Holt-Winters Forecasting Final Report

Baker, Mica; Bogdewic, Eva; Mikhail, Lara; Wang, Mengchen

Introduction

The goal of this project is to determine whether the Holt-Winters method can produce a more reliable model for forecasting U.S. retail auto sales than a generic time series model. We hypothesized two models using the simple Time Series and Holt-Winters methods using the data of automobile sales provided by the U.S. Bureau of Economic Analysis from January of 1970 to May of 1998. The models were used to predict the future auto sales and to determine which model is more accurate by comparing the 1-year forecasts generated by both models with the actual sales data from that year.

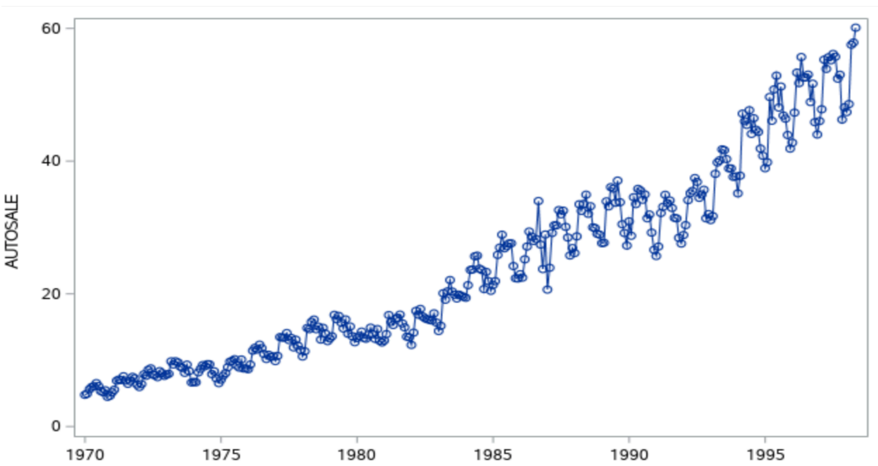
Data Summary

Data Review

The data used was published by a Duke University website titled, "Statistical forecasting: notes on regression and time series analysis." The 341 observations data were originally collected by the United States Bureau of Economic Analysis. The 341 observations were collected monthly from January of 1970 to May of 1998. Each observation contains the independent variables year and month and the dependent variable U.S. auto sales. To account for both of the independent variables, we created the variable "TIME" in SAS. "TIME" gives each observation a numerical value in the range 1 to 341, increasing by one unit for each advance of month. For example, for January 1970, TIME = 1, for February 1970 TIME = 2, through May 1998 where TIME = 341. The dependent variable is the total revenue of auto sales (in billions of dollars) made in the U.S. during that month. This variable is referred to as "U.S. auto sales" in the report and "AUTOSALE" in SAS.

Exploratory Data Analysis

A visual observation of the dot plot shows an upward trend throughout time with seemingly steady cycle of peaks and troughs throughout its course. This variance can be accredited to the seasonal variance in sales throughout the year. For this reason, an analysis to account for seasonal variation is used to predict automobile sales in future months. The use of such an analysis will yield more accurate predictions.



Theoretical model:

The theoretical relationship between TIME and the automobile sales would be an overall positive trend with variations (fluctuation) of month. The positive trend might be due to sales prices increasing over time due to inflation and a growing US economy and population, and the fluctuations might be due to the holiday seasons and nature of auto sales market, etc.

Hypothesized Regression Models:

Model 1: A simple time series model with variables time and months from January to December

Model 2: A Holt-Winters model with exponentially smoothed component, trend component, seasonal component

The hypothesized models will forecast future sales. Model 1 features variable time as a linear trend and dummy variables as difference in months, while model 2 accounts for exponential smoothing component with seasonal variance forecast values that will reflect the noted fluctuations.

Model 1: Time Series

$y_t = E(y_t) + R_t$

$E(y_t) = \beta_0 + \beta_1 {*}t + \beta_2 {*}x_1 + \beta_3 {*}x_2 + \beta_4 {*}x_3 + \beta_5 {*}x_4 + \beta_6 {*}x_5 + \beta_7 {*}x_6 + \beta_8 {*}x_7 + \beta_9 {*}x_8 + \beta_{10} {*}x_9 + \beta_{11} {*}x_{10} + \beta_{12} {*}x_{11}$

Where t = time; $x_1 \ldots x_{11} = 0$, $x_1 = 1$ if January, $x_2 = 1$ if February, $x_3 = 1$ if March, $x_4 = 1$ if April, $x_5 = 1$ if May, $x_6 = 1$ if June, $x_7 = 1$ if July, $x_8 = 1$ if August, $x_9 = 1$ if September, $x_{10} = 1$ if October, and $x_{11} = 1$ if November.

Model 2: Holt-Winters Method

$F_t = (E_n + T_n)S_{n+1-P}$, where $t = n + 1$

$= (E_n + 2T_n)S_{n+2-P}$, where $t = n + 2$

…

$= (E_n + kT_n)S_{n+k-P}$, where $t = n + k$

where

$E_t = y_2$, where $t = 2$

$= wyt + (1 - w)(Et\text{-}1 + Tt\text{-}1)$, where $t = 2, 3, \ldots P + 2$

$= w(y_t / S_{t-p}) + (1 - w)(E_{t-1} + T_{t-1})$, where $t > P + 2$

$T_t = y_2 - y_1$, where $t = 2$

$= v(E_t - E_{t-1}) + (1 - v)T_{t-1}$, where $t > 2$

$S_t = y_t/E_t$, where $t = 2, 3, \ldots P + 2$

$= u(y_t/E_t) + (1 - u)S_{t-p}$, where $t > P + 2q$

where w, v, u are smoothing constants, P is the number of periods in a cycle, F_t is the forecasting component, E_t is the exponentially smoothed component, T_t is the trend component, and S_t is the seasonal component.

Methodology

Holt-Winters is one of several smoothing methods used in statistics. There are two forms: trend component only and trend and seasonal component, and in this report, with the U.S. auto sales data, we use Holt-Winters with both trend and seasonal component. More comprehensive than the exponential smoothing forecasting method, Holt-Winters explicitly recognizes the trend and seasonal variation in a time series. As a result, it is the most useful when the observer detects a trend and seasonal variation in the time series, in the case that rapid fluctuations are observed time series data and that these fluctuations fall into a pattern based on seasonality within a period of time. When the best model is determined, it can be used to forecast future values. There are no restrictive assumptions concerning the use of this method (Gelper, S., Fried, R., & Croux, C. (2009)). However, it must be used on time series data and is logically employed when the data deems it potentially useful as previously stated.

