Forecasting Market Direction with Sentiment Indices

by

David A. Mascio

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

EDHEC Business School

May 2018

Dissertation Committee

Frank J. Fabozzi, PhD

Abraham Lioui, PhD

Turan G. Bali, PhD

EDHEC BUSINESS SCHOOL

APPROVAL FORM

CANDIDATE'S NAME: David Mascio

TITLE OF DISSERTATION: "Forecasting Market Direction with Sentiment Indices"

APPROVAL:

Abraham Lioui		*Chair*
Turan G. Bali	Turan Bali	*External Examiner*
Frank Fabozzi		*Member*

☐ with minor revisions

☐ with major revisions

DATE: May 17, 2018

Abstract

Successful market timing strategies depend on superior forecasting ability and the accuracy of market forecasts. We use six predictive models to forecast the S&P 500 Index (SPX) consisting of investor sentiment, current business conditions, economic policy uncertainty, market dislocation information, credit spreads, and financial uncertainty. These indices are combined to create two additional forecast models, a "kitchen sink logistic regression" and a "least absolute square shrinkage and selection operator." Each model and the combined models are used in a logistic regression analysis to predict the one-month ahead returns of the SPX. In order to determine how successful each strategy is at forecasting the market direction; each prediction is used to adjust the beta of the portfolio. "Beta optimization" refers to a strategy designed to create a portfolio with a beta of 1.0 when the market is expected to go up, and a beta of -1.0 when a bear market is expected. Successful application of this strategy generates returns that are consistent with a call option or an option straddle position; that is, positive returns are generated in both up and down markets. Analysis reveals that the models' forecasts have discriminatory power in identifying substantial market movements, particularly during the bursting of the tech bubble and the financial crisis. We determine the individual forecast indices and the combined portfolios consistently have higher annual returns and lower monthly drawdowns than the buy-and-hold SPX portfolio, and the benchmark index, Baker and Wurgler (2004) Value-Weighted Dividend Premium (VWDP) model.

Contents

Acknowledgments

This work would not have been possible without the professional and academic mentorship of the following four people. First, my dear friend and mentor, Dr. J. Kent Zumwalt. I will forever be grateful for his wisdom, guidence and motivation throughout this endeavor. Second, I am fortunate to have had the priviledge to work with Dr. Frank J. Fabozzi as my dissertation advisor, whose experience and insight was invaluable. Third, Dr. Abraham Lioui for challenging me to accomplish goals that I never thought were possbile. Lastly, Dr. Hong Miao, who posesses one of the best quantitative minds in finance; I am so appreciative of everything you have taught me.

Nobody has been more important to me during this endeavor than the members of my family. I would like to thank my parents, Jack and Barbara Mascio, whose love and encourgement are deep within me in whatever I decide to pursue. Most importantly, I am so thankful for my loving and supportive wife, Monica, and my five beautiful children, Marissa, Anthony, Dominic, Olivia, and Dante, who provide inspiration each and everyday of my life.

Finally, none of these of accomplishments would take place without the grace given by my Lord and Savior, Jesus Christ.

List of Figures

List of Tables

Figure 1: Perfectly Forecasted Returns of the S&P 500 Index

This figure represents the returns of perfectly forecasting the monthly returns of the S&P 500 Index from 01/1973 - 04/2014. The SPXLS portfolio represents perfect forecasting when the markets is DOWN by shorting the S&P 500 Index, and going long when the S&P 500 Index is UP. The SPXRF portfolio represents perfect forecasting when the markets is DOWN by investing in the risk-free asset (RF), and investing long in the SPX when the S&P 500 Index is UP. The SPX portfolio represents the returns on the S&P 500 Index, and the RF portfolio represents the US 3-month treasury bill.

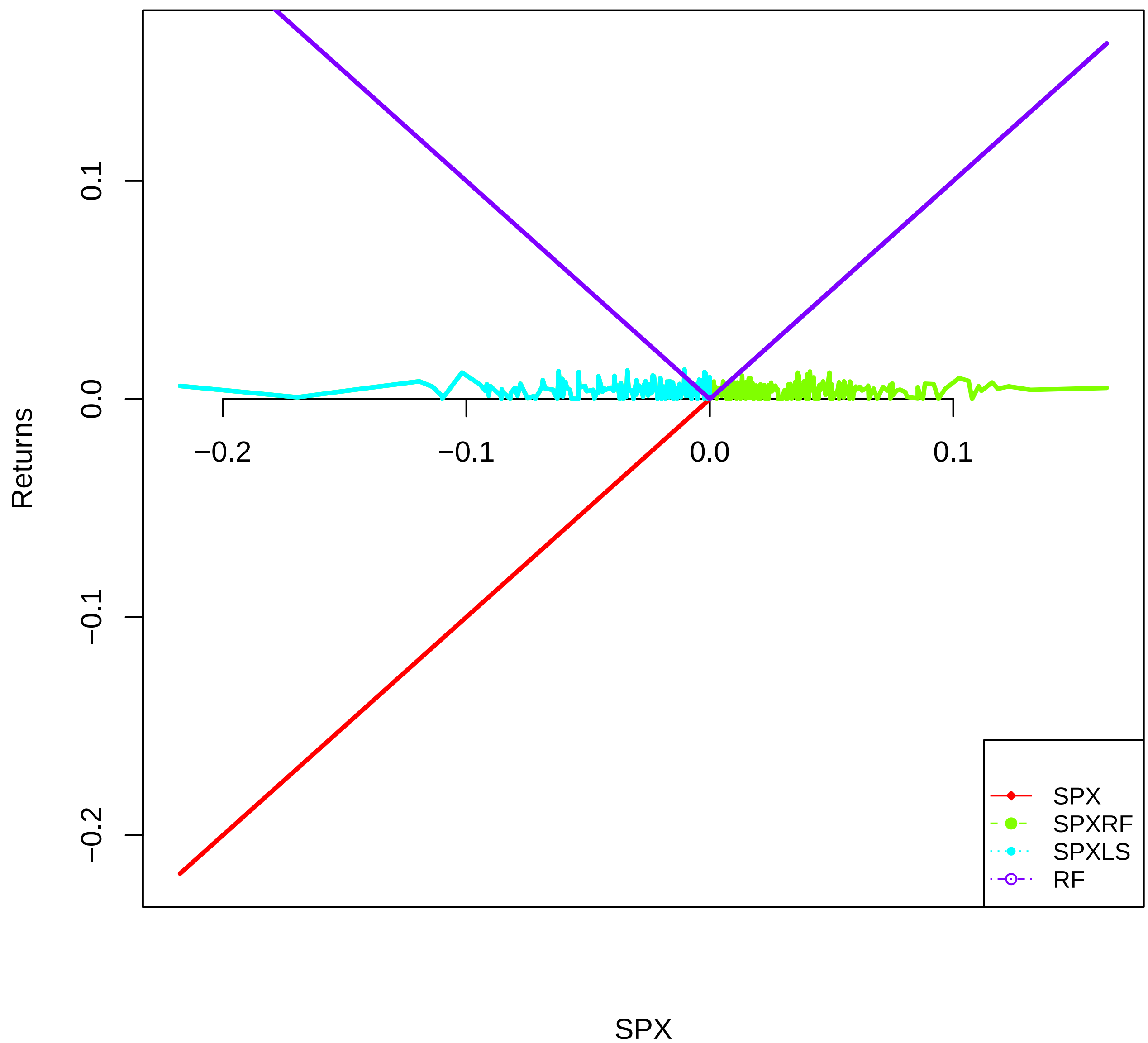

Figure 2: Box-Plot of the Probability Estimates

Figure 3 illustrates the probability estimates mean, median and range of the five forecasting model portfolios, and the model selection portfolios. The five forecasting model portfolios are represented as follows: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The benchmark model is Baker and Wurgler Value-Weighted Dividend Premium (VWDP).

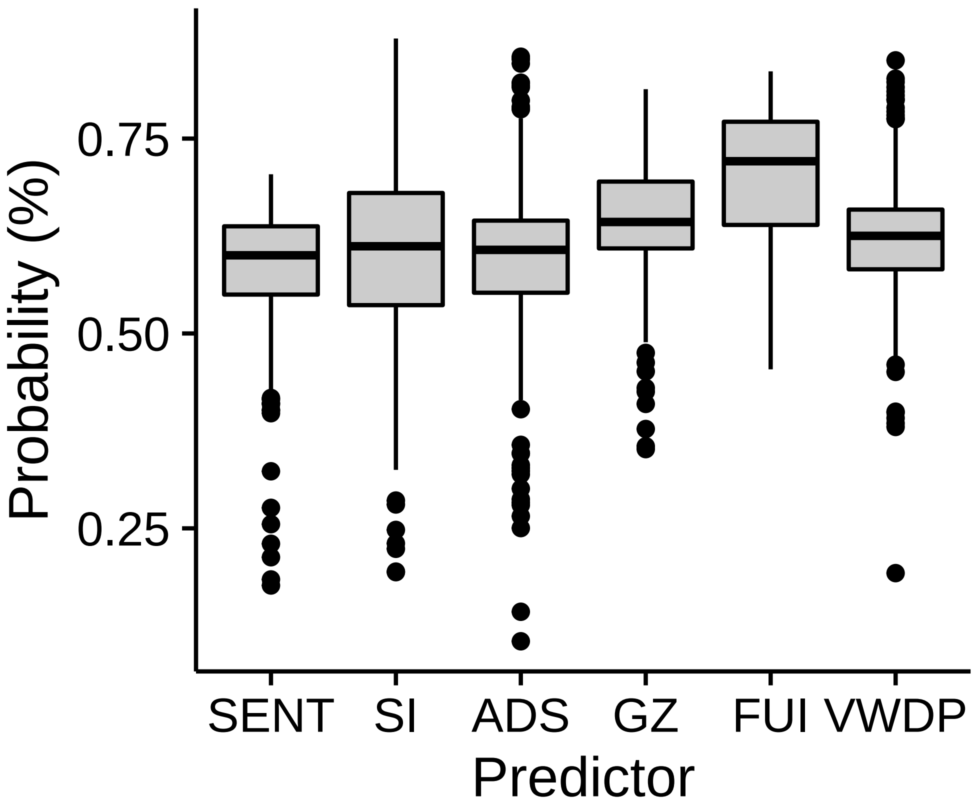

Figure 3: Distribution of the Probability Estimates

This figure represents the probability estimates density of the five forecast model portfolios, and and the benchmark index from 01/1983 - 04/2014. (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark is index is represented by (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All estimates are calculated using a 120-month rolling window.

