Assignment #6

Principal Components in Predictive Modeling

Name: Young, Brent

Predict 410 Section #: 57

Quarter: Summer 2017

Introduction

Context

The dataset that we will be working with is called stock_porfolio. The raw data consists of daily closing stock prices for twenty stocks (predictor variables) and a large-cap index fund from Vanguard (VV), which will serve as our response variable. The date is a factor variable and stock prices and index fund are numeric. The focus of this assignment and upcoming assignments in the next few weeks will be on multivariate analysis methods.

Objectives/Purpose

The overall purpose/objective of assignment 6 is to use Principal Components Analysis as a method of dimension reduction and as a remedial measure for multicollinearity in linear regression. We will also learn how to manipulate data in R and make useful R graphics. First, we will conduct some data prep by using the log-returns of the individual stocks to explain the variation in the log-returns of the market index. We will explore this concept using both linear regression and principal components analysis. Second, we will conduct an EDA by examining correlations using a correlation matrix. Third, we will create a corrplot so that we can visualize the pairwise correlations, determine its usefulness compared to the simple barplot, and discuss multicollinearity. Fourth, we will use models as tools for exploratory data analysis by fitting two naïve models (e.g., small and full model – which allows us to compute the VIF value for every predictor variable) and multicollinearity concerns will be discussed. Fifth, we will address multicollinearity by transforming the predictor variables using principal component analysis by computing the principal components for the return data. We will then plot the loadings for the first two principal components from the principal components analysis and relevant discussion will follow (e.g., what are loadings, evidence of groupings in the first two principal components, etc.). Sixth, upon computing the principal components so that we can use PCA for dimension reduction, we will decide how many principal components to keep using decision rules and tools (e.g., default scree plot and common plots associated with PCA). Seventh, we will use principal components in predictive modeling by computing and scoring the principal components on the whole data set, and then splitting it into training and test data sets. We will then using our train and test data sets to fit a linear regression model using the first eight principal components and compute the Mean Absolute Error (MAE) for that model, while also scoring that model out-of-sample on our test data set and computing the out-of-sample MAE. Eighth, we will discuss and compare the principal component regression model with Model #1 and #2 using MAE to determine the best model. Lastly, we will discuss whether the unsupervised decision rule translates to producing the best predictive model (e.g., was our eight principal components model the best predictive model?). We will then consider a supervised approach of using variable selection to select the number of principal components to keep and which principal components to keep.

Section 1: Data Prep

Figure 1: Structure of Data Frame

```
'data.frame':  502 obs. of  22 variables:
 $ Date: Factor w/ 502 levels "1-Apr-13","1-Aug-12",..: 394 381 321 303 270 253 20
6 173 155 139 ...
 $ AA  : num  10.6 10.5 10.7 10.4 10.4 ...
 $ BAC : num  15.6 15.5 15.7 15.6 15.7 ...
 $ BHI : num  55.3 54.5 54.8 54.6 54 ...
 $ CVX : num  125 124 125 125 124 ...
 $ DD  : num  65 64.7 64.2 64.2 63.8 ...
 $ DOW : num  44.4 44.6 44.6 44.9 44.4 ...
 $ DPS : num  48.7 48.8 48.8 48.6 48.5 ...
 $ GS  : num  177 176 176 176 176 ...
 $ HAL : num  50.8 50.4 51.1 51.2 50.7 ...
 $ HES : num  83 82.6 83.1 82.6 81.1 ...
 $ HON : num  91.4 91 91.1 91.1 90.5 ...
 $ HUN : num  24.6 24.3 24.2 24 24.1 ...
 $ JPM : num  58.5 58 58.1 58.2 58.2 ...
 $ KO  : num  41.3 41.1 40.7 40.5 40.2 ...
 $ MMM : num  140 139 139 138 137 ...
 $ MPC : num  91.7 88.5 89.1 89.5 89.3 ...
 $ PEP : num  82.9 82.9 82.7 82.5 82 ...
 $ SLB : num  90.1 89.2 89.9 89.4 88.3 ...
 $ WFC : num  45.4 45.5 45.5 45.5 45.4 ...
 $ XOM : num  101.2 100.3 101.5 100.9 99.2 ...
 $ VV  : num  84.8 84.4 84.5 84.5 84.1 ...
```

Observations (See Appendix for rest of output): Figure 1 shows that our raw data consists of daily closing stock prices for twenty stocks (predictor variables) and a large-cap index fund from Vanguard (VV), which will serve as our response variable. The date is a factor variable and stock prices and index fund are numeric. Additionally, dates are shown in decreasing order.

Figure 2: Log Returns for all 20 stocks and the Index fund

```
'data.frame':  501 obs. of  21 variables:
 $ AA : num  0.02356 -0.00957 -0.0216 0.02799 0.00212 ...
 $ BAC: num  0.00172 0.08256 -0.02082 0.01446 0.05583 ...
 $ BHI: num  0.00995 -0.01387 0.00862 0.00622 0.00715 ...
 $ CVX: num  -0.00172 -0.00985 -0.00727 0.01084 -0.00394 ...
 $ DD : num  0.01091 -0.00683 -0.01423 0.00844 0.01518 ...
 $ DOW: num  0.00536 0.00632 0.00595 -0.00033 0.02186 ...
 $ DPS: num  0.00546 0.00621 -0.00698 0 0.00259 ...
 $ GS : num  -0.00652 -0.00169 -0.01234 0.0135 0.03772 ...
 $ HAL: num  0.028 -0.0161 0.0121 0.0114 0.0265 ...
 $ HES: num  0.01022 -0.02401 -0.0207 0.00847 0.02876 ...
 $ HON: num  -0.0009 0.00108 -0.0074 0.0083 0.01675 ...
 $ HUN: num  -0.00807 -0.00508 0.00811 -0.00608 0.03593 ...
 $ JPM: num  -0.000858 0.020672 -0.009009 -0.001698 0.021024 ...
 $ KO : num  -0.00629 -0.00489 -0.00636 0 0.00608 ...
 $ MMM: num  0.00823 -0.00452 -0.00514 0.00598 0.00511 ...
 $ MPC: num  0.01042 -0.05604 -0.00818 -0.02236 0.02771 ...
 $ PEP: num  0.00511 -0.00782 -0.01261 0.00519 -0.00107 ...
 $ SLB: num  -0.00759 -0.02165 -0.00427 0.01523 0.02766 ...
 $ WFC: num  0.00456 0.01598 -0.00276 0.01236 0.00375 ...
 $ XOM: num  0.000233 -0.003027 -0.007491 0.004454 0.00257 ...
 $ VV : num  0.0012 0.00326 -0.00206 0.00223 0.0092 ...
```

Observations (see Appendix for rest of output): Figure 2 shows the computed log-returns for all the stocks and the index fund. The log-returns of the individual stocks helps us explain the variation in the log-returns of the market index. We conduct this step because the stock market data is auto correlated due to day to day stock prices. This step removes autocorrelation, which is important because regression models assumes that rows are independent. As a result, this issue is addressed by taking yesterday's stock price and today's stock price and then taking the log of today's stock price and the log of yesterday's stock price (e.g., log of today's price vs. log of yesterday's price or log return number). In the end, index fund will be our response variable and all other stock returns will be our predictors.

Section 2: Exploratory Data Analysis – Statistical Graphics

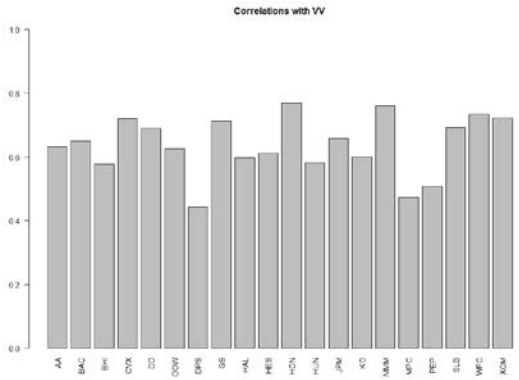
Figure 3: Correlation Matrix for Returns