After deciding to employ this method, one must calculate the values corresponding to its components. This model contains three components: the exponentially smoothed data, the trend component, and the seasonality component. These components are expressed in Model 2 as E_t, T_t, and S_t, respectively. While some programs, such as SAS, have the capacity to quickly perform these calculations, they can be calculated by "hand" using excel. As one can see, each smoothed data point is dependent upon the previous data points. Using excel allows one to easily reference the previous value and to continue this pattern for a column of values. First, after determining that Holt-Winters is capable for the data through the plot, the original data is entered into a column. The first data point is at t=1. As we see in Model 2, all of our calculations begin at t=2 and the formulas used vary based on t, so this must be accounted for when calculating the new values. The forecasting formula itself changes depending on how far in the future the statistician wishes to forecast (Mendenhall 503).

The constants w, v, and u pertaining to exponential smoothing, trend, and seasonality, are known as smoothing parameters. The smoothing parameters can be chosen in a subjective manner. For example, previous experience may allow the forecaster to specify similar values. However, a more robust method of choosing these values is performing within-sample forecasting. This means forecasting later known values based on the projections obtained from earlier known values and then comparing the predicted values with the actual values. This process is repeated until the sum of squared errors (SSE) is minimized ((Gottschling, I. R. (2013)). The smoothing constants used in this trial are the constants used in the model for forecasting unknown future data.

There are several major drawbacks to this procedure. First, as time goes on, variability increases. Much of this variability comes from the parameter estimates and the chosen constants remaining the same despite the form of the model requiring change. Forecast errors can only be computed when future values in the time series have been observed; of course, this is a drawback found with all time series smoothing methods. A second prominent drawback to this method is its inability to begin forecasting before t = P. Without a period of information, the formula cannot account for seasonality within the data. Additionally, due to the forecast formula, Holt-Winters smoothing method can only forecast the next period of time, or the seasonal component S_{n+k-P} will be out of range. Furthermore, Holt-Winters only includes the exponentially smoothed component, trend component, seasonal component, so another potential drawback is the lack of a business cycle component or other components which could further improve a model.

Analysis

Model 1: Time Series

We first generate the best model using simple Time Series, and compare it with Holt-Winters method. After coding the time variable and generate the prediction equation containing independent variable time and dummy variables representing months using SAS, we check the assumptions of time series.

Assumption Checking:

We first check the "non-correlated" assumption first since it is usually violated in the time series data. To check the "non-correlated" assumption, we check the Residual vs. x plot graphs and Durbin-Watson Test. The Durbin-Watson value for Model 1 is 0.204, which is very close to 0, which indicates a positive autocorrelation. Besides, Pr<DW value is <.001 value, which is less than 0.05, so there should be a positive autocorrelation. As a result, we add a first autoregressive term to the time series model. The model with autoregressive term is :

$$y = \beta_0 + \beta_1 * t + \beta_2 * x_1 + \beta_3 * x_2 + \beta_4 * x_3 + \beta_5 * x_4 + \beta_6 * x_5 + \beta_7 * x_6 + \beta_8 * x_7 + \beta_9 * x_8 + \beta_{10} * x_9 + \beta_{11} * x_{10} + \beta_{12} * x_{11} + \Phi * R_{t-1} + \varepsilon_t$$

Where y is the autosale (in billions of dollars), t is the time, x's are the dummy variables for months, and R_{t-1} is the first-order autoregressive term, ε_t is the white noise.

The p-value of the autoregressive term is <.0001, which is less than 0.05 (alpha), which also indicates that the autoregressive term is statistically significant.

Assumptions Checking for Time Series Model with First-Order Autoregressive Term:

To check the assumption of the residuals, we used SAS arima to generate the Residual vs. x plot, QQ plot, and histogram of residual.

To check the "equal variance" assumption, we can check the residual vs. observation graphs. It shows a mild fanning out pattern on the graph due to the increasing seasonal variation, so the residuals violate this assumption. As a result, we perform a y-transformation (on autosale).

Transformation

We try two different y-transformation: square root of y (autosale) and natural log of y, and check the assumptions again. Square root transformation seems to a better correction of the violation.

As a result, we build a new model with this transformation.

$\mathrm{sqrt}(y) = \beta_0 + \beta_1 {*}t + \beta_2 {*}x_1 + \beta_3 {*}x_2 + \beta_4 {*}x_3 + \beta_5 {*}x_4 + \beta_6 {*}x_5 + \beta_7 {*}x_6 + \beta_8 {*}x_7 + \beta_9 {*}x_8 + \beta_{10} {*}x_9 + \beta_{11} {*}x_{10} + \beta_{12} {*}x_{11} + \Phi{*}R_{t-1} + \varepsilon_t$

Since the p-value of the autoregressive term is <0.0001, which is smaller than 0.05, the first-order autoregressive term is statistically significant, so we should keep it. The p-value of independent variable t (time) is <0.0001, which is smaller than alpha, so it is statistically significant and should be included in the final prediction equation. Besides, for the dummy variables of months, 9 out of 11 have a p-value less than 0.05, and since all these dummy variables describe the same quality variable: month, we should keep all of them as long as one of them is statistically significant.

Assumptions Checking for Transformed Time Series Model with First-Order Autoregressive Term:

The "mean zero" assumption is not violated since there is no clear trend shown on the Residual vs. x plot (residual vs. observation) graph.The residuals seem to be even spread across zero.

The "equal variance" assumption is not violated since after the transformation, there is no clear pattern on the Residual vs. x plot (residual vs. observation) graph.

To check the "normality" assumption, we can check the QQ plot and the histogram of residual. The QQ plot is fairly linear with no point that deviates excessively from the general trend except the points very beginning and very end. However, it is reasonable since the the variance is likely to be larger at the each end of the time series data. The histogram does not show severe skewness either, so the residuals do not violate this assumption.

To check the "non-correlated" assumption, we check the Residual vs. x plot (residual vs. observation) graph, and since there is no evidence shown autocorrelation, the assumption is met.

Overall, there are no severe violations after adding the autoregressive term and transformation. As a result, the prediction equation with first-order autoregressive term and y-transformation of square root is the best one among the time series model. We forecast the next 12 months auto sales with the prediction equation: $\mathrm{sqrt}(\boldsymbol{\hat{y}}) = 1.932 + 0.015{*}t - 0.046{*}x_1 + 0.041{*}x_2 + 0.429{*}x_3 + 0.402{*}x_4 + 0.489{*}x_5 + 0.501{*}x_6 + 0.401{*}x_7 + 0.388{*}x_8 + 0.243{*}x_9 + 0.259{*}x_{10} + 0.065{*}x_{11} + 0.804{*}R_{t-1}$.

Model 2: Holt-Winters method

Choosing Constants

Since using Holt-Winters method to forecast needs the smoothing constants, which are between 0 and 1, with our data, which has seasonal component, we need w, v and u to calculate Et (the exponentially smoothed series), Tt (the trend component) and St (seasonal component) in order to forecast the future auto sales. To figure out the better constants, we use the Holt-Winters method to predict a small part of our data with the remaining data (Gottschling, I. R. (2013)). We use data from time = 1 to time = 324 (January 1970 to December 1996) to predict the auto sales between time = 325 and time = 336 (January 1997 to December 1997). We try different constants from 0 and 1, changing one at a time, and check if the predictions are close to the actual data we have. We calculate the sum of square error of each prediction with different constants and try to find out the smallest one. Even though we cannot try all the possibilities of constants from 0 to 1, we do try more than 30 different combinations of constants. We start from u = 0.3, v = 0.8, and u = 0.5 from the textbook question, and among all the trials, w = 0.18, v = 0.7, and u = 0.42 seems to be the best set of constants since its SSE is the smallest among all the more than 30 trials.

