10/6/2013

**12-Week Regional Sales Forecast**

# Model Selection and 12-week Forecasts for Asia, Europe and North America

Orange Team 6:

*Team Lead:* Phillip **Domschke**

*Other Contributors:* Marc Zimmerman, Steve Neola, Wes Ledebuhr, Jacob Frost

# TABLE OF CONTENTS

TABLE OF CONTENTS 1

EXECUTIVE SUMMARY 2

ANALYSIS 3

Exploratory Analysis 3

Model Selection 5

12-Week Forecast 7

CONCLUSION 8

APPENDIX 8

# 

# EXECUTIVE SUMMARY

In this report, we will show how we developed time series models to forecast weekly sales in our company’s three main regions of operation. Asia, Europe, and North America display distinctively different patterns in their historic sales data. In our analysis we found that for North America a two lag autocorrelation model is most appropriate, while for Europe and Asia a one-lag moving average model and a mixed model, respectively, explain the patterns the best.

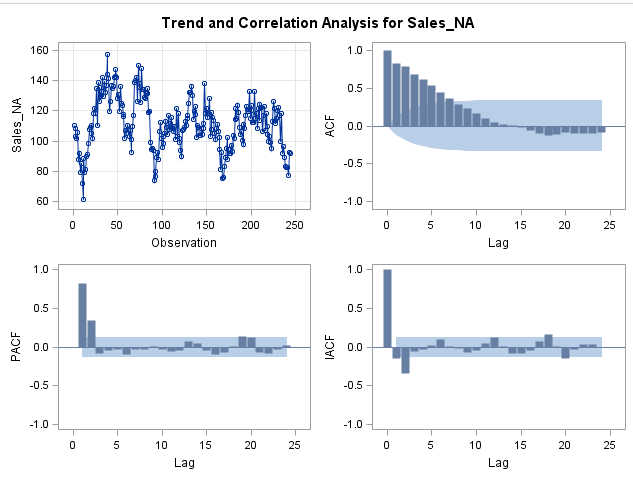
All illustrations in this report can also be found in the Appendix with much greater detail.

# ANALYSIS

## Exploratory Analysis

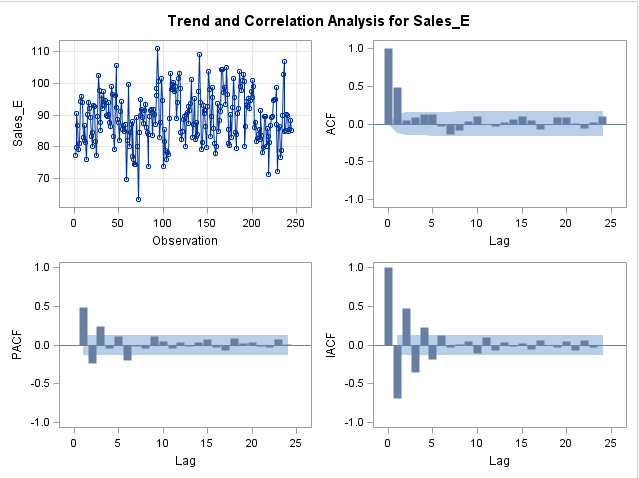


Graph 1: Historical Sales Information for Products A, E and NA. Colors represent different years.

Graph 1 illustrates the unique behaviors of all three geographic regions. In the rest of this section, we are going to analyze each of the regions individually and we will make first conclusions about, which model type explains the pattern underlying the sales data.

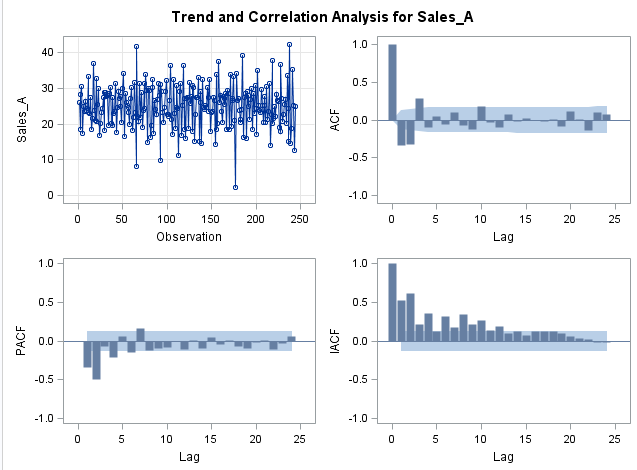
Graph 2: Trend and Correlation Analysis of Historic Sales in North America

