

**CALL CENTER DEMAND FORECASTING: IMPROVING SALES CALLS
PREDICTION ACCURACY THROUGH THE COMBINATION OF STATISTICAL
METHODS AND JUDGMENTAL FORECAST**

by

Juan Manuel Boulin

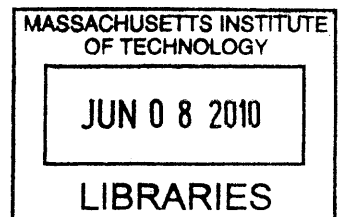
B.Sc. Electrical engineering, Universidad de Mendoza, 1999
B.Sc. Bioengineering, Universidad de Mendoza, 2000

ARCHIVES

Submitted to the MIT Sloan School of Management and the Engineering Systems Division in
Partial Fulfillment of the Requirements for the Degrees of

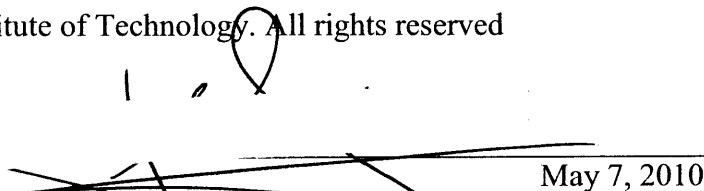
**Master of Business Administration
AND
Master of Science in Engineering Systems**

In conjunction with the Leaders for Global Operations at the
Massachusetts Institute of Technology
June 2010



© 2010 Massachusetts Institute of Technology. All rights reserved

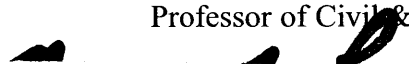
Signature of Author: _____


May 7, 2010
Department of Engineering Systems Division
MIT Sloan School of Management


Certified by _____


Roy E. Welsch
Professor of Statistics and Management Science, Director CCREMS
Thesis Supervisor


Certified by _____


David Simchi-Levi
Professor of Civil & Environmental Engineering and ESD
Thesis Supervisor

Accepted by _____


Nancy Leveson
Education Committee Chair, Engineering Systems Division
Professor, Aeronautics and Astronautics, Engineering Systems Division

Accepted by _____


Debbie Berechman
Executive Director of MBA Program, MIT Sloan School of Management

This page has been intentionally left blank.

**CALL CENTER DEMAND FORECASTING: IMPROVING SALES CALLS
PREDICTION ACCURACY THROUGH THE COMBINATION OF STATISTICAL
METHODS AND JUDGMENTAL FORECAST**

by
Juan Manuel Boulin

Submitted to the MIT Sloan School of Management and the Engineering Systems Division on
May 7th, 2010 in Partial Fulfillment of the Requirements for the Degrees for the Degrees of
Master of Business Administration and Master of Science in Engineering Systems

ABSTRACT

Call centers are important for developing and maintaining healthy relationships with customers. At Dell, call centers are also at the core of the company's renowned direct model. For sales call centers in particular, the impact of proper operations is reflected not only in long-term relationships with customers, but directly on sales and revenue. Adequate staffing and proper scheduling are key factors for providing an acceptable service level to customers. In order to staff call centers appropriately to satisfy demand while minimizing operating expenses, an accurate forecast of this demand (sales calls) is required. During fiscal year 2009, inaccuracies in consumer sales call volume forecasts translated into approximately \$1.1M in unnecessary overtime expenses and \$34.5M in lost revenue for Dell.

This work evaluates different forecasting techniques and proposes a comprehensive model to predict sales call volume based on the combination of ARIMA models and judgmental forecasting. The proposed methodology improves the accuracy of weekly forecasted call volume from 23% to 46% and of daily volume from 27% to 41%. Further improvements are easily achievable through the adjustment and projection processes introduced herein that rely on contextual information and the expertise of the forecasting team.

Thesis Supervisor: Roy E. Welsch
Title: Professor of Statistics and Management Science, Director CCREMS

Thesis Supervisor: David Simchi-Levi
Title: Professor of Civil & Environmental Engineering and ESD

This page has been intentionally left blank.

ACKNOWLEDGEMENTS

I would like to thank all the people at Dell that provided guidance and support throughout the internship. In particular, to the people without whom this work would have not been possible: Joel Drummond, David McGuire, Eric North and Damián Fernández.

To the Leaders for Global Operations staff, faculty and students for helping create an incredible learning environment and being a constant source of knowledge and support. These two years have been an amazing educational experience that I will take with me throughout my career.

To my thesis advisors, Roy Welsch and David Simchi-Levi for their invaluable assistance and mentorship. I am particularly grateful to Roy for taking time from his busy schedule to visit me in Austin.

To all the people that provided me with support, encouragement and inspiration throughout the years. To my brothers and sisters, my friends Adriana, Jeff, Gustavo and Victor and particularly to my uncle Fernando for his permanent support and motivation.

I am most grateful to my girlfriend Ana who endured the countless hours I dedicated to this work twice; first during the time I spent in Austin and then when I was home.

Finally, I would like to thank my parents for being my role models and a constant source of inspiration, especially when I most needed it. This work is dedicated to them.

This page has been intentionally left blank.

"I have seen the future and it is very much like the present, only longer."

Kehlog Albran, The Profit

"My concern is with the future since I plan to spend the rest of my life there."

Charles F. Kettering

"Prediction is very difficult, especially if it is about the future."

Nils Bohr, Nobel laureate in Physics

NOTE ON PROPRIETARY INFORMATION

In order to protect proprietary Dell information, the data presented throughout this thesis have been altered and do not represent actual values used by Dell, Inc. Any dollar values, product names and/or contact center data have been disguised, altered, or converted to percentages in order to protect competitive information.

CONTENTS

1. Introduction.....	13
1.1. Company background	13
1.2. Problem statement.....	13
1.2.1. Call forecasting process at Dell U.S. consumer sales.....	15
1.2.2. Challenges on the forecasting process	17
1.3. Focus of this thesis	19
1.3.1. Weekly volume forecasting model	19
1.3.2. Day of week distribution.....	20
1.3.3. Intra-quarter updates	20
2. Literature review	21
3. Forecasting models.....	23
3.1. Data available for modeling.....	23
3.1.1. Call data	23
3.1.2. Marketing data	26
3.1.3. Price data.....	28
3.1.4. Other data.....	29
3.2. Forecasting with Multiple Linear Regression	30
3.2.1. Multiple linear regression models.....	32
3.2.2. Regression with lagged dependent variables	37
3.2.3. Using autoregressive terms	39
3.2.4. Predicting residuals to reduce serial correlation	40
3.2.5. Summary on multiple linear regression models.....	42
3.3. Forecasting with Time Series.....	43
3.3.1. ARIMA models.....	44
3.3.2. Using SAS Forecast Studio.....	46
3.3.3. The need for judgmental adjustment.....	49
3.3.4. Adjusting time series (ARIMA) models	51

3.3.5. Summary on time series models	54
4. Day of week distribution	57
4.1. Normal weeks.....	59
4.2. Holiday weeks	60
4.3. Seasonality	62
4.4. Intra-quarter trends.....	63
4.5. Summary on day of week distribution	64
5. Intra-quarter projection process	67
6. Results of forecasting model	71
7. Future work.....	75
8. Acronyms	77
9. Bibliography	79

TABLES AND FIGURES

Table 1. Model I Regression Output.....	32
Table 2. Model II Regression Output	33
Table 3. Durbin-Watson critical values for Models I and II.....	36
Table 4. Model Ia (including lag calls term) Regression Output.....	38
Table 5. Model Ib (including autoregressive term) Regression Output.....	40
Table 6. Model Ic (including ARIMA term) Regression Output.....	41
Figure 1. Master Sales Plan (MSP) forecast accuracy, FY09.....	15
Figure 2. Live Voice Recognition (LVR) call breakdown	24
Figure 3. Sales Calls Data.....	24
Figure 4. Sales call volume by week, dataset used for modeling	25
Figure 5. Selected response decay rates.....	28
Figure 6. Model I Residual plots.....	33
Figure 7. Model II Residual plots	34
Figure 8. Models I and II forecasts, FY10 Q1	35
Figure 9. Models I and II forecast accuracy, FY10 Q1	35
Figure 10. Sales calls historical data FY02-FY08	45
Figure 11. Baseline forecast, FY10 Q1.....	48
Figure 12. Calls received as percentage of forecast, FY10 Q1.....	48
Figure 13. Adjusted Baseline Forecast, FY10 Q1	52
Figure 14. Calls received as percentage of adjusted forecast, FY10 Q1	52
Figure 15. Baseline adjustment template	54
Figure 16. Day of week distribution accuracy, Master Sales Plan	58
Figure 17. DOW Dashboard, Normal weeks	60
Figure 18. Sample call distribution on holiday weeks.....	61
Figure 19. DOW Dashboard, Holiday weeks	62
Figure 20. DOW Dashboard, Seasonality.....	63
Figure 21. DOW Dashboard, intra-quarter trends	64

Figure 22. Hit rate improvement from day of week distribution.....	65
Figure 23. Actual calls as percentage of MSP and projection	68
Figure 24. MSP forecast and projection hit rates.....	68
Figure 25. Weekly forecast accuracy comparison	71
Figure 26. Weekly forecast hit rate comparison	72
Figure 27. Daily forecast accuracy comparison.....	73
Figure 28. Daily forecast hit rate comparison.....	73

1. Introduction

1.1. Company background

Founded by a teenager in a college dorm room, Dell revolutionized and transformed an entire industry. Michael Dell's story, the development of the direct model and the company's rapid growth to a leading position in the computer industry is subject of numerous articles and books. Dell is a Fortune 50 company with global presence and revenues of approximately \$60 billion.

Dell is organized around corporate and consumer divisions. Although representing a relatively small share of revenue, around 20%, Dell's consumer group is seen as a very large growth opportunity. Dell targets consumers through three main channels, retail, online and offline. End customer relationship in the retail channel is largely managed by individual retailers like Wal-Mart or Best Buy, an interesting digression from the company's direct model.

Online and offline channels represent the majority of Dell consumer sales efforts, with offline contributing approximately 40% of the revenue. However, cost structures of online and offline groups are significantly different. In an industry known for razor-thin margins the management of these costs is particularly important.

1.2. Problem statement

Sales call centers are at the core of Dell's direct sales model. Call center operations are a delicate balance between customer experience and cost. Proper staffing on a call center insures customers are assisted promptly, avoiding long waiting times and dropped calls. At sales call centers in particular, dropped calls are critical, since each one constitutes a potential lost sale. At the same time, overstaffing might jeopardize profitability since staffing cost is one of the most important components of the cost structure. Antipov (Antipov & Meade, 2002) estimates that staffing expenses constitutes about 65% of a typical call center operating cost.

Scheduling of sales representatives at Dell's Round Rock sales call center is done two weeks in advance. This seems to be typical for the majority of Dell's call centers. Staffing however, must be done well in advance since newly hired representatives need to undergo training before being able to answer calls. The length of this training varies, but thirteen weeks is

typical for newly hired representatives with no previous experience. Dell also outsources some call volume to specialized companies. The contracts signed with these outsource providers (OSP) are based on forecasted call volume and they are typically locked between four and six weeks in advance.

Accurate call volume forecasting then, plays a major role in call center staffing and scheduling and it is of key importance to call center operations. At Dell, forecast accuracy is typically expressed as a percentage of volume forecasted in the Master Sales Plan (MSP):

$$\text{Forecast accuracy} = \frac{\text{Received calls}}{\text{Forecasted calls}} \times 100.$$

Call centers offer some flexibility regarding forecast accuracy. Some levers have been identified that help increase the call volume handled up to 110% of plan without major impact on customer satisfaction. These levers include, for example, shutting down the online chat service on days where demand exceeds expected volume. Chat services offer help to online customers but are not considered revenue-generating activities. On days of high demand, chat agents can be allocated to handle sales calls, effectively increasing the number of sales agents available. In a similar way, some activities help minimize the impact on operating cost when the call volume offered is lower than expected, up to 90% of original plan.

Thus the forecasted call volume for a given day is considered accurate if offered calls for that day fall between 90% and 110% of the forecasted value. Beyond these limits, profitability and customer experience might be compromised.

At Dell the same concept of forecast accuracy is used to measure the percentage of days, within a period of time, on which the call forecast fell within this 90% to 110% range. Thus if during a week only one day the volume of calls received fall within this band, we would call the forecast to be $1/7 = 14\%$ accurate. In order to avoid confusion, we will use a different metric to measure this accuracy. For this work, we will use hit rate to measure the percentage of days where the daily forecast was accurate. In this way, we will refer to forecast accuracy when comparing forecasted call volume to received calls and we will refer to hit rate accuracy when comparing number of days when the forecast was accurate to number of days where the call volume received fell outside the 90% to 110% band.

During the last fiscal year at Dell (FY09, Feb 08 – Jan 09) the hit rate was 27%, which means that sales calls received fell within this band only 97 days of the 362 days. Figure 1 shows daily forecast accuracy for FY09, expressed as calls received as a percentage of forecast.

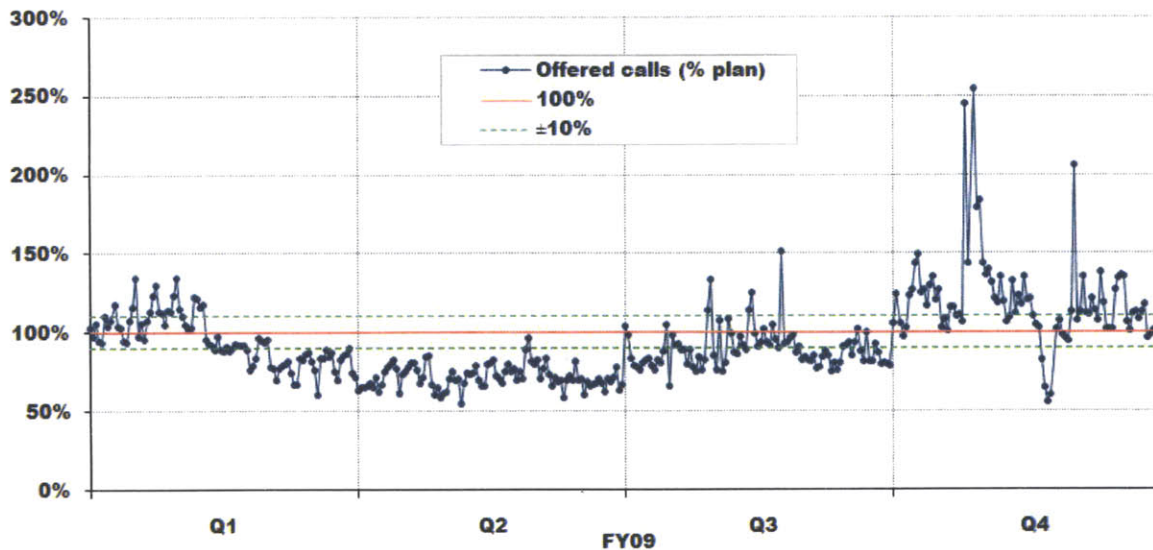


Figure 1. Master Sales Plan (MSP) forecast accuracy, FY09

An estimate of forecast accuracy related expenses shows that Dell spent \$1.1M in unnecessary overtime expenses, incurred during the 186 days where forecast accuracy was below 90%. Unanswered calls on the 79 days where forecast accuracy exceeded 110% represent an estimated lost revenue of \$34.5M¹.

1.2.1. Call forecasting process at Dell U.S. consumer sales

The last week of each quarter a Master Sales Plan (MSP) is published. This plan includes, among other things, the expected call volume, by day, for the incoming quarter. Half way through the quarter, during week 6 of a total of 13, a revised MSP is published. This revised plan includes expected call volume for the remainder of the quarter and an outlook for the incoming quarter. Operational decisions regarding call center staffing and scheduling are based on the call volumes published in the MSP.

The forecasting process described here is called Integrated Forecasting Model (MSP-IFM) at Dell. It is a relatively recent process, introduced to account for the impact of Live Voice

¹ This estimate accounts only for abandoned calls in excess of the tolerable daily 4% abandon rate used by call center operations group as performance metric.

Recognition (LVR) on the call forecasting process. Looking at this process closely, we see it is based on two different but important components. One is the expected call volume estimated from advertisement and historical trends, referred to as bottom-up forecast, and the other is the call volume required to meet corporate goals, referred to as top-down forecast.

The bottom-up call forecast is estimated based on actual and historical call, marketing and advertising data. These data usually include weekly spend and/or circulation volume on different advertisement channels like network and cable television, Sunday papers, on-line advertisement, free-standing-inserts, mailings and catalogs. Product price is also an important determinant of call volume. Historical price data without some reference to competitors' price position, however, is not highly correlated to historical sales calls volume. Total margin per unit (TMU), an average of the marginal profit per unit calculated as total profit over the total number of systems sold per unit time, has traditionally been used at Dell as a proxy for price aggressiveness.

Historical call data by marketing vehicle and marketing spend on that same vehicle are used to determine the vehicle's 'response rate', the expected number of calls per million dollars spend on advertisement. This, combined with the advertisement budget for the incoming quarter, determines the expected call volume by vehicle. After expected volume is calculated for each vehicle and aggregated, the forecast is adjusted based on expected TMU for each week of the incoming quarter. From this call volume forecast, the volume of units to be sold is calculated using another historical metric, close rate. On average, every three calls answered one unit is sold which determines a close rate of approximately 30%. This rate is forecasted for every day of the quarter to predict unit volume per day. Sales calls and units forecasts constitute the core of the Master Sales Plan. Although extremely simplified, this would be a typical forecasting methodology for call center demand at Dell.

At the same time, a top-down forecast is produced every quarter. Executives of each business unit publish an expected performance plan. This plan is typically expressed as revenue, units and margin (RUM). Having an expected unit target for the quarter indirectly dictates the number of phone calls that should be received. This RUM plan is decomposed into units per week and calls required to meet this goal are calculated using historical close rate.

These two plans rarely agree on the expected number of units to be sold, so they are compared and the one with the larger expected volume is usually published as the Master Sales Plan. The top-down plan is almost always larger than the bottom-up plan by ten to fifteen percent. Most of the time then, the top-down plan is published as the expected volume of calls to be received, even when this plan publishes a target volume rather than a forecasted volume. The difference between the two plans is referred to as risk or challenge and published along with the plan.

It should be clear then that the quarterly forecast published at Dell is not really a forecast of demand but rather a performance target plan. It does not represent the expected demand but the desired demand. It should not be surprising then that call volume received tends to be lower than the volume forecasted.