Then, we use Excel to calculate the Et, Tt, and St with these constants. The period P is 12 since we have a monthly data.

<u>Assumption Checking</u>

In our research, we do not find any rigorous assumption for the Holt-Winters method. This method is robust against almost everything, so there is no assumption checking for the Holt-Winters method (Gelper, S., Fried, R., & Croux, C., 2009).

<u>Calculation</u>

$E(2) = y(2) = 4.955$

$T(2) = y(2) - y(1) = 4.955 - 4.792 = 0.163$

$S(2) = y(2) / E(2) = 4.955/4.955 = 1$

$E(3) = wy_3 + (1 - w)(E_{3-1} + T_{3-1}) = 0.18*5.639 + (1-0.18)*(E(2) + T(2)) = 0.18*5.639 + 0.82(4.955 + 0.163) = 5.212$

$T(3) = v(E_3 - E_{3-1}) + (1- v)T_{3-1} = 0.7*(E(3) - E(2)) + (1-0.7)*T(2) = 0.7*(5.212 - 4.955) + 0.3*0.163 = 0.229$

$S(3) = y(3) / E(3) = 1.082$

<u>*... (Same calculation from t = 4 to t = P+2 = 12+2 = 14)*</u>

$E(15) = w(y_{15} /S_{15-12}) + (1 - w)(E_{15-1} + T_{15-1}) = 0.18*(6.917/1.082) + (1-0.18) *(4.628 - 0.270) = 4.725$

$T(15) = v(E_{15} - E_{15-1}) + (1- v)T_{15-1} = 0.7*(4.725 - 4.628) + 0.3 *(-0.270) = -0.014$

$S(15) = u(y_{15}/E_{15}) + (1 - u)S_{15-12} = 0.42*(6.917/4.725) + 0.58*1.082 = 1.242$

<u>*... (Same Calculation from t = 16 to t = 341)*</u>

$E(341) = 0.18(y_{341} /S_{341-12}) + (1 - 0.18)(E_{341-1} + T_{341-1}) = 51.756$

$T(341) = 0.7(E_{341} - E_{341-1}) + (1- 0.7)T_{341-1} = 0.539$

$S(341) = 0.42(y_{341}/E_{341}) + (1 - 0.42)S_{341-12} = 1.131$

<u>Forecasting</u>

$F(t) = (E_n + kT_n) S_{n+k-P}$, $t = n+k$ $(n = 341)$

$F(342) = (E_{341} + T_{341}) S_{341+1-12} = (51.756 + 0.539)*1.096 = 57.329$

$F(343) = (E_{341} + 2T_{341}) S_{341+2-12} = (51.756 + 2*0.539)*1.073 = 56.707$

…

$F(353) = (E_{341} + 12T_{341}) S_{341+12-12} = (51.756 + 12*0.539)*1.131 = 65.837$

<u>***Model Comparison***</u>

	Model 1 (Time Series)	Model 2 (Holt-Winters)
SSE	311.36	56.18

Prediction Power (Actual auto sales data are from a different source (U.S. census), which might cause slightly error.)

	Time	*Actual Auto Sales*	*Forecasting with Time Series*	*Residuals for Time Series*	*Forecasting with Holt-Winters method*	*Forecast error for Holt-Winters*
Jun. 1998	*342*	*63.375*	59.382	3.993	57.329	6.046
Jul. 1998	*343*	*58.258*	57.16	1.098	56.707	1.551
Aug. 1998	*344*	*56.24*	56.466	-0.226	57.536	-1.296
Sep. 1998	*345*	*55.143*	53.949	1.194	54.46	0.683
Oct. 1998	*346*	*57.971*	53.942	4.029	55.86	2.111
Nov. 1998	*347*	*51.195*	50.98	0.215	50.985	0.21
Dec. 1998	*348*	*53.385*	49.978	3.407	51.41	1.975
Jan. 1999	*349*	*51.229*	49.314	1.915	51.9	-0.671
Feb. 1999	*350*	*55.675*	50.565	5.11	54.4	1.275
Mar. 1999	*351*	*66.245*	56.291	9.954	64.184	2.061
Apr. 1999	*352*	*63.376*	55.98	7.396	63.288	0.088
May. 1999	*353*	*66.422*	57.406	9.016	65.837	0.585

Since the residuals of Holt-Winters forecasts are generally smaller than that of the Time Series model we generate, and also the sum of square error of Holt-Winters is much smaller than that of the Time Series as well, we conclude that Holt-Winters model has a stronger prediction power, and hence, we conclude that Holt-Winters generate a better prediction model.

Overall Evaluation

Through the comparison of prediction power, we conclude that the model using Holt-Winters method is better to forecast the U.S. auto sales, which includes exponentially smoothed component, trend component, and seasonal component.

Conclusion

Our final equation ends up being:

$$F(t) = (E_n + kT_n)\, S_{n+k-P}\,, \quad t = n+k \ (n = 341)$$

Where

$$E(341) = 0.18(y_{341} / S_{341-12}) + (1 - 0.18)(E_{341-1} + T_{341-1}) = 51.756$$

$$T(341) = 0.7(E_{341} - E_{341-1}) + (1 - 0.7)T_{341-1} = 0.539$$

$$S(341+k-12) = 0.42(y_{329+k}/E_{329+k}) + (1 - 0.42)S_{317+k}$$

1. *Interpretation of the Prediction Equation*

To interpret the prediction equation, we can see that the forecast is depended on one constant exponentially smoothed component (E(341)), constant trend component (T(341)), and seasonal component which changes as the forecast distance increases. Besides, as the forecast distance increases, the trend component becomes more influential as its coefficient k increases.

2. *Conclusion and areas for future research and improvements*

In sum, according to the comparison of the prediction power, our forecast equation developed using the Holt-Winters method yields more accurate predictions. This is an unsurprising result because, as we mentioned before, the Holt-Winters method accounts for seasonal variation very well. Looking at the scatterplot of the original data, there is very obvious seasonal variation. This is because car sales generally peak in the certain months and fall and dip during the winter and summer. This is because consumers are spending a greater percentage of their disposable income on other consumer gifts during the summer and winter. Additionally, with regards to the random error assumptions, other time series models are greatly affected by violations to these assumptions. However, the Holt-Winters method is robust to most of these same assumptions, meaning a violation of an assumption will not have significant, detrimental effects on our predictions that would make them illegitimate.