	AA	BAC	BHI	CVX	DD	DOW	DPS	GS	HAL	HES	HON	HUN	JPM	KO	MMM	MPC	PEP	SLB	WFC	XOM	VV
AA	1.0000000	0.5367187	0.5150838	0.5298038	0.5180594	0.4651004	0.2248131	0.5339638	0.5129001	0.4789793	0.5125477	0.4572748	0.4293073	0.2828796	0.5284540	0.3179508	0.2683623	0.5340605	0.5044261	0.5391202	0.6324106
BAC	0.5367187	1.0000000	0.3436988	0.4571983	0.4756738	0.4479004	0.1904652	0.6925595	0.3822280	0.4333956	0.4774514	0.4315406	0.7102490	0.3001553	0.4592117	0.3233239	0.2552220	0.4686633	0.6458383	0.4714917	0.6501877
BHI	0.5150838	0.3436988	1.0000000	0.5554974	0.4238151	0.3595952	0.2154207	0.4708514	0.7284132	0.5308592	0.4212415	0.3463544	0.3552885	0.2974471	0.4714887	0.3319386	0.2560867	0.7112089	0.4065921	0.5231291	0.5774988
CVX	0.5298038	0.4571983	0.5554974	1.0000000	0.5542110	0.4702170	0.3428003	0.5202132	0.5258214	0.6030530	0.5460861	0.3703411	0.4587828	0.4908741	0.5976374	0.3884111	0.4439708	0.6168476	0.5368363	0.7460119	0.7209041
DD	0.5180594	0.4756738	0.4238151	0.5542110	1.0000000	0.6249069	0.3242849	0.5161635	0.4464519	0.4475557	0.5753468	0.5343228	0.4753553	0.4173935	0.6408984	0.3386685	0.3736498	0.5021188	0.5482524	0.5442086	0.6895190
DOW	0.4651004	0.4479004	0.3595952	0.4702170	0.6249069	1.0000000	0.1773703	0.4716100	0.3854083	0.4173819	0.5429382	0.5196042	0.4275565	0.3207388	0.4892010	0.3399170	0.2644920	0.4360083	0.4976916	0.4666825	0.6264550
DPS	0.2248131	0.1904652	0.2154207	0.3428003	0.3242849	0.1773703	1.0000000	0.2194156	0.2397646	0.2549822	0.3635036	0.1725909	0.2561960	0.5162796	0.3804475	0.2049708	0.4567402	0.2327236	0.2445577	0.3120334	0.4435005
GS	0.5339638	0.6925595	0.4708514	0.5202132	0.5161635	0.4716100	0.2194156	1.0000000	0.5172029	0.5348167	0.5477281	0.4518345	0.7366364	0.3217882	0.5472489	0.3996328	0.3187963	0.5732592	0.6537527	0.5586785	0.7121620
HAL	0.5129001	0.3822280	0.7284132	0.5258214	0.4464519	0.3854083	0.2397646	0.5172029	1.0000000	0.5672873	0.4913323	0.3721343	0.4021608	0.3128718	0.4845256	0.3390298	0.2879595	0.7443540	0.3981402	0.5304929	0.5974989
HES	0.4789793	0.4333956	0.5308592	0.6030530	0.4475557	0.4173819	0.2549822	0.5348167	0.5672873	1.0000000	0.4909095	0.4017311	0.4123103	0.3274208	0.4645992	0.4217009	0.2746954	0.6077722	0.4495616	0.5820285	0.6107960
HON	0.5125477	0.4774514	0.4212415	0.5460861	0.5753468	0.5429382	0.3635036	0.5477281	0.4913323	0.4909095	1.0000000	0.4929541	0.5025561	0.4543087	0.6659124	0.3854155	0.4036529	0.5514632	0.5899769	0.5486838	0.7683784
HUN	0.4572748	0.4315406	0.3463544	0.3703411	0.5343228	0.5196042	0.1725909	0.4518345	0.3721343	0.4017311	0.4929541	1.0000000	0.3998628	0.2807697	0.4431757	0.3761363	0.2309619	0.4064911	0.4553336	0.3427736	0.5819449
JPM	0.4293073	0.7102490	0.3552885	0.4587828	0.4753553	0.4275565	0.2561960	0.7366364	0.4021608	0.4123103	0.5025561	0.3998628	1.0000000	0.3639307	0.4814331	0.3562568	0.3228739	0.4423912	0.6500578	0.4808664	0.6578478
KO	0.2828796	0.3001553	0.2974471	0.4908741	0.4173935	0.3207388	0.5162796	0.3217882	0.3128718	0.3274208	0.4543087	0.2807697	0.3639307	1.0000000	0.4968764	0.2743291	0.5752982	0.3645841	0.4271055	0.4754139	0.5997988
MMM	0.5284540	0.4592117	0.4714887	0.5976374	0.6408984	0.4892010	0.3804475	0.5472489	0.4845256	0.4645992	0.6659124	0.4431757	0.4814331	0.4968764	1.0000000	0.3345463	0.4583911	0.5664340	0.5459150	0.6403116	0.7608489
MPC	0.3179508	0.3233239	0.3319386	0.3884111	0.3386685	0.3399170	0.2049708	0.3996328	0.3390298	0.4217009	0.3854155	0.3761363	0.3562568	0.2743291	0.3345463	1.0000000	0.2395772	0.3754614	0.3553949	0.3841860	0.4731198
PEP	0.2683623	0.2552220	0.2560867	0.4439708	0.3736498	0.2644920	0.4567402	0.3187963	0.2879595	0.2746954	0.4036529	0.2309619	0.3228739	0.5752982	0.4583911	0.2395772	1.0000000	0.3129127	0.3573397	0.4561750	0.5075264
SLB	0.5340605	0.4686633	0.7112089	0.6168476	0.5021188	0.4360083	0.2327236	0.5732592	0.7443540	0.6077722	0.5514632	0.4064911	0.4423912	0.3645841	0.5664340	0.3754614	0.3129127	1.0000000	0.4886284	0.6053182	0.6928534
WFC	0.5044261	0.6458383	0.4065921	0.5368363	0.5482524	0.4976916	0.2445577	0.6537527	0.3981402	0.4495616	0.5899769	0.4553336	0.6500578	0.4271055	0.5459150	0.3553949	0.3573397	0.4886284	1.0000000	0.5364858	0.7335731
XOM	0.5391202	0.4714917	0.5231291	0.7460119	0.5442086	0.4666825	0.3120334	0.5586785	0.5304929	0.5820285	0.5486838	0.3427736	0.4808664	0.4754139	0.6403116	0.3841860	0.4561750	0.6053182	0.5364858	1.0000000	0.7211079
VV	0.6324106	0.6501877	0.5774988	0.7209041	0.6895190	0.6264550	0.4435005	0.7121620	0.5974989	0.6107960	0.7683784	0.5819449	0.6578478	0.5997988	0.7608489	0.4731198	0.5075264	0.6928534	0.7335731	0.7211079	1.0000000

Figure 4: Correlations for Stock Returns vs. VV

AA	BAC	BHI	CVX	DD	DOW	DPS	GS	HAL
0.6324106	0.6501877	0.5774988	0.7209041	0.6895190	0.6264550	0.4435005	0.7121620	0.5974989
HES	**HON**	**HUN**	**JPM**	**KO**	**MMM**	**MPC**	**PEP**	**SLB**
0.6107960	0.7683784	0.5819449	0.6578478	0.5997988	0.7608489	0.4731198	0.5075264	0.6928534
WFC	**XOM**	**VV**						
0.7335731	0.7211079	1.0000000						

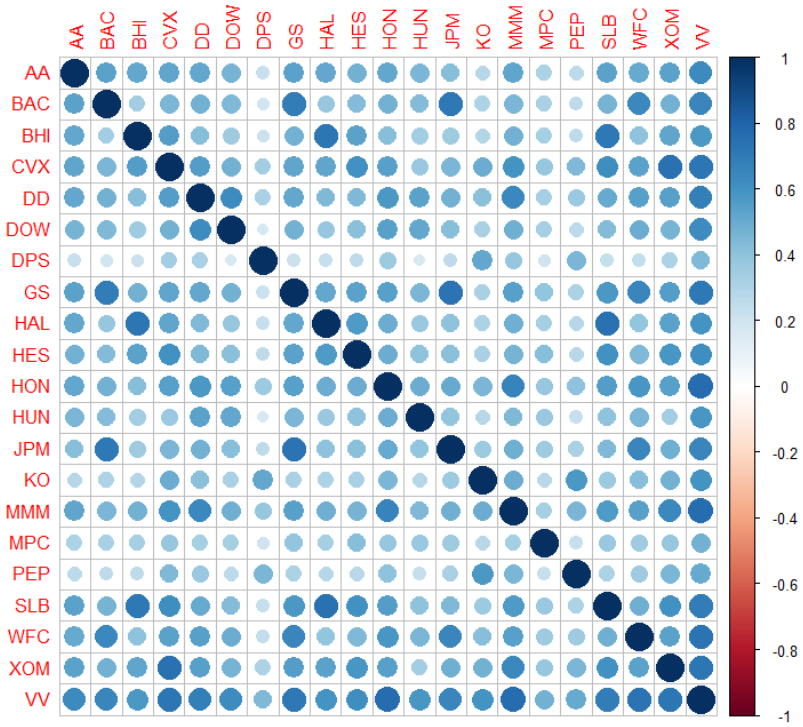
Figure 5: Barplot of Correlations with VV



Observations: Figure 3 and 5 shows a correlation matrix for returns and a barplot of correlation with VV so that we can understand the correlations of the index fund VV with all the other potential predictor stock returns. As a result, we are most interested in the VV column/row of the correlation matrix because it's our response variable. The results in figure 4 show that DPS, MPC, and PEP have the smallest correlations to VV, while MMM and HON have the highest correlations to VV.

Section 3: Exploratory Data Analysis – Data Visualization

Figure 6: Correlation Plot for Returns



Observations: Figure 6 shows a correlation plot for returns so that we can visualize the pairwise correlations in the data. The corrplot is more useful or insightful than our simple barplot because it shows the "full" picture and looks at all possible factors that affect the process (comprehensive), whereas the simple barplot shows a "limited" view and does not look at all factors. Additionally, it's important to note that a statistical graphic is meant to "uncover" and obtain answers that are being sought, whereas data visualization has to do with tailoring the visualization to the dataset and the narrative. Furthermore, data visualization allows for higher levels of visual abstraction which can be accomplished by presenting data pictorially and/or graphically using color, form, and positioning that are both "visually appealing" and allows the user to see and grasp patterns/trends quickly. For example, the corrplot uses color and form (e.g., size) to distinguish between higher and lower correlations. This is seen in the corrplot where the dark circles and bigger circles indicate higher correlation, while the lighter and smaller circles indicate lower correlation between the predictor variables and the index fund. This gives us an idea on which potential predictors (e.g., higher correlated variables) we may want to include in our model. The results show that DPS, MPC, and PEP have the smallest correlations to VV, while MMM and HON have the highest correlations to VV. Additionally, we can also visually see the stocks that should have low VIF values and stocks that should have high VIF values. For instance, the three stocks that should have low VIF values are MPC, DPS, and PEP since there are a lot of lighter and smaller circles along their respective rows/columns in the corrplot. On the other hand, three stocks that should have high VIF values are SLB, GS, and XOM since there are a lot of darker and bigger circles along their respective rows/columns in the corrplot, which indicates that a lot of them have high correlations between two or more predictor variables.