As we can see in the autocorrelation function (ACF) in the top right of graph 2, there is a continuously decreasing effect over time. The partial autocorrelation function (PACF, bottom left) and the inverse autocorrelation function (IACF, bottom right) display a sudden drop into insignificance after lag two and three. Therefore, it is a fair evaluation to expect only three time periods to be important to explain what is happening in the current time period. The behavior illustrated in these graphs is associated with autoregressive model types. We will keep these findings in mind, when we select our candidate models to explain the pattern.



Graph 3: Trend and Correlation Analysis for Historic Sales in Europe

Graph 3 illustrates the same four charts for historic sales in Europe. However, three autocorrelation charts show a different behavior for the sales in Europe, when compared to sales in North America and Asia. The sudden drop after lag two in the ACF and the exponentially tailing off of the PACF and IACF are telling signs that a moving average model would be the best fit for the European sales data. The ACF also tells us that most likely only one lag is necessary to model the effect in the current time period. Time periods that are more than one time period in the past seem to have no effect on the current period’s behavior. The alternating bars between negative and positive effects indicate the underlying moving average function contains a negative term at the second position.



Graph 4: Trend and Correlation Analysis for Historic Sales in Asia

The historic sales pattern of sales in the Asian market are analyzed in graph four. The three autocorrelation charts make it difficult to identify a specific pattern that can be explained solely by a moving average model or an autoregressive model. All plots are somewhat tailing off exponentially, but there is a lot more going on in the data that needs to be explained. It is also difficult to identify, which lags or past time periods have an effect on the current time period. It looks like lags 2, 3, 4, 7 and 9 could play a role. Therefore, we decided to use mixed models, which are a combination of autoregressive and moving average models, also known as ARMA models.

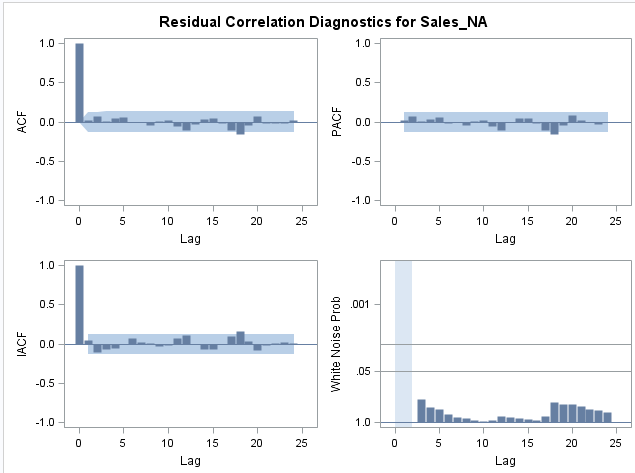
## Model Selection

As discussed in the exploratory analysis of this report, the sales pattern in all three regions mandate different types of time series models in order to explain the underlying trends and patterns. In order to find the most appropriate model for each of these patterns we used a variety of selection techniques. First of all, we identified probable models based on the autocorrelation function charts and made first conclusions about how many time periods are significant to explain the behavior of the current time period. To add to our model selection process, we deployed three different automatic selection techniques. The Minimum Information Criterion (MINIC) determines what it thinks is the best model by minimizing the Bayesian information criterion (BIC) in order to determine the best combination of moving average lags and autoregressive lags to estimate the error series. The smallest canonical correlation (SCAN) and the extended autocorrelation function (ESACF) are both algorithms that rank different MA and AT model combinations by analyzing the eigenvalues of the ARMA correlation matrix or making use of the estimated least squares estimates of the autoregressive parameters, respectively.

After using the before mentioned techniques, we selected a few candidate models that had only significant lags left for each of three regions and evaluated them by using the last 16 time periods as a benchmark for the models. The mean absolute percentage error (MAPE) as well as Schwarz Criterion and the Aikake information criterion can help to compare different models. The MAPE identifies how much error lies within our predictions when compared to the actual recorded values of the 16 time periods in the holdout sample.

