank Sum	1	6	4	5	2	3
	K-W Multi-Comparison Count*	5	5	4	5	5	4
15	Average	2.87	3.71	3.66	3.57	3.26	3.57
	Rank	1	6	5	3	2	3
	Geometric Mean	2.68	3.53	3.47	3.35	3.13	3.19
	Rank	1	6	5	4	2	3
	Average Rank by Series	3.08	3.90	3.58	3.68	3.05	3.73
	Rank of Average Rank	2	6	3	4	1	5
	Kruskal-Wallis Rank Sum	894.5	1355.5	1320.0	1298.0	1123.0	1282.5
	Rank of K-W Rank Sum	1	6	5	4	2	3
	K-W Multi-Comparison Count*	5	4	2	3	5	3

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Historical Level Shift (Restricted) Table: 13B-2

Period: Scenario 13b Range of Percent Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive
1	Average	8.60%	8.46%	8.88%	8.56%	8.23%	9.46%
	Rank	4	2	5	3	1	6
	Geometric Mean	4.71%	4.86%	4.97%	4.75%	4.58%	4.84%
	Rank	2	5	6	3	1	4
	Average Rank by Series	3.40	3.88	3.78	3.00	3.15	3.80
	Rank of Average Rank	3	6	4	1	2	5
	Kruskal-Wallis Rank Sum	1213.0	1213.0	1222.5	1216.0	1175.0	1228.0
	Rank of K-W Rank Sum	2	2	5	4	1	6
	K-W Multi-Comparison Count*	1	0	1	0	3	1
5	Average	16.16%	18.08%	18.00%	16.79%	17.64%	21.11%
	Rank	1	5	4	2	3	6
	Geometric Mean	7.37%	7.64%	7.80%	7.55%	7.83%	7.25%
	Rank	2	4	5	3	6	1
	Average Rank by Series	3.25	3.48	3.43	3.65	3.50	3.70
	Rank of Average Rank	1	3	2	5	4	6
	Kruskal-Wallis Rank Sum	1193.0	1199.0	1212.5	1219.0	1227.0	1188.0
	Rank of K-W Rank Sum	2	3	4	5	6	1
	K-W Multi-Comparison Count*	1	0	0	0	1	0
10	Average	17.33%	21.64%	19.99%	19.36%	20.39%	16.29%
	Rank	2	6	4	3	5	1
	Geometric Mean	7.99%	8.78%	8.61%	8.50%	8.82%	8.36%
	Rank	1	5	4	3	6	2
	Average Rank by Series	3.10	3.48	2.98	3.50	3.85	4.10
	Rank of Average Rank	2	3	1	4	5	6
	Kruskal-Wallis Rank Sum	1181.0	1215.0	1210.5	1174.0	1235.0	1209.0
	Rank of K-W Rank Sum	2	5	4	1	6	3
	K-W Multi-Comparison Count*	1	0	0	1	2	0
15	Average	12.75%	17.62%	15.19%	15.72%	14.26%	13.21%
	Rank	1	6	4	5	3	2
	Geometric Mean	8.07%	7.98%	7.76%	7.46%	7.11%	7.08%
	Rank	6	5	4	3	2	1
	Average Rank by Series	3.05	3.98	3.48	3.75	3.35	3.40
	Rank of Average Rank	1	6	4	5	2	3
	Kruskal-Wallis Rank Sum	1231.0	1241.0	1237.5	1178.0	1164.0	1174.0
	Rank of K-W Rank Sum	4	6	5	3	1	2
	K-W Multi-Comparison Count*	3	3	3	3	3	3

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Table: 13B-3

Historical Level Shift (Restricted)