That being said, even though our model improved the predictions auto sales for each month, they still differed from the actual car sales. This is a consequence of business cycle factors that were not taken into consideration. Holt-Winters does not account for the cyclical trend. We know this because there are other forces at play, other than general consumer, seasonal trends, in the auto sale industry. We must consider supply and demand; for example, Japanese cars entered the US Market in 1958, which increased the overall supply of cars in the US, but these cars really only gained traction in the 1960's and 1970's (increase in demand). Another huge force is just how the economy overall is doing. If the economy has declined, people will be more stingy with their money--consequently buying less cars. In the time period accounted for in our model, the 1973 Oil Crisis definitely decreased car sales because of how hard it was to get gas. Even though this is a special, extreme case, it is still shaping our predictions. In the 1990's the US had a major economic boom; the entire decade, GDP was continuously increasing. From 1995-2000 specifically, there was a series of economic downturns in other parts of the globe which allowed the US economy to rise up. This is why for our prediction of 1998-1999, the actual car sales value was generally higher than we could've predicted. There are constant forces at play, both political and economic. If we tried to predict car sales in 2019 using this model, we would be unable to.

That brings us to the next drawback, which is how far into the future we are able to predict. Our Holt-Winters model can only predict for one period after the range, which is why our last prediction is for May 1999. Because of the external factors mentioned in the previous paragraph and the fact that each period is more or less related to the performance in the previous period, it is risky to predict further into the future. In general, the equation for this is S_{n+k-P}, Sn being the number of seasons, P being the length of the period, and k is the number of months we want to predict for after Sn. In the context of our data, Sn = 341 months and P = 12 months. If we wanted to predict more than 13 months (k = 13), we would be out of range. Besides, the smoothing constants we find out might not be the best to conduct the calculation, and an improvement might be conducted using other software that calculates the optimal smoothing constant automatically.

The main improvement for our method would be to somehow include cyclical trends, because while Holt-Winters is generally better than other methods when dealing with seasonally-varying data, the lack of business-cycle accountability is a major drawback. There is research being done on how to possibly combine different methods. We could have also attempted to use the multiplicative Holt-Winters method rather than simply the additive one. The multiplicative Holt-Winters model is preferred when the seasonal variations will change based on the progression of the series. In our case, it appears that there is increasing seasonal variation, meaning that seasonal variation increases as time progresses, which would make using the multiplicative model appropriate.

Works Cited:

The Arima Procedure (2017). *SAS/ETS® 14.3 User's Guide*. Cary, NC: SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/ets/143/arima.pdf

Annual Benchmark Report for Retail Trade (2000). *U.S. Census*, U.S. CENSUS BUREAU, www2.census.gov/retail/releases/benchmark/annpub99.pdf

Gelper, S., Fried, R., & Croux, C. (2009). Robust forecasting with exponential and Holt-Winters smoothing. *Journal of Forecasting*. doi:10.1002/for.1125 https://onlinelibrary.wiley.com/doi/epdf/10.1002/for.1125

Goodwin, Paul. (2018). The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong. *Foresight: The International Journal of Applied Forecasting*. 30-33. https://www.researchgate.net/publication/227439091_The_Holt-Winters_Approach_to_Exponential_Smoothing_50_Years_Old_and_Going_Strong

Gottschling, I. R. (2013). Determining The Optimal Values Of Exponential Smoothing Constants – Does Solver Really Work? *American Journal Of Business Education*. doi:10.3897/bdj.4.e7720.figure2f https://files.eric.ed.gov/fulltext/EJ1054363.pdf

Mendenhall, William, and Terry Sincich (2014). *A Second Course in Statistics Regression Analysis*. Pearson Education.

Appendix A: SAS Code

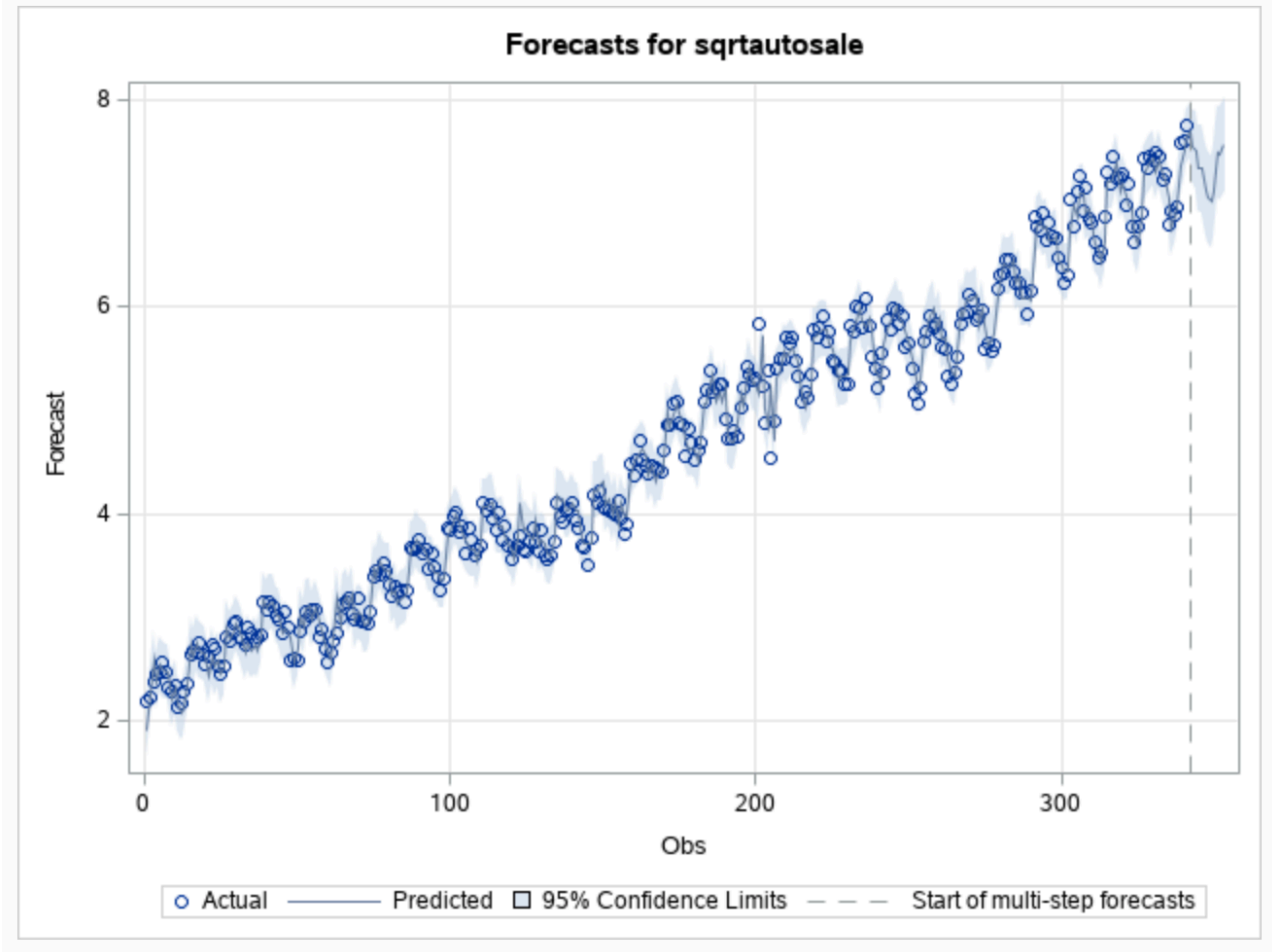
```
data sales;
infile '/folders/myfolders/Final Project.txt' dlm='09'X firstobs=2;
input INDEX YEAR MONTH DATE AUTOSALE;
run;

* plot;
proc sgplot data=sales  pad=(bottom=15%);
series x=date y=autosale / markers;
run;

* Variable time;
data sales2;
set sales;
TIME=_n_;
run;

*Dummy variables with seasonal component;
data sales3;
set sales2;
month1 = 0;
if month = '1' then month1 = 1;
```

```
month2 = 0;
if month = '2' then month2 = 1;
month3 = 0;
if month = '3' then month3 = 1;
month4 = 0;
if month = '4' then month4 = 1;
month5 = 0;
if month = '5' then month5 = 1;
month6 = 0;
if month = '6' then month6 = 1;
month7 = 0;
if month = '7' then month7 = 1;
month8 = 0;
if month = '8' then month8 = 1;
month9 = 0;
if month = '9' then month9 = 1;
month10 = 0;
if month = '10' then month10 = 1;
month11 = 0;
if month = '11' then month11 = 1;
run;
* base level: month 12 (December)

*Timer series;
proc reg data=sales3 plots=none;
model autosale = time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11;
run;

*Check Assumptions;
proc reg data=sales3 plots(only)=(residualbypredicted residualplot qqplot
 residualhistogram);
model autosale = time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11;
run;

*Check autocorrelation (Durbin-Watson);
proc reg data=sales3 plots=none;
model autosale = time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11 / dwprob;
run;

*Model with autogressive term (positive autocorrelation);
```