Section 4: Exploratory Data Analysis – Naïve Models

Figure 7: Model.1 (Small Model) VV ~ GS + DD + DOW + HON + HUN + JPM + KO + MMM + XOM

```
Call:
lm(formula = VV ~ GS + DD + DOW + HON + HUN + JPM + KO + MMM +
    XOM, data = returns.df)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0139179 -0.0016005 -0.0000926  0.0016690  0.0172703

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0001008  0.0001331   0.757 0.449290
GS          0.0784765  0.0138277   5.675 2.37e-08 ***
DD          0.0354057  0.0177154   1.999 0.046204 *
DOW         0.0406763  0.0116993   3.477 0.000552 ***
HON         0.1449817  0.0170837   8.487 2.53e-16 ***
HUN         0.0385118  0.0077371   4.978 8.93e-07 ***
JPM         0.0505123  0.0132262   3.819 0.000151 ***
KO          0.1419686  0.0176282   8.054 6.14e-15 ***
MMM         0.1336002  0.0239378   5.581 3.96e-08 ***
XOM         0.1480728  0.0213601   6.932 1.31e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002951 on 491 degrees of freedom
Multiple R-squared:  0.8518,   Adjusted R-squared:  0.849
F-statistic: 313.5 on 9 and 491 DF,  p-value: < 2.2e-16
```

Observations: Figure 7 shows a summary of Model.1 (Small Model) VV ~ GS + DD + DOW + HON + HUN + JPM + KO + MMM + XOM. The equation of the regression line is: VV = 0.0001008 + 0.0784765*GS + 0.0354057*DD + 0.0406763*DOW + 0.1449817*HON + 0.0385118*HUN + 0.0505123*JPM + 0.1419686*KO + 0.1336002*MMM + 0.1480728*XOM. Since the t-test of all the predictor variables are statistically significant, we can use this equation. The residual standard error of 0.002951, shows us that when predicting VV, one standard error = 0.003 units. The multiple R-squared value of 0.8518, indicates that 85.18% of the variation in VV is explained by the predictor variables.

Figure 8: VIF Values of Model.1

```
      GS       DD      DOW      HON      HUN      JPM       KO
2.705795 2.368257 1.919773 2.261397 1.633336 2.324600 1.473202
     MMM      XOM
2.590177 2.073721
```

Observations: Figure 8 shows us the VIF values for Model.1. A VIF of 1 would mean that no multicolinearity exists at all, while a large VIF number (e.g., 10) would indicate serious multicolinearity issues. GS has the largest VIF value of 2.7. However, since the VIF for all the predictors above are low, this concludes that we don’t have serious multicolinearity issues.

Figure 9: Model.2 (Full Model) VV ~ BAC+GS+JPM+WFC+BHI+CVX+DD+DOW+DPS+HAL+HES+HON+HUN+KO+MMM+MPC+PEP+SLB+XOM

```
Call:
lm(formula = VV ~ BAC + GS + JPM + WFC + BHI + CVX + DD + DOW +
    DPS + HAL + HES + HON + HUN + KO + MMM + MPC + PEP + SLB +
    XOM, data = returns.df)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0138057 -0.0015122  0.0000384  0.0014777  0.0143463

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.846e-05  1.213e-04   0.730 0.466020
BAC          3.037e-02  9.474e-03   3.205 0.001440 **
GS           3.528e-02  1.358e-02   2.598 0.009675 **
JPM          2.019e-02  1.323e-02   1.526 0.127769
WFC          7.829e-02  1.581e-02   4.952 1.02e-06 ***
BHI          1.834e-02  1.152e-02   1.593 0.111884
CVX          5.925e-02  2.067e-02   2.866 0.004337 **
DD           1.148e-02  1.624e-02   0.707 0.480021
DOW          3.671e-02  1.070e-02   3.431 0.000652 ***
DPS          5.722e-02  1.495e-02   3.828 0.000146 ***
HAL         -5.837e-04  1.208e-02  -0.048 0.961476
HES          4.589e-03  9.699e-03   0.473 0.636297
HON          1.085e-01  1.607e-02   6.751 4.25e-11 ***
HUN          2.988e-02  7.184e-03   4.160 3.78e-05 ***
KO           9.194e-02  1.843e-02   4.990 8.45e-07 ***
MMM          1.117e-01  2.198e-02   5.080 5.41e-07 ***
MPC          1.059e-02  7.032e-03   1.506 0.132741
PEP          2.024e-02  2.036e-02   0.994 0.320703
SLB          4.807e-02  1.454e-02   3.306 0.001019 **
XOM          6.115e-02  2.294e-02   2.665 0.007947 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002669 on 481 degrees of freedom
Multiple R-squared:  0.8812,   Adjusted R-squared:  0.8765
F-statistic: 187.8 on 19 and 481 DF,  p-value: < 2.2e-16
```

Observations: Figure 9 shows a summary of Model.2 (Full Model) VV ~ BAC + GS + JPM + WFC + BHI + CVX + DD + DOW + DPS + HAL + HES + HON + HUN + KO + MMM + MPC + PEP + SLB + XOM. The results show that some of the predictor variables are not significant such as JPM, BHI, DD, HAL, HES, MPC, and PEP. The residual standard error of 0.002669, shows us that when predicting VV, one standard error = 0.0026 units. This is lower than what we saw in Model.1 (Small Model). The multiple R-squared value of 0.8812, indicates that 88.12% of the variation in VV is explained by the predictor variables. The adjusted-R squared value (87.65%) is also better than Model.1 (Small Model) of 84.9%.

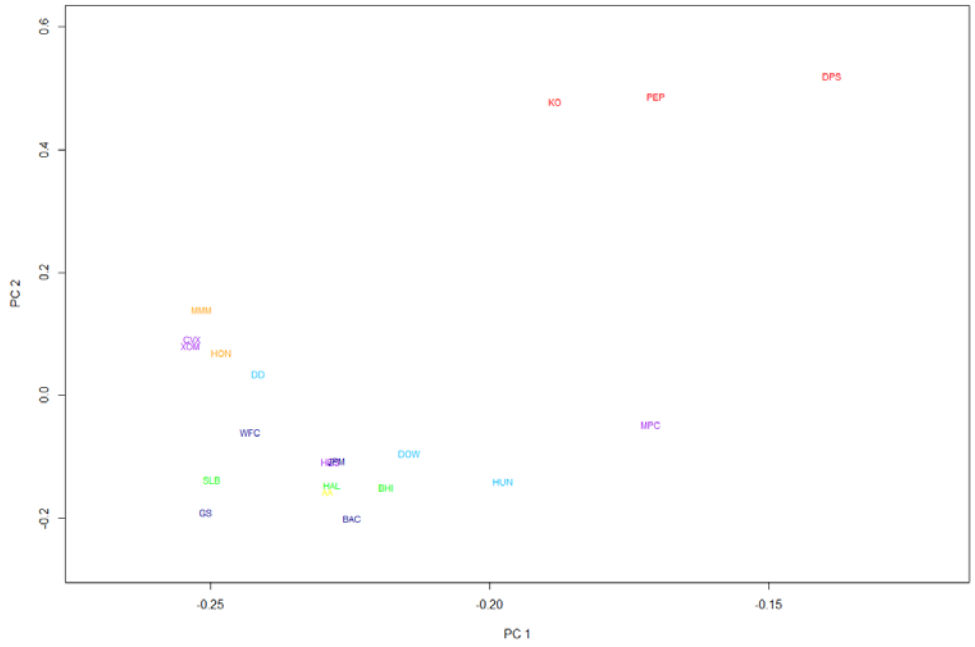
Figure 10: VIF Values of Model.2

```
     BAC       GS      JPM      WFC      BHI      CVX       DD
2.558097 3.190808 2.844537 2.528808 2.603510 2.909686 2.432674
     DOW      DPS      HAL      HES      HON      HUN       KO
1.961953 1.524399 2.902240 2.095666 2.447013 1.721319 1.967512
     MMM      MPC      PEP      SLB      XOM
2.670404 1.376185 1.719788 3.257595 2.924084
```

Observations: Figure 10 shows us the VIF values for Model.2. A VIF of 1 would mean that no multicolinearity exists at all, while a large VIF number (e.g., 10) would indicate serious multicolinearity issues. CVX has the largest VIF value of 2.9. However, since the VIF for all the predictors above are low, this concludes that we don't have serious multicolinearity issues.

Section 5: Principal Component Analysis

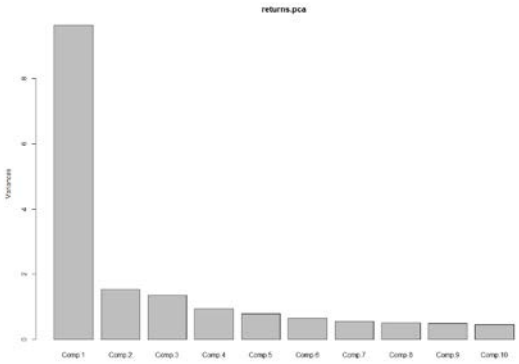
Figure 11: Plot of Loadings for First Two Principal Components from the PCA



Observations: Figure 11 shows us a plot of the loadings for the first two principal components from the PCA. The plot shows eigenvector values so that we can see which individual stocks have a higher correlation with the principal components. The loadings are the correlation coefficients between the original variables (rows) and the factors (columns). Yes, interestingly there are 5 primary clusters/groupings. KO, PEP, and DPS are all part of the soda industry and are clustered in the top right hand corner (positive correlation). Additionally, companies in the banking industry such as JPM, BAC, WFC, and GS are located close together as well (bottom left hand corner, negative correlation). There also seems to be clusters with companies in the oil field services industry such as SLB, HAL, and BHI (bottom left hand corner, negative correlation). Furthermore, companies in the industrial – chemical industry such as DD, DOW, and HUN are also somewhat close together as well (negative correlation). Companies in the manufacturing industry such as HON and MMM are also clustered together as well (negative correlation). The biggest surprise is that companies in the oil refining industry seem to be spread out. For instance, CVX, and XOM are clustered together, but HES and MPC (outlier) far away from the primary cluster.