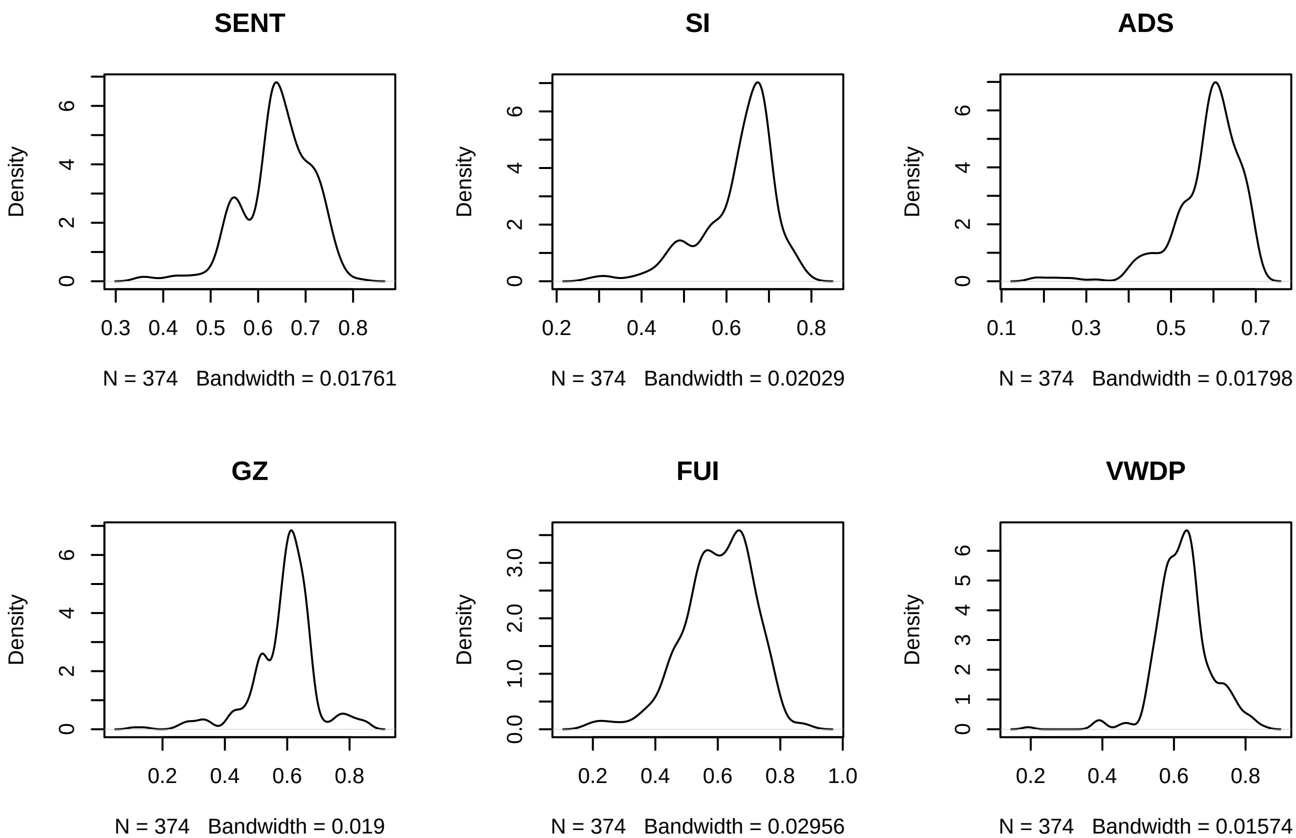

Figure 4: Normality Test of the Probability Estimates

This figure represents the results of the Shapiro normality tests using ggplots on the probability estimates of the five forecast model portfolios, and and the benchmark index from 01/1983 - 04/2014. (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark is index is represented by (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All estimates are calculated using a 120-month rolling window.

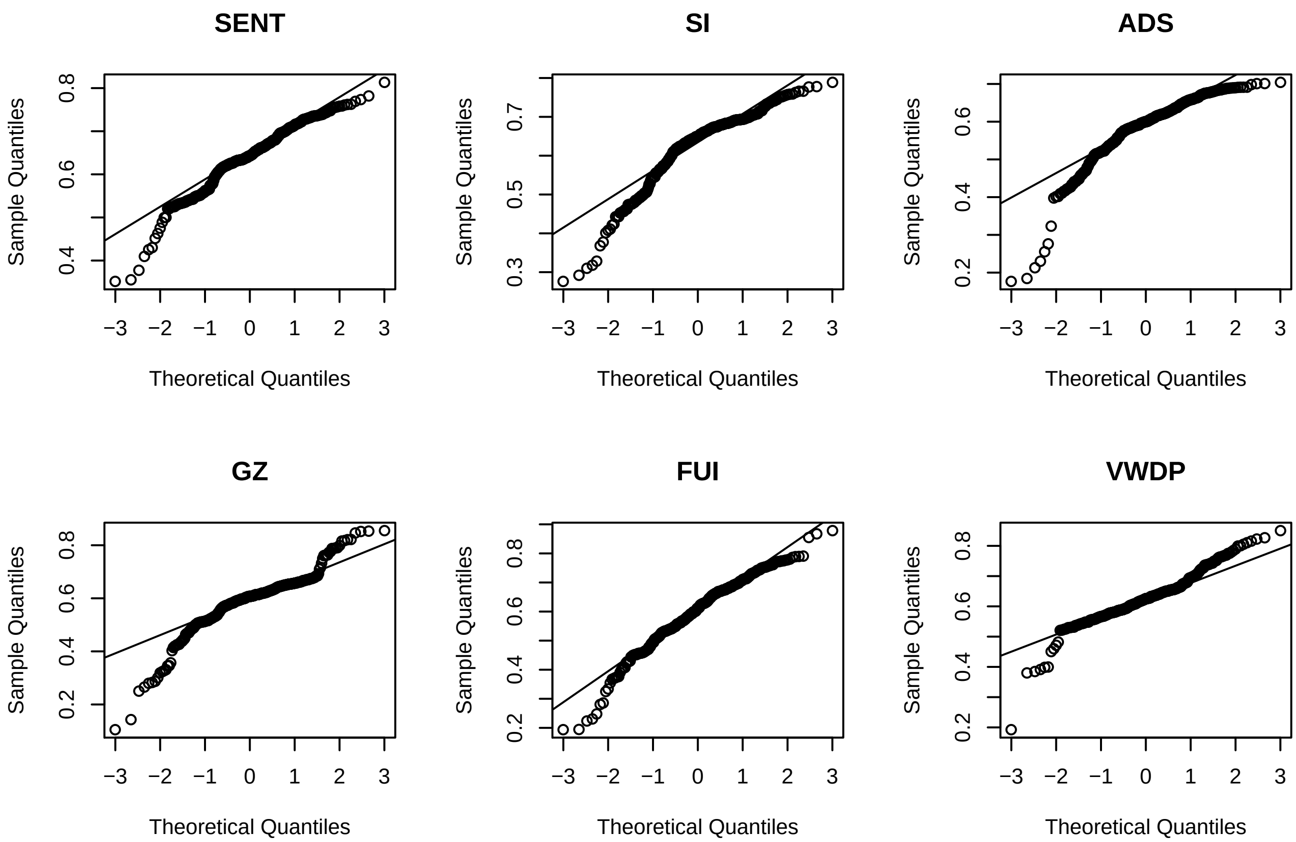

Figure 5: Wealth Index of the Five Model Portfolios vs. S&P 500 Index

This figure represents the cumulative wealth index of monthly returns of the S&P 500 Index (SPX) and the five forecast Model Portfolios from 01/1983 - 04/2014. (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable returns are represented by (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All returns are calculated using a 120 month rolling estimation window.