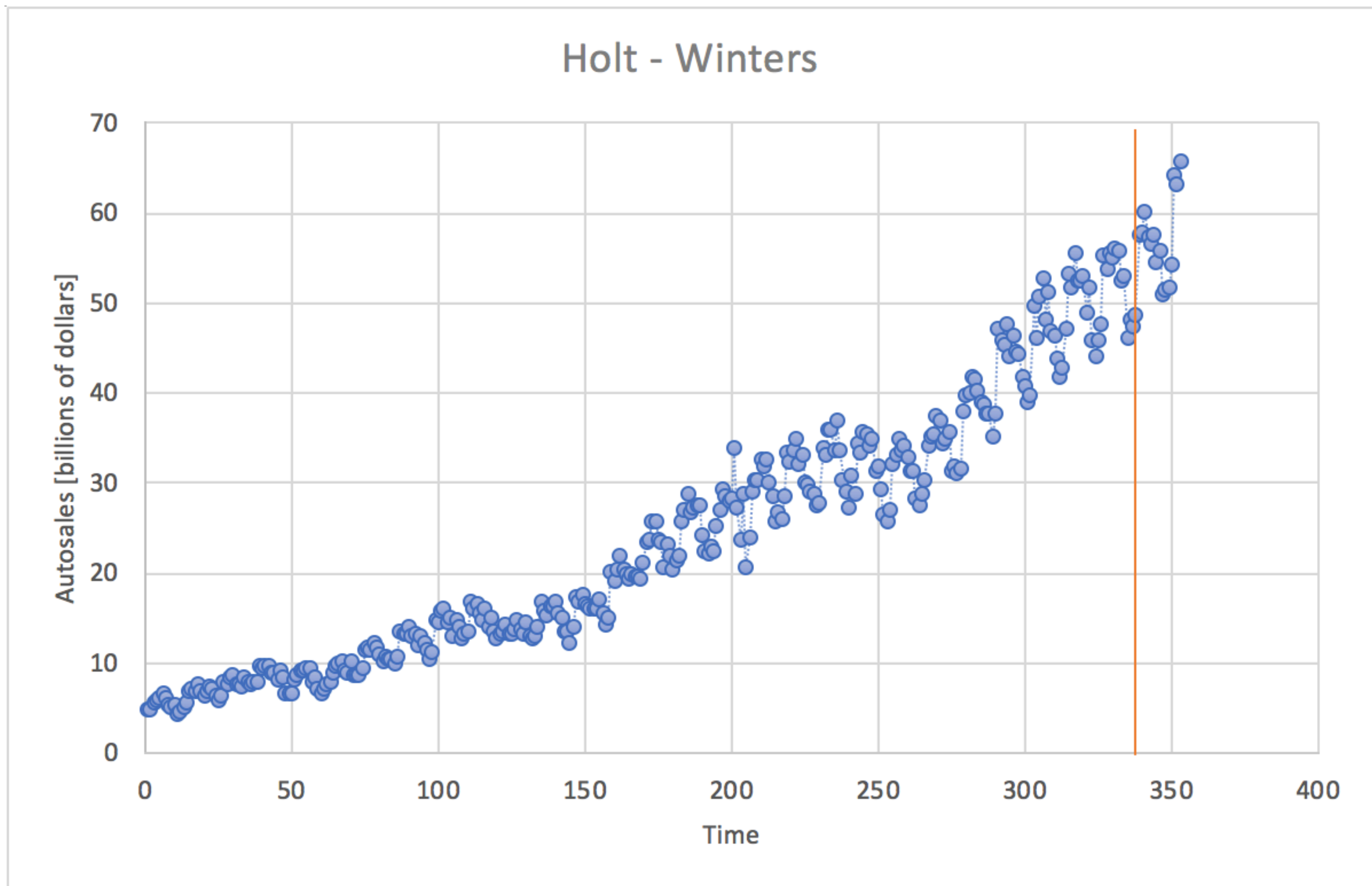
```
proc arima data=sales3 plots=none;
identify var=autosale crosscor=(time month1 month2 month3 month4 month5 month6 month7 month8
month9 month10 month11) noprint;
estimate input=(time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11) p=(1);
run;

*Check Assumption (code written with the help from SAS/ETS User's Guide ("The Arima Procedure");
proc arima data=sales3 plots(only)=(series(corr crosscorr) residual(hist normal smooth));
identify var=autosale crosscor=(time month1 month2 month3 month4 month5 month6 month7 month8
month9 month10 month11) noprint;
estimate input=(time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11) p=(1);
run;

*Transformation;
data sales4;
set sales3;
sqrtautosale=sqrt(autosale);
lnautosale = log(autosale);
run;

*Check assumptions (code written with the help from SAS/ETS User's Guide ("The Arima Procedure");
proc arima data=sales4 plots(only)=(series(corr crosscorr) residual(hist normal smooth));
identify var=sqrtautosale crosscor=(time month1 month2 month3 month4 month5 month6 month7
month8 month9 month10 month11) noprint;
estimate input=(time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11) p=(1);
run;

*forecasting;
proc arima data=sales4 plots(only)=forecast(forecast);
identify var=sqrtautosale crosscor=(time month1 month2 month3 month4 month5 month6 month7
month8 month9 month10 month11) noprint;
estimate input=(time month1 month2 month3 month4 month5 month6 month7 month8 month9
month10 month11) p=(1);
forecast lead=12;
run;
```