Lastly, the call forecast published at Dell is more than just a call forecast. The Master Sales Plan includes sales calls volume, total incoming calls volume, online visits, units sold by online and offline channels, etc. Most of these metrics are typically reviewed and evaluated at weekly, rather than daily, levels and this has dictated that most of them are forecasted as weekly volumes and then decomposed into daily volumes if necessary.

1.2.2. Challenges on the forecasting process

As the quarter progresses performance against the plan is evaluated in terms of revenue, units and margin. As performance falls short of expectations, actions are taken to reshape demand and improve one of these three metrics. Different actions have greater impact on one of them than on the other two. These metrics, however, are not completely independent and any action that affects them would impact the sales calls volume in some degree. Usually, intra-quarter actions include promotions at reduced prices, reallocation of marketing funds to vehicles with a better short-term response rate or reschedule of future promotions.

There are several important considerations regarding intra-quarter changes to plan. The first one is that the call and unit forecasting model is a key component of the decision making process when the marketing budget might be reallocated. Not only the incremental volume of calls resulting from the reallocation of funds must be estimated, but also different scenarios must be compared to find the optimal allocation of these funds.

The second consideration is that the actions to take are normally determined by the marketing and sales departments. At the beginning of my work at Dell, decisions were made and incremental call volume estimated. However, the Call Center Operations (CCO) group was not involved in the decision making process and sometimes resources were not available to handle this incremental call volume. In some other cases, incremental demand was not even communicated to the CCO so scheduling of agents was not modified to account for this change. After a few weeks, a process to review marketing actions and publish updated sales calls volume was put in place by the CCO team. As a result, a projection of the expected call volume for the next two weeks is now published every week.

In light of this issue, an ideal forecasting model will permit evaluating scenarios to help in the decision making process for budget allocation. At the same time, it will produce updated forecasts in a reasonable amount of time so as to encourage its use for weekly projections.

As of the beginning of FY09, sales agents were handling the total incoming call volume attributable to sales phone numbers. Although this incoming call volume is expected to be sales-related, a significant portion is not and agents spent a considerable amount of time re-routing calls to other queues such as customer care or technical support.

A new call routing process was gradually introduced during the first two quarters of FY09. In order to optimize call handling by the appropriate queue, incoming calls were routed through a Live Voice Recognition (LVR) system that pre-screens and routes calls appropriately. Existing forecasting models at Dell were designed to predict incoming call volume, estimate a daily filter rate and calculate the volume of calls ultimately routed to sales queues. Forecasting errors on each step got propagated through the process. Since CCO's costs are mainly associated with sales agents scheduling and staffing, this process was being revisited to focus on sales calls accuracy rather than total incoming volume. The deployment of this new process and the shift on focus introduces challenges regarding the analysis of historical data, which are described further on section 3.1.

Finally, it is of interest to note that we might be in presence of a phenomenon of inflated demand that could present some challenges similar to the lost sales described by Nahmias (Nahmias, 2009). Demand is typically forecasted based on sales data and it is assumed sales are a

good indication of true demand. Stock outs can lead to lost sales and present a challenge since true demand is not really known for the period. In a similar way, we forecast demand based on received calls and we assume that calls received are a good indication of true demand. On days where demand is higher than expected, holding times are longer and abandon rates higher. It is normally assumed that customers not willing to wait hang up and call again until they can talk to a representative. If this were the case, we should assume that calls received during those days are not really representative of true demand. Unfortunately, we could not find data to verify that this is indeed the case and thus we will not explore this issue as part of this work.

1.3. Focus of this thesis

This work will focus on finding a forecasting model for sales calls for Dell's U.S. consumer sales division. This model would be used to predict sales calls, by week, for the incoming quarter. Once the weekly volume has been determined, it would be distributed among days of the week to produce a breakdown of sales calls expected by day for the incoming quarter. Finally, we will explore how this model could be used to produce weekly projections for the following two weeks.

1.3.1. Weekly volume forecasting model

The model to forecast weekly sales call volume is the most important component of the forecasting process. This work will evaluate statistical methods like multiple regression analysis and time series along with judgmental forecasting methods to determine the most appropriate model for Dell U.S. consumer sales.

A multiple regression model would be ideal to univocally link marketing spend and price aggressiveness to sales calls. By making the relationship between marketing spend and sales call explicit and quantifiable, these models would allow evaluating scenarios and support decision-making. Regression models will face several challenges. Historical sales calls data is in the form of a time series, potentially introducing autocorrelation problems. Also, the relationship between different forms of advertisement and demand response is usually very complex and it might be difficult to describe using the available data. Lastly, marketing spend and circulation tends to be highly correlated, which might cause multicollinearity problems.

Time series models could provide simplicity and speed, particularly if the process can be automated using software. Time series models present some challenges as well. They require a large amount of historical data, particularly when seasonal effects are present. Also, a model based solely on historical sales calls data might not be useful for evaluating scenarios and helping on the decision making process. Judgmental forecasting might help alleviate this problem by incorporating the vast undocumented knowledge of promotions impact and market analysis available at Dell in order to improve the accuracy of statistical forecasts.

1.3.2. Day of week distribution

This work will analyze historical data on day of week distribution, in order to assess the adequacy of the current methodology and to proposed a revised method if necessary. The analysis of these data will reveal if there are seasonality effects on this distribution or if some holidays are becoming significantly more important in the last few years due to increases in marketing activity. The lack of daily marketing data might make these effects hard to model mathematically, but a methodical analysis of historical data might prove very useful in judgmental estimation of day of week distribution.

1.3.3. Intra-quarter updates

A weekly review and projection process was put in place during the first weeks of this work. This process reviews intra-quarter changes to the published marketing schedule and assesses their impact on the forecasted call volume. While the process is outside the scope of this work, its impact on call forecasting methods and requirements might be significant. At the same time, if the forecasting model proposed could be used to produce or facilitate these weekly projections, it would provide the projection process with a methodical approach that, over time, would help improve accuracy of the projections.

2. Literature review

Demand forecasting has been the subject of copious research and its use in industry is well documented. In his survey about sales forecasting practices in the United States, Dalrymple (Dalrymple, 1987) finds that 99% of the companies produce a formal forecast as part of their marketing plans and that 93% consider them ‘one of the most critical’ or a ‘very important’ aspect for company success. One field where demand forecasting research has been particularly prevalent is electricity demand. Electricity demand research usually focuses on very short-term predictions, typically demand for the next day, but some findings can be generalized and applied to call center demand forecasting. In particular, Zhou (Zhou, Yan, Ni, & Li, 2004) introduces an interesting concept of using a second autoregressive integrated moving average (ARIMA) model to improve the accuracy of the predictions of the ARIMA model use to forecast electricity demand.

Andrews and Cummings (Andrews & Cunningham, 1995) developed ARIMA models for L.L. Bean call centers. In their work, they show ARIMA models can be improved by the use of independent variables and that the impact of these variables is different on customer care and sales call centers. They rely heavily, however, on the forecast of weekly sales as a predictor for sales calls; a forecast that in their case was accurate and useful as a predictor. Antipov (Antipov & Meade, 2002) presents the idea of improving predictions by taking advantage of known information that helps shape demand, primarily marketing spend and advertisement. This idea is very applicable to this work, and it is following a similar approach that we will explore the idea of using marketing spend as independent variables in a multiple regression model. Despite these and other efforts, Gans et al. (Gans, Koole, & Mandelbaum, 2003) deem the state of the art in forecasting call volume as ‘rudimentary’ and state that improving forecast accuracy on call arrival rate is ‘perhaps the most pressing practical need on call center research’.

One of the major challenges on predicting demand from advertisement activities is to capture the complex interactions among these activities. We cannot assume, for example, that a sales transaction is the result of a single advertisement campaign or a particular ad, but rather the cumulative result of a series of these campaigns. Koyck (Koyck, 1954) introduced the concept of ‘distributed lags’ into econometrics analysis. The term ‘distributed lags’ is used to describe a

phenomenon in which a stimulus evokes a full reaction only after some passage of time (Palda, 1965). Palda uses this concept to measure the effect of lagged advertisement.

The variables that affect demand are sometimes too numerous or too complex to model. Companies sometimes rely on judgmental or expert forecasting and the use of contextual information to try to incorporate the effects of these variables into a forecast. In some of the cases, the use of contextual information can be important in reducing forecasting error (Edmundson, Lawrence, & O'Connor, 1988). However, judgmental forecast alone is not better than simple quantitative methods (Lawrence, O'Connor, & Edmundson, 2000) and experts are no better than chance when they use their unaided judgment to forecast decisions made by people in conflict situations (Green & Armstrong, 2007). Combining judgmental and quantitative models can help improve the accuracy of forecasts (Lawrence, Edmundson, & O'Connor, 1986) and management knowledge can help improve forecast accuracy if it is incorporated into the forecast process in a methodical way (Armstrong & Green, 2005). Methodical combination of forecasts is preferred, since individuals tend to over-react to noise and changes in the demand signal (Watson & Zheng, 2008) and to be bad predictors of data with downward trends (O'Connor, Remus, & Griggs, 1997).

Fellows at MIT's Leaders for Global Operations (LGO) program have studied call centers and demand forecasting, but there is no previous LGO work that tackles demand forecasting for call centers. Sen (Sen, 2009) and Gill (Gill, 2008) studied call centers at Dell and their work provides great background for this thesis. Both of them, however, focused on improving customer experience. Einhorn (Einhorn, 2007) explores the idea of using ARIMA models to replace the primarily judgmental models in use at Dell for component demand forecasting. His study showed that ARIMA models were not consistently better than the existing process which is based primarily on contextual information and individuals' expertise. Einhorn's models suffered from limited availability of historical data on component demand, something to be expected given the short life cycles of electronic components. The consistently better forecasts produced by Dell planning teams highlight the importance of contextual information and expertise. Bankston (Bankston, 2008) improved the forecasting process at a consumer goods company to help reduce retailers' stock-outs. Her work, however, dealt with the challenge of very limited data available from retailers and a very large number of products to forecast.

3. Forecasting models

This chapter will explore statistical methods like multiple regression and time series analysis in order to assess the adequacy and accuracy of these methodologies for weekly sales call volume forecasting. Before examining these methodologies, we will review the data available for modeling.

3.1. Data available for modeling.

Gathering data and creating a dataset for modeling proved to be one of the most challenging tasks through this work. In some cases, not enough historical data were available. In other cases there was no means to validate datasets. Other times, the data available did not contain all the information, or the granularity, we would have liked to have. This section will present a quick description of the data available and the datasets used, as well as a description of assumptions made, steps taken to create datasets and opportunities for future improvement as they are relevant to modeling purposes.

Data used for modeling can be separated into four categories: call data, marketing data, price information and dummy variables.

3.1.1. Call data

The introduction of the Live Voice Recognition (LVR) process during FY09 changed the definition of sales calls and the way call data are presented. Traditionally, sales agents at Dell handled the total sales attributable volume of calls. But as a significant percentage of these calls was eventually redirected to customer care or technical support services, the LVR process was introduced to pre-screen and route calls to the appropriate queue in order to maximize the time sales agents spend on sales calls. As of the beginning of FY10, a typical distribution of the call volume handled by the LVR can be schematized as follows:



Figure 2. Live Voice Recognition (LVR) call breakdown

Two things should be expected from the introduction of this process. First, the total volume of incoming calls should go up since some additional volume (SOAR, Lost and Spanish)² is now routed through the LVR system. Second, the sales calls volume should drop since some calls are being filtered away. This is exactly what we see in the data.

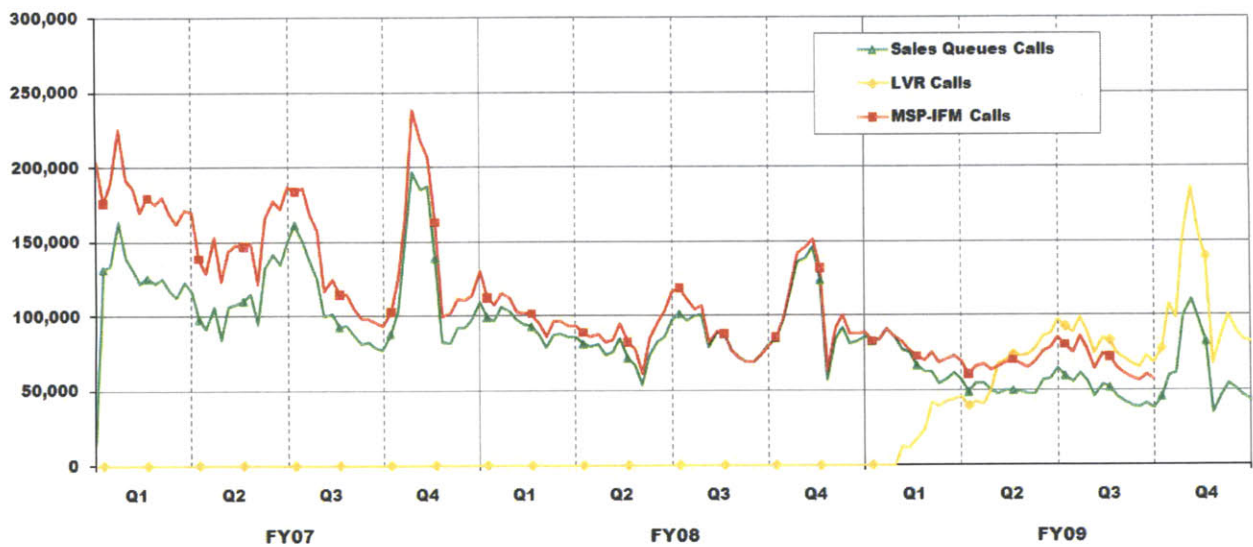


Figure 3. Sales Calls Data³

As the main goal of this project is to create a model that forecasts sales calls volume instead of total incoming calls volume, we need to create a dataset of sales calls. The first challenge we encountered was that no information regarding historical transfer out rate (the volume of calls rerouted to customer care or technical support queues) was available. We estimated this rate using data from FY09 Q3, the first quarter for which the entire volume of

² SOAR (Sales Outbound Agent Reach), Lost callers and Spanish are legacy phone queues with relatively low volume. The calls received on these phone lines are typically handled as generic calls.

³ Sales queues calls don't include volume handled by outsource providers. This volume accounts for the difference between sales queues and MSP-IFM calls for the years FY07 and FY08.

sales calls was routed through the LVR. Assuming that this rate, which was around 30%, was constant for the entire period FY07-FY09 data, we created a sales calls dataset that we could use for modeling.

The second challenge was to expand this dataset. Three years of historical data might not be enough for modeling and the presence of seasonal features in the data increases these requirements. According to Hanke (Hanke & Wichern, 2004), 6 to 10 years of historical data are required for ARIMA models if seasonal features are present. Two other call data sources are available to expand this dataset:

- Historical data kept by the Master Sales Plan team for their Integrated Forecasting Model (MSP-IFM) provides sales-attributable weekly call volume back to FY02.
- Calls by Vehicle (CBV) database tracks sales attributable call volume and the contribution of each marketing vehicle. It provides data on calls received back to FY05, but the data is spotty and apparently not reliable.

Since the volume reported for the three data sources for FY08 was very similar, we used the historical MSP-IFM data to create a dataset for the period FY02-FY08. We assume a constant transfer out rate of 30% for the entire period. This provided us with total of 8 years of daily and weekly sales calls data for modeling. Figure 4 shows sales call volume data, by week, for the entire period.

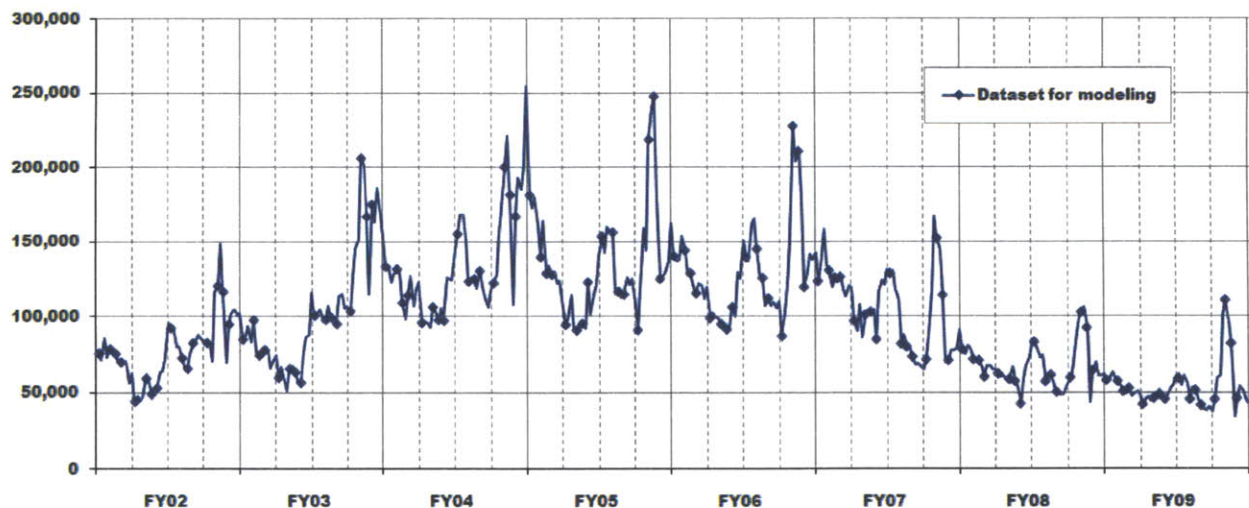


Figure 4. Sales call volume by week, dataset used for modeling

The assumption on historical transfer out rate is critical. A constant transfer out rate is likely to disguise, either exaggerating or dampening, features in the data that might be important for the accuracy of the forecasting models. The assessment of seasonal or periodic events, or the importance of a particular promotion, relies heavily on such features. Not having better information available, this methodology provides us with an acceptable dataset for modeling. But improving the accuracy of the sales calls dataset by finding reliable data on transfer out rate will have a direct impact on the accuracy of any of the models developed throughout this work and it should be the first improvement to focus on if any of the models developed here is adopted at Dell.

3.1.2. Marketing data

Marketing data available for modeling includes total spend by week and by vehicle for the last five years, this is, the period FY05-FY09. The advertisement vehicles for which we have data available include:

- Catalog
- Direct mail (DM)
- Free standing inserts (FSI)
- Sunday supplements (SS)
- Cable TV
- Network TV
- Online advertisement
- Online search
- Other print (magazines)

For some of these vehicles, particularly catalog and direct mail, total circulation volume by week is also available. Spend and circulation data for a given vehicle are highly correlated, so including both in any model is very likely to cause multicollinearity problems. Dell's Marketing and Communication (Marcom) team considers spend data to be more reliable than circulation data. Since there is no other data source available to validate the accuracy of these marketing data, we decided to use spend rather than circulation for modeling purposes.