Period: Scenario 13b Mean Absolute Percent Error

		Adjusted	HWW	HW	Adaptive	Auto	Naive
1	Average	5.12%	5.71%	5.50%	5.00%	5.41%	4.59%
	Rank	3	6	5	2	4	1
	Geometric Mean	2.87%	3.00%	2.96%	2.90%	2.87%	2.65%
	Rank	2	6	5	4	3	1
	Average Rank by Series	3.80	3.93	3.73	4.05	2.85	2.65
	Rank of Average Rank	4	5	3	6	2	1
	Kruskal-Wallis Rank Sum	1224.0	1232.0	1229.5	1246.0	1186.0	1148.0
	Rank of K-W Rank Sum	3	5	4	6	2	1
	K-W Multi-Comparison Count*	2	2	1	2	3	4
5	Average	6.90%	11.51%	11.09%	8.12%	11.69%	8.56%
	Rank	1	5	4	2	6	3
	Geometric Mean	4.37%	5.64%	5.44%	4.99%	5.90%	5.18%
	Rank	1	5	4	2	6	3
	Average Rank by Series	2.70	4.03	3.68	4.15	3.30	3.15
	Rank of Average Rank	1	5	4	6	3	2
	Kruskal-Wallis Rank Sum	1117.0	1229.0	1202.5	1180.0	1254.0	1222.0
	Rank of K-W Rank Sum	1	5	3	2	6	4
	K-W Multi-Comparison Count*	5	2	2	3	3	1
10	Average	8.53%	18.93%	18.30%	11.19%	18.84%	11.19%
	Rank	1	6	4	2	5	3
	Geometric Mean	5.42%	8.15%	8.01%	6.50%	7.96%	7.59%
	Rank	1	6	5	2	4	3
	Average Rank by Series	2.75	4.03	3.83	3.75	3.00	3.65
	Rank of Average Rank	1	6	5	4	2	3
	Kruskal-Wallis Rank Sum	1068.0	1251.0	1230.5	1149.0	1222.0	1269.0
	Rank of K-W Rank Sum	1	5	4	2	3	6
	K-W Multi-Comparison Count*	5	2	2	5	3	3
15	Average	8.85%	24.55%	23.78%	12.73%	24.59%	12.11%
	Rank	1	5	4	3	6	2
	Geometric Mean	6.17%	9.26%	9.32%	7.41%	8.67%	8.86%
	Rank	1	5	6	2	3	4
	Average Rank by Series	2.95	4.08	3.63	3.55	3.05	3.75
	Rank of Average Rank	1	6	4	3	2	5
	Kruskal-Wallis Rank Sum	1064.0	1225.0	1228.5	1143.0	1203.0	1306.0
	Rank of K-W Rank Sum	1	4	5	2	3	6
	K-W Multi-Comparison Count*	5	3	3	5	3	5

*K-W Multi-Comparison Count valid only if Kruskal-Wallis statistic is significant.

Historical Level Shift (Restricted) Table: 13B-4

Period:		Scenario 13b Adjusted	HWW	Root Mean Square Error HW	Adaptive	Auto	Naive
1	Geometric Mean	32.35	34.63	34.68	33.65	32.95	31.75
	Rank	2	5	6	4	3	1
	Average Rank by Series	3.60	3.93	3.93	3.90	2.65	3.00
	Rank of Average Rank	3	5	5	4	1	2
5	Geometric Mean	50.87	64.51	63.67	57.23	66.90	58.68
	Rank	1	5	4	2	6	3
	Average Rank by Series	2.60	3.93	3.53	4.15	3.15	3.65
	Rank of Average Rank	1	5	3	6	2	4
10	Geometric Mean	63.07	93.33	92.08	75.33	92.20	86.73
	Rank	1	6	4	2	5	3
	Average Rank by Series	2.50	3.98	3.73	3.80	3.20	3.80
	Rank of Average Rank	1	6	3	4	2	4
15	Geometric Mean	73.10	108.94	110.51	86.87	100.71	101.90
	Rank	1	5	6	2	3	4
	Average Rank by Series	2.80	4.18	3.73	3.55	3.00	3.75
	Rank of Average Rank	1	6	4	3	2	5

Historical Level Shift (Restricted) Table: 13B-5

Period:		Scenario 13b Adjusted	HWW	Geometric Root Mean Square Error HW	Adaptive	Auto	Naive
1	Geometric Mean	22.09	22.19	21.36	20.94	22.45	19.21
	Rank	4	5	3	2	6	1
	Average Rank by Series	3.80	3.53	3.38	3.70	3.50	3.10
	Rank of Average Rank	6	4	2	5	3	1
5	Geometric Mean	32.03	42.73	37.36	38.87	43.82	0.00
	Rank	2	5	3	4	6	1
	Average Rank by Series	3.05	3.98	3.53	4.00	3.05	3.40
	Rank of Average Rank	1	5	4	6	1	3
10	Geometric Mean	40.26	62.60	61.72	50.74	59.01	0.00
	Rank	2	6	5	3	4	1
	Average Rank by Series	3.15	3.725	3.675	3.9	3.1	3.45
	Rank of Average Rank	2	5	4	6	1	3
15	Geometric Mean	51.59	76.35	76.10	64.49	75.22	80.18
	Rank	1	5	4	2	3	6
	Average Rank by Series	3.05	3.93	3.63	3.55	3.10	3.75
	Rank of Average Rank	1	6	4	3	2	5