Appendix B: Figures

1. Time Series Plot with Predictions (With autoregressive term and transformation)



2. Holt - Winters Plot with Predictions



Appendix C: Holt-Winters Calculation

w=0.18; v=0.7; u = 0.42						
INDEX (Time)	YEAR	MONTH	AUTOSALE	E(t)	T(t)	S(t)
1	1970	1	4.792			
2	1970	2	4.955	4.955	0.163	1.000
3	1970	3	5.639	5.212	0.229	1.082
4	1970	4	5.975	5.537	0.296	1.079
5	1970	5	6.076	5.876	0.327	1.034
6	1970	6	6.548	6.265	0.370	1.045
7	1970	7	6.105	6.540	0.303	0.934
8	1970	8	5.365	6.577	0.117	0.816
9	1970	9	5.171	6.420	-0.075	0.805
10	1970	10	5.48	6.189	-0.184	0.885
11	1970	11	4.485	5.732	-0.375	0.782
12	1970	12	4.65	5.229	-0.464	0.889
13	1971	1	5.164	4.837	-0.414	1.068
14	1971	2	5.567	4.628	-0.270	1.203
15	1971	3	6.917	4.725	-0.014	1.242
16	1971	4	7.098	5.047	0.222	1.217
17	1971	5	7.015	5.541	0.413	1.131
18	1971	6	7.583	6.188	0.577	1.121
19	1971	7	6.932	6.884	0.660	0.964
20	1971	8	6.448	7.609	0.705	0.829
21	1971	9	6.87	8.353	0.732	0.813
22	1971	10	7.498	8.974	0.655	0.864
23	1971	11	7.293	9.573	0.616	0.774
24	1971	12	6.333	9.637	0.229	0.792
25	1972	1	5.968	9.097	-0.309	0.895
26	1972	2	6.399	8.163	-0.746	1.027
27	1972	3	7.882	7.224	-0.882	1.179
28	1972	4	7.656	6.333	-0.888	1.213
29	1972	5	8.568	5.829	-0.620	1.274
30	1972	6	8.778	5.681	-0.289	1.299
31	1972	7	7.819	5.881	0.053	1.118
32	1972	8	7.748	6.548	0.483	0.978
33	1972	9	7.453	7.416	0.753	0.893
34	1972	10	8.361	8.439	0.942	0.917
35	1972	11	8.06	9.568	1.072	0.803
36	1972	12	7.643	10.463	0.948	0.766
37	1973	1	7.843	10.935	0.615	0.820
38	1973	2	7.956	10.865	0.136	0.903
39	1973	3	9.832	10.522	-0.199	1.076
40	1973	4	9.368	9.854	-0.527	1.103
41	1973	5	9.845	9.040	-0.728	1.196
42	1973	6	9.668	8.155	-0.838	1.251
43	1973	7	9.064	7.459	-0.738	1.159
44	1973	8	8.859	7.142	-0.443	1.088
45	1973	9	8.104	7.126	-0.145	0.996
46	1973	10	9.337	7.556	0.258	1.051
47	1973	11	8.365	8.284	0.587	0.890
48	1973	12	6.652	8.837	0.563	0.760
49	1974	1	6.744	9.188	0.415	0.784
50	1974	2	6.684	9.206	0.137	0.829
51	1974	3	8.175	9.029	-0.083	1.004
52	1974	4	8.704	8.756	-0.216	1.057
53	1974	5	9.267	8.398	-0.316	1.157
54	1974	6	9.105	7.937	-0.417	1.208
55	1974	7	9.412	7.628	-0.341	1.190
56	1974	8	9.389	7.528	-0.172	1.155
57	1974	9	7.909	7.462	-0.098	1.023
58	1974	10	8.332	7.465	-0.027	1.078
59	1974	11	7.277	7.571	0.066	0.920
60	1974	12	6.553	7.814	0.190	0.793
61	1975	1	7.072	8.186	0.318	0.818

62	1975	2	7.654	8.636	0.410	0.853
63	1975	3	8.063	8.862	0.282	0.965
64	1975	4	8.937	9.020	0.195	1.029
65	1975	5	9.786	9.078	0.099	1.124
66	1975	6	9.912	9.003	-0.023	1.163
67	1975	7	10.157	8.899	-0.079	1.170
68	1975	8	9.225	8.670	-0.184	1.117
69	1975	9	8.896	8.524	-0.157	1.032
70	1975	10	10.117	8.550	-0.030	1.122
71	1975	11	8.758	8.700	0.097	0.956
72	1975	12	8.771	9.204	0.381	0.860
73	1976	1	8.651	9.765	0.507	0.846
74	1976	2	9.354	10.397	0.595	0.873
75	1976	3	11.446	11.149	0.705	0.991
76	1976	4	11.88	11.797	0.665	1.020
77	1976	5	11.614	12.079	0.397	1.056
78	1976	6	12.344	12.141	0.163	1.101
79	1976	7	11.824	11.909	-0.114	1.095
80	1976	8	10.944	11.436	-0.365	1.050
81	1976	9	10.195	10.857	-0.515	0.993
82	1976	10	10.83	10.217	-0.602	1.096
83	1976	11	10.501	9.861	-0.430	1.002
84	1976	12	10.586	9.948	-0.068	0.946
85	1977	1	9.875	10.202	0.157	0.897
86	1977	2	10.653	10.692	0.390	0.925
87	1977	3	13.492	11.539	0.710	1.066
88	1977	4	13.407	12.410	0.823	1.045
89	1977	5	13.417	13.139	0.757	1.041
90	1977	6	14.098	13.698	0.619	1.071
91	1977	7	13.046	13.883	0.315	1.030
92	1977	8	13.363	13.934	0.130	1.012
93	1977	9	11.938	13.698	-0.127	0.942
94	1977	10	13.104	13.280	-0.330	1.050
95	1977	11	12.193	12.809	-0.429	0.981
96	1977	12	11.543	12.349	-0.451	0.941
97	1978	1	10.567	11.876	-0.466	0.894
98	1978	2	11.352	11.566	-0.357	0.948
99	1978	3	14.857	11.701	-0.013	1.151
100	1978	4	14.686	12.113	0.285	1.115
101	1978	5	15.791	12.896	0.634	1.118
102	1978	6	16.127	13.805	0.826	1.112
103	1978	7	14.679	14.562	0.778	1.021
104	1978	8	15.064	15.260	0.722	1.001
105	1978	9	13.113	15.611	0.462	0.899
106	1978	10	14.883	15.731	0.223	1.007
107	1978	11	14.059	15.662	0.019	0.946
108	1978	12	12.887	15.322	-0.232	0.899
109	1979	1	13.273	15.046	-0.263	0.889
110	1979	2	13.617	14.706	-0.317	0.939
111	1979	3	16.839	14.432	-0.287	1.158
112	1979	4	16.189	14.211	-0.241	1.125
113	1979	5	16.675	14.140	-0.122	1.144
114	1979	6	15.671	14.031	-0.112	1.114
115	1979	7	14.796	14.023	-0.040	1.035
116	1979	8	16.151	14.369	0.231	1.053
117	1979	9	14	14.775	0.353	0.919
118	1979	10	15.079	15.102	0.335	1.003
119	1979	11	13.627	15.251	0.205	0.924
120	1979	12	12.724	15.221	0.040	0.873
121	1980	1	13.328	15.213	0.006	0.884
122	1980	2	13.668	15.099	-0.077	0.925
123	1980	3	14.311	14.543	-0.413	1.085
124	1980	4	13.369	13.725	-0.696	1.062