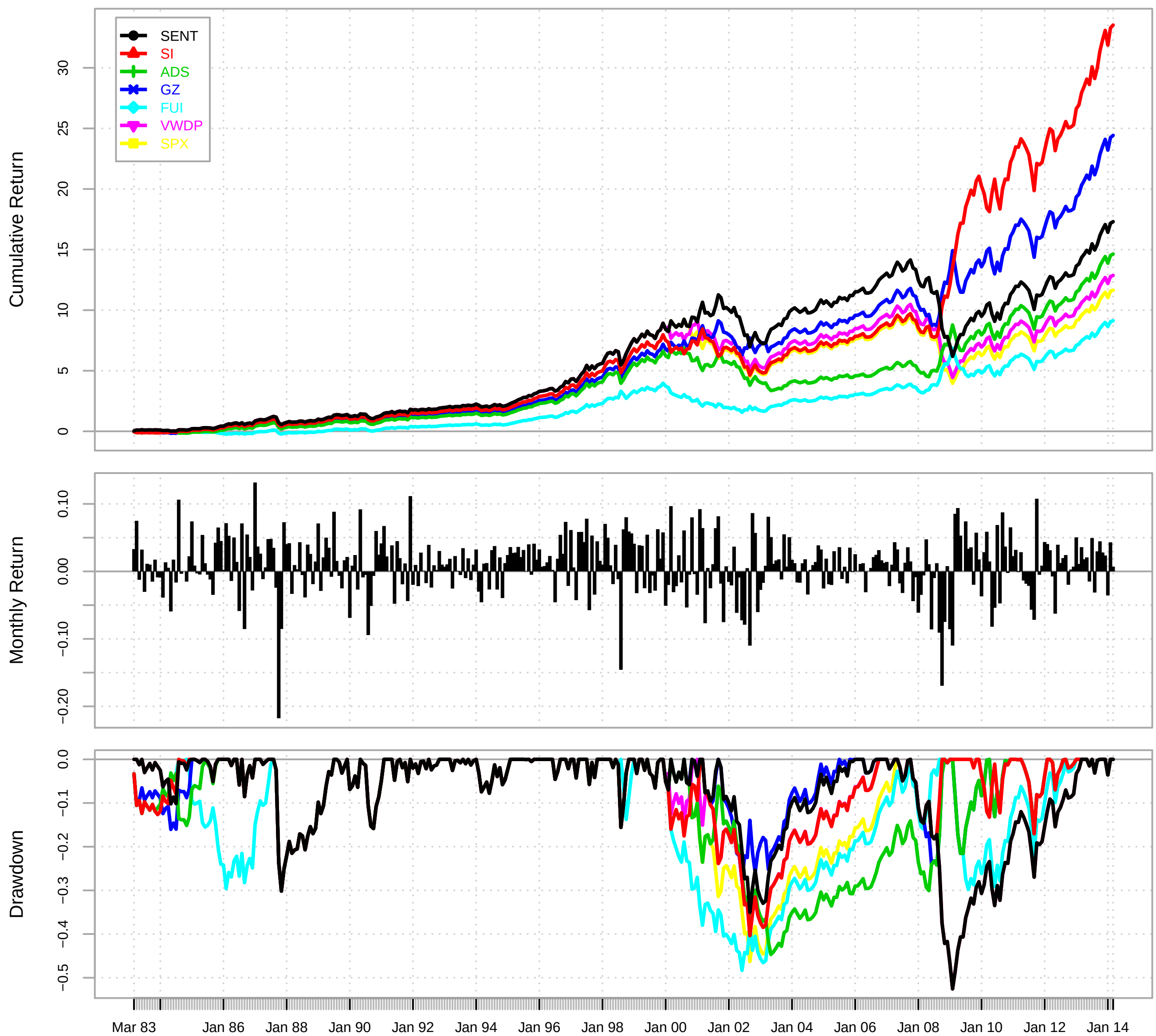

Figure 6: Probability of an UP Market (SENT)

This figure illustrates between 01/1983 - 04/2014 the monthly probability from the (SENT) Baker and Wurgler Sentiment Index forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. The gold colored line represents the statistical significance (p-value) of the monthly forecast. Each probability is based on a 120-month rolling estimation window.

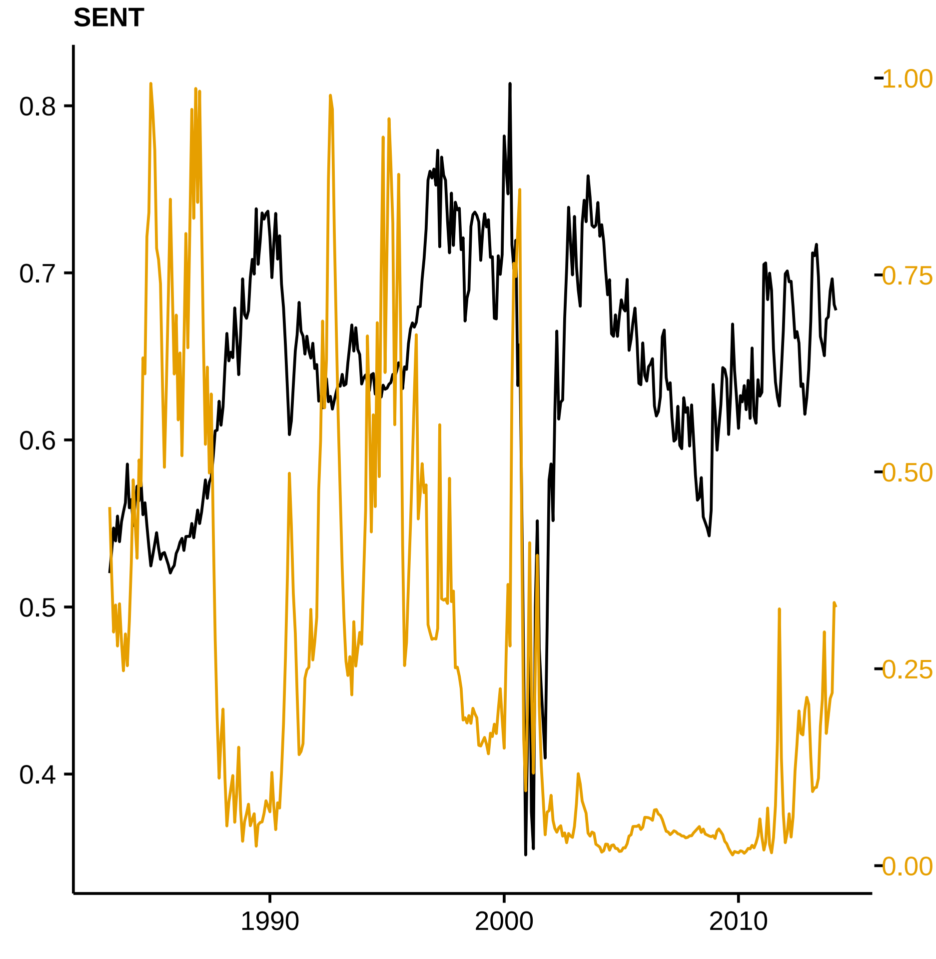

Figure 7: Probability of an UP Market (SI)

This figure illustrates between 01/1983 - 04/2014 the monthly probability from the (SI) Huang, Jiang, and Tu Sentiment Index forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. The gold colored line represents the statistical significance (p-value) of the monthly forecast. Each probability is based on a 120-month rolling estimation window.

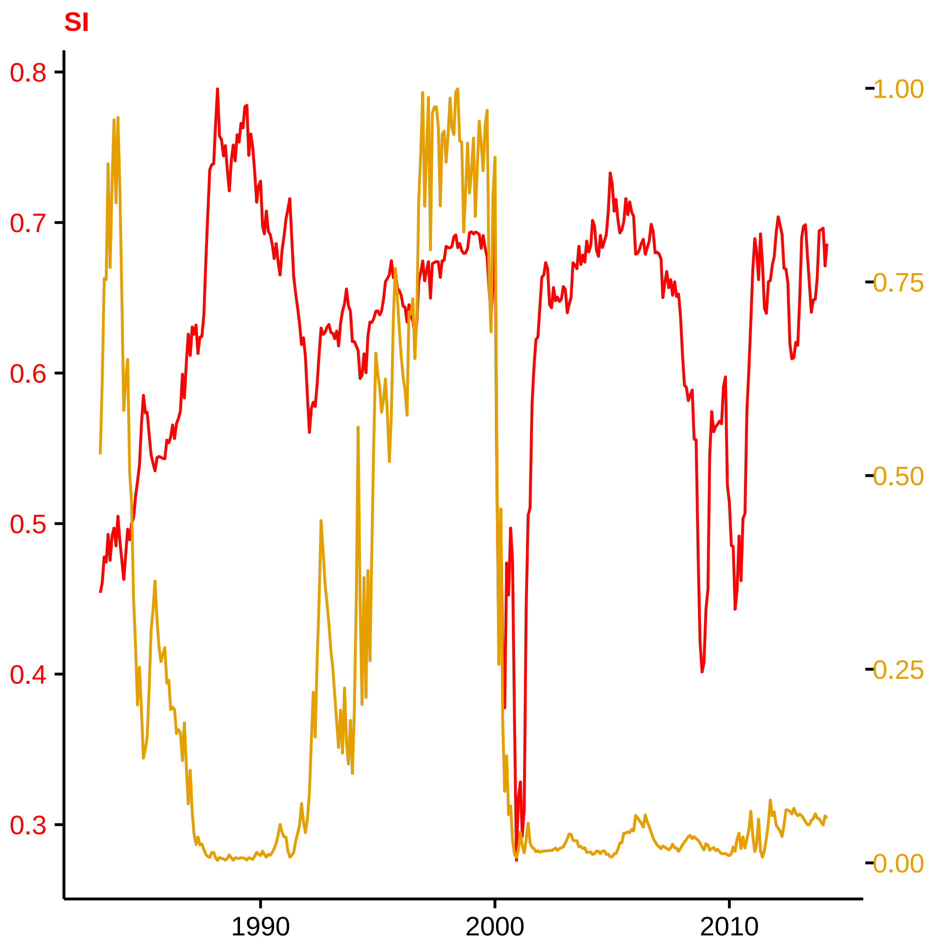

Figure 8: Probability of an UP Market (ADS)

This figure illustrates between 01/1983 - 04/2014 the monthly probability from the (ADS) Arubold-Diebold-Scotti Business Conditions Index forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. The gold colored line represents the statistical significance (p-value) of the monthly forecast. Each probability is based on a 120-month rolling estimation window.