Table 1: Final models. P=number of autoregressive lags, Q=number of moving average lags

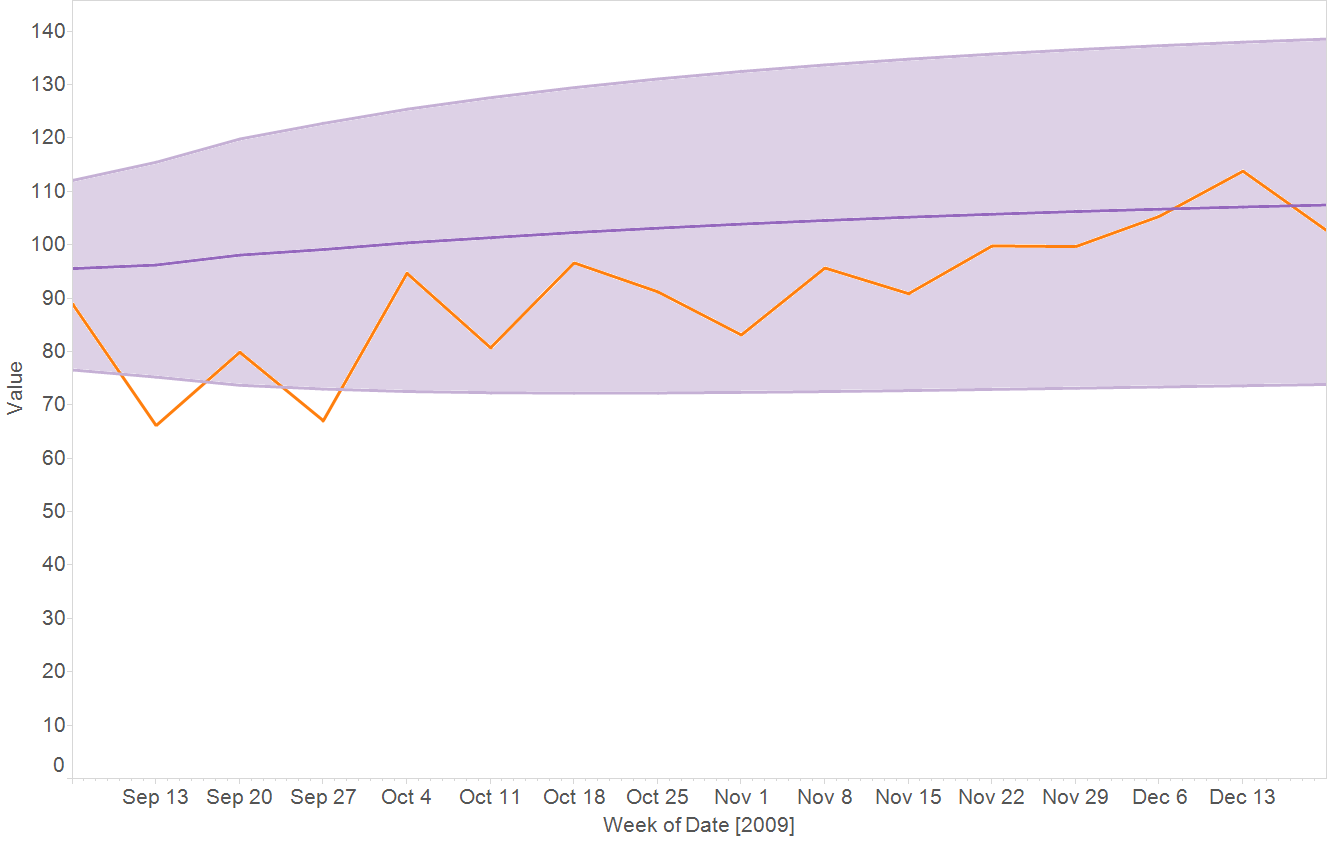
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Region | P | Q | MAPE | Type of Model |
| Asia | 1 | 2, 3 | 21.88% | Mixed Model (ARMA) |
| North America | 2 | 0 | 15.53% | Autoregressive Model |
| Europe | 0 | 1 | 10.07% | Moving Average Model |