125	1980	5	13.24	12.767	-0.880	1.099
126	1980	6	13.832	11.983	-0.813	1.131
127	1980	7	14.911	11.752	-0.405	1.133
128	1980	8	13.93	11.686	-0.168	1.111
129	1980	9	13.224	12.033	0.193	0.995
130	1980	10	14.7	12.664	0.499	1.069
131	1980	11	12.952	13.317	0.607	0.944
132	1980	12	12.684	14.034	0.684	0.886
133	1981	1	12.975	14.711	0.680	0.883
134	1981	2	13.939	15.334	0.639	0.918
135	1981	3	16.809	15.887	0.579	1.074
136	1981	4	15.772	16.176	0.376	1.025
137	1981	5	15.314	16.080	0.046	1.037
138	1981	6	16.284	15.816	-0.172	1.088
139	1981	7	16.411	15.435	-0.318	1.104
140	1981	8	16.872	15.128	-0.310	1.113
141	1981	9	15.548	14.964	-0.208	1.013
142	1981	10	14.961	14.619	-0.304	1.050
143	1981	11	13.556	14.322	-0.299	0.945
144	1981	12	13.462	14.234	-0.151	0.911
145	1982	1	12.308	14.057	-0.169	0.880
146	1982	2	14.164	14.165	0.025	0.953
147	1982	3	17.44	14.560	0.284	1.126
148	1982	4	16.875	15.134	0.487	1.063
149	1982	5	17.731	15.885	0.672	1.070
150	1982	6	16.593	16.321	0.507	1.058
151	1982	7	16.331	16.462	0.251	1.057
152	1982	8	16.178	16.321	-0.024	1.062
153	1982	9	16.082	16.220	-0.078	1.004
154	1982	10	15.991	15.978	-0.193	1.029
155	1982	11	17.068	16.194	0.093	0.991
156	1982	12	15.679	16.454	0.210	0.929
157	1983	1	14.407	16.612	0.173	0.875
158	1983	2	15.187	16.634	0.067	0.936
159	1983	3	20.099	16.908	0.213	1.152
160	1983	4	19.129	17.278	0.323	1.082
161	1983	5	20.381	17.860	0.504	1.100
162	1983	6	22.086	18.815	0.820	1.107
163	1983	7	20.35	19.566	0.772	1.050
164	1983	8	19.87	20.045	0.567	1.032
165	1983	9	19.314	20.364	0.393	0.981
166	1983	10	19.892	20.499	0.213	1.005
167	1983	11	19.696	20.562	0.107	0.977
168	1983	12	19.568	20.742	0.158	0.935
169	1984	1	19.419	21.135	0.323	0.893
170	1984	2	21.313	21.694	0.488	0.955
171	1984	3	23.614	21.879	0.276	1.122
172	1984	4	23.688	22.109	0.244	1.077
173	1984	5	25.665	22.528	0.367	1.117
174	1984	6	25.764	22.964	0.415	1.113
175	1984	7	23.805	23.252	0.326	1.039
176	1984	8	23.605	23.451	0.237	1.021
177	1984	9	20.734	23.229	-0.084	0.944
178	1984	10	23.321	23.157	-0.075	1.006
179	1984	11	21.928	22.967	-0.156	0.968
180	1984	12	20.464	22.646	-0.272	0.922
181	1985	1	21.349	22.649	-0.079	0.914
182	1985	2	21.922	22.637	-0.032	0.961
183	1985	3	25.87	22.688	0.026	1.129
184	1985	4	26.98	23.134	0.320	1.115
185	1985	5	28.918	23.893	0.628	1.156
186	1985	6	26.871	24.452	0.580	1.107
187	1985	7	27.195	25.238	0.724	1.055

188	1985	8	27.606	26.153	0.858	1.036
189	1985	9	27.626	27.419	1.143	0.971
190	1985	10	24.19	27.750	0.575	0.949
191	1985	11	22.353	27.385	-0.083	0.904
192	1985	12	22.319	26.746	-0.472	0.885
193	1986	1	22.968	26.068	-0.616	0.900
194	1986	2	22.456	25.077	-0.879	0.933
195	1986	3	25.209	23.860	-1.115	1.099
196	1986	4	27.132	23.032	-0.914	1.141
197	1986	5	29.392	22.714	-0.497	1.214
198	1986	6	28.625	22.871	-0.039	1.168
199	1986	7	27.968	23.494	0.424	1.112
200	1986	8	28.269	24.525	0.849	1.085
201	1986	9	33.995	27.112	2.065	1.090
202	1986	10	27.434	29.127	2.030	0.946
203	1986	11	23.743	30.276	1.413	0.854
204	1986	12	28.947	31.872	1.541	0.895
205	1987	1	20.648	31.528	0.222	0.797
206	1987	2	23.939	30.651	-0.547	0.869
207	1987	3	29.171	29.464	-0.995	1.053
208	1987	4	30.331	28.128	-1.234	1.115
209	1987	5	30.329	26.550	-1.474	1.184
210	1987	6	32.631	25.592	-1.113	1.213
211	1987	7	32.004	25.253	-0.571	1.177
212	1987	8	32.544	25.639	0.099	1.162
213	1987	9	30.103	26.078	0.337	1.117
214	1987	10	28.498	27.082	0.804	0.991
215	1987	11	25.793	28.304	1.097	0.878
216	1987	12	26.905	29.521	1.181	0.902
217	1988	1	26.155	31.081	1.447	0.816
218	1988	2	28.667	32.608	1.503	0.873
219	1988	3	33.504	33.697	1.213	1.028
220	1988	4	32.507	33.875	0.488	1.050
221	1988	5	33.567	33.282	-0.269	1.110
222	1988	6	34.952	32.258	-0.797	1.159
223	1988	7	32.068	30.701	-1.329	1.121
224	1988	8	33.215	29.229	-1.429	1.151
225	1988	9	30.031	27.636	-1.544	1.104
226	1988	10	29.922	26.832	-1.026	1.043
227	1988	11	29.055	27.118	-0.108	0.959
228	1988	12	28.927	27.922	0.531	0.958
229	1989	1	27.652	29.433	1.217	0.868
230	1989	2	27.704	30.842	1.351	0.884
231	1989	3	33.945	32.340	1.454	1.037
232	1989	4	33.192	33.403	1.180	1.026
233	1989	5	36.086	34.209	0.918	1.087
234	1989	6	35.864	34.376	0.393	1.110
235	1989	7	33.738	33.925	-0.198	1.068
236	1989	8	37.068	33.451	-0.391	1.133
237	1989	9	33.809	32.621	-0.698	1.076
238	1989	10	30.497	31.440	-1.037	1.012
239	1989	11	29.171	30.405	-1.035	0.959
240	1989	12	27.285	29.209	-1.148	0.948
241	1990	1	30.925	29.425	-0.193	0.945
242	1990	2	28.762	29.827	0.224	0.918
243	1990	3	34.571	30.641	0.637	1.076
244	1990	4	33.53	31.529	0.813	1.042
245	1990	5	35.827	32.453	0.891	1.094
246	1990	6	35.571	33.109	0.727	1.095
247	1990	7	34.19	33.507	0.496	1.048
248	1990	8	34.988	33.440	0.102	1.097
249	1990	9	31.393	32.758	-0.447	1.026
250	1990	10	31.963	32.178	-0.540	1.004