Validating and improving the accuracy of this dataset will help improve the accuracy of the models. We identified at least one feature on these data that we believe should be corrected, for some vehicles on some periods of time, spend on the first week of the quarter is consistently large. This might indicate that the dataset includes not only advertising spend but probably some kind of agency fee. While tracking fee spending is important, it is fair to assume that money spent on fees will not generate demand directly and it should be removed from the dataset.

In order to account for lagged advertisement effect, each marketing vehicle spend dataset was modified to create a second dataset that simulates this effect. If we assume that the response to one week spend will be seen across several weeks, we should distribute this spend accordingly so the spend data closer matches the customer response data. We applied different 'decay rates' to the weekly spend of each vehicle and we compared correlations of each decay rate to data on sales calls received. For each vehicle we chose the decay rate that provides with the highest correlation between calls and spend. Some of the decay rates used are shown in Figure 5. Both datasets, modified using these decay rates and original data, were used for modeling in order to determine which one better explains the demand observed.

Marketing comprises a variety of actions. These actions may have very different nature and provoke very different responses from customers. Some advertisement is designed to provide an immediate response, call-for-action advertisement that will spike demand for a short period of time. Brand awareness advertisement, on the other hand, will produce a subtler response for a longer period of time. Sometimes advertisement is directed to different customer segments over time. At Dell, it is typical to segment customers into acquisition, returning and current for marketing purposes and design and focus advertisement to acquire new customers, gain back customers that left or maintain loyalty of current customers. Each one of these segments provides a very different response.

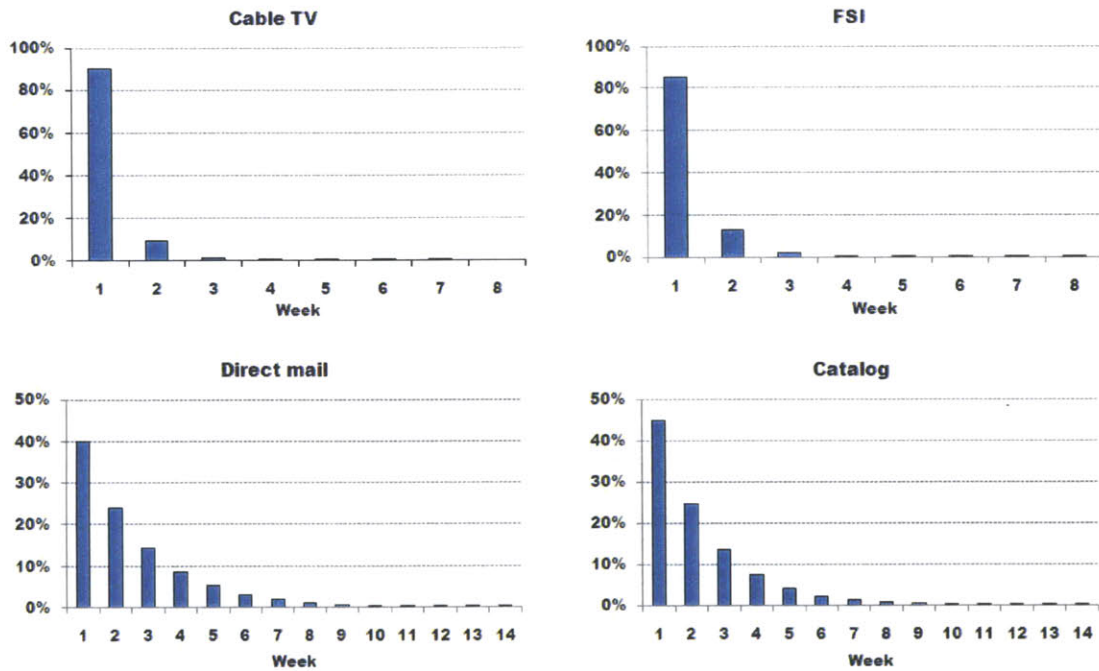


Figure 5. Selected response decay rates⁴

Spend information might not be enough to capture and characterize these subtleties and it might not, by itself, be enough description of marketing activity. Employees' expertise and knowledge of past campaigns might be the best source of information regarding these effects. A subjective process to adjust model outputs might be necessary to incorporate this knowledge into the forecast.

3.1.3. Price data

Another factor that might have a significant effect on demand shaping is price aggressiveness. Including price information into our forecasting model could be important. Price information alone, however, won't be useful unless we also have a reference to compare against. Competitors pricing information, for instance, could be a good reference point. At Dell, two indexes are typically used to measure price aggressiveness. Weighted Average Price Parity (WAPP) and Total Margin per Unit (TMU).

⁴ Charts show typical time response of advertisement expenditure on different channels. For cable TV, where advertisement tends to be of direct action, most of the response is observed in the first week, decaying rapidly. For catalogs, issued monthly, the response decay rate is much longer.

WAPP is a price parity index developed and maintained by the pricing team. It provides a normalized index that compares pricing of Dell and competitors products. Unfortunately, the methodology used to calculate this index has changed several times in the last few years, thus making the historical data difficult to interpret and use in regression models. WAPP is perhaps the most appropriate index to use for decisions that don't need to rely on historical information or decisions that are relevant at a single point in time. For our analysis however, WAPP is of very little help.

TMU is an average of the gross margin earned by unit sold, simply calculated as total gross margin over total number of units sold in a period of time. A more aggressive price strategy implies selling units at lower margins, the lower the margin the more aggressive the strategy. Although competitor's prices are not included in the calculation, it is implicit that Dell is pricing units according to competitor's behavior and thus their actions are reflected in TMU. If we assume competitors' cost structure is similar to Dell's, then TMU seems to be a reasonable proxy for price aggressiveness.

However, there are some concerns regarding this metric. TMU is calculated based on cumulative volume and price and, similarly to our concerns regarding marketing spend, it might lack the granularity or specificity we would desire. The impact on TMU of a high volume, low margin product might be very similar to the impact of a low volume, high margin one. The impact of these two products of sales calls can be quite different though. There is also no indication that TMU includes any compensation for price deflation over time. A \$10 increase in unit margin in 2002 might be less significant than the same increase in 2009.

Despite these concerns, TMU seems to be the most appropriate proxy for price for our models. TMU data is easily accessible, the methodology has been consistent over time and it has been used as a proxy for price aggressiveness at Dell for many years. We will be using TMU as a proxy for price aggressiveness in our models.

3.1.4. Other data

Lastly, a set of dummy variables was created in order to help models explain some events and abnormalities in the data. The use of dummy variables in regression models to explain seasonal features can help mitigate the effect of serial correlation on the residuals, a problem that

might be prevalent when data is in a time series form. Periodical events like President's day or one-time events like a battery defect, which caused the recall of a large number of batteries and was followed by a decrease in demand, might be better explained by using dummy variables as well.

The dummy variables created include:

- Dell's fiscal years, quarters and months.
- All major U.S. holidays.
- All holidays that could be commercially significant, including Valentine's day, Mothers' day and Fathers' day.
- Pre- Back to School period.
- Back to School period.
- Holiday season.
- One-time events that are believed to have affected demand. The most significant events are a battery recall, followed by a drop in demand, and a paint supply issue that caused stock outs.

3.2. Forecasting with Multiple Linear Regression

If we assume sales calls, as a demand signal, are generated as a response to or largely influenced by, marketing activities then a multiple regression model seems to be an appropriate method to describe and predict their behavior. Given the data available on marketing spend and pricing, a regression model that explains our dependent variable (sales calls) from a linear combination of independent variables (marketing spend and pricing) would allow us to predict call volume directly from marketing activity.

The explicit relationship between action and response would be extremely helpful to assess the impact of marketing vehicles and promotions. Furthermore, it would allow us to evaluate different scenarios and determine the most effective way of allocating, or re-allocating, funds and, in this way, help with the decision making process for intra-quarter changes to plan.

In its general form, a multiple linear regression model can be expressed as:

$$E(y_t) = B_0 + B_1x_1 + \dots + B_nx_n + \varepsilon_t$$

Regression models are based on four basic assumptions regarding the probability distribution of the random component ϵ (Mendenhall & Sincich, 2003):

1. The mean of the probability distribution of ϵ is 0.
2. The variance of the probability distribution of ϵ is constant for all settings of the independent variables x .
3. The probability distribution of ϵ is normal
4. The errors associated with any two different observations are independent

Errors of two different observations are considered independent if the error associated with one value of y has no effect on the errors associated with other y values. Data recorded over time, like the sales call data we will model, tend to violate this last assumption since residuals of different observations are not independent. Analysis of residuals and Durbin-Watson's test help determine if serial correlation of residuals is present. Serial correlation on residuals reduces estimates for the standard errors and artificially improves t statistics and p -values of regression parameters (Mendenhall & Sincich, 2003). If serial correlation is present, R^2 values might be artificially high and some independent variables might be erroneously assumed significant. We will pay particular attention to these issues when determining the adequacy of regression models.

Regression models are good predictors of responses when the response being predicted is within the region of observation of the independent variable. It is risky to use regression models to predict responses outside this region. Since forecasting always involves prediction of future values of a time series, this problem is unavoidable. However, it is important to recognize the dangers of this prediction (Mendenhall & Sincich, 2003). A regression model might help explain the existing data and their relationships with the independent variables while still fail to predict future responses if these responses are not contained within the range of the responses modeled. While overall adequacy of a regression model can be determined by its F ratio, one rule of thumb suggests using regression models with F values of at least four times the critical value for forecasting purposes (Hanke & Wichern, 2004).

Data from FY05 until FY09 seems to be a good calibration dataset for modeling, since marketing spend data is not available before FY05. We will use the data corresponding to the first quarter of FY10 to validate the predictions of the models created.

We developed a large number of multiple regression models. We have selected two models to illustrate the adequacy of the methodology for sales calls prediction. Both models rely heavily on dummy variables to explain seasonal features. Model I includes most of the available marketing vehicles as independent variables and Model II incorporates some interaction terms. We used interaction terms to try to model the effect of advertisement during periods like back to school, where demand response might be particularly sensitive.

3.2.1. Multiple linear regression models

Table 1 and Table 2 show the SAS JMP results of the two models described. Figure 6 and Figure 7 show the residual plots for both models.

Summary of fit		Analysis of Variance				
R ²	0.906762	Source	DF	SumSquare	MeanSquare	F Ratio
R ² adj.	0.899841	Model	17	3.63E+11	2.14E+10	131.0049
Root MSE	12773.87	Error	229	3.74E+10	163171771	Prob > F
Observations	247	C. Total	246	4.01E+11		<.0001*

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	206619.78	10498.2	19.68	<.0001*
FY05[0]	-37219.61	2426.446	-15.34	<.0001*
FY06[0]	-33092.14	1721.386	-19.22	<.0001*
FY07[0]	-21441.69	1640.577	-13.07	<.0001*
FY08[0]	-9122.993	1424.228	-6.41	<.0001*
February flag[0]	-12871.34	1895.858	-6.79	<.0001*
March flag[0]	-5542.57	1869.448	-2.96	0.0033*
October flag[0]	4809.9884	1592.871	3.02	0.0028*
November flag[0]	14325.312	1787.951	8.01	<.0001*
Pre-back to school[0]	5336.2358	1362.097	3.92	0.0001*
Back_school[0]	-7478.541	1583.575	-4.72	<.0001*
Holidays[0]	-17029.79	2240.27	-7.6	<.0001*
Q4[0]	-5204.414	1545.116	-3.37	0.0009*
POS TMU	-176.3908	23.79521	-7.41	<.0001*
FSI Spend	0.0050717	0.001934	2.62	0.0093*
Catalog Spend	0.0014209	0.000734	1.94	0.0541
Cable TV Spend	0.0040637	0.00181	2.24	0.0257*
Network TV Spend (TV4)	-0.004606	0.001538	-3	0.0030*

Durbin-Watson	
Durbin-Watson	0.9177493
Autocorrelation	0.5273

Table 1. Model I Regression Output

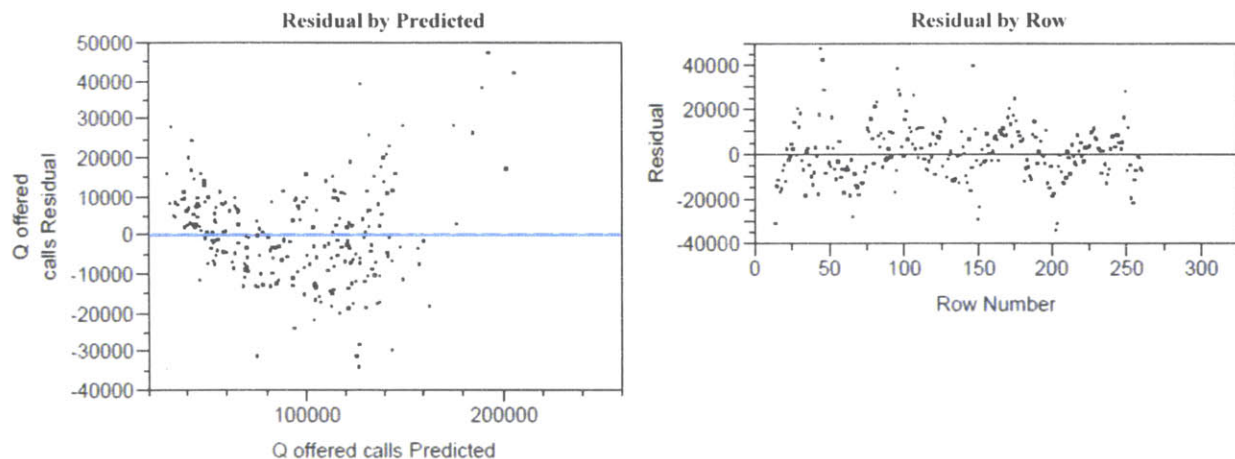


Figure 6. Model I Residual plots

Summary of fit	
R ²	0.925421
R ² adj.	0.918822
Root MSE	11499.97
Observations	247

Analysis of Variance				
Source	DF	SumSquare	MeanSquare	F Ratio
Model	20	3.71E+11	1.85E+10	140.2181
Error	226	2.99E+10	132249375	Prob > F
C. Total	246	4.01E+11		<.0001*

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	188955.14	9782.578	19.32	<.0001*
FY05[0]	-35208.14	2324.102	-15.15	<.0001*
FY06[0]	-30945.78	1619.576	-19.11	<.0001*
FY07[0]	-18646.53	1573.378	-11.85	<.0001*
FY08[0]	-8246.327	1385.17	-5.95	<.0001*
February flag[0]	-12195.06	1698.245	-7.18	<.0001*
March flag[0]	-5065.351	1684.964	-3.01	0.0029*
October flag[0]	4661.7381	1433.507	3.25	0.0013*
November flag[0]	13306.538	1640.368	8.11	<.0001*
Pre-back to school[0]	4210.5427	1239.728	3.4	0.0008*
Back_school[0]	-8354.14	1429.942	-5.84	<.0001*
Holidays[0]	-12075.89	2060.333	-5.86	<.0001*
Q4[0]	-4818.459	1364.682	-3.53	0.0005*
POS TMU Co19	-169.398	21.47019	-7.89	<.0001*
Cable TV Spend	0.0108535	0.002434	4.46	<.0001*
Network TV Spend	-0.006392	0.001558	-4.1	<.0001*
Catalog Spend	0.0017502	0.000667	2.63	0.0092*
FSI Spend	0.0047534	0.001743	2.73	0.0069*
(Cable TV Spend-1246996)*Holidays[0]	-0.006462	0.001739	-3.72	0.0003*
(Network TV Spend-708184)*Back_school[0]	0.0051253	0.001455	3.52	0.0005*

Durbin-Watson	
Durbin-Watson	1.1282566
Autocorrelation	0.4234

Table 2. Model II Regression Output

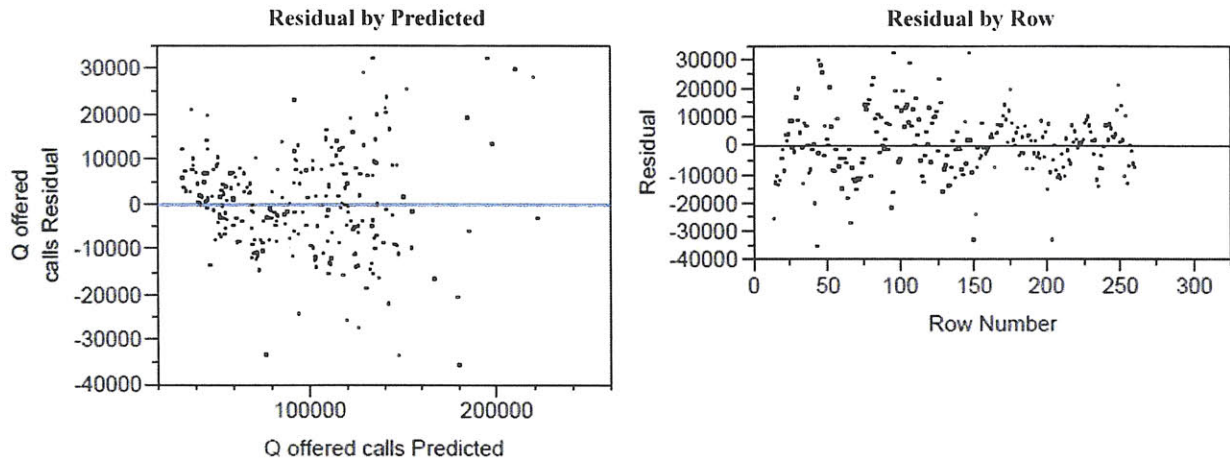


Figure 7. Model II Residual plots

A quick analysis of these models highlights some possible modifications and things to examine in detail:

- Coefficient for Network TV spend is negative on both models. This is, at least, counterintuitive since it would imply that increments on Network TV advertisement would decrease sales calls volume.
- Catalog spend p-value is very close to the critical value of 0.05. A different decay rate used for data preparation might help improve this statistic.
- The residuals by predicted plots (Figure 6 and Figure 7) show a well-defined U shape, particularly for Model I. This would suggest that data transformation would help improve the model's fit. In fact, a square root transformation of the dependent variable significantly improved the fit of Model I, increasing R^2 to 0.94.
- There seems to be indications of heteroscedasticity on both models.

We will assess the adequacy of regression models to call forecasting before addressing the attempting to address these issues and fine-tune the models. Models I and II were created using data for FY05-FY09, data on FY10 Q1 was withheld as validation dataset in order to verify the accuracy of these models when predicting future calls. Figure 8 and Figure 9 below show prediction results for Models I and II both in absolute call volume predicted and as actual

calls as a percentage of forecasted values. Prediction accuracy is far from our expectations and far from anything that could be considered acceptable.