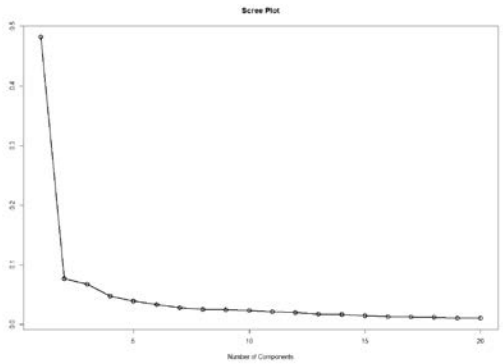
Section 6: PCA for Dimension Reduction

Figure 12: Default Scree Plot



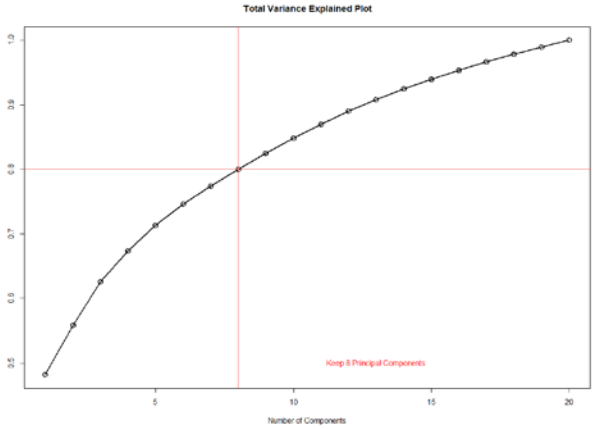
Observations: Figure 12 shows us a default scree plot to help us decide the number of principal components to keep. The plot shows the variances for each principal component. Principal component 1 has the highest variance, but drops off significantly afterwards (smaller variances for principal components 2 through 10). However, the default scree plot is not that great, so we will create two additional common plots associated with PCA.

Figure 13: Scree Plot (Custom)



Observations: Figure 13 shows a scree plot that was custom built and is similar to figure 12. We see similar results in this plot compared to figure 12. For example, variance of the first principal component is 0.5, but then for principal component #2 there's a huge drop off to 0.1. This trend continues for the remainder of the principal components. When using screen plots, we are generally looking for the value corresponding to an "elbow" in the curve (change of slope from steep to shallow). Additionally, looking for a point on the plot beyond which the screen plot defines more or less straight line…the first point on the straight line is taken as the last component to be retained. Given this context, there are two inflection points: one at component 1 (variance 0.5) and the other at component 8. As a result, we should keep the first 8 principal components since the first component would only explain 50% of the total variation.

Figure 14: Total Variance Explained Plot



Observations: Figure 14 shows the total variance explained plot, which shows cumulative variance explained by these components. The plot shows that if we take the first 8 principal components, 80% of the data is explained by the first 8 principal components, whereas if we keep only 1 principal component, only 50% of the total variation would be explained. I agree that we should keep the first 8 principal components because my cut off is 80%, which falls in the 70 – 90% range. There are also other decision rules to consider. For example, we could keep the principal components whose eigenvalues were greater than or equal to the average. Another decision rule that we could have used is when R is used and average eigenvalue is 1, we can keep the principal components with eigenvalues greater than or equal to 1 (sometimes 0.7 is also used rather than 1).

Section 7: Principle Components in Predictive Modeling

Figure 15: Principle Component Scores & Train/Test Data Sets

```
      Comp.1     Comp.2     Comp.3      Comp.4      Comp.5      Comp.6
1 -1.1360839 -0.1001746 -0.8674072  0.42752665 -0.29967832 -0.44555232
2  1.1571296 -0.8364333  3.8310239 -2.03778795 -2.44561893 -1.11738629
3  2.6730619 -1.2621717 -0.8505104  0.69508835 -0.05100329 -0.81982193
4 -2.1318679 -0.2549437 -0.4895088 -0.60836423 -1.88400278  0.26777730
5 -4.4748787 -1.7704163  0.9632157 -0.09346543  1.00668797 -1.33335849
6  0.6710031 -3.1592357  3.7817738  0.95104343  1.34743815 -0.01974151
      Comp.7     Comp.8      Comp.9    Comp.10     Comp.11     Comp.12
1 -0.6721856  0.7073485 -0.62081617 -0.8726573 -0.80377898 -0.32746435
2 -0.8900266 -0.5325490  0.92505952 -0.2789242  0.63166518 -0.45785493
3  0.9347629 -0.8921977  0.07286909 -0.2904567  0.85908742  0.03687000
4 -0.4889668  0.7285611 -0.16296811  0.2960505 -0.09423674 -0.46011533
5  0.3955532 -1.0283541  0.54516084  0.2009467 -0.63714894 -0.17138083
6 -0.8407598  1.1026857 -0.30564378 -0.2369338 -0.47143933 -0.09881718
     Comp.13    Comp.14    Comp.15     Comp.16     Comp.17    Comp.18
1 -0.3354962 -1.0771942 -0.3248166  0.16406639  0.05713498  1.2326901
2 -0.1718655  0.9144362  1.3878807 -1.01061327 -0.37774595  1.6338578
3 -0.3671263  1.0450753 -0.4637268 -0.25824197  0.26429039  0.4277640
4 -0.3864692 -0.5174733  0.3778572  0.19644908  0.19461057 -0.6601559
5  1.5089513 -0.3190476  0.9321220 -0.28359470  0.35969953 -0.2062964
6  0.1976405  0.1037615  0.4921323 -0.09505413  0.45454584 -0.3616146
     Comp.19    Comp.20           VV          u
1 -0.7506313 -0.4718081  0.001202439 0.88493071
2  0.6839883 -0.5780093  0.003256494 0.07501205
3 -0.3951312 -0.3290371 -0.002055499 0.95396299
4 -0.1726132 -0.2297121  0.002226600 0.49131144
5  0.7707300 -0.3233901  0.009196250 0.04452786
6 -0.1454178  0.5307264  0.001185938 0.73712242
```

```
> dim(train.scores)
[1] 354  22
> dim(test.scores)
[1] 147  22
> dim(train.scores)+dim(test.scores)
[1] 501  44
> dim(return.scores)
[1] 501  22
```

Observations: Figure 15 shows the principle component scores for the 20 stock returns along with the train and test data sets. The principle component scores represent the transformed variable values corresponding to a particular data point. In other words, the 20 original raw data variables were transformed to 20 principal component scores. The letter u allows us to generate a uniform random number between 0 and 1. We can then use this to split the data into train and test data sets (principle component score data) as seen above.

Figure 17: Model pca1.lm - VV ~ Comp.1+Comp.2+Comp.3+Comp.4+Comp.5+Comp.6+Comp.7+Comp.8

```
Call:
lm(formula = VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 +
    Comp.6 + Comp.7 + Comp.8, data = train.scores)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0134195 -0.0015478 -0.0000415  0.0015667  0.0148726

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.493e-04  1.474e-04   5.083 6.12e-07 ***
Comp.1      -2.249e-03  4.858e-05 -46.306  < 2e-16 ***
Comp.2       4.261e-04  1.211e-04   3.518 0.000492 ***
Comp.3       5.807e-04  1.307e-04   4.442 1.20e-05 ***
Comp.4       1.627e-04  1.528e-04   1.065 0.287747
Comp.5      -2.578e-04  1.609e-04  -1.602 0.110181
Comp.6       3.049e-05  1.773e-04   0.172 0.863588
Comp.7       9.543e-05  1.931e-04   0.494 0.621404
Comp.8      -2.759e-04  2.064e-04  -1.337 0.182236
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002769 on 345 degrees of freedom
Multiple R-squared:  0.8635,   Adjusted R-squared:  0.8603
F-statistic: 272.7 on 8 and 345 DF,  p-value: < 2.2e-16
```

Observations: Figure 17 shows a summary of model pca1.lm: VV ~ Comp.1+Comp.2+Comp.3+Comp.4+Comp.5+Comp.6+Comp.7+Comp.8. The results show that components 1 through 3 are significant. However, components 4 through 8 are not significant. The residual standard error of 0.002769, shows us that when predicting VV, one standard error = 0.0027 units. The multiple R-squared value of 0.8635, indicates that 86.35% of the variation in VV is explained by the predictor variables. This is almost comparable to the Model.2 (Full Model), but we were able to accomplish this with fewer predictors (8 principle components).

Figure 18: Mean Absolute Error on the Training Sample (See Appendix for Score)

```
pca1.mae.train          0.00199310370181127
```

Observations: Figure 18 shows the MAE for model pca1.lm using the training data set. The MAE is really low at 0.00199.

Figure 19: VIF Values of pca1.lm

```
  Comp.1   Comp.2   Comp.3   Comp.4   Comp.5   Comp.6   Comp.7   Comp.8
1.020430 1.002519 1.007520 1.009464 1.008865 1.007684 1.007435 1.011134
```

Observations: Figure 19 shows us the VIF values for model pca1.lm. A VIF of 1 would mean that no multicolinearity exists at all, while a large VIF number (e.g., 10) would indicate serious multicolinearity issues. The VIF values associated with every predictor variable in the principal components regression model are all one because by definition, principle component scores are uncorrelated. Therefore, we never have to worry about multicollinearity because VIF will always be equal to 1.

Figure 20: MAE of out-of-sample

pca1.mae.test	0.00212560571144729

Observations: Figure 19 shows the MAE for model pca1.lm using the test data set. The MAE is really also really low at 0.00212, but slightly higher than the MAE for model pca1.lm using the training data set.

Section 8: Regression Model Comparisons

Figure 21: Train/Test Split of the Returns Data Set

```
> dim(train.returns)
[1] 354  22
> dim(test.returns)
[1] 147  22
> dim(train.returns)+dim(test.returns)
[1] 501  44
> dim(returns.df)
[1] 501  22
```

Observations: Figure 21 shows the train and test split of the returns data set so that we can match the scores data set.