251	1990	11	29.236	31.429	-0.686	0.947
252	1990	12	26.649	30.268	-1.018	0.920
253	1991	1	25.712	28.884	-1.274	0.922
254	1991	2	27.131	27.962	-1.028	0.940
255	1991	3	32.188	27.473	-0.651	1.116
256	1991	4	33.126	27.717	-0.024	1.106
257	1991	5	34.934	28.456	0.510	1.150
258	1991	6	33.656	29.284	0.732	1.118
259	1991	7	34.07	30.464	1.046	1.078
260	1991	8	32.943	31.246	0.861	1.079
261	1991	9	31.469	31.846	0.678	1.010
262	1991	10	31.378	32.293	0.517	0.991
263	1991	11	28.451	32.312	0.168	0.919
264	1991	12	27.589	32.033	-0.145	0.895
265	1992	1	28.87	31.786	-0.216	0.916
266	1992	2	30.36	31.702	-0.124	0.947
267	1992	3	34.131	31.400	-0.249	1.104
268	1992	4	35.17	31.267	-0.168	1.114
269	1992	5	35.518	31.060	-0.195	1.147
270	1992	6	37.453	31.339	0.137	1.150
271	1992	7	36.872	31.970	0.482	1.109
272	1992	8	34.48	32.363	0.420	1.073
273	1992	9	34.973	33.113	0.651	1.030
274	1992	10	35.695	34.173	0.937	1.013
275	1992	11	31.388	34.937	0.816	0.910
276	1992	12	32.025	35.757	0.819	0.895
277	1993	1	31.128	36.108	0.492	0.893
278	1993	2	31.711	36.038	0.098	0.919
279	1993	3	38.094	35.844	-0.106	1.087
280	1993	4	39.812	35.737	-0.106	1.114
281	1993	5	40.131	35.513	-0.189	1.140
282	1993	6	41.778	35.503	-0.064	1.161
283	1993	7	41.657	35.819	0.202	1.132
284	1993	8	40.32	36.300	0.397	1.089
285	1993	9	38.917	36.895	0.536	1.040
286	1993	10	38.909	37.606	0.658	1.022
287	1993	11	37.645	38.819	1.047	0.935
288	1993	12	37.695	40.268	1.328	0.912
289	1994	1	35.152	41.191	1.045	0.877
290	1994	2	37.817	42.040	0.908	0.911
291	1994	3	47.167	43.032	0.966	1.091
292	1994	4	46.016	43.513	0.627	1.090
293	1994	5	45.476	43.375	0.091	1.102
294	1994	6	47.676	43.031	-0.213	1.139
295	1994	7	44.114	42.126	-0.698	1.096
296	1994	8	46.475	41.653	-0.540	1.100
297	1994	9	44.743	41.455	-0.301	1.057
298	1994	10	44.413	41.567	-0.012	1.042
299	1994	11	41.905	42.139	0.397	0.960
300	1994	12	40.814	42.931	0.673	0.929
301	1995	1	38.938	43.751	0.776	0.882
302	1995	2	39.837	44.385	0.677	0.905
303	1995	3	49.676	45.150	0.738	1.095
304	1995	4	46.089	45.237	0.283	1.060
305	1995	5	50.82	45.630	0.360	1.107
306	1995	6	52.887	46.070	0.416	1.143
307	1995	7	48.088	46.014	0.085	1.075
308	1995	8	51.25	46.186	0.146	1.104
309	1995	9	46.883	45.979	-0.101	1.041
310	1995	10	46.404	45.638	-0.269	1.031
311	1995	11	43.943	45.441	-0.219	0.963
312	1995	12	41.893	45.204	-0.232	0.928
313	1996	1	42.794	45.608	0.214	0.906

314	1996	2	47.315	46.982	1.026	0.948
315	1996	3	53.348	48.139	1.118	1.100
316	1996	4	51.765	49.178	1.063	1.057
317	1996	5	55.694	50.256	1.073	1.107
318	1996	6	52.595	50.375	0.405	1.101
319	1996	7	52.632	50.454	0.177	1.062
320	1996	8	53.039	50.163	-0.150	1.085
321	1996	9	48.914	49.468	-0.532	1.019
322	1996	10	51.677	49.147	-0.384	1.040
323	1996	11	45.877	48.561	-0.526	0.955
324	1996	12	44.017	47.929	-0.600	0.924
325	1997	1	46.045	47.960	-0.158	0.929
326	1997	2	47.822	48.277	0.175	0.966
327	1997	3	55.303	48.777	0.403	1.114
328	1997	4	53.906	49.507	0.631	1.070
329	1997	5	55.65	50.159	0.646	1.108
330	1997	6	55.208	50.684	0.561	1.096
331	1997	7	56.158	51.544	0.770	1.073
332	1997	8	55.742	52.149	0.655	1.078
333	1997	9	52.438	52.561	0.485	1.010
334	1997	10	53.028	52.677	0.227	1.026
335	1997	11	46.27	52.100	-0.336	0.927
336	1997	12	48.112	51.820	-0.297	0.926
337	1998	1	47.397	51.437	-0.357	0.926
338	1998	2	48.606	50.943	-0.453	0.961
339	1998	3	57.545	50.697	-0.308	1.123
340	1998	4	57.883	51.053	0.156	1.097
341	1998	5	60.12	51.756	0.539	1.131
342	1998	6	57.329			
343	1998	7	56.707			
344	1998	8	57.536			
345	1998	9	54.460			
346	1998	10	55.860			
347	1998	11	50.985			
348	1998	12	51.410			
349	1999	1	51.900			
350	1999	2	54.400			
351	1999	3	64.184			
352	1999	4	63.288			
353	1999	5	65.837			

Pledge: On my honor as a student, I have neither given nor received aid on this project.
Baker, Mica; Bogdewic, Eva; Mikhail, Lara; Wang, Mengchen