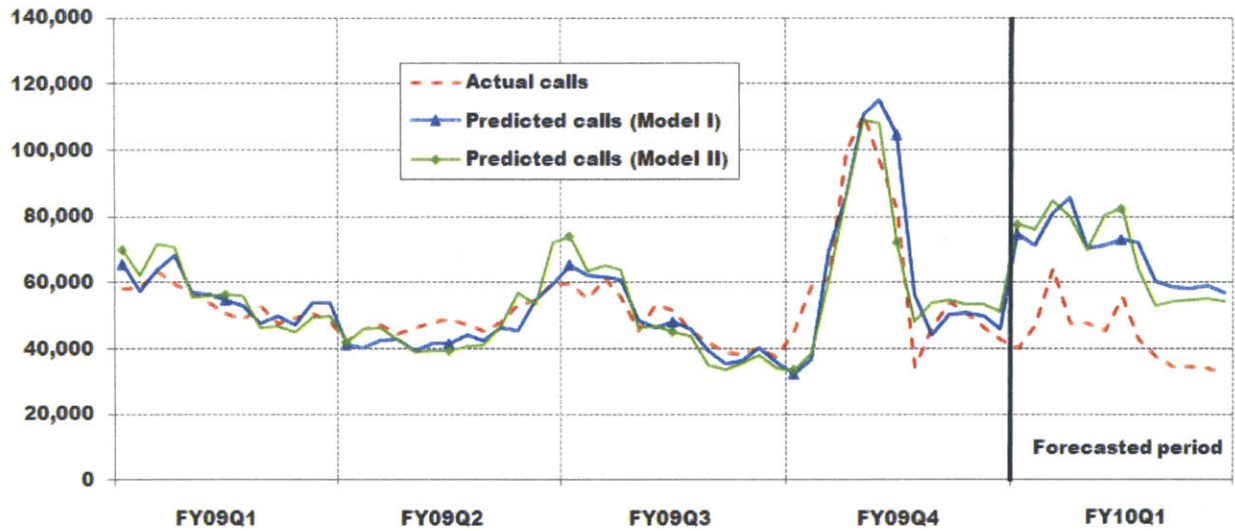


Figure 8. Models I and II forecasts, FY10 Q1

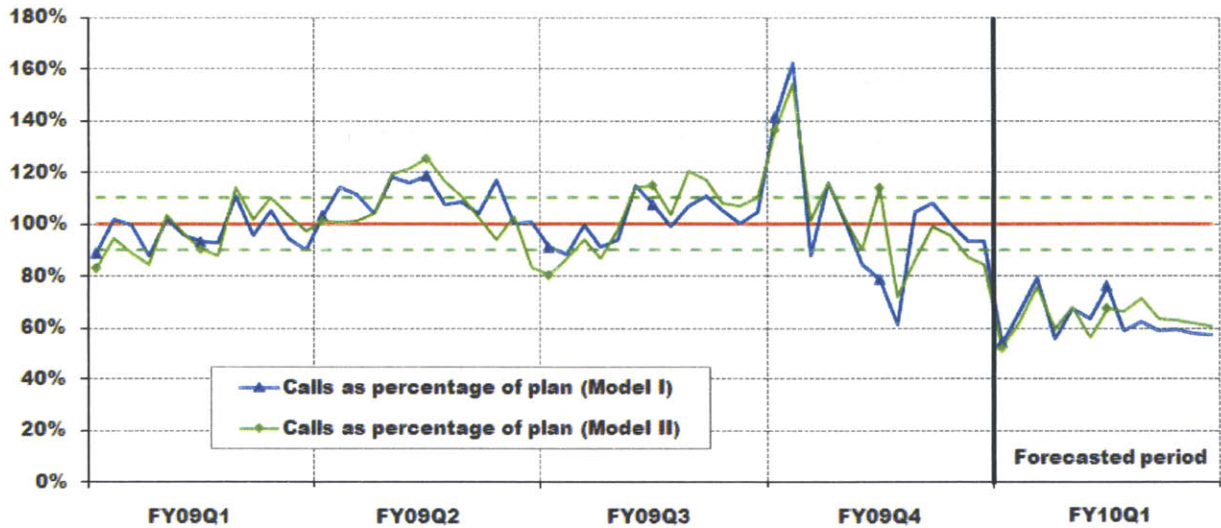


Figure 9. Models I and II forecast accuracy, FY10 Q1

Critical F values for models with 250 observations, 20 parameters to be estimated and significance level of 0.05 are approximately 1.6. This is several times smaller than the F ratios reported for these models, certainly more than the four times recommended by Hanke (Hanke & Wichern, 2004).

Looking at the residuals by row plots (Figure 6 and Figure 7), we observe a sinusoidal pattern on the residuals. Since each row on this dataset is a different weekly observation, we can consider this a plot of residuals over time. A pattern on the residuals suggests that errors among observations are not independent. In other words, we are violating the fourth assumption of linear regression models. This, in turns, might indicate that t statistics and p-values might be inflated and the model fit might not be as good as we would surmise from the statistics on the tables above.

The presence of serial correlation among the residuals is verified by the low values of the Durbin Watson’s statistics for both models. Critical upper and lower limit values for Durbin-Watson statistics, dL and dU, for models with these characteristics are shown in Table 3.

Durbin-Watson Critical Values					
	DW	n	K (parameters)	dL	dU
Model I	0.917	247	18	1.64959	1.93904
Model II	1.128	247	21	1.62293	1.96690

Table 3. Durbin-Watson critical values for Models I and II⁵

A Durbin-Watson’s (DW) value over the upper limit dU indicates there is no serial correlation among residuals. A value below the lower limit dL indicates evidence of serial correlation among residuals and a value between dU and dL indicates the test is inconclusive. Given the DW statistics for Models I and II, 0.917 and 1.128 respectively, we should assume serial correlation is present on these models. Thus the statistics are inflated and the model’s fit is not as good as presumed. We will deal with this issue before attempting to fine-tune them.

There are several different ways of dealing with serial correlation. The first alternative is to explore the use of additional independent variables that can help explain this correlation among residuals. Typically, the serial correlation among residuals implies there is a variable, significant to the process, which has been left out of the model. Unfortunately, no other data series could be identified to try to explain this correlation.

With no more independent variables available, we explore some alternative methods for dealing with serial correlation. These methods include using lagged sales calls as an independent

⁵ <http://www-leland.stanford.edu/~clint/bench/dw05c.htm> (accessed March 26th, 2010)

variable, fitting an autoregressive term and using an ARIMA model to help explain the errors in the observations⁶.

3.2.2. Regression with lagged dependent variables

If the prediction errors, or residuals, on consecutive observations are not independent past expressions of the dependent variable could help explain this relationship and thus reduce serial correlation. One explanation for serial correlation is the delayed effect of advertisement on sales calls. Palda (Palda, 1965) studied the delayed effect of advertisement in sales revenue and found that using lagged sales revenue as an independent variable will not only capture the delayed effects of advertisement but it will also reduce autocorrelation in the residuals. Furthermore, since only one lag variable is used, sales revenue, as opposed to many marketing vehicles at different time intervals, this model would be less exposed to issues with multicollinearity.

The use of lagged sales calls in our models presents us with a challenge. Since forecasts at Dell are done for an entire quarter, this is a period of thirteen weeks; the information of last week call volume is only available for the first week of the quarter. After this first week, forecasted values will have to be used as lagged calls. Since forecasted values are not direct observations and might include an error, this error might propagate throughout the thirteen weeks to be forecasted.

Table 4 shows the results of fitting Model I including sales calls lagged one week as an independent variable, which we will call Model Ia. Since lagged calls help explain the effects of lagged advertisement, we should also modify this model so marketing data used is not modified using decay rates. We did fit such model, with very similar results to the ones shown here. For ease of comparison, we show in Table 4 Model I with the addition of lagged calls and no changes to marketing data.

Durbin-Watson's statistic has improved significantly. However, we should not rely on this statistic since it is biased towards 2 when a lagged dependent variable is included as a predictor (Hanke & Wichern, 2004).

⁶ This list is not exhaustive. Several other methods were considered, with similar results. We have selected the three methods mentioned as they are good examples of the issues encountered.

Summary of fit	
R ²	0.925912
R ² adj.	0.920063
Root MSE	11411.69
Observations	247

Analysis of Variance					
Source	DF	SumSquare	MeanSquare	F Ratio	
Model	18	3.71E+11	2.06E+10	158.3017	
Error	228	2.97E+10	130226669	Prob > F	
C. Total	246	4.01E+11		<.0001*	

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	154385.16	11586.95	13.32	<.0001*
FY05[0]	-27567.05	2505.969	-11.00	<.0001*
FY06[0]	-22942.56	2028.023	-11.31	<.0001*
FY07[0]	-14883.65	1696.421	-8.77	<.0001*
FY08[0]	-6732.908	1309.889	-5.14	<.0001*
February flag[0]	-8524.875	1785.817	-4.77	<.0001*
March flag[0]	-3316.021	1695.091	-1.96	0.0517
October flag[0]	2448.4263	1455.881	1.68	0.0940
November flag[0]	5495.1133	1968.348	2.79	0.0057*
Pre-back to school[0]	4066.9995	1228.027	3.31	0.0011*
Back_school[0]	-5739.803	1432.721	-4.01	<.0001*
Holidays[0]	-11701.21	2118.322	-5.52	<.0001*
Q4[0]	-2031.727	1440.889	-1.41	0.1599
POS TMU Co19	-159.7151	21.36843	-7.47	<.0001*
FSI Spend	0.0026286	0.001757	1.50	0.1359
Catalog Spend	0.0009724	0.000658	1.48	0.1411
Cable TV Spend	-0.000717	0.001733	-0.41	0.6793
Network TV Spend (TV4)	-0.003644	0.00138	-2.64	0.0088*
Q offered calls (lag 1 week)	0.3800966	0.049512	7.68	<.0001*

Durbin-Watson	
Durbin-Watson	1.4483377
Autocorrelation	0.2667

Table 4. Model Ia (including lag calls term) Regression Output

More importantly, the use of lagged calls as an independent variable causes some marketing vehicles predictors to be non-significant. Looking at the regression results, FSI, Catalog and Cable TV spend are no longer significant for predicting sales calls. The only significant marketing input is now Network TV spend but its coefficient continues to be negative.

We could interpret the results in two ways. We could assume that lagged calls are explaining the effect of lagged advertisement on these vehicles, thus these vehicles are no longer significant. Indications of multicollinearity problems are typical when more than one predictor is used to explain the same effect. Since there seems to be no clear indications of multicollinearity in this model, this might not explain why some marketing vehicles are no longer relevant.

We can surmise then that t statistics and p-values on Model I were inflated for these three marketing vehicles, using lagged calls reduced serial correlation, reduced the error on the estimation and produced more reliable statistics. These three vehicles were never significant predictors of calls and serial correlation might have been artificially inflating the statistics.

In any event, a regression model of this form explains and predicts sales calls based on historical call information rather than relying on marketing spend information. There are no significant marketing predictors on this model. This significantly reduces the usefulness of the model to evaluate scenarios and assess the impact of individual marketing vehicles on sales call volume.

3.2.3. *Using autoregressive terms*

The general multiple linear regression model is of the form

$$E(y_t) = B_0 + B_1x_1 + \dots + B_nx_n + \varepsilon_t$$

Where the random errors ε_t are assumed to be independent and follow a normal distribution with mean 0 and variance σ^2 . Since this assumption is broken in the presence of serial correlation, we can modify the general model to include an autoregressive term that would account for the correlation in the residuals. If we assume a model of the form

$$y_t = E(y_t) + R_t$$

Where $E(y_t)$ is the standard regression model

$$E(y_t) = B_0 + B_1x_1 + \dots + B_nx_n$$

And R_t is a first-order autoregressive term

$$R_t = \phi R_{t-1} + \varepsilon_t$$

Thus the value of the residual R_t is equal to a constant multiple of the previous residual R_{t-1} plus random error.

Using SAS Enterprise Guide, we created a new model (Model Ib) by fitting Model I including an autoregressive term. As expected, including an autoregressive term reduces serial correlation on the residuals and improves DW statistic significantly. The new DW value of 1.7245 is closer to 2, although it is not close enough to provide evidence that there is not serial correlation among residuals. Durbin-Watson's test is now inconclusive.

Reducing serial correlation improves the accuracy of the t statistics and p-values helping us assess the significance of the independent variables used in the model. As we can see from

Table 5, once the statistics are corrected three of the marketing vehicles, catalog, cable TV and network TV spend, are no longer significant. FSI spend is now the only marketing vehicle that provides information regarding sales calls volume.

Estimates of Autocorrelations						
Lag	Covariance	Correlation	-1	9	8	7
0	1.51E+08	1				
1	79764991	0.527265				

Estimates of Autoregressive Parameters			
Lag	Coefficient	St Error	t Value
1	-0.527265	0.056273	-9.37

Variable	DF	Estimated	Std Error	t Value	Prob > t
Intercept	1	93822	6265	14.98	<.0001
FY05	1	82230	5966	13.78	<.0001
FY06	1	69033	4818	14.33	<.0001
FY07	1	45390	4677	9.71	<.0001
FY08	1	19533	4230	4.62	<.0001
February_flag	1	22498	4633	4.86	<.0001
March_flag	1	10566	4256	2.48	0.0138
October_flag	1	-8584	3746	-2.29	0.0228
November_flag	1	-27518	4029	-6.83	<.0001
Pre_back_to_school	1	-9927	3090	-3.21	0.0015
Back_school	1	12087	3440	3.51	0.0005
Holidays	1	24777	4051	6.12	<.0001
Q4	1	11565	3895	2.97	0.0033
Cable_TV_Spend	1	0.001291	0.001779	0.73	0.4686
FSI_Spend	1	0.005068	0.001483	3.42	0.0008
Network_TV_Spend_TV4	1	-0.001852	0.001809	-1.02	0.3069
Catalog_Spend	1	0.000448	0.000548	0.82	0.4145
POS_TMU_Co19	1	-222.3494	22.8201	-9.74	<.0001

Yule-Walker Estimates			
SSE	2.48E+10	DFE	228
MSE	108888193	Root MSE	10435
SBC	5357.12975	AIC	5290.45137
Regress R-Square	0.8013	Total R-Square	0.9381
Durbin-Watson	1.7245		

Table 5. Model Ib (including autoregressive term) Regression Output

Once again, trying to correct for serial correlation in the residuals causes most marketing vehicles to be non-significant, leaving us with a model that explains the existing data and predicts future call volume without relying on information from marketing activities, thus limiting the usefulness of the model.

3.2.4. Predicting residuals to reduce serial correlation

On another attempt to deal with the serial correlation among residuals, we tried to use a time-series model to explain and predict the error terms of Model I. Similarly to the concept introduced by Zhuo (Zhou, Yan, Ni, & Li, 2004) where he uses one ARIMA model to predict

demand and a second model to predict the error of the first model, thus improving accuracy, we tried to use an ARIMA model to predict the error of our multiple regression model.

An ARIMA model was fit to the residuals of Model I and used to predict the error for the next thirteen time intervals (weeks). The output of this model was then included as one additional independent variable into Model I and the model refitted as Model Ic. The ARIMA prediction of the residuals was intended to reduce the regression error, thus improving accuracy of the model, and to reduce the correlation among residuals, thus improving DW and t statistics and p-values for individual parameters. Results of this new model are summarized in Table 6 below.

Summary of fit		Analysis of Variance				
R ²	0.940233	Source	DF	SumSquare	MeanSquare	F Ratio
R ² adj.	0.935514	Model	18	3.77E+11	2.09E+10	199.2668
Root MSE	10249.63	Error	228	2.40E+10	105054837	Prob > F
Observations	247	C. Total	246	4.01E+11		<.0001*

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	220194.76	8508.883	25.88	<.0001*
FY05[0]	-40456.21	1967.913	-20.56	<.0001*
FY06[0]	-34466.16	1386.565	-24.86	<.0001*
FY07[0]	-22290.54	1318.525	-16.91	<.0001*
FY08[0]	-9923.703	1144.981	-8.67	<.0001*
February flag[0]	-12246.05	1522.224	-8.04	<.0001*
March flag[0]	-5789.296	1500.185	-3.86	0.0001*
October flag[0]	4576.3632	1278.271	3.58	0.0004*
November flag[0]	12856.196	1440.513	8.92	<.0001*
Pre-back to school[0]	5251.8365	1092.959	4.81	<.0001*
Back_school[0]	-7126.297	1271.027	-5.61	<.0001*
Holidays[0]	-13966.64	1817.895	-7.68	<.0001*
Q4[0]	-5277.6	1239.802	-4.26	<.0001*
POS TMU Co19	-212.9971	19.36592	-11.00	<.0001*
Cable TV Spend	0.0021825	0.001462	1.49	0.1369
Network TV Spend (TV4)	-0.003124	0.001241	-2.52	0.0125*
FSI Spend	0.0051633	0.001552	3.33	0.0010*
Catalog Spend	0.0005692	0.000594	0.96	0.3388
Predicted Residual calls (using ARIMA)	1.0615803	0.093947	11.30	<.0001*

Durbin-Watson	
Durbin-Watson	1.8797815
Autocorrelation	0.0399

Table 6. Model Ic (including ARIMA term) Regression Output

This method helped improve the overall fit of the model. It improved R² and adjusted R² significantly and improved DW almost to the critical value that would have let us ascertain there is no serial correlation among residuals. After improving t statistics and p-values for parameters, once again some marketing inputs are no longer significant. FSI spend is the only marketing input we could reliably include in this model.

Once again, we are left with a model that explains existing data without relying on marketing information.

3.2.5. Summary on multiple linear regression models

At first glance, multiple linear regression seemed the logical choice for creating a model to explain and predict sales calls volume. Conceptually, linear regression models are similar to the forecasting model currently used at Dell. The concept of evaluating the response to each marketing vehicle and then aggregate these responses to estimate the expected call volume is not different in essence from what a multiple regression model would do.

In this sense, it would have made very easy the adoption of regression models for call forecasting by employees and management at Dell. It would have also introduced statistical rigor and methodology to the process, as well as the ability to calculate probabilities and prediction intervals. These improvements would have improved the credibility of the forecasting process at Dell's U.S. consumer sales group, which has been questioned for a long time.

As we have seen, however, the relationship among sales calls and marketing activity is not easy to model, at least with the data available at this time. Several different regression models were evaluated and none of them proved useful to forecast sales calls volume with acceptable accuracy. After some scrutiny, most marketing data proved not significant in these regression models. Unfortunately, a forecasting model that does not rely on marketing information to predict demand is prone to be mistrusted by management since it is assumed to convey the message that marketing actions don't have a direct impact on call volume. We don't think this is the case and we are certain that marketing activity impacts demand directly. However, this does not imply that we can explain the relationship through a statistical model.