Figure 22: Table of MAE values from models pca1.lm, model.1, and model.2

Model Name	MAE Value (Train)	Rank	MAE Value (Test)	Rank
pca1.lm	0.00199	2	0.00212	1
model.1 (Small)	0.00213	3	0.00232	3
model.2 (Full)	0.00191	1	0.00215	2

Observations: Figure 22 shows a table of MAE values from models pca1.lm, model.1, and model.2. For the train data set, model.2 had the best MAE value, followed by pca1.lm, and then model.1. For the test data set, pca1.lm and model.2 were virtually tied, while model.1 had the largest MAE value. As a result, in my opinion, pca1.lm is the best model because it is a simpler model. For instance, we are making a tradeoff – settling for less accuracy/more bias, but more precision. Furthermore, this was accomplished using 8 predictor variables versus 20 predictor variables in model.2 (full). Additionally, the MAE values for the train and test data set and residual standard error (0.0027 vs. 0.0026) were comparable to model.2 and better than model.1 (0.0029). Additionally, the adjusted r-squared value for pca1.lm of 86% was also comparable to model.2 (87.7%) and better than model.1 (84.9%).

Section 9: Supervised Approach of Using Variable Selection to Select # of Principal Components

Figure 23: Model backward.lm – VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 + Comp.9 + Comp.10 + Comp.11

```
Call:
lm(formula = VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 +
    Comp.9 + Comp.10 + Comp.11, data = train.scores)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0139929 -0.0014466  0.0000648  0.0014682  0.0141487

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.192e-04  1.423e-04   5.055 7.01e-07 ***
Comp.1      -2.240e-03  4.661e-05 -48.059  < 2e-16 ***
Comp.2       4.497e-04  1.167e-04   3.854 0.000139 ***
Comp.3       5.724e-04  1.258e-04   4.551 7.41e-06 ***
Comp.5      -2.622e-04  1.551e-04  -1.690 0.091853 .
Comp.8      -2.845e-04  1.984e-04  -1.434 0.152485
Comp.9      -5.125e-04  2.030e-04  -2.524 0.012037 *
Comp.10      6.609e-04  2.157e-04   3.063 0.002361 **
Comp.11      7.949e-04  2.133e-04   3.727 0.000226 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002665 on 345 degrees of freedom
Multiple R-squared:  0.8735,   Adjusted R-squared:  0.8706
F-statistic: 297.7 on 8 and 345 DF,  p-value: < 2.2e-16
```

Observations (see appendix for variable selction): Figure 23 shows a summary of backward.lm: VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 + Comp.9 + Comp.10 + Comp.11. The results show that all the components are significant except component 5 (p-value greater than 0.05) and component 8. The results show that we should keep 8 principal components (specifically, 1, 2,3,5,8,9,10, and 11). However, an argument can be made to drop component 5 and 8 since they are insignificant (p-value greater than 0.05). This differs from our previous discussion about the number of principal components to keep because the variable selection approach “chooses” the components whereas in pca1.lm we used the first 8 principle components. The residual standard error of 0.002665, shows us that when predicting VV, one standard error = 0.0026 units. The multiple R-squared value of 0.8735, indicates that 87.35% of the variation in VV is explained by the predictor variables. Additionally, the adjusted R-squared of 87% is comparable to pca1.lm (86%) and model.2 (87.7%).

Figure 24: VIF Values of backward.lm

```
  Comp.1   Comp.2   Comp.3   Comp.5   Comp.8   Comp.9  Comp.10  Comp.11
1.013666 1.004187 1.006236 1.011192 1.008004 1.005927 1.012893 1.003629
```

Observations: Figure 24 shows us the VIF values for backward.lm. A VIF of 1 would mean that no multicolinearity exists at all, while a large VIF number (e.g., 10) would indicate serious multicolinearity issues. The VIF values associated with every predictor variable in the principal components regression backward model are all one because by definition, principle component scores are uncorrelated. Therefore, we never have to worry about multicollinearity because VIF will always be equal to 1.

Figure 25: Table of MAE values from models pca1.lm, model.1, model.2, and backward.lm

Model Name	MAE Value (Train)	Rank	MAE Value (Test)	Rank
pca1.lm	0.00199	2	0.00212	1
model.1 (Small)	0.00213	3	0.00232	3
model.2 (Full)	0.00191	1	0.00215	2
backward.lm	0.00191	1	0.00215	2

Observations: Figure 25 shows a table of MAE values from models pca1.lm, model.1, model.2, and backward.lm. For the train data set, model.2 and backward.lm had the best MAE value, followed by pca1.lm, and then model.1. For the test data set, pca1.lm, model.2, and backward.lm were virtually tied, while model.1 had the largest MAE value. As a result, in my opinion, backward.lm is the best model because not only is it a simpler model (less accuracy/more bias, but more precision), but the variable selection was more accurate at choosing the 8 predictor variables than pca1.lm. Additionally, 6 out of the 8 variables had strong statistical significance compared to only 3 out of the 8 for pca1.lm and 12 out of the 19 for model.2. Furthermore, backward.lm also had comparable MAE values for the train and test data set and comparable or better residual standard error (0.0026) compared to pca1.lm (residual standard error: 0.0027) and model.2 (residual standard error: 0.0026) and model.1 (residual standard error: 0.0029). Additionally, the adjusted r-squared value of 87% was also comparable or better than model.2 (87.7%), pca1.lm (86%), and model.1 (84.9%).

Summary/Conclusion

In section 1, we conducted some data prep by using the log-returns of the individual stocks to explain the variation in the log-returns of the market index. In section 2, we conducted an EDA by examining correlations using a correlation matrix. The results showed that DPS, MPC, and PEP have the smallest correlations to VV, while MMM and HON have the highest correlations to VV.

In section 3, we created a corrplot so that we can visualize the pairwise correlations, determine its usefulness compared to the simple barplot, and discuss multicollinearity. The results showed that the corrplot is more useful or insightful than our simple barplot because it shows the "full" picture and looks at all possible factors that affect the process (comprehensive), whereas the simple barplot shows a "limited" view and does not look at all factors. The results showed that DPS, MPC, and PEP have the smallest correlations to VV, while MMM and HON have the highest correlations to VV. Additionally, we can also visually see the stocks that should have low VIF values and stocks that should have high VIF values. For instance, the three stocks that should have low VIF values are MPC, DPS, and PEP since there are a lot of lighter and smaller circles along their respective rows/columns in the corrplot. On the other hand, three stocks that should have high VIF values are SLB, GS, and XOM since there are a lot of darker and bigger circles along their respective rows/columns in the corrplot, which indicates that a lot of them have high correlations between two or more predictor variables.

In Section 4, we used models as tools for exploratory data analysis by fitting two naïve models (e.g., Model.1 (small) and Model.2 (full). Model.1's predictor variables were all statistically significant. The residual standard error of 0.002951, shows us that when predicting VV, one standard error = 0.003 units. The multiple R-squared value of 0.8518, indicates that 85.18% of the variation in VV is explained by the predictor variables. Model.2 has some variables that were significant and some that were not. The residual standard error of 0.002669, shows us that when predicting VV, one standard error = 0.0026 units. This is lower than what we saw in Model.1 (Small Model). The multiple R-squared value of 0.8812, indicates that 88.12% of the variation in VV is explained by the predictor variables. The adjusted-R squared value (87.65%) is also better than Model.1 (Small Model) of 84.9%. Additionally, using VIF, we saw no evidence of multicollinearity in both models.

In section 5, we addressed multicollinearity by transforming the predictor variables using principal component analysis by computing the principal components for the return data. We then plotted the loadings for the first two principal components from the principal components analysis. The results showed 5 primary clusters/groupings. KO, PEP, and DPS are all part of the soda industry and are clustered in the top right hand corner (positive correlation). Additionally, companies in the banking industry such as JPM, BAC, WFC, and GS are located close together as well (bottom left hand corner, negative correlation). There also seems to be clusters with companies in the oil field services industry such as SLB, HAL, and BHI (bottom left hand corner, negative correlation). Furthermore, companies in the industrial – chemical industry such as DD, DOW, and HUN are also somewhat close together as well (negative correlation). Companies in the manufacturing industry such as HON and MMM are also clustered together as well (negative correlation). The biggest surprise is that companies in the oil refining industry seem to be spread out. For instance, CVX, and XOM are clustered together, but HES and MPC (outlier) far away from the primary cluster.

In section 6, we used a scree plot and total variance explained plot to determine how many principal components to keep. The scree plot revealed that we should keep the first 8 principal components. Additionally, the total variance explained plot revealed that if we take the first 8 principal components, 80% of the data is explained by the first 8 principal components, whereas if we keep only 1 principal component, only 50% of the total variation would be explained. As a result, this confirmed that we should keep the first 8 principal components because my cut off is 80%, which falls in the 70 – 90% range.

In section 7, we used principal components in predictive modeling by computing and scoring the principal components on the whole data set, and then splitting it into training and test data sets. We then used our train and test data sets to fit a linear regression model using the first eight principal components and computed the Mean Absolute Error (MAE) for that model, while also scoring that model out-of-sample on our test data set and computing the out-of-sample MAE. The results show that components 1 through 3 are significant om model pca1.lm. However, components 4 through 8 are not significant. The residual standard error of 0.002769, shows us that when predicting VV, one standard error = 0.0027 units. The multiple R-squared value of 0.8635, indicates that 86.35% of the variation in VV is explained by the predictor variables. This is almost comparable to the Model.2 (Full Model), but we were able to accomplish this with fewer predictors (8 principle components).

Additionally, the VIF values associated with every predictor variable in the principal components regression model were all one because by definition, principle component scores are uncorrelated. Additionally, the calculation of MAE on the training set for model pca1.lm showed 0.001993. In comparison, the calculation of MAE on the test set showed 0.00212, which was slightly higher than the training set.

In Section 8, we compared the principal component regression model with Model #1 and #2 using MAE to determine the best model. The results revealed that pca1.lm is the best model because it is a simpler model. For instance, we are making a tradeoff – settling for less accuracy/more bias, but more precision. Furthermore, this was accomplished using 8 predictor variables versus 20 predictor variables in model.2 (full). Additionally, the MAE values for the train and test data set and residual standard error (0.0027 vs. 0.0026) were comparable to model.2 and better than model.1 (0.0029). Additionally, the adjusted r-squared value for pca1.lm of 86% was also comparable to model.2 (87.7%) and better than model.1 (84.9%).