Historical Level Shift (R)		Average Rank of Absolute Error			Table 13B-6
Period Scenario 13b		Chi Square	DF	p Value	Period
1	RANK ANOVA	2.95	19	1.0000	
	KRUSKAL-WALLIS	5.61	5	0.3457	
5	RANK ANOVA	3.60	19	1.0000	
	KRUSKAL-WALLIS	7.25	5	0.2028	
10	RANK ANOVA	3.50	19	1.0000	
	KRUSKAL-WALLIS	7.85	5	0.1647	
15	RANK ANOVA	2.60	19	1.0000	
	KRUSKAL-WALLIS	6.32	5	0.2759	

Historical Level Shift (R)		Symmetry Adjusted MAPE			Table 13B- 7
	Scenario 13b	Chi Square	DF	p Value	
1	RANK ANOVA	4.77	19	0.9996	
	KRUSKAL-WALLIS	0.25	5	0.9985	
5	RANK ANOVA	4.13	19	0.9999	
	KRUSKAL-WALLIS	0.44	5	0.9942	
10	RANK ANOVA	3.83	19	0.9999	
	KRUSKAL-WALLIS	1.32	5	0.9331	
15	RANK ANOVA	3.04	19	1.0000	
	KRUSKAL-WALLIS	1.54	5	0.9084	

Historical Level Shift (R)		Range of Percent Error			Table 13B-8
	Scenario 13b	Chi Square	DF	p Value	
1	RANK ANOVA	2.71	19	1.0000	
	KRUSKAL-WALLIS	0.07	5	0.9999	
5	RANK ANOVA	1.65	19	1.0000	
	KRUSKAL-WALLIS	0.05	5	1.0000	
	RANK ANOVA	3.15	19	1.0000	
10	KRUSKAL-WALLIS	0.11	5	0.9998	
	RANK ANOVA	2.40	19	1.0000	
15	KRUSKAL-WALLIS	0.27	5	0.9982	

Historical Level Shift (R)		Geometric Root Mean Square Error			Table 13B- 9
	Scenario 13b	Chi Square	DF	p Value	
1	RANK ANOVA	1.98	19	1.0000	
5	RANK ANOVA	3.10	19	1.0000	
10	RANK ANOVA	2.40	19	1.0000	
15	RANK ANOVA	2.59	19	1.0000	

Historical Level Shift (R)		Log Mean Square Error Ratio			Table 13B- 10
Scenario 13b		Chi Square	DF	p Value	
1	RANK ANOVA	1.98	19	1.0000	
	KRUSKAL-WALLIS	1.84	5	0.8706	
5	RANK ANOVA	3.19	19	1.0000	
	KRUSKAL-WALLIS	3.06	5	0.6907	
10	RANK ANOVA	2.24	19	1.0000	
	KRUSKAL-WALLIS	2.43	5	0.7863	
15	RANK ANOVA	2.59	19	1.0000	
	KRUSKAL-WALLIS	2.52	5	0.7733	

Historical Level Shift (R)		Mean Absolute Percent Error			Table 13B- 11
Scenario 13b		Chi Square	DF	p Value	
1	RANK ANOVA	4.77	19	0.9996	
	KRUSKAL-WALLIS	0.28	5	0.9980	
5	RANK ANOVA	4.32	19	0.9998	
	KRUSKAL-WALLIS	0.48	5	0.9928	
10	RANK ANOVA	3.84	19	0.9999	
	KRUSKAL-WALLIS	1.20	5	0.9448	
15	RANK ANOVA	3.15	19	1.0000	
	KRUSKAL-WALLIS	1.43	5	0.9205	

Historical Level Shift (R)		Median Absolute Percent Error			Table 13B- 12
Scenario 13b		Chi Square	DF	p Value	
1	RANK ANOVA	2.45	19	1.0000	
	KRUSKAL-WALLIS	0.47	5	0.9931	
5	RANK ANOVA	2.13	19	1.0000	
	KRUSKAL-WALLIS	0.25	5	0.9985	
10	RANK ANOVA	3.84	19	0.9999	
	KRUSKAL-WALLIS	1.34	5	0.9307	
15	RANK ANOVA	3.57	19	1.0000	
	KRUSKAL-WALLIS	1.66	5	0.8945	

Historical Level Shift (R)		Root Mean Square Error			Table 13B- 13
Scenario 13b		Chi Square	DF	p Value	
1	RANK ANOVA	4.26	19	0.9998	
5	RANK ANOVA	4.37	19	0.9998	
10	RANK ANOVA	4.35	19	0.9998	
15	RANK ANOVA	3.90	19	0.9999	

Appendix V Forecast Evaluation Statistics

Descriptive Statistics

Following are the statistics that are used in this analysis. Citations are in the text.