Some models were able to relate marketing activity and sales calls. While their accuracy predicting calls was not enough to be useful for this work, such models could still be of use for budgeting allocation decisions since they provide a general sense on the relative magnitude of responses to marketing vehicles. Since these models could not forecast the magnitude of this response with acceptable accuracy, they were not included in this work.

More sophisticated regression models could be developed to further explore the ability to forecast sales call volume. Including more information on marketing activities, beyond spend by

week, might help capture effects that the presented models could not. However, a more complex model does not ensure better accuracy and the principle of parsimony leads us to pick simpler models over complex ones.

Very late in my internship, issues with the Calls by Vehicle (CBV) database were resolved and sales calls by vehicle and TMU data by vehicle became available. This would allow us to create datasets of sales calls by marketing vehicle that could be better correlated to marketing spend by vehicle. If these data were available, it might be worth exploring the idea of developing one regression model per vehicle to forecast sales calls by vehicle. The aggregate volume then could be used as total sales calls volume forecast.

Considering sales calls data is in the form of a time series and regression models failed to explain the relationship between marketing information and sales calls, we should explore the idea of using time series models to predict future call volumes.

3.3. Forecasting with Time Series

When the data to forecast is in the form of a time series, time series models are the natural choice. Time series models are able to forecast future responses solely from a collection of past responses, thus they would allow us to forecast future sales calls from historical sales calls data without relying on marketing activities and past spend.

A wide variety of models fit into the time series category. Models appropriate for stationary series, like moving average models; models that better handle trends like exponential smoothing and models that can deal with seasonal effects like Holt-Winters, all fit within this category. Perhaps the most sophisticated and adaptable among time series models are autoregressive integrated moving average (ARIMA) models.

Time series models are widely used at Dell, particularly exponential smoothing and Holt-Winters models. They have not been used, however, for call forecasting. Since forecasts created using time series models wouldn't normally rely on marketing information to forecast demand, they would not be useful to evaluate scenarios and help on the decision making process. Furthermore, since they seem to break the link between marketing activities and demand response, they are traditionally rejected by management since they are seen as forecasting

models whose output does not change when marketing plans change. This is, in a way, like saying that marketing actions will not have an impact on future demand.

This perception about time series models is mistaken and misleading. It is true that time series models could be used to forecast sales calls without information from marketing activities, but this does not mean they assume there is no relationship between them. In fact, and as Palda (Palda, 1965) described it, historical demand data does contain information about lagged effects of advertisement. The trends imposed by these lagged effects might be difficult to change on a short period of time, thus the marketing activities plan for the period being forecasted might not have as large an influence on demand as expected.

We are not proposing here to ignore marketing activities, but as we have seen from regression models, the effect of marketing activities on demand shaping might be difficult to model mathematically. In any case, and before addressing the perception issue around time series models, we should assess the technical feasibility of these models for forecasting sales calls.

3.3.1. ARIMA models

A quick analysis of historical sales calls data (Figure 10) reveals a long-term downward trend and the presence of, at least, two important seasonal features, back to school and holiday periods. Series with these characteristics are better explained by more sophisticated time series models; models that account, at least, for seasonal patterns on the data. ARIMA models seem to be a reasonable modeling choice for sales calls.

Autoregressive integrated moving average (ARIMA) models can produce accurate forecasts based on information contained in the series itself. They rely heavily on autocorrelation patterns in the data. The methodology for identifying and checking ARIMA models was greatly advanced by G. Box and G. Jenkins and it is often referred to as the Box-Jenkins methodology.

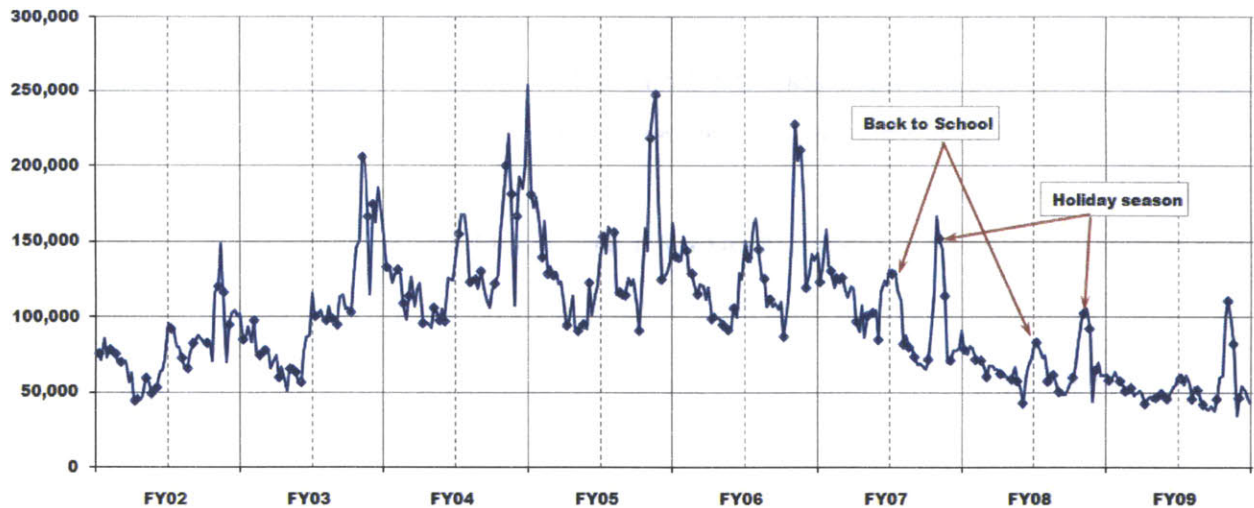


Figure 10. Sales calls historical data FY02-FY08

Since the introduction of this methodology, ARIMA models have become common. They are widely used to replace large and complex multiple regression models. We have seen that relatively simple regression models failed to predict sales calls data. Since more complex regression models might be required, evaluating the feasibility of ARIMA models for sales calls forecasting seems appropriate.

ARIMA models require a relatively large dataset in order to produce accurate forecasts. When trying to forecast variables that contain yearly seasonal features, each year of historical data represents one data point. ARIMA models require several years of historical data in order to predict seasonal features. This is a very prevalent problem of ARIMA models. In his study of ARIMA models for component forecasting at Dell, Einhorn (Einhorn, 2007) found that the very nature of the data made ARIMA models not more accurate than the models currently used. When the entire life cycle of a component is less than five years, as is commonly the case on electronic components, gathering five years of demand data is not even possible.

Fortunately, we have several years of sales calls data available. If we assume a constant transfer out rate for FY02-FY09, as described at the beginning of this chapter, we can create a dataset that covers eight years of sales calls. Although a constant transfer out rate assumption might impact the accuracy of the data and present some challenges when producing forecast, the resulting dataset should be large enough to model seasonal features on the data and help us determine the feasibility of these models.

In fact, the dataset created might be too large for properly forecasting future data. Forecasts are created assuming that ‘the future is like the past’ (Hanke & Wichern, 2004) and this assumption might not hold true for the entire period covered by this dataset. Some external events, like the gradual adoption of online sales or the introduction of retail sales, might violate this assumption. How many years of historical data should be use for modeling is one question that we will need to address as part of the model creation process.

Probably the most important concern about ARIMA models is the fact that the methodology for building, evaluating and selecting models, albeit significantly simplified by Box-Jenkins, is time consuming and requires significant expertise. Furthermore, models might need to be recreated every time a new forecast is to be produced, adding yet more complexity to the process. This undermines the adoption of ARIMA models at some companies and it would difficult the adoption of ARIMA models by the forecasting team at Dell’s U.S. consumer sales group.

Fortunately, software packages can facilitate the model creation process. At Dell, SAS Forecast Studio – and SAS Forecast Server – are already adopted and being used by procurement and planning teams. Choosing software already in use at Dell and supported by Dell’s IT department will not only accelerate the proof of concept but it will facilitate the adoption of ARIMA model for sales calls forecasting.

3.3.2. Using SAS Forecast Studio

SAS Forecast Studio (FS) automates the process of data evaluation, model creation, comparison and selection. SAS FS evaluates the dataset and creates several models that would explain and forecast the data. These models are then compared and ranked using a user-selected metric. By default, Mean Absolute Percentage Error (MAPE) is used as comparison metric. The best fitting model is then used to produce forecasts.

When new data become available and new forecasts are required, the user has the option of simply updating forecasts using last used models or recreating models entirely and going through the model comparison and selection process again. In any event, the process is entirely automated and fast. There is very little, or almost no, expertise required for generating forecasts.

Making the model creation process simpler and quicker facilitates the adoption of ARIMA models by individuals with no previous experience with these models. Given the limited time available for implementing any changes proposed by this work at Dell, having to train people on creating ARIMA models would have undermined their adoption.

There are several additional features of Forecast Studio that might improve the fit of traditional ARIMA models even further. Some of these features seem particularly applicable for forecasting sales calls at Dell:

- **Events:** FS supports the creation of events that can be used to flag abnormalities or particular features of the data. These events can be used to explain one-time abnormalities or to forecast recurrent features that we do expect to see in the future. When going through the model creation process, FS evaluates the significance of each event and whether it helps improve the fit of the model. For this project we used events for flagging US holidays, back to school period and holidays season.
- **Independent variables:** independent variables can also be used to improve the fit of a time series model. FS evaluates whether an independent variable improves the fit of the model and includes them if they do. Since regression models showed that marketing inputs were not significant to predict sales calls, we should not expect them to be significant here. TMU however, was significant on every regression model and it could be used as an independent variable.
- **Hierarchies:** while not useful given the dataset available for this project, hierarchical forecasting could be useful for sales calls forecasting at Dell. FS could simultaneously forecast, for example, sales calls for a region and for each one of the countries within that region. This could be particularly useful in regions like Latin America where close to forty five different countries must be individually forecasted every quarter.

Figure 11 and Figure 12 present the baseline forecast for FY10 Q1 created using Forecast Studio. We will refer to forecasts created using SAS Forecast Studio as ‘Baseline Forecasts’ throughout this work. Five years of data, FY05-FY09, were used as training data while FY10 Q1 was withheld as validation data. The events dataset mentioned above was made available and no independents variables were used.

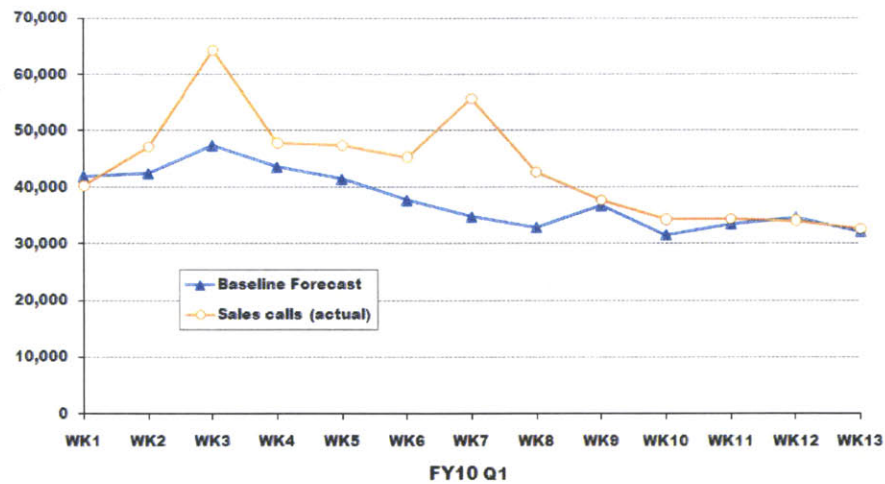


Figure 11. Baseline forecast, FY10 Q1

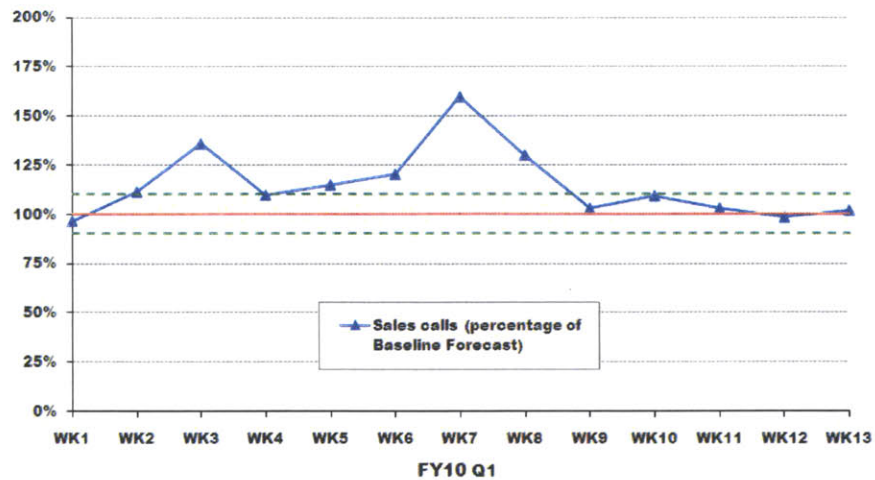


Figure 12. Calls received as percentage of forecast, FY10 Q1

These charts show that ARIMA models can be accurate predictors of calls. The results of this forecast are much better than the results obtained using multiple regression models. There are two weeks within this quarter however, where the call volume received was significantly higher than the volume forecasted by Forecast Studio. These are weeks 3 and 7.

Week 3 is the week of President's day holiday. The call volume received during this week was significantly higher than the call volume received on adjacent weeks, primarily based on very aggressive deals offered for new models of portable computers (netbooks) recently launched. An aggressive deal, combined with a large marketing spend for this week, could help explain the incremental volume.

Week 7 of this quarter held a ‘Days of Deals’ event. During a ‘Days of Deals’ event different products are offered at reduced prices on different days. Due to the non-periodical nature of this event, it is particularly difficult for a time series model to forecast. This is a typical example of how defined events in forecast studio could help improve the accuracy of the forecasts produced. For our particular case however, events did not help improve the forecast. When looking at the limited data available for past ‘Days of Deals’ we see that only some of them have a significant impact on demand. Not having consistent responses for past events, it is difficult to forecast response for future events. ‘Days of Deals’ generate a significant increase in demand when the prices offered are aggressive, when the products offered are attractive and when the events are not scheduled too often or too close to other promotions. Finding more historical information for ‘Days of Deals’ might help set up a combination of event flags that could help improve the predictions. Again, the combination of events that determine the increase in demand on these weeks might simply prove to be too complex to model mathematically.

Weeks 3 and 7 are excellent examples of weeks that should be carefully looked into by forecasting teams at Dell. There is not enough information in the historical data for an ARIMA model to forecast these weeks accurately. However, some of this information is known before producing the forecast. The information regarding planned increase in marketing spend and aggressive deals offered during President’s day was not available to the time series model but it was available to the forecasting team at Dell. Leveraging this knowledge and their expertise might help us improve the accuracy of forecasts on weeks where special events take place.

3.3.3. The need for judgmental adjustment

The baseline forecast is created using time series models not based on any known information about the period being forecasted. Embedded in historical calls data there is information regarding marketing spend, market conditions, price aggressiveness, seasonality, etc. The baseline is then created assuming the future is like the past, thus information about marketing spend, for example, is contained in the form of expected spend.

Information regarding future spend, future activities or promotions is not available to the time series model directly. It is, however, available to the forecasting team. Since future marketing spend or activities are designed to influence demand, we should assume they will

affect the accuracy of the baseline. We should make use of known information regarding marketing spend and promotions schedule to adjust baseline forecast.

Time series models assume that marketing spend on different vehicles will continue to follow trends similar to the trends experienced in the past. Large increments or large decreases on spend can be expected to cause a reaction in the demand signal. Having advance information on changes in spend might help us predict their impact and adjust our forecasts.

Sometimes similar changes on spend might produce very different responses. Action-driven advertisement and campaigns available for a very limited period of time are known to produce an important change in demand for a short period of time. Brand-building advertisement will produce a moderate response for a longer period of time. On a similar note, spend on Direct Mail advertisement is expected to have very different response rates depending on the customer segment or geography being targeted.

In a similar way, events that are not periodical or that don't follow a deterministic pattern might be particularly challenging for SAS Forecast Studio to predict. This is the case, for example, of 'Days of Deals'; promotions that usually last from 2 to 5 days and offer reduced prices. The response to these promotions has varied widely over time, and it is not just linked to the event but also to how aggressive prices are, how long the promotion lasts and how wide the variety of products offered is. As mentioned before, simply defining a dataset of events for 'Days of Deals' did not help improve the accuracy of the forecast. Events datasets are not only used by Forecast Studio to explain past data but also to forecast future occurrences of non-periodical events. However, the magnitude of the response is hard to predict when historical responses, as in this case, are not consistent. Again, the information is available to the forecasting team and it should be used to adjust the baseline forecast.

We could also expect demand to drop on the week, or weeks, following a high demand week. A time series model will not predict this well-known effect of demand being pulled forward if it couldn't predict the high-demand week. Scheduling a 'Days of Deals' event on week 7, for example, is likely to reduce the demand on week 8. It may also, if the promotion schedule is known to customers in advance, reduce demand on the previous week.

Adjusting a baseline forecast to incorporate all this information is a process that requires knowledge of contextual information, expertise on market response and experience on forecasting demand. We propose an adjustment process that would be based on judgment and expertise rather than on a mathematical algorithm. This adjustment process will be similar to the process used today to forecast demand and we will improve this process by facilitating access to information that is normally difficult to obtain.

3.3.4. *Adjusting time series (ARIMA) models*

Figure 11 shows the baseline forecast for FY10 Q1 created using SAS Forecast Studio. As we have seen, forecasts are accurate except for week 3, President's day week, and week 7, 'Days of Deals' week. These are weeks for which we can expect a response different from other weeks.

Before the forecast is produced, the forecasting team has information regarding schedules, featured deals, marketing spend, etc. for the entire quarter. This information should be enough to critically review the forecast baseline and identify weeks that should be carefully examined and adjusted accordingly. Known information about the incoming quarter, combined with the expertise of people at Dell, could help improve the forecast accuracy.

In order to examine how this adjustment can improve the accuracy of forecasts, we will try to adjust sales calls volume forecasted for FY10 Q1 weeks 3 and 7. One way of adjusting this forecast without introducing our own judgment or after-the-fact knowledge is to simply rely on the forecast published by the MSP team. While the accuracy of the MSP forecast is what this work is trying to improve, the MSP team has the information, knowledge and expertise required to evaluate the impact of the promotions and marketing activities mentioned. Furthermore, their work for this particular quarter was completely independent of our efforts and thus completely unbiased.