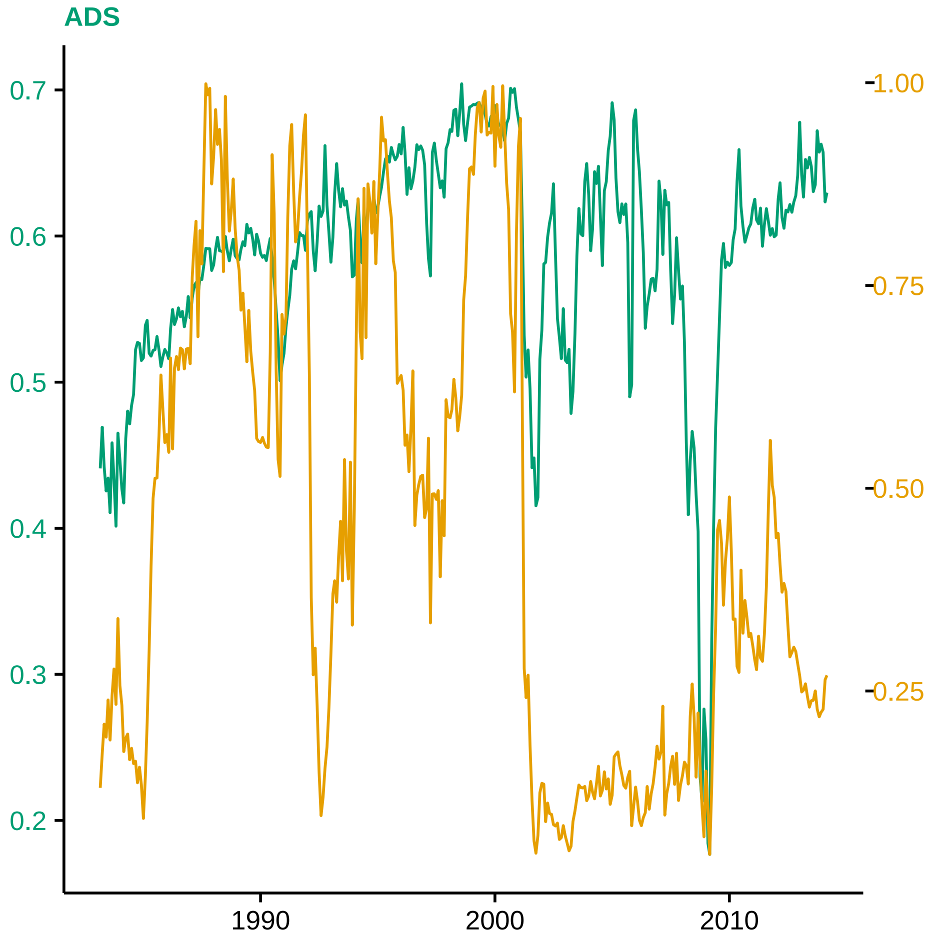

Figure 9: Probability of an UP Market (GZ)

This figure illustrates between 01/1983 - 04/2014 the monthly probability from the (GZ) Gilchrist-Zakrajsek Credit Spread Index forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. The gold colored line represents the statistical significance (p-value) of the monthly forecast. Each probability is based on a 120-month rolling estimation window.

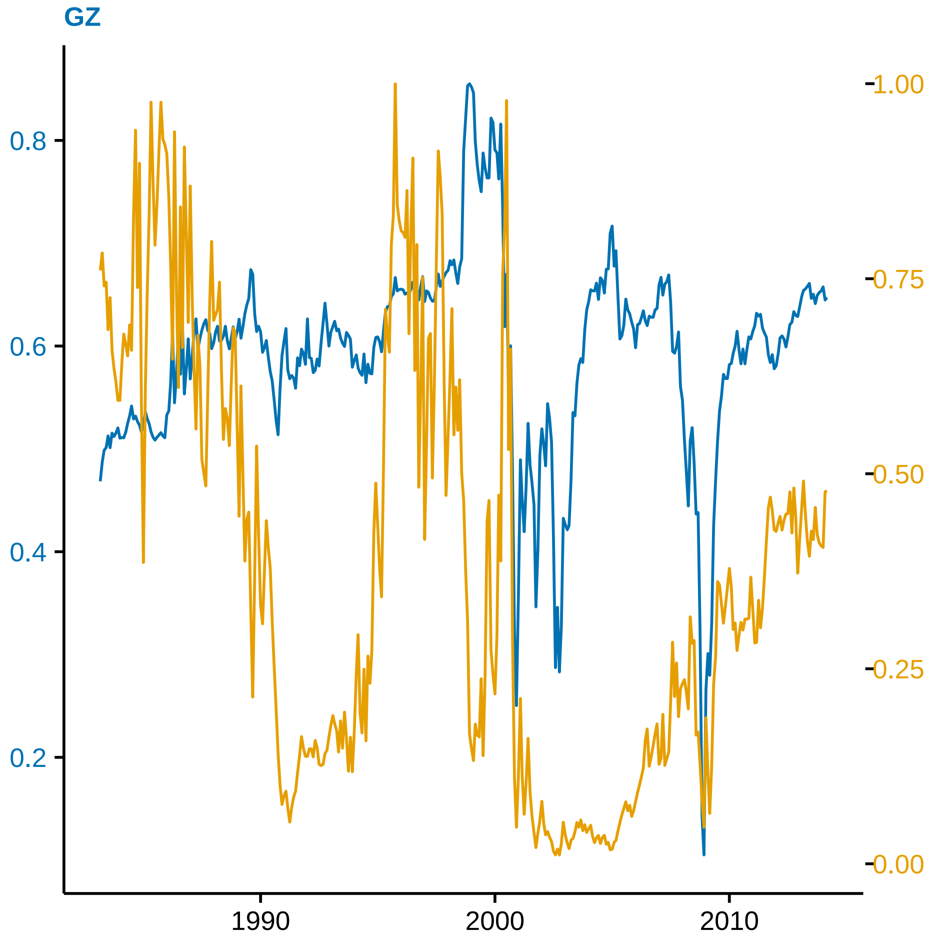

Figure 10: Probability of an UP Market (FUI)

This figure illustrates between 01/1983 - 04/2014 the monthly probability from the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. The gold colored line represents the statistical significance (p-value) of the monthly forecast. Each probability is based on a 120-month rolling estimation window.

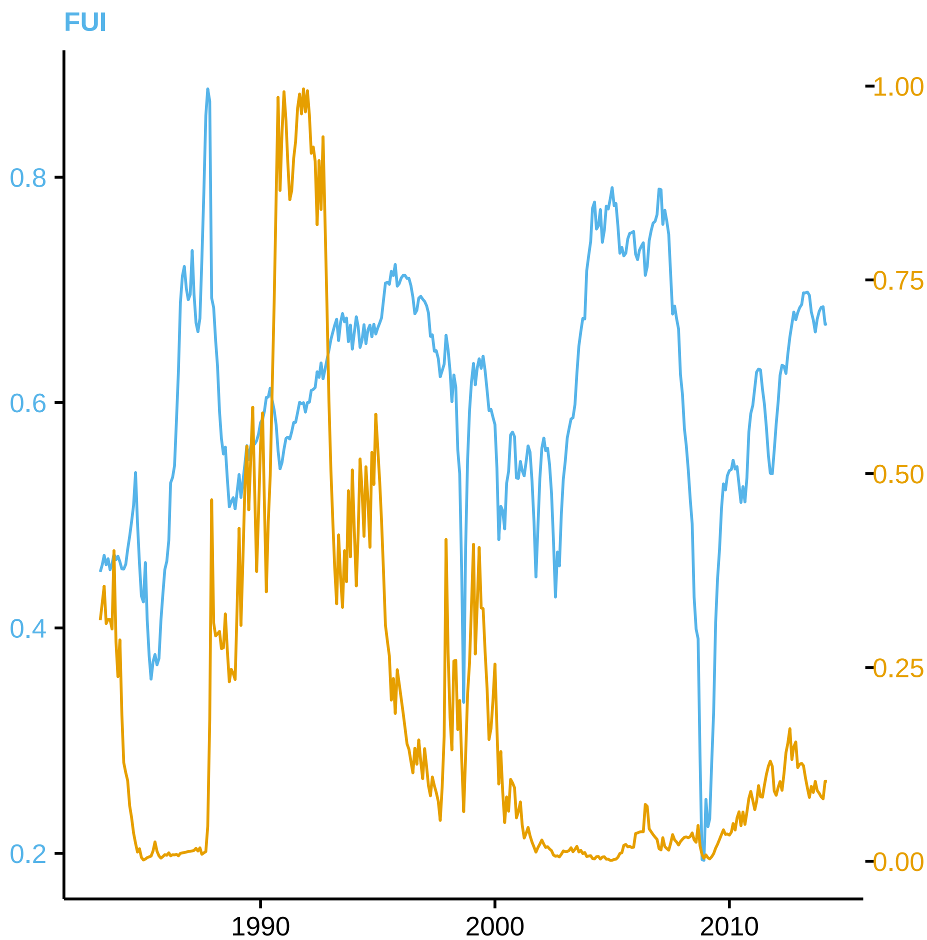

Figure 11: Probability of an UP Market (VWDP)