Furthermore, we took into consideration the White Noise probability plots shown in the appendix for each sales region. The objective of the white noise probability plot is to determine whether or not there is any pattern left unexplained in the data after using the model. We are looking for not to reject the null hypothesis, which is there is only white noise left and no recurring pattern. Also, at this point we reevaluate the autocorrelation functions in order to detect any lag periods that could add to the model’s predictive capabilities. An example is given in graph 5, which displays the final results for our autoregressive model evaluation that we used for North America. Lastly, we tested the residuals for their normal distribution. If residuals are not normally distributed, there is still a pattern left in the data that the model has not captured in its current form. As with the other diagnostic graphs, these can be found in the appendix.

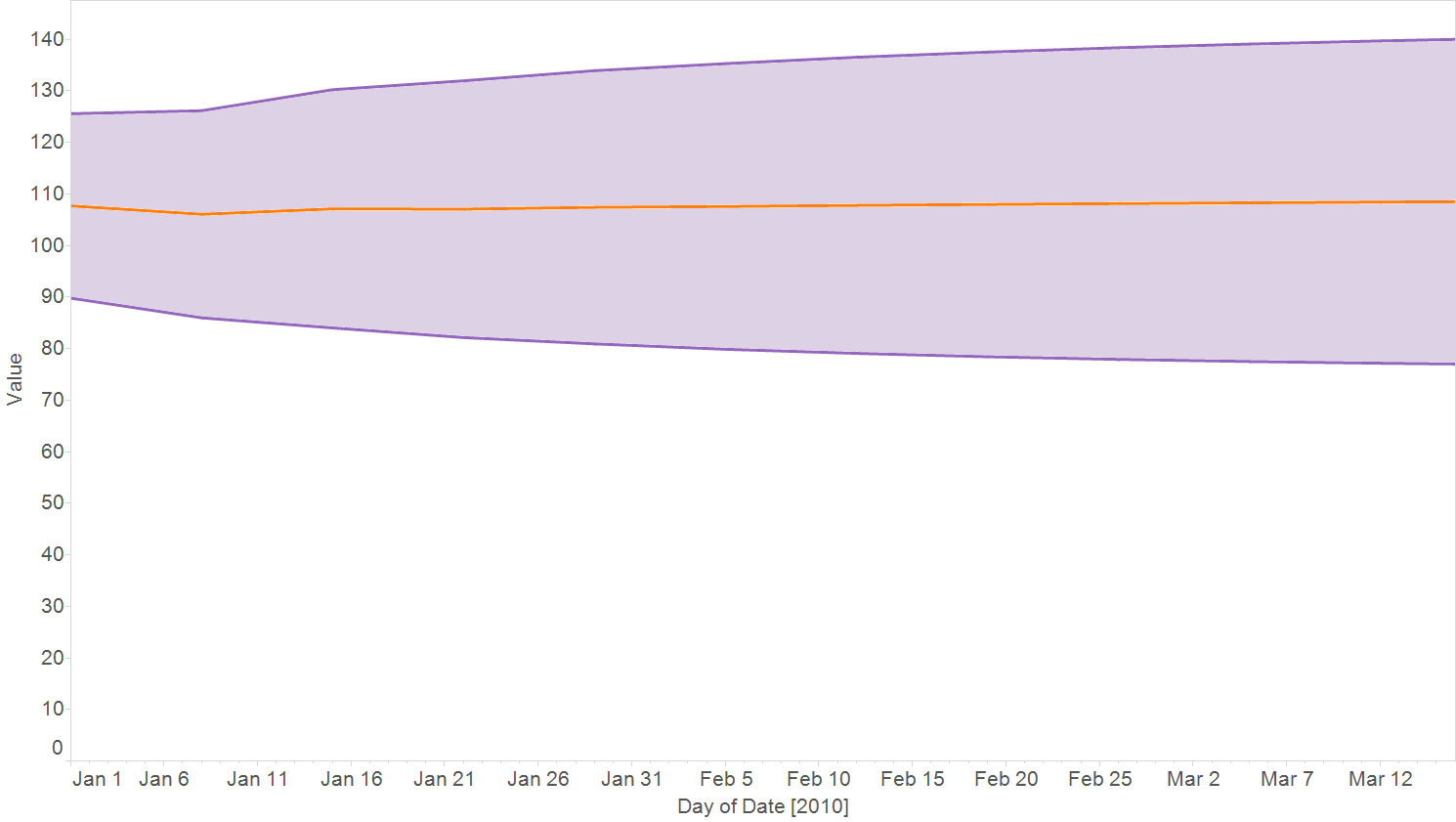
Graph 5: Residual Correlation Diagnostics for Historic Sales in North America