Figure 13 shows an adjusted forecast for FY10 Q1. To create this adjusted forecast we simply replace the call volumes predicted by Forecast Studio for weeks 3 and 7 with the volumes originally forecasted by the MSP team. This is a good example of how the use of contextual information, knowledge about future marketing actions and judgment can help improve the

baseline forecast created by a time series model. Figure 14 shows forecast accuracy expressed as calls received as a percentage of adjusted forecast.

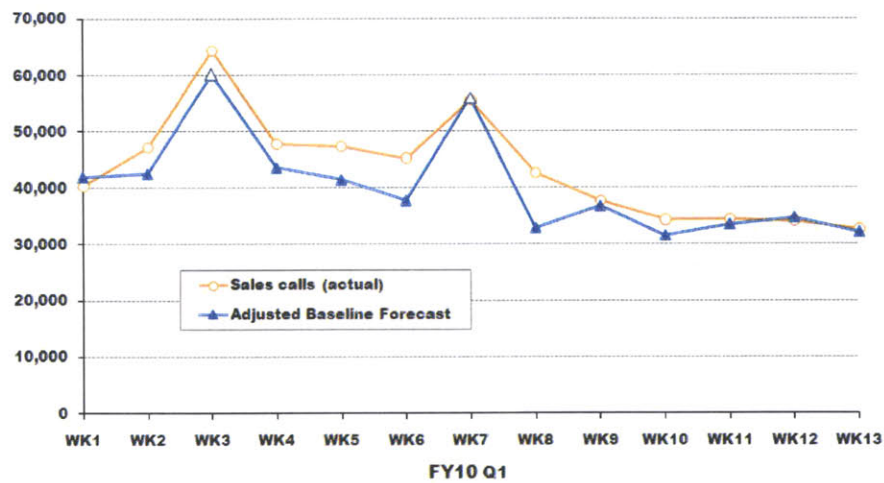


Figure 13. Adjusted Baseline Forecast, FY10 Q1

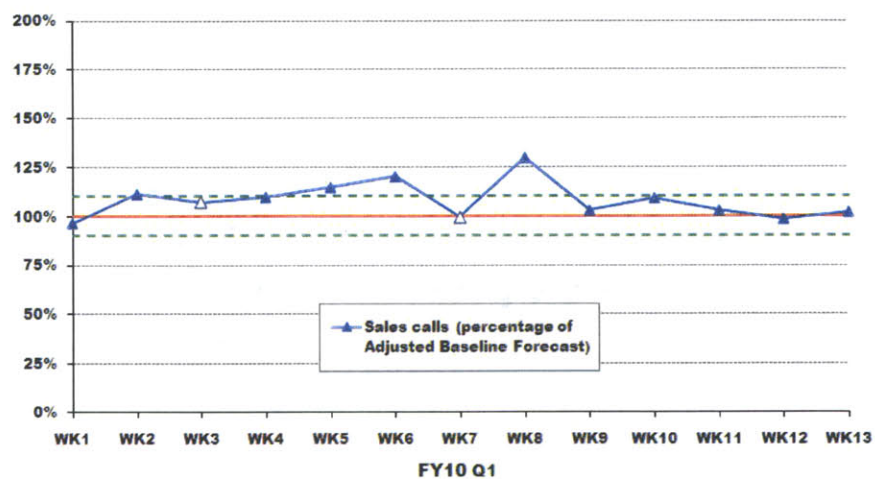


Figure 14. Calls received as percentage of adjusted forecast, FY10 Q1

This work will not propose a particular methodology for adjusting baseline forecasts, since this methodology should be largely dependent on the forecasting team’s expertise and judgment. In order to facilitate this adjustment process, we have created an adjustment template and a SAS Enterprise Guide (EG) procedure that automates the access to most of the information relevant to adjustments on demand baseline forecasts. Easy access to information and standard reporting of this information might accelerate the process and provide it with methodology and

repeatability. As discussed in chapter 2, research shows that judgmental forecasting can be improved by the incorporation of methodology into the forecasting process.

The SAS EG procedure created will gather information from several data sources, arrange this information according to the period to be forecasted and export it into an Excel spreadsheet. The adjustment template displays in an easy to see format the following information:

- **Baseline call volume forecast by week.** Forecasted sales calls volume, as produced by a time series model using SAS Forecast Studio, will be displayed here. This is the baseline forecast we are trying to evaluate and adjust.
- **Total margin per unit (TMU).** As a proxy for price aggressiveness, TMU can have a significant impact on expected call volume. Template displays the plan TMU, by week, for the entire quarter and the correspondent changes over the previous week, previous quarter and previous year, both in dollar figures and percent increase (or decrease).
- **Marketing spend.** Marketing plan, by week, for the entire quarter and the correspondent changes over the previous week, previous quarter and previous year, both in dollar figures and percent increase (or decrease).
- **Call volume contribution from each vehicle.** In order to assess the impact of marketing changes on a determined vehicle, we need an estimation of the relative contribution of the vehicle to the total call volume. We estimate and display the expected contribution of each vehicle by week for the entire quarter.⁷
- **Estimated response rate from each marketing vehicle.** Similarly, we estimate the expected response rate of each vehicle in order to help assess the impact of a marketing spend change on the forecasted call volume.⁸
- **Overrides.** Template offers the ability to override the total or vehicle contribution call volume for any week and calculates the new total forecasted volume.

⁷ Despite concerns with accuracy, we have used Call by Vehicle data to provide an estimation of call contribution by vehicle. This estimation is not assumed to be accurate but to only provide some guidance on estimating the impact of marketing changes to call volume.

⁸ Without relying on contextual information, the adjustment template provides an estimation of the number of additional calls expected by any additional dollar spent in advertisement.

An example template is shown in Figure 15:

First week to forecast		1/31/2009												
		1/31/2009	2/7/2009	2/14/2009	2/21/2009	2/28/2009	3/7/2009	3/14/2009	3/21/2009	3/28/2009	4/4/2009	4/11/2009	4/18/2009	4/25/2009
		FW1	FW2	FW3	FW4	FW5	FW6	FW7	FW8	FW9	FW10	FW11	FW12	FW13
Baseline calls		41,950	42,466	47,409	49,558	41,400	37,726	34,770	32,895	36,754	31,542	33,528	34,741	32,160
Adjusted call forecast		41,950	42,466	47,409	49,558	41,400	37,726	34,770	32,895	36,754	31,542	33,528	34,741	32,160
Manual override														
Total adjusted calls		41,950	42,466	47,409	49,558	41,400	37,726	34,770	32,895	36,754	31,542	33,528	34,741	32,160

Catalog adjustment														
		FW1	FW2	FW3	FW4	FW5	FW6	FW7	FW8	FW9	FW10	FW11	FW12	FW13
Expected calls by vehicle		5,420	5,486	6,125	5,627	5,349	4,874	4,492	4,250	4,748	4,075	4,332	4,488	4,155
Spend by vehicle - Plan		\$335,280	\$559,520	\$684,112	\$760,232	\$820,160	\$873,400	\$904,640	\$926,760	\$628,232	\$815,162	\$940,500	\$1,014,343	\$1,051,612
Delta Spend W/W		-\$238,700	\$218,240	\$130,592	\$76,120	\$59,928	\$53,240	\$31,240	\$32,120	-\$308,528	\$186,930	\$125,338	\$75,843	\$37,269
		-42%	65%	24%	11%	8%	6%	4%	4%	-33%	30%	15%	8%	4%
Delta Spend Q/Q		-\$674,678	-\$597,489	-\$547,494	-\$682,647	-\$767,767	-\$789,077	-\$211,712	\$182,818	-\$159,508	-\$8,201	\$104,177	\$206,386	\$477,631
		-67%	-52%	-44%	-47%	-48%	-47%	-19%	24%	-20%	-1%	12%	26%	83%
Delta Spend Y/Y		-\$230,173	-\$348,529	-\$479,528	-\$525,230	-\$480,028	-\$506,566	-\$569,056	-\$579,142	-\$839,860	-\$581,524	-\$281,310	-\$314,639	-\$320,938
		-41%	-39%	-41%	-41%	-37%	-37%	-39%	-38%	-57%	-42%	-23%	-24%	-23%
Estimated response rate		0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075	0.0075
Calls by vehicle - Override														
Total calls by vehicle		5,420	5,486	6,125	5,627	5,349	4,874	4,492	4,250	4,748	4,075	4,332	4,488	4,155

Figure 15. Baseline adjustment template

Finally, this template will facilitate evaluating different marketing spend scenarios. By facilitating the access to spend, call volume and response rate information, we facilitate assessing the impact of marketing spend changes on the expected call volume. In this way, the call forecasting model could be used as part of the marketing plan decision making process.

3.3.5. Summary on time series models

Time series models are widely used for forecasting purposes. Even within Dell, time series models are common. The misconception regarding the irrelevance of marketing activity has prevented their use in call forecasting models.

We have shown how time series models, particularly ARIMA models, can produce fairly accurate forecasts. Furthermore, we have shown how we can leverage the knowledge regarding marketing information and schedule to adjust and improve these forecasts. The use of SAS Forecast Studio greatly simplifies the model selection, evaluation and checking process while the adjustment process can facilitate the evaluation of scenarios and the incorporation of known information to the forecast.

Forecast accuracy and hit rate improvements are very promising. Chapter 6 will examine the accuracy of the overall model in detail. The potential improvements offered by time series models are evident from the data shown.

Furthermore, the time series models explored in this work could be greatly improved by leveraging additional data and hierarchies. Chapter 7 will review some ideas regarding future work in detail, including leveraging Forecast Studio to produce:

- Hierarchical forecast for regions.
- Calls by vehicle forecast.
- Sales calls and total incoming calls.

Finally, we should mention again that the sales calls dataset was created assuming a constant transfer out rate for the entire period of eight years. We believe this disguises features in the data, features that are particularly important when using ARIMA models. Improving the accuracy of the existing dataset will certainly improve the accuracy of the forecasts produced.

This page has been intentionally left blank.

4. Day of week distribution

So far, we have been examining and comparing different methodologies to forecast weekly call volume. In order for this forecast to be useful to the operations group for scheduling sales agents, we need to produce daily call volumes. In other words, we need to distribute the weekly call volume into daily volume.

This is probably the point where sales call volumes differ the most from all other magnitudes forecasted as part of the Master Sales Plan (MSP). For other magnitudes, like units to be sold or visits to Dell's website, the accuracy of daily volumes is not critical. A unit not sold today can be sold tomorrow and, for as long as the total units sold over the entire period meet the target, variance from daily forecasted volumes is not critical. Accuracy of daily sales calls forecasts, on the other hand, is critical to properly schedule agents to handle these calls. We can't make up for an agent not scheduled today by scheduling extra staff tomorrow. For this reason, it is important that the weekly forecasted volume of calls is accurately distributed among days every week. Even a perfectly accurate weekly forecast can produce inaccurate daily forecasts if the day of week distribution is not correct.

In the past, when the MSP was being produced and the weekly call volume has been determined, the day of week (DOW) distribution of the previous 5 to 10 weeks was assumed to be identical to the expected DOW distribution of the incoming quarter and it was used to predict weekly volume for the next thirteen weeks. The 'not-normal' weeks were not part of this process and were considered in isolation. This means that for periods like Q4 finding 10 weeks to average could take us back several months since all the weeks around Thanksgiving and other holidays were not considered 'normal'. Also, when trying to predict DOW distribution on a holiday week, last year's distribution was used rather than the average of last few weeks. No historical or seasonal trends were used when forecasting DOW distribution.

Despite its simplicity, this model is fairly accurate. To test its accuracy independently from the accuracy of weekly forecasts, we replaced the weekly forecasts produced by the MSP team with a perfectly accurate forecast for the entire year FY09. If the forecasted weekly volume had been 100% accurate for every single week throughout the year, the DOW distribution used during FY09 would have produced the following distribution of daily hit rate.

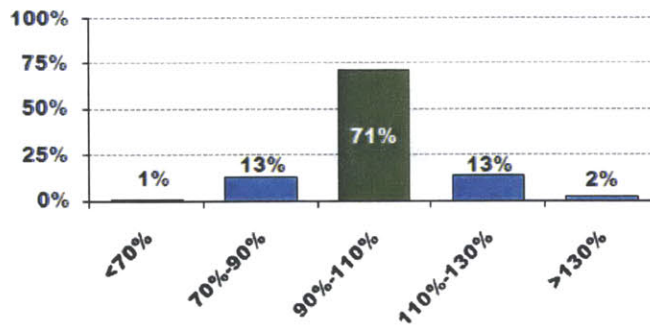


Figure 16. Day of week distribution accuracy, Master Sales Plan

Where accuracy is expressed as actual calls received, by day, as a percentage of calls forecasted. Thus if the weekly forecasted volume would have been perfectly accurate, the daily forecasted volume would have been within the acceptable range 71% of the days.

The simplicity of the methodology has raised questions among Dell employees and some were convinced that improving just this part of the forecasting process would improve the accuracy of the overall forecast. Some of these questions are worth exploring. Is DOW distribution in summer and winter months identical? Some employees at Dell believe that call volume during summer weekends is lower than during winter weekends. Are there any intra-quarter trends in DOW distribution? Are all weeks with a Monday holiday the same? Are there any trends in holidays? Some, like President's day, seem to be becoming more and more important over the years. How are the weeks of Memorial day, Labor day and President's day different?

Unfortunately, available data for sales calls only go back three years. This is not enough to analyze the presence of yearly trends. If we can assume that DOW distribution is similar before and after the Live Voice Recognition (LVR) filter, we can attempt to answer some of these questions and draw conclusions applicable to sales calls distribution.

This work will not propose a methodology for forecasting DOW distribution. Based on our experience at Dell, we realize that proper interpretation of features in daily distribution requires significant expertise and knowledge about past events. At the same time, daily distribution is highly impacted by changes in marketing schedule and activity. The schedule of catalog drops, promotional emails or direct mailing can affect the distribution of calls on a given week. Changes to these schedules are not uncommon. We did not find information on these

schedules that allowed us to try to create a model for prediction. Schedule information for incoming weeks, however, is usually available to the forecasting team. At the same time, there are some interesting trends developing around some holidays, Valentine's day and President's day in particular. For the last few years, demand has been increasing around these two days. This increase in demand was followed by an increase in marketing activity that boosted demand even more. We could not find a model to link demand and marketing activity, but the forecasting team at Dell seems to be particularly good at quantifying this kind of response.

We focused then on providing access to information that would help evaluate past distributions, identify trends and analyze historical data. We have created a DOW dashboard that quickly and easily provides access to information on distribution during normal and holiday weeks as well as seasonal and intra-quarter trends. We will analyze each one of these four aspects separately.

4.1. Normal weeks

As we can see from FY09 data, the yearly average, excluding some high demand weeks, of DOW distribution seems to be a good predictor of future distributions. As a baseline, this seems to be a good starting point to try to predict what normal weeks are going to look like.

We will first focus on normal or typical weeks. For the purpose of this work, normal weeks are weeks that don't include a holiday. Of course, the definition of 'normal' is critical and highly subjective. There are several holidays that might or might not impact the distribution of calls. At the same time, periods of high demand like back-to-school or the period between Thanksgiving and Christmas might not behave like the rest of the year. For this reason, the dashboard was designed to allow for an easy redefinition of 'normality'. The dashboard, shown in Figure 17, allows the user to easily include or exclude from the statistics being calculated the weeks including the following days:

- Valentine's day
- President's day
- Mother's day
- Memorial day
- Father's day

- 4th of July
- Labor day
- Back to school season⁹
- Thanksgiving day
- Christmas eve
- Holiday season¹⁰
- New year's eve

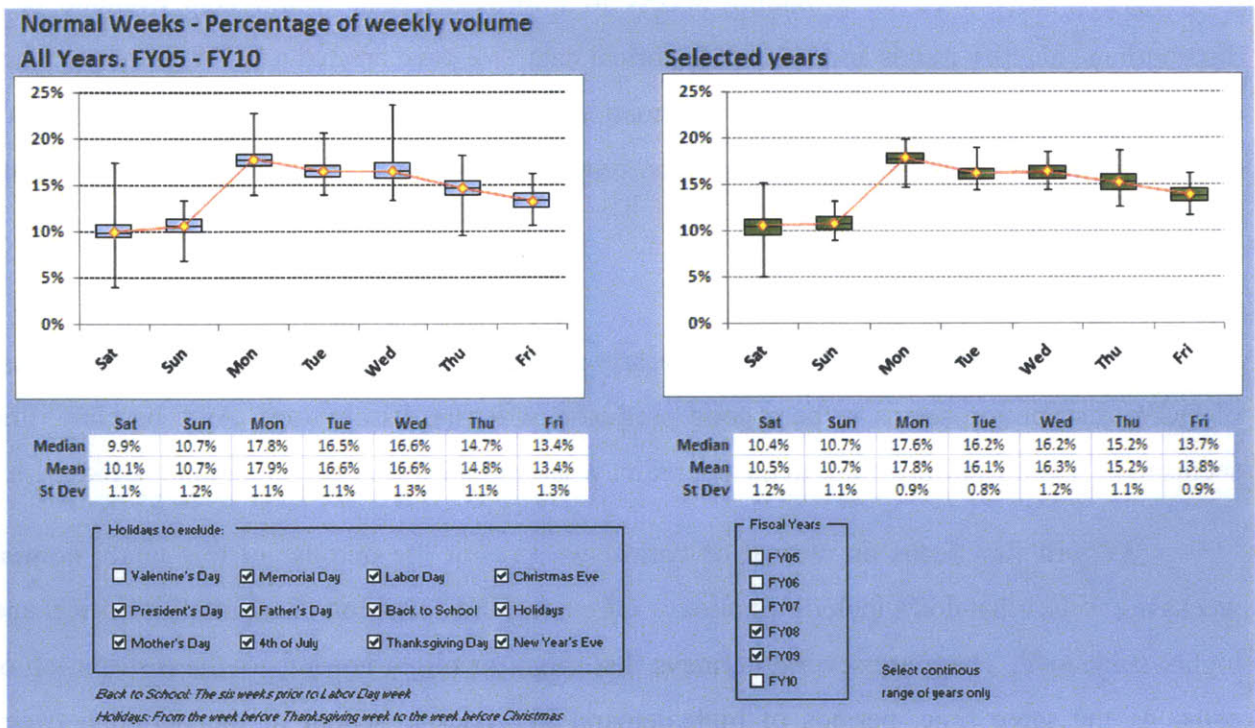


Figure 17. DOW Dashboard, Normal weeks

At the same time, the dashboard allows for selecting any year, or any continuous range of years, from FY05 to FY10 in order to analyze the DOW distribution of a particular year or exclude from the statistics data that might not reflect current conditions any longer.