Lastly, we used variable selection (backward) to see if it translated to producing the best predictive model. The results revealed that backward.lm is the best model because not only is it a simpler model (less accuracy/more bias, but more precision), but the variable selection was more accurate at choosing the 8 predictor variables than pca1.lm. Additionally, 6 out of the 8 variables had strong statistical significance compared to only 3 out of the 8 for pca1.lm and 12 out of the 19 for model.2. Furthermore, backward.lm also had comparable MAE values for the train and test data set and comparable or better residual standard error (0.0026) compared to pca1.lm (residual standard error: 0.0027) and model.2 (residual standard error: 0.0026) and model.1 (residual standard error: 0.0029). Additionally, the adjusted r-squared value of 87% was also comparable or better than model.2 (87.7%), pca1.lm (86%), and model.1 (84.9%).

Appendix

Section 1

```
> str(my.data)
'data.frame':  502 obs. of  22 variables:
 $ Date: Factor w/ 502 levels "1-Apr-13","1-Aug-12",..: 394 381 321 303 270 253 20
6 173 155 139 ...
 $ AA  : num  10.6 10.5 10.7 10.4 10.4 ...
 $ BAC : num  15.6 15.5 15.7 15.6 15.7 ...
 $ BHI : num  55.3 54.5 54.8 54.6 54 ...
 $ CVX : num  125 124 125 125 124 ...
 $ DD  : num  65 64.7 64.2 64.2 63.8 ...
 $ DOW : num  44.4 44.6 44.6 44.9 44.4 ...
 $ DPS : num  48.7 48.8 48.8 48.6 48.5 ...
 $ GS  : num  177 176 176 176 176 ...
 $ HAL : num  50.8 50.4 51.1 51.2 50.7 ...
 $ HES : num  83 82.6 83.1 82.6 81.1 ...
 $ HON : num  91.4 91 91.1 91.1 90.5 ...
 $ HUN : num  24.6 24.3 24.2 24 24.1 ...
 $ JPM : num  58.5 58 58.1 58.2 58.2 ...
 $ KO  : num  41.3 41.1 40.7 40.5 40.2 ...
 $ MMM : num  140 139 139 138 137 ...
 $ MPC : num  91.7 88.5 89.1 89.5 89.3 ...
 $ PEP : num  82.9 82.9 82.7 82.5 82 ...
 $ SLB : num  90.1 89.2 89.9 89.4 88.3 ...
 $ WFC : num  45.4 45.5 45.5 45.5 45.4 ...
 $ XOM : num  101.2 100.3 101.5 100.9 99.2 ...
 $ VV  : num  84.8 84.4 84.5 84.5 84.1 ...

> head(my.data)
       Date    AA   BAC   BHI    CVX    DD   DOW   DPS     GS   HAL
1 31-Dec-13 10.63 15.57 55.26 124.91 64.97 44.40 48.72 177.26 50.75
2 30-Dec-13 10.53 15.54 54.45 124.23 64.65 44.60 48.84 175.73 50.40
3 27-Dec-13 10.69 15.67 54.77 125.23 64.25 44.60 48.80 176.35 51.08
4 26-Dec-13 10.43 15.65 54.58 124.81 64.25 44.86 48.59 176.45 51.21
5 24-Dec-13 10.36 15.70 54.00 123.51 63.83 44.42 48.50 176.16 50.68
6 23-Dec-13 10.13 15.69 53.64 122.80 62.74 43.88 48.19 176.47 50.30
    HES   HON   HUN   JPM    KO    MMM   MPC   PEP   SLB   WFC    XOM
1 83.00 91.37 24.60 58.48 41.31 140.25 91.73 82.94 90.11 45.40 101.20
2 82.61 91.00 24.33 57.95 41.09 139.42 88.50 82.91 89.17 45.50 100.31
3 83.07 91.14 24.24 58.14 40.66 139.35 89.13 82.71 89.90 45.50 101.51
4 82.60 91.10 24.01 58.20 40.49 138.29 89.54 82.45 89.39 45.54 100.90
5 81.14 90.45 24.14 58.25 40.19 136.99 89.35 82.04 88.31 45.39  99.22
6 80.22 89.72 23.94 58.24 40.16 136.80 88.77 81.86 87.32 45.21  98.51
     VV
1 84.80
2 84.37
3 84.45
4 84.45
5 84.07
6 84.31

> names(my.data)
 [1] "Date" "AA"   "BAC"  "BHI"  "CVX"  "DD"   "DOW"  "DPS"  "GS"
[10] "HAL"  "HES"  "HON"  "HUN"  "JPM"  "KO"   "MMM"  "MPC"  "PEP"
[19] "SLB"  "WFC"  "XOM"  "VV"
```

```
> # Note Date is a string of dd-Mon-yy in R this is '%d-%B-%y';
> head(sorted.df)
         Date   AA  BAC   BHI    CVX    DD   DOW   DPS    GS   HAL
502  3-Jan-12 9.23 5.80 51.02 110.37 46.51 29.79 38.34 95.36 34.15
501  4-Jan-12 9.45 5.81 51.53 110.18 47.02 29.95 38.55 94.74 35.12
500  5-Jan-12 9.36 6.31 50.82 109.10 46.70 30.14 38.79 94.58 34.56
499  6-Jan-12 9.16 6.18 51.26 108.31 46.04 30.32 38.52 93.42 34.98
498  9-Jan-12 9.42 6.27 51.58 109.49 46.43 30.31 38.52 94.69 35.38
497 10-Jan-12 9.44 6.63 51.95 109.06 47.14 30.98 38.62 98.33 36.33
      HES   HON   HUN   JPM    KO   MMM   MPC   PEP   SLB   WFC   XOM
502 58.40 55.58  9.95 34.98 35.07 83.49 33.41 66.40 70.09 28.43 86.00
501 59.00 55.53  9.87 34.95 34.85 84.18 33.76 66.74 69.56 28.56 86.02
500 57.60 55.59  9.82 35.68 34.68 83.80 31.92 66.22 68.07 29.02 85.76
499 56.42 55.18  9.90 35.36 34.46 83.37 31.66 65.39 67.78 28.94 85.12
498 56.90 55.64  9.84 35.30 34.46 83.87 30.96 65.73 68.82 29.30 85.50
497 58.56 56.58 10.20 36.05 34.67 84.30 31.83 65.66 70.75 29.41 85.72
       VV      RDate
502 58.18 2012-01-03
501 58.25 2012-01-04
500 58.44 2012-01-05
499 58.32 2012-01-06
498 58.45 2012-01-09
497 58.99 2012-01-10

> # Type cast the array as a data frame;
> str(returns.df)
'data.frame':  501 obs. of  21 variables:
 $ AA : num  0.02356 -0.00957 -0.0216 0.02799 0.00212 ...
 $ BAC: num  0.00172 0.08256 -0.02082 0.01446 0.05583 ...
 $ BHI: num  0.00995 -0.01387 0.00862 0.00622 0.00715 ...
 $ CVX: num  -0.00172 -0.00985 -0.00727 0.01084 -0.00394 ...
 $ DD : num  0.01091 -0.00683 -0.01423 0.00844 0.01518 ...
 $ DOW: num  0.00536 0.00632 0.00595 -0.00033 0.02186 ...
 $ DPS: num  0.00546 0.00621 -0.00698 0 0.00259 ...
 $ GS : num  -0.00652 -0.00169 -0.01234 0.0135 0.03772 ...
 $ HAL: num  0.028 -0.0161 0.0121 0.0114 0.0265 ...
 $ HES: num  0.01022 -0.02401 -0.0207 0.00847 0.02876 ...
 $ HON: num  -0.0009 0.00108 -0.0074 0.0083 0.01675 ...
 $ HUN: num  -0.00807 -0.00508 0.00811 -0.00608 0.03593 ...
 $ JPM: num  -0.000858 0.020672 -0.009009 -0.001698 0.021024 ...
 $ KO : num  -0.00629 -0.00489 -0.00636 0 0.00608 ...
 $ MMM: num  0.00823 -0.00452 -0.00514 0.00598 0.00511 ...
 $ MPC: num  0.01042 -0.05604 -0.00818 -0.02236 0.02771 ...
 $ PEP: num  0.00511 -0.00782 -0.01261 0.00519 -0.00107 ...
 $ SLB: num  -0.00759 -0.02165 -0.00427 0.01523 0.02766 ...
 $ WFC: num  0.00456 0.01598 -0.00276 0.01236 0.00375 ...
 $ XOM: num  0.000233 -0.003027 -0.007491 0.004454 0.00257 ...
 $ VV : num  0.0012 0.00326 -0.00206 0.00223 0.0092 ...
```