This figure illustrates between 01/1983 - 04/2014 the monthly probability from the benchmark Index (VWDP) Baker and Wurgler (2004) Value Weighted Dividend Premium Index forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. The gold colored line represents the statistical significance (p-value) of the monthly forecast. Each probability is based on a 120-month rolling estimation window.

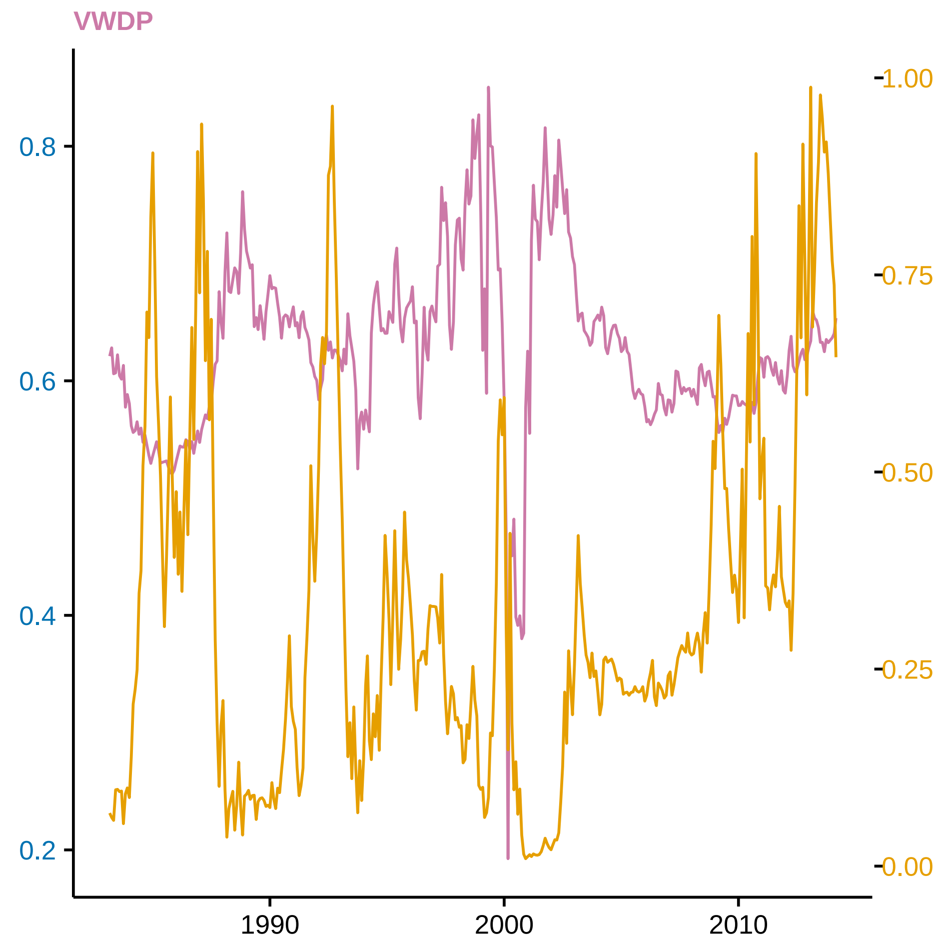

Figure 12: Persistence of Probability Forecast using Autocorrelation

The following six plots represent the persistence of probability forecasts among the five forecast model portfolios, and the benchmark model from 01/1983 - 04/2014. The (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable returns are represented by (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. The blue horizontal dashes represent the 95% confidence interval. In addition, each plot illustrates the autocorrelation corresponding to each monthly lag based on the following AR(1) model: $X_k = \rho X_{k-1} + \epsilon$ where, k= [0, 1, 2,...,20]

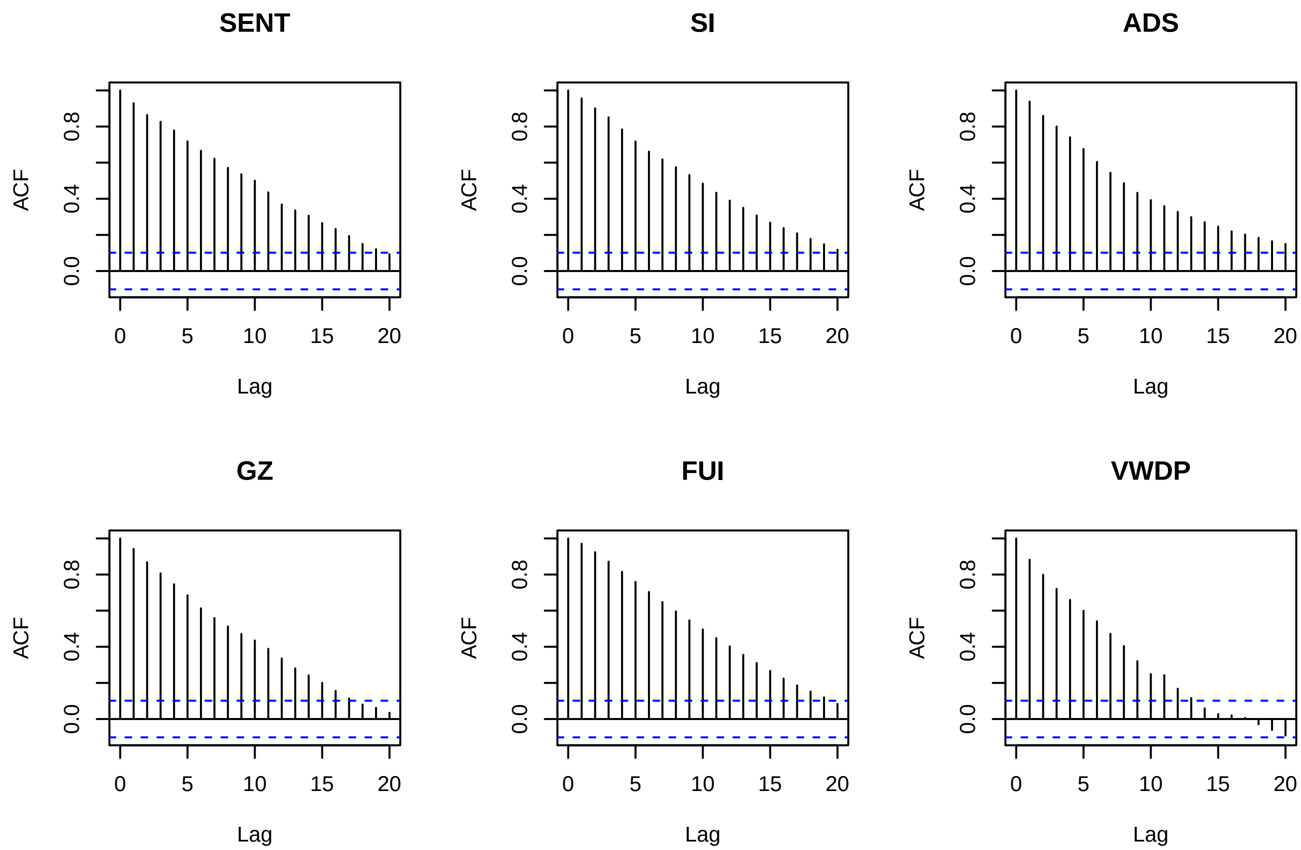

Figure 13: Box-Plot of Probability Estimates

This figure is a Box-Plot of the probability estimates of the five forecasting model portfolios, and the model selection portfolios. The model selection portfolio is the lasso model (LASSO). The five forecasting model portfolios are represented as follows: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model.