4.2. Holiday weeks

Embedded in the concept of normal weeks then, is the idea that some weeks are different. It is easy to imagine that holiday weeks will behave differently from normal weeks. But, how

⁹ Defined as the six-week period before the week of Labor Day.

¹⁰ Defined as the period from the week before Thanksgiving through and including the week before Christmas.

different? Are they just different from normal weeks or from one another as well? Are there year-over-year trends that we should consider when predicting distribution for future weeks?

These are important questions to answer in order to predict DOW distribution for each one of these weeks. A closer look to the data reveals that these weeks are very different from normal weeks and very different from each other. Some of them are stable over the years but some other present significant trends. Each one of these weeks then, should be analyzed separately. Figure 18 shows a sample of the different distribution found on holiday weeks.

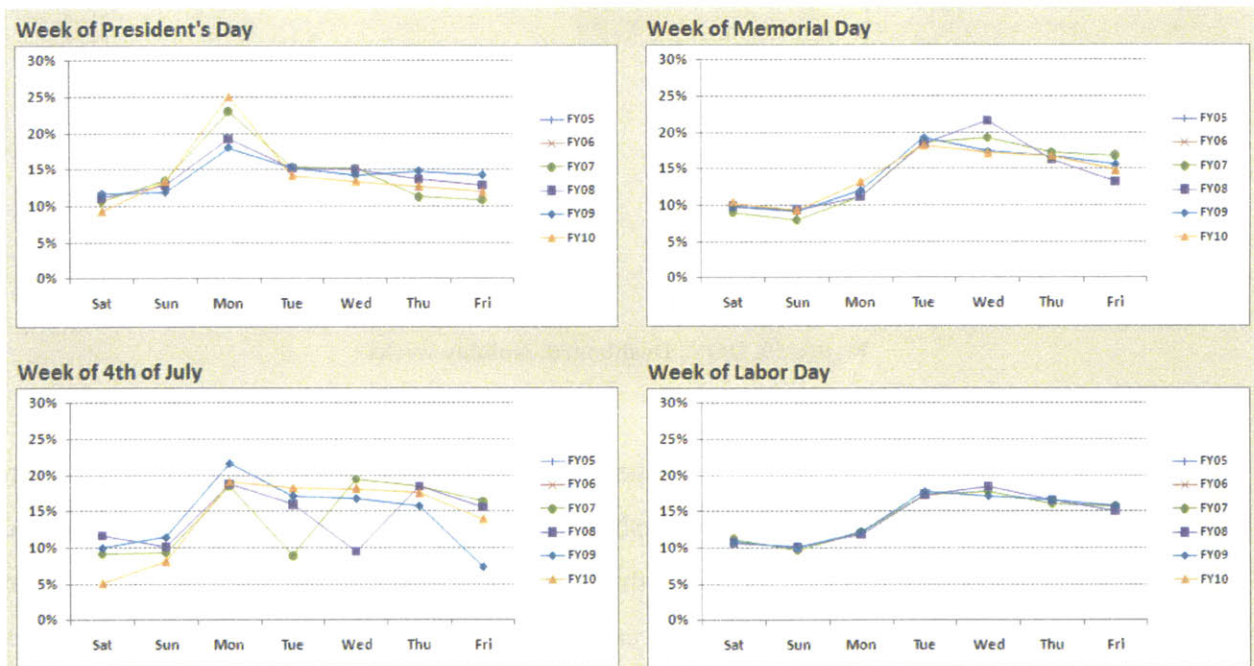


Figure 18. Sample call distribution on holiday weeks

The analysis of each one of these weeks and their historical trends along with planned marketing information will help forecasters estimate the expected distribution better. The dashboard created allows for the individual analysis of each one of these weeks over any period of time between FY05 and FY10. Figure 19 below shows a snapshot of the dashboard.

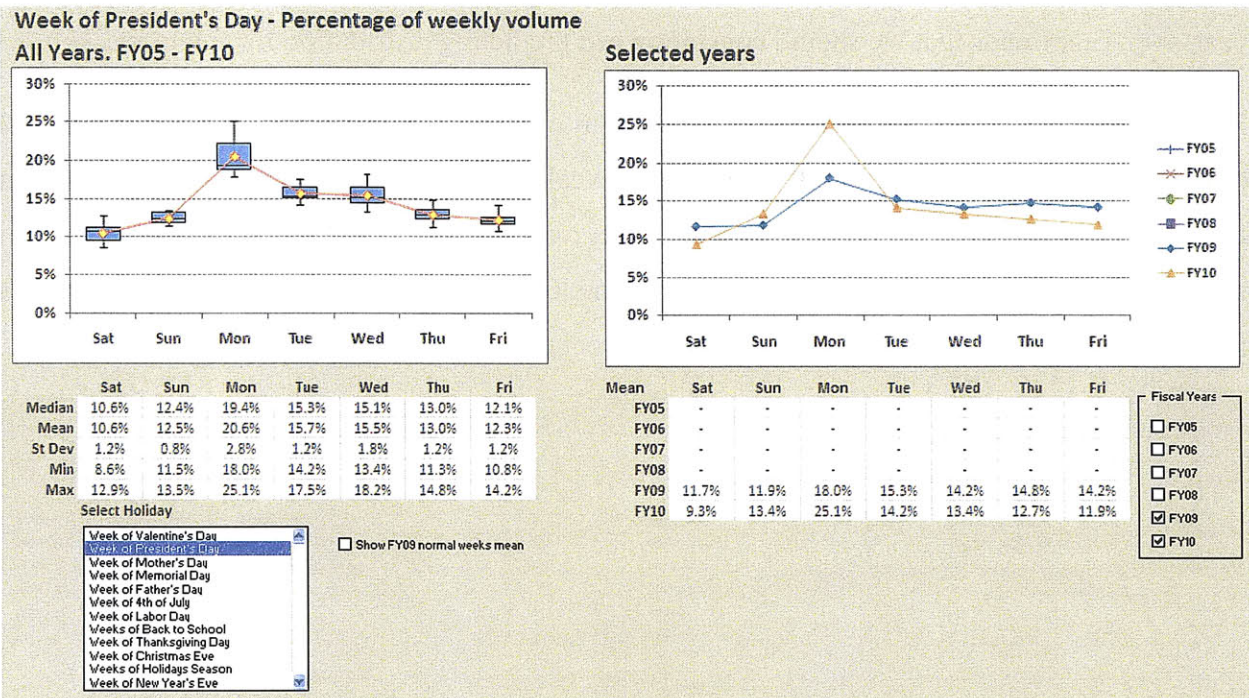


Figure 19. DOW Dashboard, Holiday weeks

4.3. Seasonality

We have shown that yearly averages are reasonable good predictors of future distributions for normal weeks. We have also shown how distribution for not-normal weeks can be estimated. There are however, questions that can't be answered yet. One of the recurrent questions at Dell involves the seasonality of calls volume and call distribution. As mentioned before, there is a strong suspicion that summer and winter months are fairly different.

The seasonality screen of the dashboard displays weekly distribution of calls by quarter and by year, facilitating comparisons among different quarters and different years. Figure 20 shows quarterly distributions for FY08 and FY09.

Using these data we can easily compare yearly averages to quarterly averages in order to determine whether there are seasonal effects on DOW distribution. More importantly, we can use this dashboard to assess the impact of any event on the weekly distribution. For example, did the introduction of retail sales channels or the change on catalog drop schedule affect the distribution of calls? Did the change on catalog drop schedule affect the call volume distribution? Questions like these are typical of the forecasting process at Dell and some need to be evaluated on a

recurrent basis. The dashboard provides access to information in order to evaluate the impact of events like these.

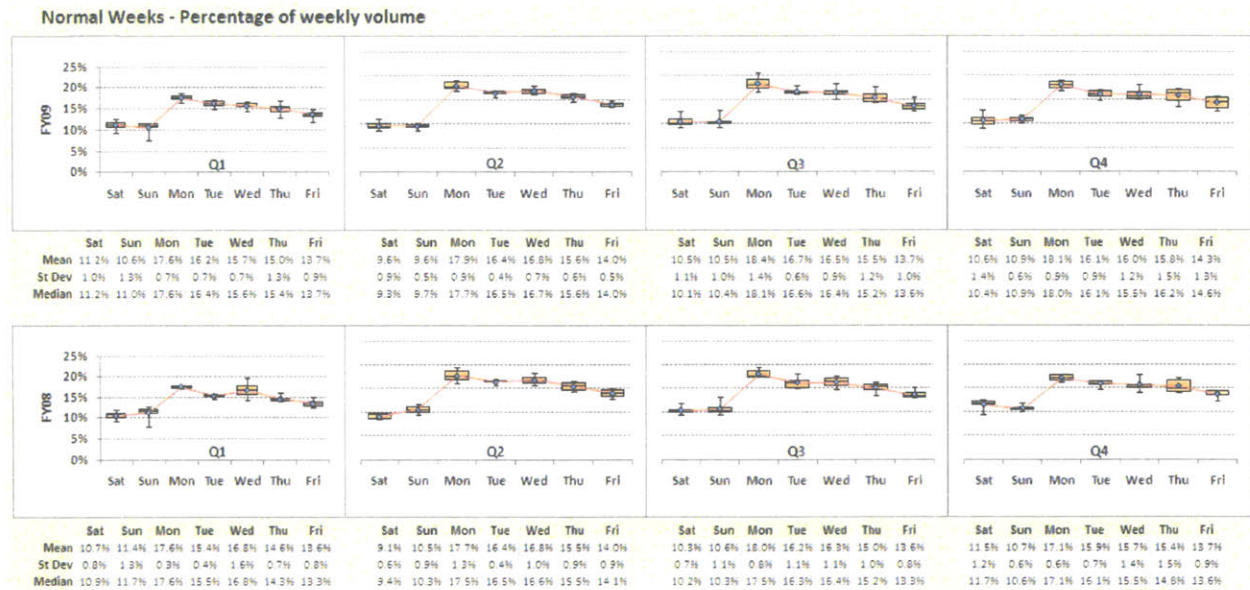


Figure 20. DOW Dashboard, Seasonality

4.4. Intra-quarter trends

Finally, we have included in the dashboard the ability to check for intra-quarter trends in DOW distribution. We must assume that if effects like seasonality have an impact on the distribution of calls, this impact would be gradual over several weeks rather than a sudden step between quarters. The intra-quarter trends screen on the dashboard (Figure 21) will display trends for any day of the week over any year or combination of years.

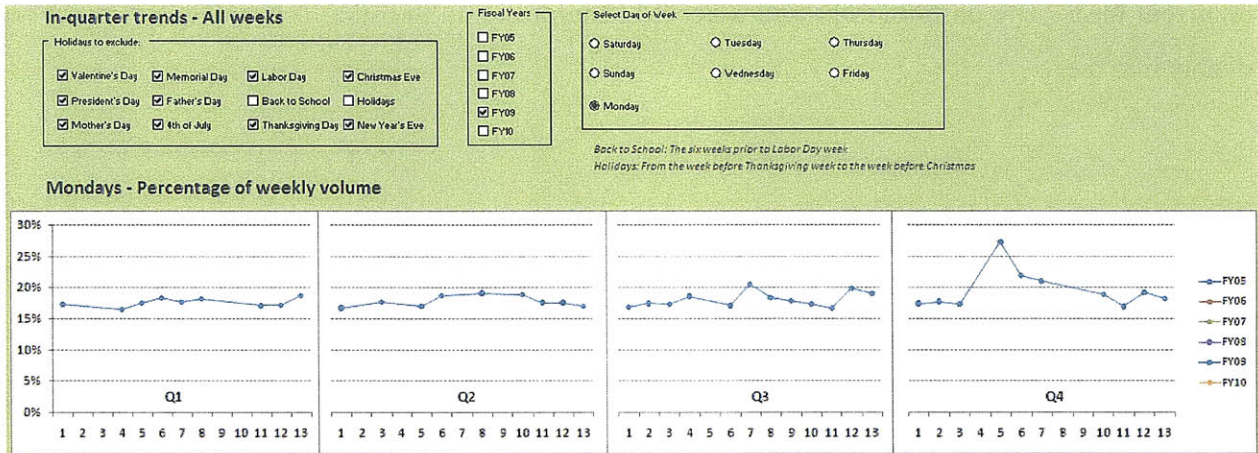


Figure 21. DOW Dashboard, intra-quarter trends

4.5. Summary on day of week distribution

We have described in this chapter the importance of accurately distributing the weekly call volume forecast into daily volumes. As opposed to other forecasts in the MSP, the daily accuracy of sales calls is critical. Agents scheduling heavily relies on this forecast so its accuracy translates to costs at Dell.

The process followed at Dell was too simple to capture all the subtleties of the day of week distribution of calls. We have created and introduced here a dashboard that allows users to display historical DOW distribution information in a myriad of ways. The analysis of this information will facilitate the prediction daily distribution in future weeks. This prediction is dependent on expertise and knowledge about past events and their impact. Since this information is not easily quantifiable, it is difficult to assess the benefits of the dashboard here presented.

However, we can attempt to evaluate the benefits by quantifying one aspect of the DOW distribution that has not been considered at Dell before, seasonal effects on DOW distribution. Starting from a perfectly accurate weekly forecast for FY09, assuming that weekly forecasted call volume was exactly the volume received on a given week, we can compare the DOW distribution used by the MSP team (MSP DOW) with a distribution that considers seasonal effects (Seasonal DOW). To create this seasonal DOW we assumed normal weeks distribution was identical to the distribution observed during the same quarter the previous year, this is FY09 Q1 distribution was assumed to be identical to FY08 Q1 distribution. Holiday weeks were assumed to be identical to the same week the previous year. For weeks like 4th of July week,

where the intra-week pattern varies since the holiday does not always fall on Monday, a DOW distribution was estimated based on previous year distribution around the holiday.

This simple change on DOW distribution improved the daily hit rate from 71% to 77%. This is a significant improvement, 6 percentage points represent about three weeks on a year. Hit rates distributions, using MSP and seasonal DOW, are shown in Figure 22.

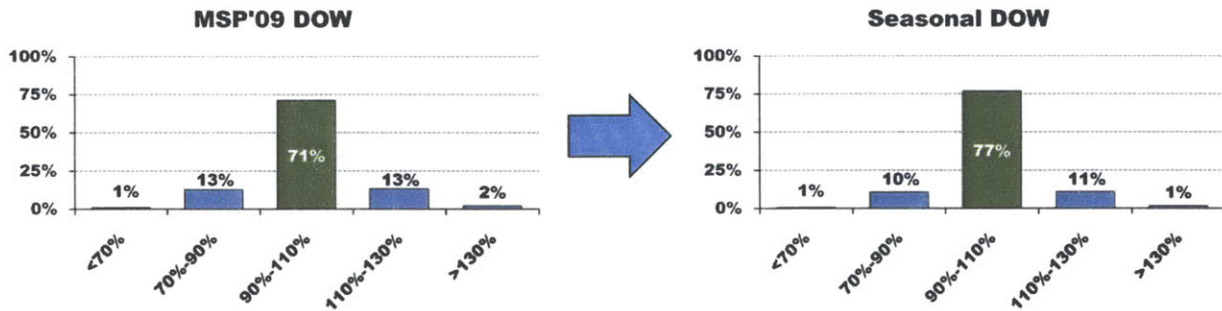


Figure 22. Hit rate improvement from day of week distribution

The benefits of using historical information are clear but we should remember that DOW distribution prediction, as weekly call volume, has a judgmental component that we should not ignore. The averages and trends revealed through this dashboard should be used carefully and good judgment should be used to override some of them. A significant increase in marketing activity during FY10 boosted demand on President’s day. This increase in demand would have not been predicted from historical trends alone (Figure 18). Fortunately, increases in marketing activity are known to members of the Master Sales Plan (MSP) team and they can be incorporated into the forecast.

The dashboard and the information it reveals don’t replace expertise and judgment, they complement it and facilitate dialog and agreement among the parties involved in the forecasting process. At the same time, it automates the access to data so the review and analysis of DOW distributions can be quickly incorporated into any existing process. Over time, it might become the first step taken towards the development of the DOW knowledge required for DOW prediction.

This page has been intentionally left blank.

5. Intra-quarter projection process

An interesting situation developed during my first week at Dell. A few days prior to my arrival, the call center experienced one day of abnormally high demand. During the month of January, when call variability is normally low, such event was particularly interesting and unexpected. The problem is that not being prepared for a sudden increase in demand normally means a large number of calls can't be answered and potential sales are lost.

In very little time, the situation became clear. The high demand day was not unexpected at all. Noticing the observed demand had been lower than expected, the marketing group ran some promotions valid for a single day.

It is not difficult to imagine the situation. On the one hand, the marketing group had spent time and resources on a campaign to boost demand only to see that the additional demand generated was lost due to the inability of the call center to handle the volume. On the other hand, the operations group scheduled sales agents based on the latest information available to them, a quarterly forecast that had not been revised to incorporate any changes to the original plan.

It was clear that at some point the communication link was broken. This episode helped reveal some flaws in processes that had to be resolved. In order to address this problem, a new projection process was put in place. The MSP team was made responsible for issuing an outlook of expected demand for the next two weeks every Wednesday. In order to generate this projection, a weekly meeting among MSP, Marcom and CCO team members was scheduled. During the meeting the previous week projection accuracy was reviewed and all incoming promotions and marketing activities were discussed, particularly if they were different to the original marketing plan used to produce the quarterly MSP forecast. After each meeting, the MSP team published the two-week projection mentioned.

Figure 23 and Figure 24 show sample results of the accuracy and hit rate improvements of these projections over the quarterly (MSP) forecast after the projection process was instituted.