## 12-Week Forecast

After selecting the best models to explain the patterns in the historic work data, we utilized these models to predict the next 12 weeks of sales. As a reminder, we show our forecast and its 95% confidence interval (purple) compared to our holdout sample (orange). This shows that we can be approximately 95% certain that the real value falls into our forecast confidence interval.



Graph 6: Forecast and 95% confidence interval vs. Actual Recorded Sales for the last 16 weeks (orange):



Graph 7: Forecast and 95% confidence interval vs. Actual Recorded Sales for the next 12 weeks (orange)

In the above displayed forecast for North America, it is important to keep in mind that based on the nature of the model with only two significant lags that predictions for more than two periods into the future will become more and more imprecise. However, the model will update with new predictions as soon as new information becomes available. The displayed forecast is the best possible based on the nature of the data. Below you will find a table with the exact numbers for the 12-week forecast for all three regions. Furthermore, the graphs are all accessible in the appendix for your convenience.

Table 2: 12-week Forecast Table for North America, Europe, and Asia

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | North America | | | Europe | | | Asia | | |
| Date | **Forecast** | **95 % Confidence Interval** | | **Forecast** | **95 % Confidence Interval** | | **Forecast** | **95 % Confidence Interval** | |
| 1/1/2010 | 107.698 | 89.8087 | 125.587 | 91.0669 | 78.0625 | 104.071 | 20.4666 | 11.4364 | 29.4969 |
| 1/8/2010 | 106.079 | 85.9863 | 126.172 | 88.8752 | 72.906 | 104.845 | 26.9497 | 16.9626 | 36.9367 |
| 1/15/2010 | 107.137 | 84.0288 | 130.246 | 88.8752 | 72.906 | 104.845 | 24.5654 | 13.7771 | 35.3537 |
| 1/22/2010 | 107.069 | 82.166 | 131.971 | 88.8752 | 72.906 | 104.845 | 25.2945 | 13.9569 | 36.6321 |
| 1/29/2010 | 107.432 | 80.938 | 133.926 | 88.8752 | 72.906 | 104.845 | 24.9501 | 13.4936 | 36.4066 |
| 2/5/2010 | 107.592 | 79.8817 | 135.302 | 88.8752 | 72.906 | 104.845 | 25.1128 | 13.6299 | 36.5957 |
| 2/12/2010 | 107.811 | 79.0883 | 136.533 | 88.8752 | 72.906 | 104.845 | 25.0359 | 13.5471 | 36.5247 |
| 2/19/2010 | 107.983 | 78.4381 | 137.527 | 88.8752 | 72.906 | 104.845 | 25.0722 | 13.5821 | 36.5623 |
| 2/26/2010 | 108.153 | 77.9262 | 138.38 | 88.8752 | 72.906 | 104.845 | 25.0551 | 13.5647 | 36.5455 |
| 3/5/2010 | 108.305 | 77.5131 | 139.097 | 88.8752 | 72.906 | 104.845 | 25.0632 | 13.5727 | 36.5536 |
| 3/12/2010 | 108.447 | 77.1838 | 139.71 | 88.8752 | 72.906 | 104.845 | 25.0594 | 13.5689 | 36.5498 |
| 3/19/2010 | 108.577 | 76.9201 | 140.233 | 88.8752 | 72.906 | 104.845 | 25.0612 | 13.5707 | 36.5516 |

Graph 8: 12-week Forecast Table for North America, Europe, and Asia

# CONCLUSION

This report laid out the methodology of the time series analysis of three different sales regions. It showed that each region displayed a distinct pattern that had to be handled by a different type of model. Furthermore, we predicted sales for each region for the next 12-weeks to support the management’s decision making and planning processes. Our forecasts show that based on the provided historic data all three markets have reached saturation and growth will remain stagnant, if nothing changes.