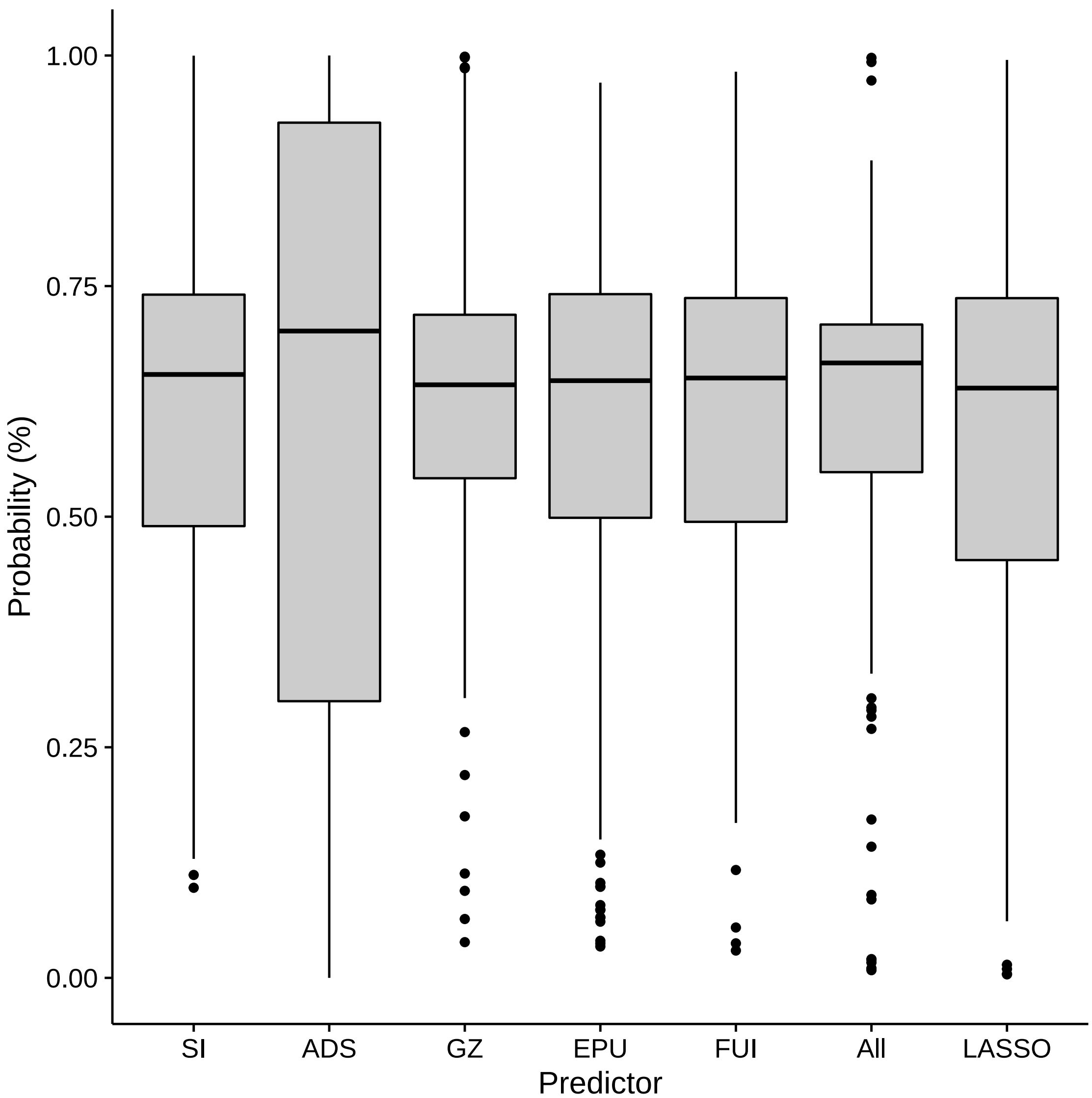

Figure 14: Distribution of the Probability Estimates

This figure represents the distribution of probability estimates for the five forecast model portfolios, and the combined models from 01/1985 - 04/2014. (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices ”kithcen sink” is represented as the (ALL) model, and the model selection index (LASSO). All estimates are calculated using a 24-month rolling window.

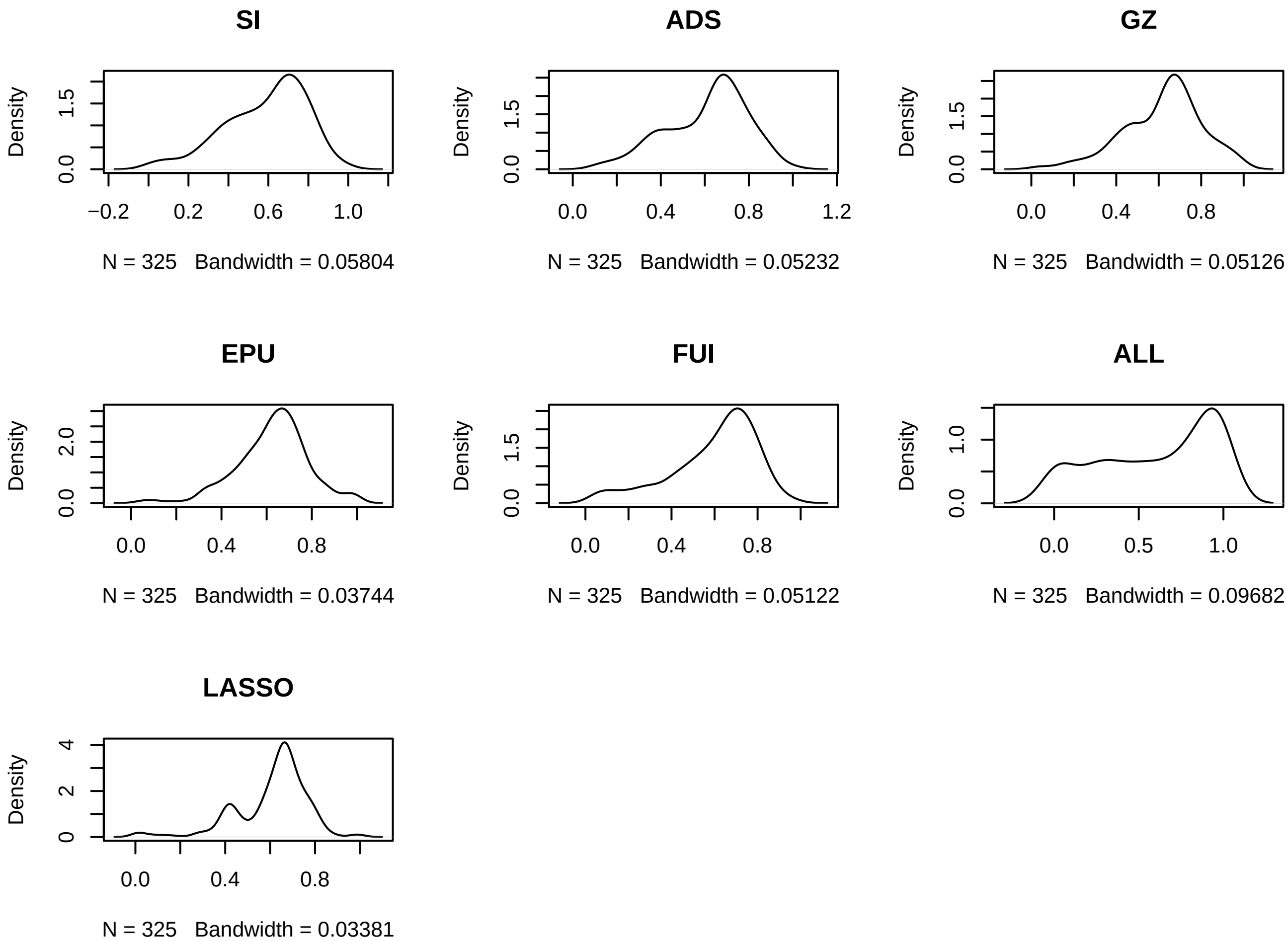

Figure 15: Normality Test of the Probability Estimates

This figure represents the results of the Shapiro normality tests using ggplots on the probability estimates of the five forecast model portfolios, and the combined indices from 01/1985 - 04/2014. (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices ”kitchen sink” is represented as the (ALL) model, and the model selection index (LASSO). All estimates are calculated using a 24-month rolling window.