Let:

X = Actual observation, (sometimes subscripted to time period i)

F = Forecast (sometimes subscripted to time period i)

$E_i = X_i - F_i$

Mean Squared Error

$MSE = \Sigma_t{}^n{}_{=1} E_i{}^2/n$

Root Mean Squared Error

$RMSE = \sqrt{MSE}$

Mean Absolute Percent Error

$MAPE = \Sigma_t{}^n{}_{=1} |E_i/X_i|/n * 100$

Percent Error

$PE_i = E_i/X_i * 100$

Mean Percent Error

$MPE = \Sigma_{i=1}^{n} PE_i/n$

Geometric Root Mean Squared Error

$GRMSE = [\Pi_t E^2{}_{it}(L)]^{1/2n}$

Where L = the number of steps ahead from T to the observation from which the error is calculated, n = T - L + 1, T = the t index value for the last actual observation, and i is the index of the series.

Log Mean Error Ratio (as compared with Naive 2).

Let m_{ij} denotes the mean squared forecast error of techniques j on series i. For this series, define the log mean squared error ratio as $lmr_{it} = \log(m_{i0}/m_{ij})$, where m_{i0} is the mean [squared] forecast error of some benchmark technique. Computed with the benchmark MSE in the numerator, a positive LMR indicates that technique j had a smaller forecast MSE on this series than the benchmark.

Symmetrical Mean Absolute Percent Error

$$SMAPE = \Sigma_{i=1}^{n}\{|E_i|/(F_t+X_t)\}/n$$

Symmetrical Percent Error

$$SPE_i = E_i/(F_t+X_t) * 100$$

Symmetrical Mean Percent Error

$$SMPE = \Sigma_{i=1}^{n} SPE_i/n$$

Median Absolute Percent Error (Median APE)

MdAPE = Observation (S+1)/2 for an odd number of observations or the average of S/2 and (S+1)/2 for an even number of observations, where the observations are the rank ordered average percent errors.

Average Rank

AR = (Summed Rank/Number of Observed Ranks).

Range of Percent Error

RPE = largest positive percent error minus largest negative percent error.

Inferential Statistics

Rank ANOVA

Let the most accurate prediction = rank 1 and the least accurate = rank n,
m = time periods (updates), and

R_{it} = the rank of the tth prediction for method i.

Summed ranks are calculated as:

$$S_i = \Sigma_{t=1}^{m} R_{it}, i = 1, 2, \ldots, n$$

Analysis of Variance by Rank (Friedman Test or Rank ANOVA)
The summed ranks can be compared using a chi-squared goodness of fit test with a prior expectation of 1/2 m(n+1) use chi-squared with using n-1 degrees of freedom:

$$\chi 2 = \Sigma i = {}_{i=1}^{n} [S_i - 1/2m(n+1)]^2 / [(n)m(n+1)/12]$$

Kruskal-Wallis Test

$$H = [12/(N(N+1))] * \Sigma_{i=1}^{k} [(1/n_i) * [W_i - .5n_i(n_i+1)]]^2$$

Where, $N = \Sigma_{k}^{i=1}(n_i)$ *and* n_i are the number of predictions for the i methods and W_i is the sum of the ranks for method i. Ranks are based on the absolute value of the forecast errors. H is a chi-squared variable with k-1 degrees of freedom.

Appendix VI Kruskal-Wallis and Analysis of Variance by Rank

The two non-parametric statistical tests produced overwhelmingly significant results except with scenario 13 where they proved not significant with all statistics and all trials. This last result led me to suspect that there might be something wrong with scenario 13 and after some investigation I came to suspect either (a) the models in scenario 13 were allowed to fit to excessive β parameters, or (b) that the statistics were sensitive to the number of treatments (models) considered. I examined the first suspicion by reducing the possible β parameter range and rerunning this trial. Results have been presented as scenario 13b and are not significantly different from scenario 13. I examined the second suspicion by excluding 5 non-*ad hoc* models from scenario 1 and calculating the Kruskal-Wallis and Rank ANOVA statistics for the rank of absolute error comparison (equivalent to Tables 1-1 and 1-7 in Appendix IV). These results are shown in tables 1 and 2.