Section 7

```
> pca1.test
            1             3             6            12            14
 0.0026323511 -0.0058588322 -0.0004876966  0.0040271226 -0.0002045138
           15            17            18            28            31
 0.0056041327  0.0009025588 -0.0045568417  0.0054776637  0.0109180792
           38            43            44            48            49
 0.0010593534 -0.0168056950  0.0064530429  0.0200978371 -0.0024585177
           50            52            53            56            57
 0.0076589297  0.0029010334 -0.0024280202  0.0053406670  0.0105258449
           59            63            68            74            76
-0.0035158291 -0.0066484106  0.0080640353 -0.0046462809 -0.0056336968
           81            83            89            96            97
-0.0015069767 -0.0040586007  0.0035060995  0.0107347040  0.0017193297
           98           101           102           104           105
 0.0044385311  0.0107472570 -0.0168396932 -0.0211313164 -0.0032726799
          106           107           109           112           114
 0.0042472248  0.0205302903  0.0069517165 -0.0054169023  0.0107603995
          129           132           134           139           146
-0.0046658942 -0.0073117497 -0.0030901015 -0.0046488602  0.0023419604
          148           175           176           177           178
 0.0171465195 -0.0009668418  0.0178686700  0.0075478079 -0.0063140595
          180           185           191           194           198
-0.0009834671 -0.0036565550  0.0079543069 -0.0061194493  0.0055666178
          204           208           210           215           218
-0.0253467923 -0.0009758065 -0.0070239790  0.0002819542 -0.0131929237
          219           221           224           225           229
 0.0001332164  0.0176427261  0.0122142231 -0.0041484951 -0.0002795313
          230           235           236           239           243
-0.0078185083  0.0008792274  0.0058717338 -0.0009428811  0.0077572661
          244           245           249           251           255
-0.0083862686 -0.0023688597  0.0109558207 -0.0008775223  0.0026748675
          256           276           278           281           285
 0.0062409782  0.0032642129  0.0011861885 -0.0006036935  0.0086874172
          286           287           290           293           299
-0.0170028697  0.0057213476  0.0011582739  0.0047664296  0.0069437776
          300           303           304           307           308
 0.0022243580  0.0042785742 -0.0065349367  0.0050981319 -0.0033477539
          314           316           317           318           319
-0.0042552379  0.0042611095  0.0082254934  0.0030382501 -0.0049046769
          325           328           334           336           342
 0.0064306356  0.0024692288  0.0094114093  0.0069899804  0.0056145034
          347           348           351           357           359
-0.0063988592 -0.0044966453 -0.0046935204  0.0053248183  0.0002644908
          363           365           366           368           371
-0.0074491997  0.0055601180 -0.0105926970  0.0033440698  0.0084555758
          375           381           387           389           397
-0.0031034668  0.0090100702  0.0043948891  0.0014250692  0.0008572295
          400           401           402           404           405
-0.0022224424  0.0020791116 -0.0003327937  0.0020478372 -0.0035728599
          414           418           419           422           428
-0.0115311742  0.0033346799  0.0065928129  0.0079143851  0.0026243349
          429           431           433           434           439
 0.0108801302 -0.0056779540 -0.0033460012  0.0006935554 -0.0009788894
          441           443           446           451           453
 0.0068159357 -0.0073221792  0.0047869275  0.0016715219  0.0073099989
```

```
          454           456           464           465           468
-0.0029486120  0.0026202915  0.0078729613 -0.0106248653 -0.0033222936
          472           475           480           481           485
-0.0010901449  0.0089686116 -0.0011847084 -0.0037440446  0.0115596894
          494           498
 0.0014656477  0.0057471166
```