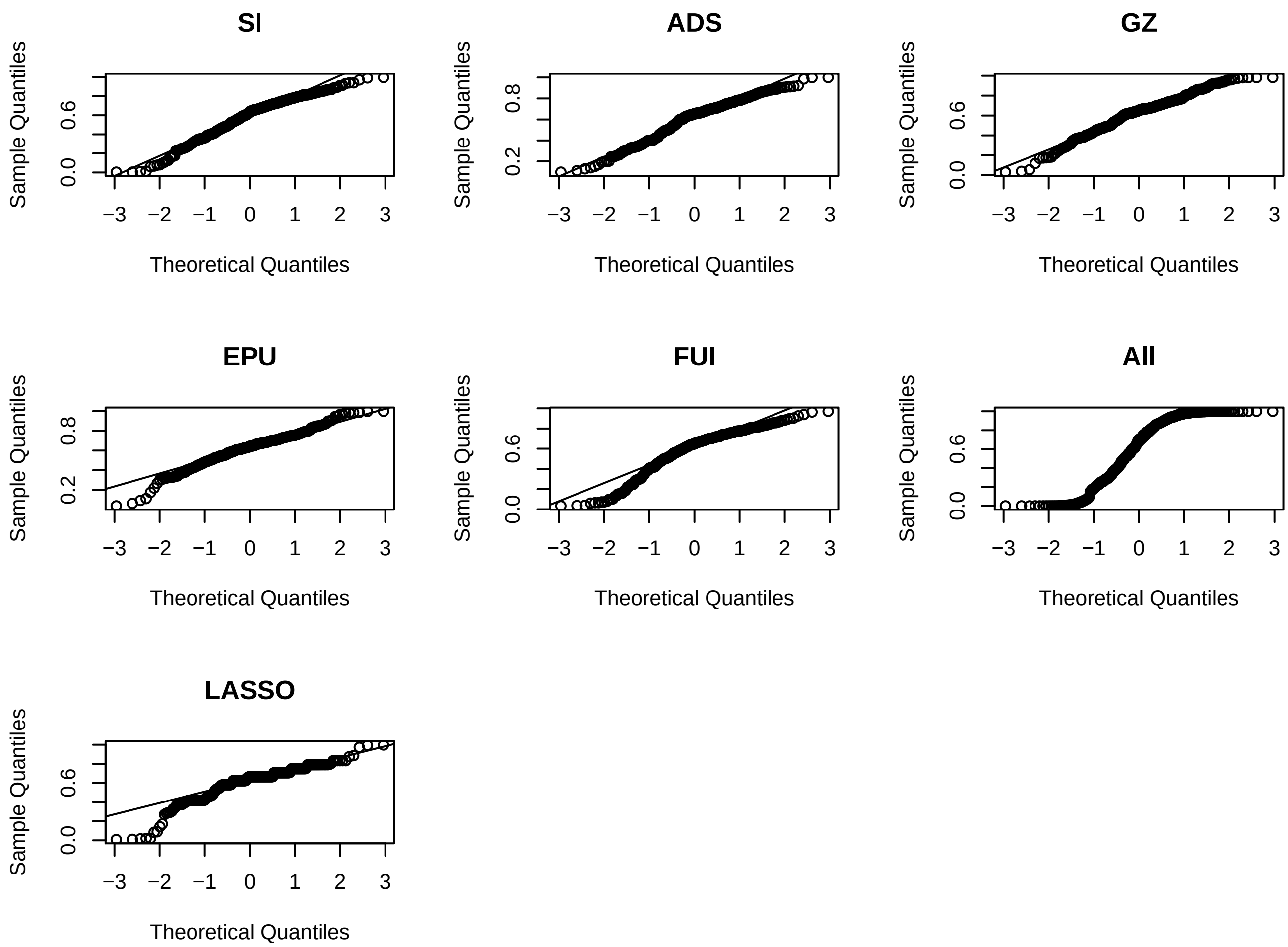

Figure 16: Wealth Index: S&P 500 vs. Forecasted Portfolios

This figure illustrates the cumulative wealth index of the buy and hold portfolio, the five forecasting model portfolios, and the model selection portfolio. The buy and hold portfolio is represented by the S&P 500 Index (SPX), and the model selection portfolio is the lasso model (LASSO). The five forecasting model portfolios are represented as follows: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model.

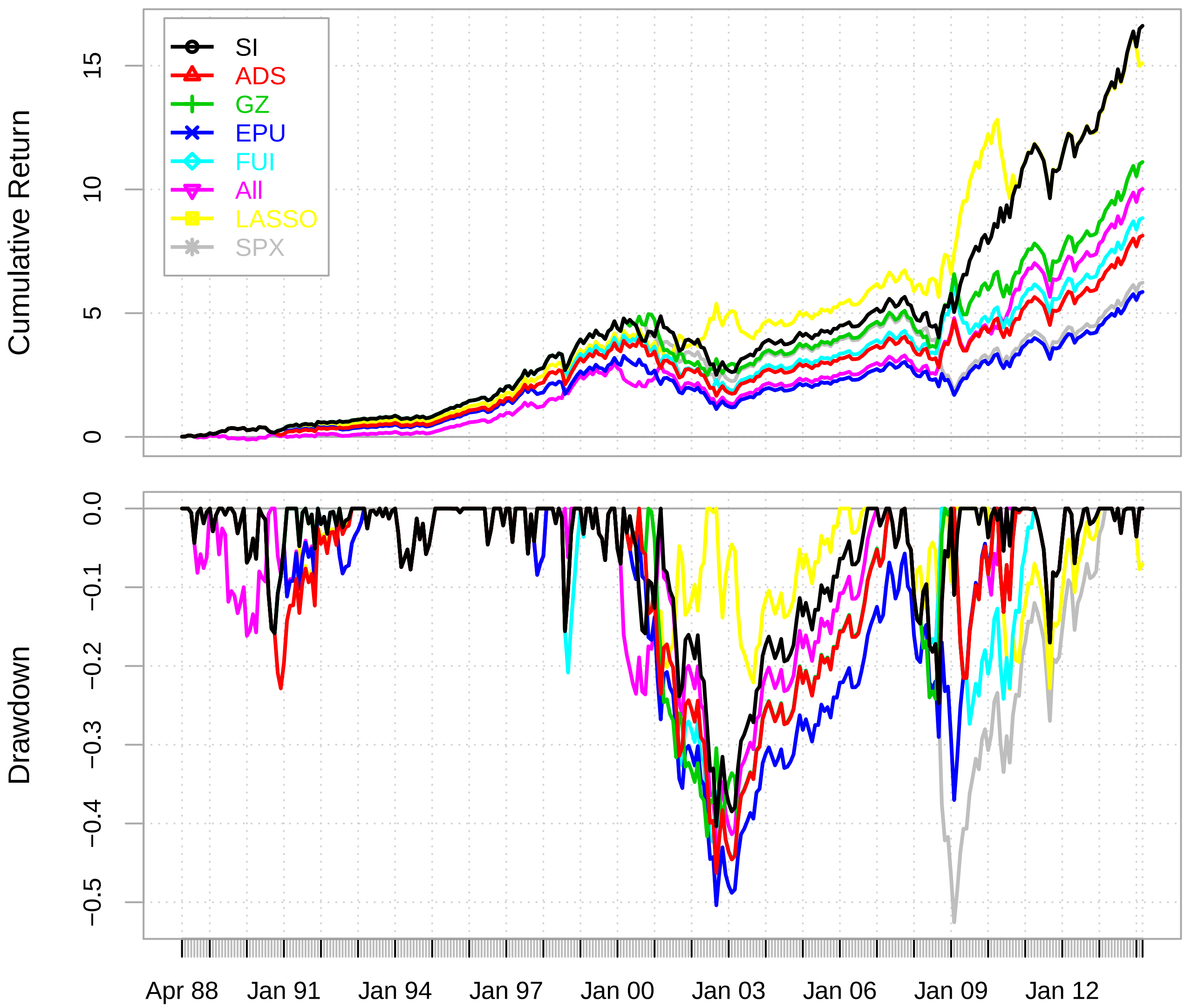

Figure 17: Wealth Index: S&P 500 vs. ALL & Lasso Portfolios

This figure illustrates the cumulative wealth index of the buy and hold portfolio, the combination of the five forecast model portfolios, and the model selection portfolio. The buy and hold portfolio is represented by the S&P 500 Index (SPX), and the model selection portfolio is the lasso model (LASSO). The combination of all five forecasting models is represented as the all portfolio (ALL) consisting of the following: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index.