Table 1 Inferential Statistics with Fewer Options

Period		Adjust	HWW*	HW*	Adapt*	Auto*	Naive*
1	Average Rank by Series	1.60	3.43	3.10	3.40	4.88	4.60
	Rank of Average Rank	1	4	2	3	6	5
	Kruskal-Wallis Rank Sum	372	1283	1069	1129	1709	1698
	Rank of K-W Rank Sum	1	4	2	3	6	5
	K-W Multi-Comparison Count*	5	5	5	5	4	4
5	Average Rank by Series	1.75	2.83	3.80	3.98	4.80	3.85
	Rank of Average Rank	1	2	3	5	6	4
	Kruskal-Wallis Rank Sum	415.5	903.5	1322.5	1402	1782.5	1434
	Rank of K-W Rank Sum	1	2	3	4	6	5
	K-W Multi-Comparison Count*	5	5	5	4	5	4
10	Average Rank by Series	1.55	2.975	3.825	4.175	4.075	4.4
	Rank of Average Rank	1	2	3	5	4	6
	Kruskal-Wallis Rank Sum	397.5	946.5	1348.5	1494	1531.5	1542
	Rank of K-W Rank Sum	1	2	3	4	5	6
	K-W Multi-Comparison Count*	5	5	5	4	3	4
15	Average Rank by Series	1.58	2.75	3.70	3.98	4.68	4.33
	Rank of Average Rank	1	2	3	4	6	5
	Kruskal-Wallis Rank Sum	383	802	1340.5	1497.5	1712	1525
	Rank of K-W Rank Sum	1	2	3	4	6	5
	K-W Multi-Comparison Count*	5	5	5	4	5	4

Table 2 Rank Anova and Kruskal-Wallis Results

Period		Chi Squared	DF	p value
1	RANK ANOVA	14.52	19	0.7528
	KRUSKAL-WALLIS	50.46	5	0.0000
5	RANK ANOVA	12.15	19	0.8789
	KRUSKAL-WALLIS	47.63	5	0.0000
10	RANK ANOVA	12.41	19	0.8674
	KRUSKAL-WALLIS	43.10	5	0.0000
15	RANK ANOVA	13.96	19	0.7859
	KRUSKAL-WALLIS	53.77	5	0.0000

For the Kruskal-Wallis statistic, the values of the statistics change (which should be expected), but the general results do not, that is, the statistics remained significant at the $\alpha = 0.05$ level. However, for the Rank ANOVA test, the statistics are no longer significant. This suggests that the significance of the previous results may be partly attributable to the use of a large number of treatments (forecast models). It is not clear whether this arises because of an increased number of observations or because of some unidentified bias that the tests bring into the analysis.

Following these explorations, I again reviewed the results of scenario 13 and found another reasonable explanation, which is that the actual summarized statistical results in scenario 13 did not vary very much. So, it seems that the lack of significance in scenario 13 as compared with the fairly strong statistical results in the other scenarios could result from the obvious statistical reason, that the different treatments in scenario 13 do not produce particularly different results.

Another problem with these statistics is that in some of the trials rank order was strong, but **inconsistent** between the various descriptive statistics. The non-parametric tests were not sensitive to these inconsistent results. The rank order results were statistically significant with extremely low p values both when the results were consistent between various descriptive statistics and when they were not. This suggests that these rank tests are not sufficient to distinguish superior and inferior forecast models by themselves, but that they may be useful as a supplement to the application of a **battery of**

descriptive statistics as presented in this dissertation. If the results are consistent across a battery of descriptive statistics **and** test significant with these tests, the researcher has reason to accept that the treatments are different. Statistical significance is a weaker result while significance without consistency is uninterpretable.

In this study, the examination of possible statistical testing of forecast treatments through non-parametric rank order tests was a secondary objective. These results should be considered exploratory. However, it appears that the application of either of these statistical tests in the manner described in this section has some promise when applied across a battery of descriptive statistics. Where results are consistent across the battery of descriptive statistics and the results are significant with one or both of these tests, as occurs with scenarios 1, 2, 8 and 9, it appears that the tests support each other and strengthen the conclusion that the differences in forecast treatments are more than just incidental. Where the results are less consistent across the battery of descriptive statistics, as with scenario 5, or where the statistical tests are not significant, as with scenario 13, results are not firmly supported by the study.

Vita

[redacted]