Section 9

```
> backward.lm <- stepAIC(full.lm,direction=c('backward'))
Start:  AIC=-4168.46
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.6 + Comp.7 +
    Comp.8 + Comp.9 + Comp.10 + Comp.11 + Comp.12 + Comp.13 +
    Comp.14 + Comp.15 + Comp.16 + Comp.17 + Comp.18 + Comp.19 +
    Comp.20 + u

          Df Sum of Sq       RSS     AIC
- Comp.20  1 0.0000000 0.0024047 -4170.5
- Comp.6   1 0.0000001 0.0024048 -4170.5
- Comp.18  1 0.0000003 0.0024051 -4170.4
- Comp.12  1 0.0000006 0.0024054 -4170.4
- Comp.7   1 0.0000024 0.0024072 -4170.1
- Comp.13  1 0.0000027 0.0024074 -4170.1
- Comp.15  1 0.0000037 0.0024084 -4169.9
- Comp.19  1 0.0000040 0.0024087 -4169.9
- Comp.4   1 0.0000052 0.0024100 -4169.7
- Comp.17  1 0.0000054 0.0024101 -4169.7
- u        1 0.0000066 0.0024113 -4169.5
- Comp.16  1 0.0000073 0.0024120 -4169.4
- Comp.14  1 0.0000079 0.0024127 -4169.3
<none>                 0.0024047 -4168.5
- Comp.5   1 0.0000177 0.0024224 -4167.9
- Comp.8   1 0.0000191 0.0024238 -4167.7
- Comp.9   1 0.0000410 0.0024457 -4164.5
- Comp.10  1 0.0000704 0.0024751 -4160.2
- Comp.11  1 0.0000985 0.0025032 -4156.3
- Comp.2   1 0.0001016 0.0025063 -4155.8
- Comp.3   1 0.0001439 0.0025487 -4149.9
- Comp.1   1 0.0159005 0.0183052 -3451.9

Step:  AIC=-4170.46
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.6 + Comp.7 +
    Comp.8 + Comp.9 + Comp.10 + Comp.11 + Comp.12 + Comp.13 +
    Comp.14 + Comp.15 + Comp.16 + Comp.17 + Comp.18 + Comp.19 +
    u

          Df Sum of Sq       RSS     AIC
- Comp.6   1 0.0000001 0.0024048 -4172.5
- Comp.18  1 0.0000003 0.0024051 -4172.4
- Comp.12  1 0.0000006 0.0024054 -4172.4
- Comp.7   1 0.0000024 0.0024072 -4172.1
- Comp.13  1 0.0000027 0.0024074 -4172.1
- Comp.15  1 0.0000037 0.0024085 -4171.9
- Comp.19  1 0.0000039 0.0024087 -4171.9
- Comp.4   1 0.0000052 0.0024100 -4171.7
- Comp.17  1 0.0000054 0.0024101 -4171.7
```

```
- u        1 0.0000066 0.0024113 -4171.5
- Comp.16  1 0.0000073 0.0024121 -4171.4
- Comp.14  1 0.0000080 0.0024127 -4171.3
<none>                0.0024047 -4170.5
- Comp.5   1 0.0000177 0.0024224 -4169.9
- Comp.8   1 0.0000191 0.0024238 -4169.7
- Comp.9   1 0.0000410 0.0024457 -4166.5
- Comp.10  1 0.0000704 0.0024752 -4162.2
- Comp.11  1 0.0000994 0.0025042 -4158.1
- Comp.2   1 0.0001016 0.0025063 -4157.8
- Comp.3   1 0.0001448 0.0025495 -4151.8
- Comp.1   1 0.0159010 0.0183058 -3453.9

Step:  AIC=-4172.45
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.7 + Comp.8 +
    Comp.9 + Comp.10 + Comp.11 + Comp.12 + Comp.13 + Comp.14 +
    Comp.15 + Comp.16 + Comp.17 + Comp.18 + Comp.19 + u

          Df Sum of Sq       RSS     AIC
- Comp.18  1 0.0000003 0.0024051 -4174.4
- Comp.12  1 0.0000006 0.0024054 -4174.4
- Comp.7   1 0.0000024 0.0024072 -4174.1
- Comp.13  1 0.0000027 0.0024074 -4174.1
- Comp.15  1 0.0000037 0.0024085 -4173.9
- Comp.19  1 0.0000039 0.0024087 -4173.9
- Comp.4   1 0.0000052 0.0024100 -4173.7
- Comp.17  1 0.0000054 0.0024102 -4173.7
- u        1 0.0000066 0.0024113 -4173.5
- Comp.16  1 0.0000074 0.0024122 -4173.4
- Comp.14  1 0.0000080 0.0024128 -4173.3
<none>                0.0024048 -4172.5
- Comp.5   1 0.0000177 0.0024225 -4171.9
- Comp.8   1 0.0000190 0.0024238 -4171.7
- Comp.9   1 0.0000409 0.0024457 -4168.5
- Comp.10  1 0.0000705 0.0024753 -4164.2
- Comp.11  1 0.0000998 0.0025046 -4160.1
- Comp.2   1 0.0001015 0.0025063 -4159.8
- Comp.3   1 0.0001447 0.0025495 -4153.8
- Comp.1   1 0.0159848 0.0183896 -3454.3

Step:  AIC=-4174.41
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.7 + Comp.8 +
    Comp.9 + Comp.10 + Comp.11 + Comp.12 + Comp.13 + Comp.14 +
    Comp.15 + Comp.16 + Comp.17 + Comp.19 + u

          Df Sum of Sq       RSS     AIC
- Comp.12  1 0.0000006 0.0024057 -4176.3
- Comp.7   1 0.0000023 0.0024074 -4176.1
- Comp.13  1 0.0000026 0.0024077 -4176.0
- Comp.19  1 0.0000038 0.0024089 -4175.8
- Comp.15  1 0.0000038 0.0024089 -4175.8
- Comp.4   1 0.0000053 0.0024104 -4175.6
- Comp.17  1 0.0000054 0.0024105 -4175.6
- u        1 0.0000067 0.0024118 -4175.4
- Comp.16  1 0.0000074 0.0024125 -4175.3
- Comp.14  1 0.0000083 0.0024134 -4175.2
<none>                0.0024051 -4174.4
```

```
- Comp.5   1 0.0000177 0.0024228 -4173.8
- Comp.8   1 0.0000191 0.0024242 -4173.6
- Comp.9   1 0.0000408 0.0024459 -4170.5
- Comp.10  1 0.0000703 0.0024754 -4166.2
- Comp.11  1 0.0001000 0.0025051 -4162.0
- Comp.2   1 0.0001019 0.0025070 -4161.7
- Comp.3   1 0.0001446 0.0025497 -4155.7
- Comp.1   1 0.0160036 0.0184088 -3455.9

Step:  AIC=-4176.32
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.7 + Comp.8 +
    Comp.9 + Comp.10 + Comp.11 + Comp.13 + Comp.14 + Comp.15 +
    Comp.16 + Comp.17 + Comp.19 + u

          Df Sum of Sq       RSS     AIC
- Comp.7   1 0.0000022 0.0024079 -4178.0
- Comp.13  1 0.0000024 0.0024081 -4178.0
- Comp.15  1 0.0000037 0.0024094 -4177.8
- Comp.19  1 0.0000038 0.0024095 -4177.8
- Comp.4   1 0.0000051 0.0024109 -4177.6
- Comp.17  1 0.0000055 0.0024113 -4177.5
- u        1 0.0000067 0.0024124 -4177.3
- Comp.16  1 0.0000073 0.0024130 -4177.3
- Comp.14  1 0.0000084 0.0024141 -4177.1
<none>                 0.0024057 -4176.3
- Comp.5   1 0.0000175 0.0024232 -4175.8
- Comp.8   1 0.0000185 0.0024242 -4175.6
- Comp.9   1 0.0000408 0.0024465 -4172.4
- Comp.10  1 0.0000706 0.0024763 -4168.1
- Comp.11  1 0.0001007 0.0025064 -4163.8
- Comp.2   1 0.0001015 0.0025072 -4163.7
- Comp.3   1 0.0001450 0.0025507 -4157.6
- Comp.1   1 0.0160083 0.0184140 -3457.8

Step:  AIC=-4177.99
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.8 + Comp.9 +
    Comp.10 + Comp.11 + Comp.13 + Comp.14 + Comp.15 + Comp.16 +
    Comp.17 + Comp.19 + u

          Df Sum of Sq       RSS     AIC
- Comp.13  1 0.0000025 0.0024104 -4179.6
- Comp.19  1 0.0000035 0.0024114 -4179.5
- Comp.15  1 0.0000037 0.0024116 -4179.4
- Comp.17  1 0.0000050 0.0024130 -4179.3
- Comp.4   1 0.0000052 0.0024132 -4179.2
- u        1 0.0000065 0.0024145 -4179.0
- Comp.16  1 0.0000072 0.0024151 -4178.9
- Comp.14  1 0.0000085 0.0024165 -4178.7
<none>                 0.0024079 -4178.0
- Comp.8   1 0.0000177 0.0024257 -4177.4
- Comp.5   1 0.0000179 0.0024258 -4177.4
- Comp.9   1 0.0000419 0.0024499 -4173.9
- Comp.10  1 0.0000702 0.0024782 -4169.8
- Comp.11  1 0.0001004 0.0025083 -4165.5
- Comp.2   1 0.0001027 0.0025107 -4165.2
- Comp.3   1 0.0001465 0.0025544 -4159.1
- Comp.1   1 0.0160062 0.0184141 -3459.8
```

```
Step:  AIC=-4179.63
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.8 + Comp.9 +
    Comp.10 + Comp.11 + Comp.14 + Comp.15 + Comp.16 + Comp.17 +
    Comp.19 + u

          Df Sum of Sq       RSS     AIC
- Comp.15  1 0.0000033 0.0024137 -4181.1
- Comp.19  1 0.0000034 0.0024138 -4181.1
- Comp.17  1 0.0000051 0.0024155 -4180.9
- Comp.4   1 0.0000052 0.0024156 -4180.9
- u        1 0.0000066 0.0024171 -4180.7
- Comp.16  1 0.0000073 0.0024177 -4180.6
- Comp.14  1 0.0000080 0.0024184 -4180.5
<none>                 0.0024104 -4179.6
- Comp.8   1 0.0000178 0.0024282 -4179.0
- Comp.5   1 0.0000182 0.0024286 -4179.0
- Comp.9   1 0.0000419 0.0024523 -4175.5
- Comp.10  1 0.0000698 0.0024802 -4171.5
- Comp.2   1 0.0001019 0.0025123 -4167.0
- Comp.11  1 0.0001020 0.0025124 -4167.0
- Comp.3   1 0.0001482 0.0025586 -4160.5
- Comp.1   1 0.0160494 0.0184598 -3461.0

Step:  AIC=-4181.14
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.8 + Comp.9 +
    Comp.10 + Comp.11 + Comp.14 + Comp.16 + Comp.17 + Comp.19 +
    u

          Df Sum of Sq       RSS     AIC
- Comp.19  1 0.0000033 0.0024171 -4182.7
- Comp.4   1 0.0000052 0.0024189 -4182.4
- Comp.17  1 0.0000052 0.0024190 -4182.4
- u        1 0.0000066 0.0024203 -4182.2
- Comp.16  1 0.0000073 0.0024210 -4182.1
- Comp.14  1 0.0000078 0.0024215 -4182.0
<none>                 0.0024137 -4181.1
- Comp.8   1 0.0000167 0.0024304 -4180.7
- Comp.5   1 0.0000186 0.0024323 -4180.4
- Comp.9   1 0.0000419 0.0024556 -4177.1
- Comp.10  1 0.0000695 0.0024832 -4173.1
- Comp.2   1 0.0001008 0.0025146 -4168.7
- Comp.11  1 0.0001025 0.0025162 -4168.4
- Comp.3   1 0.0001474 0.0025611 -4162.2
- Comp.1   1 0.0160624 0.0184761 -3462.6

Step:  AIC=-4182.65
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.8 + Comp.9 +
    Comp.10 + Comp.11 + Comp.14 + Comp.16 + Comp.17 + u

          Df Sum of Sq       RSS     AIC
- Comp.17  1 0.0000051 0.0024222 -4183.9
- Comp.4   1 0.0000052 0.0024222 -4183.9
- u        1 0.0000069 0.0024239 -4183.7
- Comp.16  1 0.0000071 0.0024241 -4183.6
- Comp.14  1 0.0000082 0.0024252 -4183.5
<none>                 0.0024171 -4182.7
```

```
- Comp.8   1 0.0000163 0.0024334 -4182.3
- Comp.5   1 0.0000180 0.0024350 -4182.0
- Comp.9   1 0.0000409 0.0024580 -4178.7
- Comp.10  1 0.0000702 0.0024873 -4174.5
- Comp.2   1 0.0001022 0.0025193 -4170.0
- Comp.11  1 0.0001034 0.0025204 -4169.8
- Comp.3   1 0.0001472 0.0025642 -4163.7
- Comp.1   1 0.0160881 0.0185052 -3464.1

Step:  AIC=-4183.91
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.4 + Comp.5 + Comp.8 + Comp.9 +
    Comp.10 + Comp.11 + Comp.14 + Comp.16 + u

          Df Sum of Sq       RSS     AIC
- Comp.4   1 0.0000049 0.0024270 -4185.2
- u        1 0.0000058 0.0024280 -4185.1
- Comp.16  1 0.0000074 0.0024296 -4184.8
- Comp.14  1 0.0000083 0.0024304 -4184.7
<none>                 0.0024222 -4183.9
- Comp.8   1 0.0000159 0.0024380 -4183.6
- Comp.5   1 0.0000191 0.0024413 -4183.1
- Comp.9   1 0.0000409 0.0024631 -4180.0
- Comp.10  1 0.0000702 0.0024924 -4175.8
- Comp.11  1 0.0001036 0.0025258 -4171.1
- Comp.2   1 0.0001039 0.0025260 -4171.0
- Comp.3   1 0.0001465 0.0025687 -4165.1
- Comp.1   1 0.0160882 0.0185104 -3466.0

Step:  AIC=-4185.19
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 + Comp.9 + Comp.10 +
    Comp.11 + Comp.14 + Comp.16 + u

          Df Sum of Sq       RSS     AIC
- u        1 0.0000054 0.0024324 -4186.4
- Comp.16  1 0.0000079 0.0024350 -4186.0
- Comp.14  1 0.0000091 0.0024362 -4185.9
<none>                 0.0024270 -4185.2
- Comp.8   1 0.0000161 0.0024431 -4184.9
- Comp.5   1 0.0000193 0.0024463 -4184.4
- Comp.9   1 0.0000417 0.0024687 -4181.2
- Comp.10  1 0.0000722 0.0024992 -4176.8
- Comp.11  1 0.0001039 0.0025310 -4172.3
- Comp.2   1 0.0001046 0.0025316 -4172.3
- Comp.3   1 0.0001481 0.0025751 -4166.2
- Comp.1   1 0.0161545 0.0185815 -3466.6

Step:  AIC=-4186.41
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 + Comp.9 + Comp.10 +
    Comp.11 + Comp.14 + Comp.16

          Df Sum of Sq       RSS     AIC
- Comp.14  1 0.0000082 0.0024407 -4187.2
- Comp.16  1 0.0000089 0.0024414 -4187.1
<none>                 0.0024324 -4186.4
- Comp.8   1 0.0000154 0.0024479 -4186.2
- Comp.5   1 0.0000192 0.0024516 -4185.6
- Comp.9   1 0.0000439 0.0024764 -4182.1
```

```
- Comp.10  1 0.0000719 0.0025043 -4178.1
- Comp.2   1 0.0001015 0.0025339 -4173.9
- Comp.11  1 0.0001036 0.0025360 -4173.6
- Comp.3   1 0.0001462 0.0025786 -4167.8
- Comp.1   1 0.0162418 0.0186742 -3466.9

Step:  AIC=-4187.21
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 + Comp.9 + Comp.10 +
    Comp.11 + Comp.16

          Df Sum of Sq       RSS     AIC
- Comp.16  1 0.0000097 0.0024503 -4187.8
<none>                 0.0024407 -4187.2
- Comp.8   1 0.0000142 0.0024548 -4187.2
- Comp.5   1 0.0000207 0.0024613 -4186.2
- Comp.9   1 0.0000461 0.0024868 -4182.6
- Comp.10  1 0.0000694 0.0025100 -4179.3
- Comp.2   1 0.0001018 0.0025425 -4174.7
- Comp.11  1 0.0001021 0.0025428 -4174.7
- Comp.3   1 0.0001470 0.0025877 -4168.5
- Comp.1   1 0.0163719 0.0188125 -3466.3

Step:  AIC=-4187.81
VV ~ Comp.1 + Comp.2 + Comp.3 + Comp.5 + Comp.8 + Comp.9 + Comp.10 +
    Comp.11

          Df Sum of Sq       RSS     AIC
<none>                 0.0024503 -4187.8
- Comp.8   1 0.0000146 0.0024649 -4187.7
- Comp.5   1 0.0000203 0.0024706 -4186.9
- Comp.9   1 0.0000453 0.0024956 -4183.3
- Comp.10  1 0.0000666 0.0025170 -4180.3
- Comp.11  1 0.0000987 0.0025490 -4175.8
- Comp.2   1 0.0001055 0.0025558 -4174.9
- Comp.3   1 0.0001471 0.0025974 -4169.2
- Comp.1   1 0.0164046 0.0188549 -3467.5
```