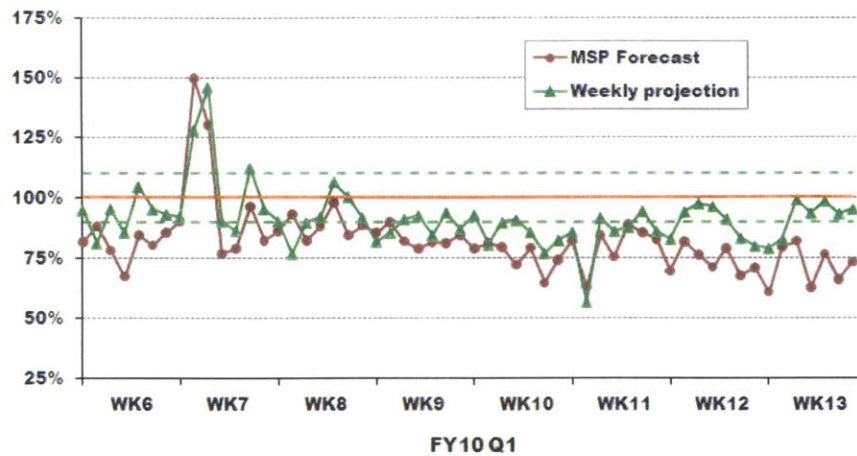


Figure 23. Actual calls as percentage of MSP and projection

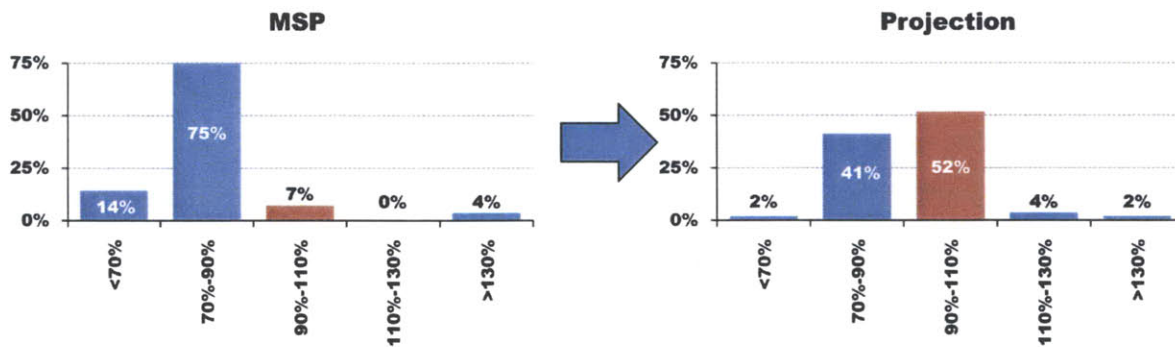


Figure 24. MSP forecast and projection hit rates

After validating this process for a few weeks, sales agents were scheduled based on projection volumes rather than forecast volumes. While this process was originally designed to address issues on the intra-quarter changes to forecast and plan, it indirectly relaxed the daily accuracy requirements on the quarterly forecast by introducing a buffer between the quarterly forecast and the daily agent scheduling.

The forecast model proposed in this work was designed with this projection process in mind. We have tried to create tools that could be used every week to produce projections if this were necessary. Whether it is to re-run time series models including recent data, or to update and revisit data used for adjustments, or to simply rethink day of week distributions in light of new information, the model and tools created should aid in the process. By using these tools, the entire process will benefit from consistency and methodology, which, over time, will contribute to the expertise development process.

This new process provides numerous benefits beyond the improvements in accuracy. Sharing information among groups and agreeing on steps to take will avoid surprises and lost opportunities. It will improve the communication among groups and reduce finger pointing. The usefulness of this process however, relies in the end on the accuracy of the projections it produces. Dell's management will trust and use these projections for as long as they are reasonably accurate. We hope that the use of the tools presented in this work will facilitate the process and improve the accuracy of these projections even further.

This page has been intentionally left blank.

6. Results of forecasting model

In section 3.3 we evaluated ARIMA models and determined they can create sales calls forecasts of acceptable accuracy, at least for the validation dataset, FY10 Q1, we withheld. In order to test these models further, we will produce a forecast for the entire fiscal year 2009 and compare the results again the MSP plan published for the same year.

From five years of historical data, FY04-FY08, we produced several models using SAS Forecast Studio. The model that best fitted the historical data was then used to forecast the entire year following a schedule identical to the one used by the MSP team. During the last week of the quarter a forecast for thirteen weeks was published. During week 6 a new forecast, this one for 7 weeks was published. Following this schedule we produce a forecast, which we will call baseline, for an entire year that we can compare to the published plan.

The baseline forecast is significantly more accurate than the published plan for FY09. Weekly hit rate improves from 23% to 46%, Figure 25 and Figure 26 show the improvements on forecast accuracy and hit rate. Once again, five years of historical data seemed to provide the best results for the dataset available. Models created using less than five years of data failed to predict seasonal features (back to school and holidays) properly, models using more than five years tend to give more weight to the downward trend and underestimate call volume. We also tried to use TMU as an independent variable. While TMU is statistically significant, the forecasts produced by models including TMU were not as promising as the model chosen. Total weekly hit rate for FY09 when using TMU as independent variable improved from 23% to 41%.

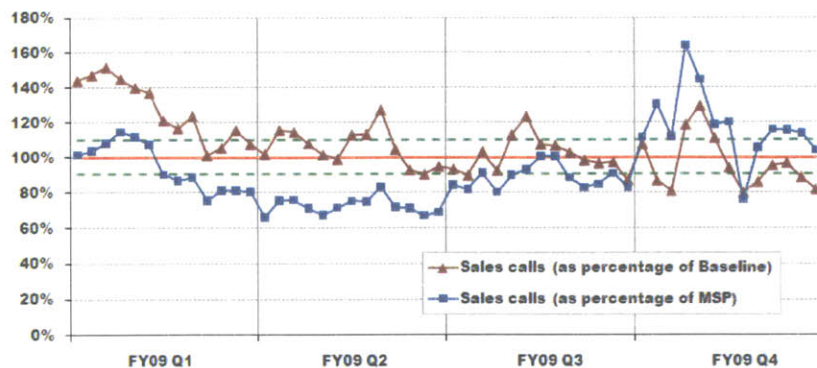


Figure 25. Weekly forecast accuracy comparison

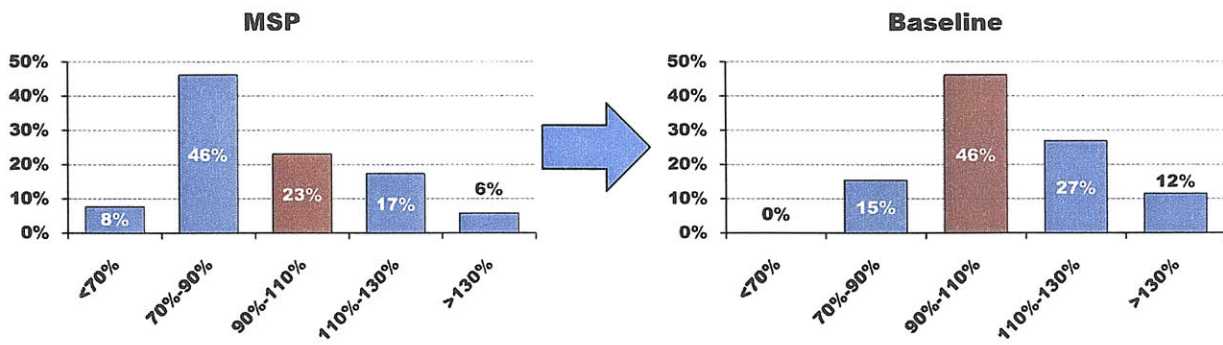


Figure 26. Weekly forecast hit rate comparison

It is worth repeating here that improving the accuracy of the historical dataset might help improve the accuracy of these predictions. If we can find real sales calls data instead of assuming a constant 30% transfer out rate, the accuracy of these models should increase.

The accuracy of this baseline is significantly better than the previous model used. Time series models can provide the increase in forecasting accuracy that we are looking for. We should, however, consider them a first step on a new forecasting process. The output of any time series model, the baseline, should be carefully examined to identify weeks that the model is not predicting accurately. These weeks can then be adjusted to incorporate information available, like incremental marketing spend, offers, promotions or ‘Days of Deals’, that are not available to the time series model. Incorporating this information into the forecast requires knowledge and expertise.

We have seen how this adjustment process can improve the accuracy of forecasts. However, we will not attempt to quantify the benefits of this process here. The adjustment process is highly subjective, based on several different sources of contextual information and extremely difficult to do after-the-fact without introducing bias. It would be difficult for Dell’s forecasting team to adjust last year’s forecast in an objective and unbiased manner and it would be almost impossible for us to do so.

Similarly, determining the day of week distribution is a subjective process. However, and as we did in Chapter 4, we can make some assumptions and propose a DOW distribution in order to assess the impact of the new process. To the baseline forecast for FY09 created using Forecast Studio we can apply the same DOW distribution mentioned in section 4.5. This means we

assume seasonal effects are present so distribution for normal weeks would be identical to the average DOW distribution of normal weeks during the same quarter the previous year and holiday weeks are identical to the same week the year before.

Using this DOW distribution, we can create a daily forecast for sales calls for the entire year FY09. We compare these results against the accuracy and hit rate obtained by the MSP forecast used in Figure 27 and Figure 28.

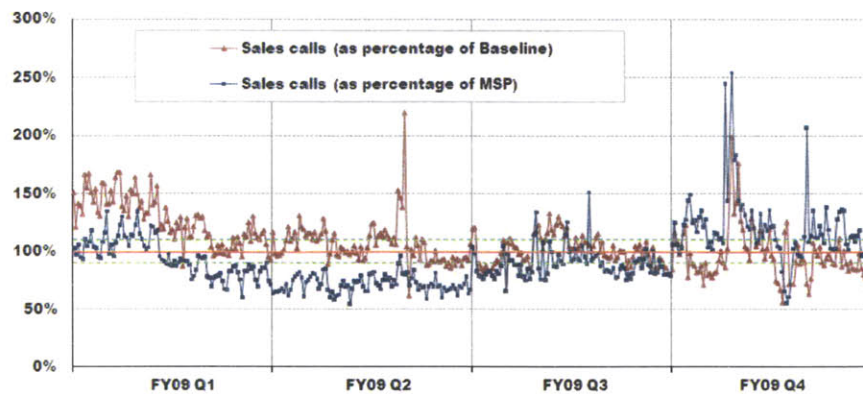


Figure 27. Daily forecast accuracy comparison

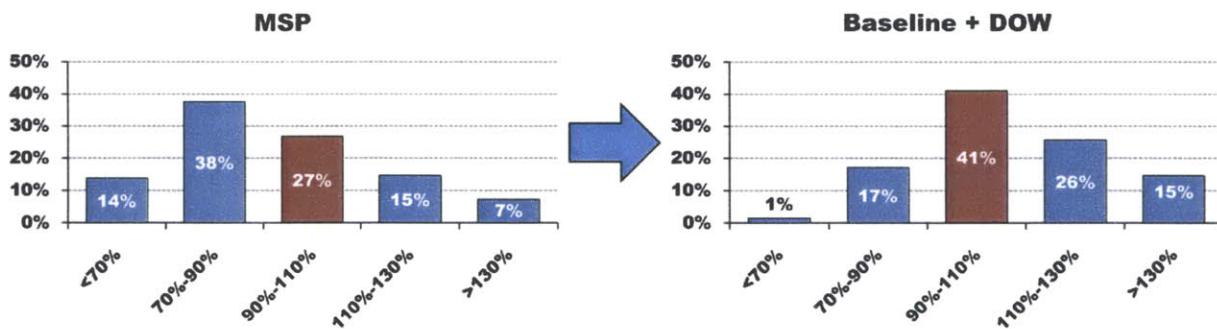


Figure 28. Daily forecast hit rate comparison

There is a significant improvement in daily hit rate for FY09, from 27% to 41%. We believe this is very conservative estimate of the potential impact of the proposed model and process on forecast accuracy. Further improvements should be easily achievable by the introduction of judgmental adjustments both to the weekly volume and the day of week distribution.

This page has been intentionally left blank.

7. Future work

This work has surveyed statistical methods for forecasting sales calls at Dell's U.S. consumer sales and it has shown the benefits of ARIMA models, particularly when combined with judgmental forecasting. There are several opportunities to build on this work to improve the accuracy of the predictions and the ease of forecasting processes.

The time series models explored in this work could be greatly improved by leveraging additional data and hierarchies. As mentioned by Nahmias (Nahmias, 2009), aggregate forecasts are always more accurate. SAS Forecast Studio greatly simplifies the process of hierarchical forecasting and future work could leverage this to modify the model presented here to:

- Produce hierarchical forecast for regions. The consumer sales group in Latin America needs to forecast sales calls volume for the entire region as well as for each one of the more than forty countries in it.
- Forecast sales and total incoming calls. Hierarchies could be leveraged to forecast sales calls along total incoming call volume. While less critical, total call volume is still very important for negotiating OSP contracts.

Very late during my internship the issues with the Calls by Vehicle (CBV) database were resolved. Sales calls by vehicle and total margin per unit (TMU) data apparently became available. These data would provide several opportunities to improve the models mentioned in this work:

- Multiple linear regression models for each vehicle could be developed and the aggregated volume used to forecast calls. Marketing spend and TMU by vehicle might be better correlated to calls by vehicle than to total sales calls.
- Hierarchical time series models could be developed to forecast sales calls. Marketing spend and TMU by vehicle could be used as independent variables to improve the accuracy of each sub-model. Forecast Studio can be used to create hierarchical forecast and aggregate the total sales call volume.

Finally, the model proposed herein for sales calls forecasting could easily be adapted to forecast other demand variables included in the Master Sales Plan. Volume of units to be sold,

visits to Dell's website, etc are being forecasted using methods similar to the one used for calls. Standardizing the process across variables, business units and/or regions could help develop forecasting expertise and improve models.

8. Acronyms

ARIMA: AutoRegressive Integrated Moving Average.

CBV: Calls By Vehicle.

CCO: Call Center Operations.

DM: Direct Mail.

DOW: Day Of Week.

DW: Durbin-Watson.

FSI: Free Standing Insert.

IFM: Integrated Forecasting Model.

LVR: Live Voice Recognition.

MAPE: Mean Absolute Percentage Error.

MARCOM: Marketing and Communications.

MSP: Master Sales Plan.

OSP: OutSource Provider.

RUM: Revenue, Units and Margin.

SOAR: Sales Outbound Agent Reach

SS: Sunday Supplement.

TMU: Total Margin per Unit.

WAPP: Weighted Average Price Parity.

This page has been intentionally left blank.

9. Bibliography

Adams, A. (1984). Using the calendar to improve sales forecasts. *Journal of the Academy of Marketing Science* , 12 (3), 103-112.

Aldor-Noiman, S. (2006). *Forecasting demand for a telephone call center: Analysis of desired versus attainable precision (Master Thesis)*. Technion - Israel Institute of Technology.

Andrews, B., & Cunningham, S. (1995). LL Bean improves call-center forecasting. *Interfaces* , 1-13.

Antipov, A., & Meade, N. (2002). Forecasting call frequency at a financial services call centre. *Journal of the Operational Research Society* , 953-960.

Arahal, M., Berenguel, M., Camacho, E., & Pavon, F. (2009). Selección de variables en la predicción de llamadas en un centro de atención telefónica. *Revista iberoamericana de automática e informática industrial (RIAI)* , 6 (1), 94.

Armstrong, J., & Green, K. (2005). Demand Forecasting: Evidence-based Methods. *Working Paper - Monash University, Department of Econometrics and Business Statistics* .

Bankston, S. (2008). *Improving the consumer demand forecast to generate more accurate suggested orders at the store-item level (Master Thesis)*. Massachusetts Institute of Technology.

Dalrymple, D. (1987). Sales forecasting practices: Results from a United States survey. *International Journal of Forecasting* , 3 (3-4), 379-391.

Edmundson, B., Lawrence, M., & O'Connor, M. (1988). The use of non-time series information in sales forecasting: a case study. *Journal of Forecasting* , 7 (3), 201-211.

Einhorn, M. (2007). *Managing forecast variability in a build-to-order environment (Master Thesis)*. Massachusetts Institute of Technology.

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* , 5 (2), 79-141.

Gill, K. (2008). *Lean concepts in customer care: adding value and reducing waste with proactive order status messaging (Master Thesis)*. Massachusetts Institute of Technology.

Green, K., & Armstrong, J. (2007). The Ombudsman: Value of Expertise for Forecasting Decisions in Conflicts. *Interfaces* , 37 (3), 287-299.

Hanke, J., & Wichern, D. (2004). *Business Forecasting*. Pearson - Prentice-Hall.

Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* , 22 (4), 679-688.

Kalekar, P. (2004). Time series forecasting using Holt-Winters exponential smoothing. *Seminar paper - Kanwal Rekhi School of Information Technology* .

Kohn, R., & Ansley, C. (1985). Efficient estimation and prediction in time series regression models. *Biometrika* , 72 (3), 694-697.

Koyck, L. (1954). *Distributed lags and investment analysis*. North-Holland Pub. Co.

Lawrence, M., & O'Connor, M. (2000). Sales forecasting updates: how good are they in practice? *International Journal of Forecasting* , 16 (3), 369-382.

Lawrence, M., Edmundson, R., & O'Connor, M. (1986). The accuracy of combining judgemental and statistical forecasts. *Management Science* , 32 (12), 1521-1532.

Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research* , 122 (1), 151-160.

Makridakis, S., & Hibon, M. (1997). ARMA models and the Box-Jenkins methodology. *Journal of Forecasting* , 16 (3), 147-163.

Mendenhall, W., & Sincich, T. (2003). *A Second Course in Statistics : Regression Analysis*. Prentice-Hall.

Nahmias, S. (2009). *Production and operations analysis*. McGraw-Hill.

Neter, J., Wasserman, W., & Kutner, M. (1990). *Applied linear statistical models : regression, analysis of variance and experimental design*. Irwin.

O'Connor, M., Remus, W., & Griggs, K. (1997). Going up-going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting* , 16 (3), 165-176.

Palda, K. (1965). The measurement of cumulative advertising effects. *The Journal of Business* , 38 (2), 162-179.

Papalexopoulos, A., & Hesterberg, T. (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems* , 5 (4), 1535-1547.

Sen, A. (2009). *Identifying system-wide contact center cost reduction opportunities through lean, customer-focused IT metrics (Master Thesis)*. Massachusetts Institute of Technology.

Shen, H., & Huang, J. (2008). Interday forecasting and intraday updating of call center arrivals. *Manufacturing and Service Operations Management* , 10, 391–410.

Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E., & Shankar, R. (2003). *Designing and managing the supply chain*. McGraw-Hill.

Taylor, J. (2008). A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science* , 54, 253-265.

Veney, J., & Luckey, J. (1983). A comparison of regression and ARIMA models for assessing program effects: An application to the mandated highway speed limit reduction of 1974. *Social Indicators Research* , 12 (1), 83-105.

Vining, G., & Kowalski, S. (2005). *Statistical Methods For Engineers*. Duxbury Press.

Watson, N., & Zheng, Y. (2008). Over-reaction to Demand Changes due to Subjective and Quantitative Forecasting. *Working Paper - Harvard University* .

Zhou, M., Yan, Z., Ni, Y., & Li, G. (2004). An ARIMA approach to forecasting electricity price with accuracy improvement by predicted errors. *IEEE Power Engineering Society General Meeting, 2004* , 233-238.