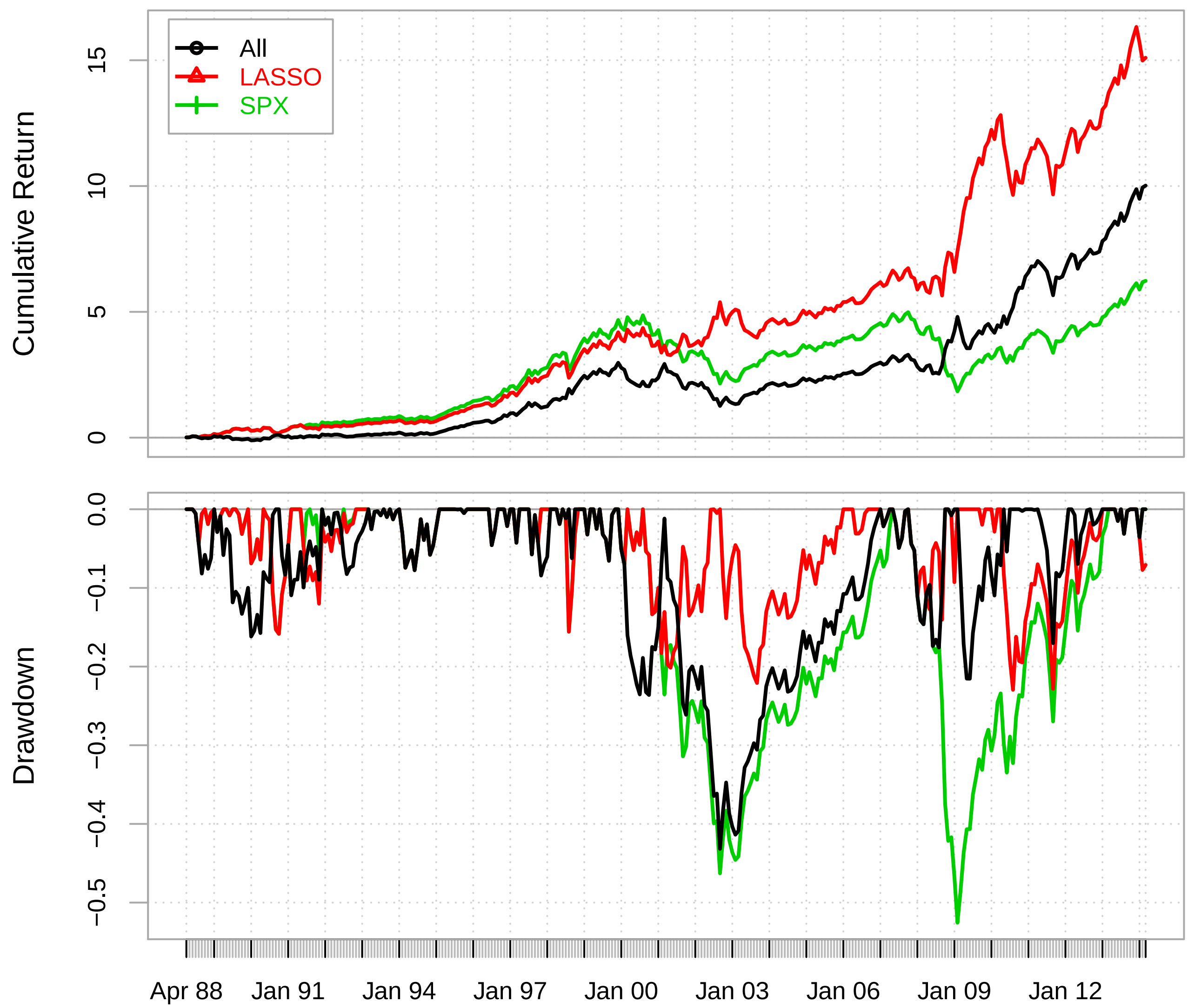

Figure 18: Forecasted Probability of an Up Movement

This figure illustrates the probability of an Up market (S&P 500 Index) prediction from each of the five forecasting models, the combination of the forecasting models, and the model selection portfolio. The model selection portfolio is the lasso (LASSO), and the five forecasting model portfolios are represented as: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting models is represented as the all portfolio (ALL). A value greater the 0.50 indicates the prediction of an UP market. Any value less than 0.50 indicates a prediction of a down market.

Figure 19: Probability of an UP Market (ALL SI)

This figure illustrates between 01/1983 - 04/2014 the monthly probability of the combined model (ALL) forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. In addition, the colored line represents the statistical significance (p-value) of the monthly forecast of the ((SI) Huang, Jiang, and Tu Sentiment Index individually represented in the ALL model. Each probability and p-value is calculated on a 24-month rolling estimation window.

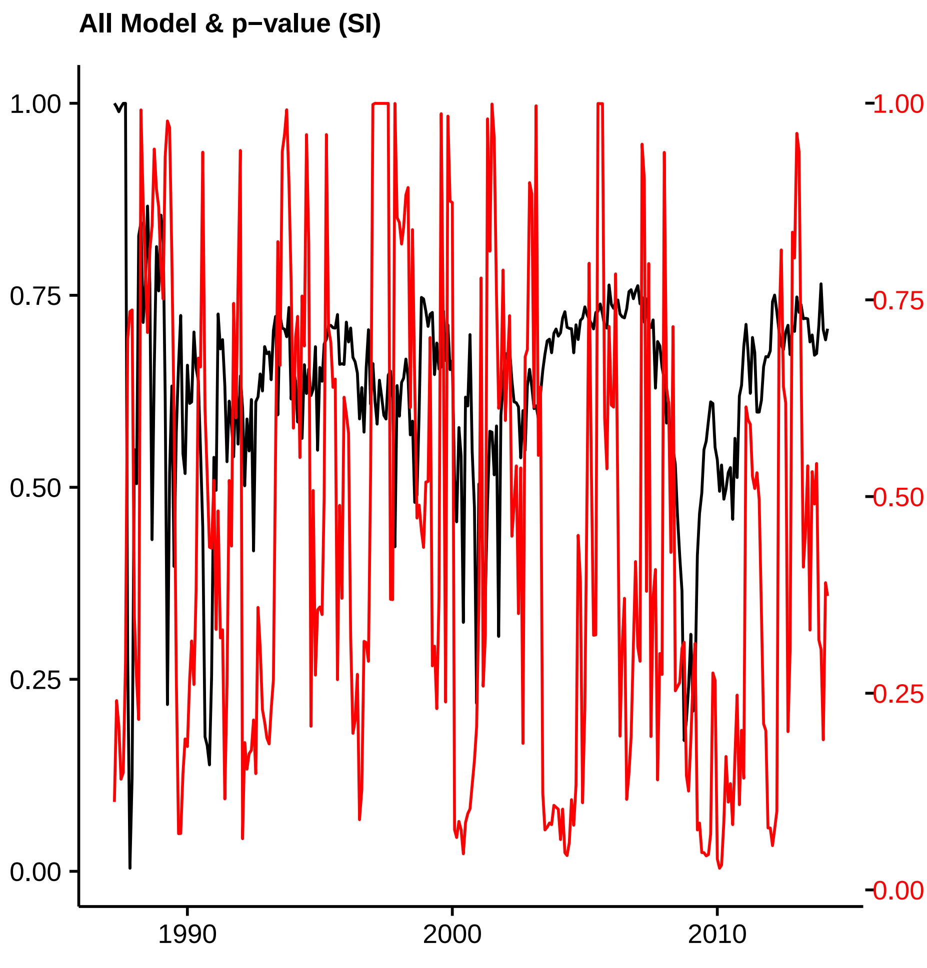

Figure 20: Probability of an UP Market (ALL ADS)

This figure illustrates between 01/1983 - 04/2014 the monthly probability of the combined model (ALL) forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. In addition, the colored line represents the statistical significance (p-value) of the monthly forecast of the (ADS) Arubold-Diebold-Scotti Business Conditions Index individually represented in the ALL model. Each probability and p-value is calculated on a 24-month rolling estimation window.

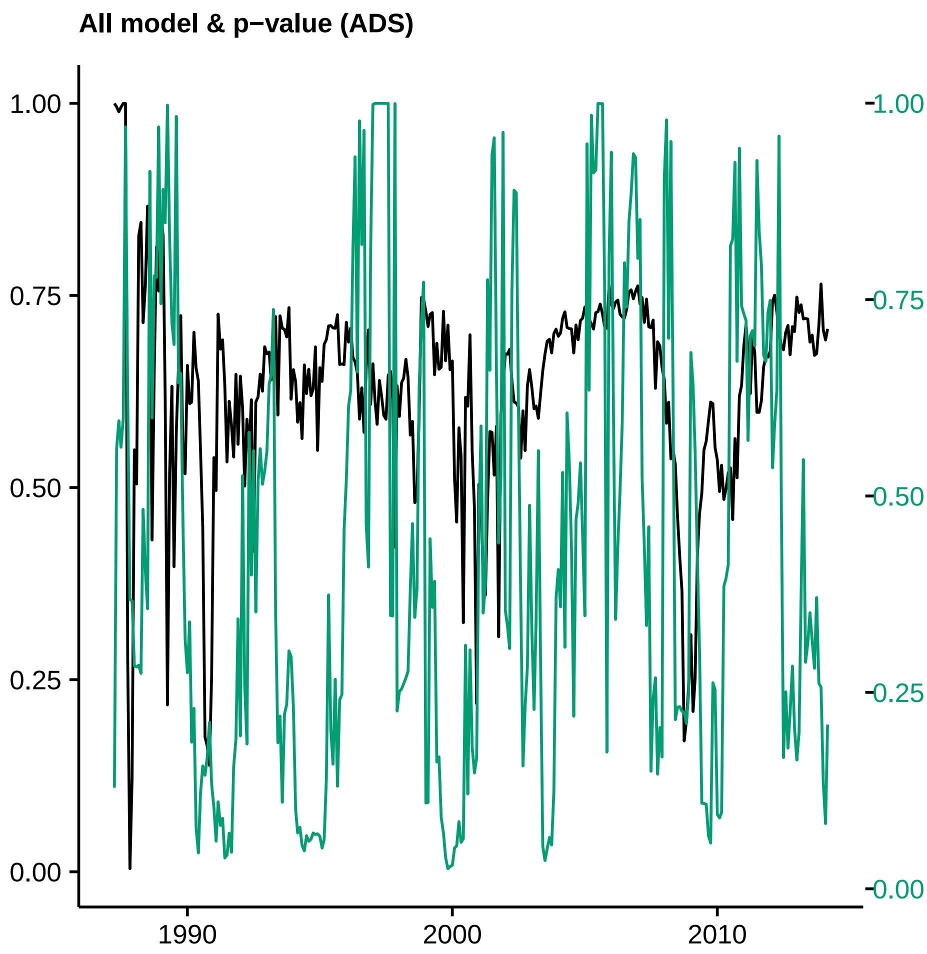

Figure 21: Probability of an UP Market (ALL GZ)

This figure illustrates between 01/1983 - 04/2014 the monthly probability of the combined model (ALL) forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. In addition, the colored line represents the statistical significance (p-value) of the monthly forecast of the (GZ) Gilchrist-Zakrajsek Credit Spread Index individually represented in the ALL model. Each probability and p-value is calculated on a 24-month rolling estimation window.

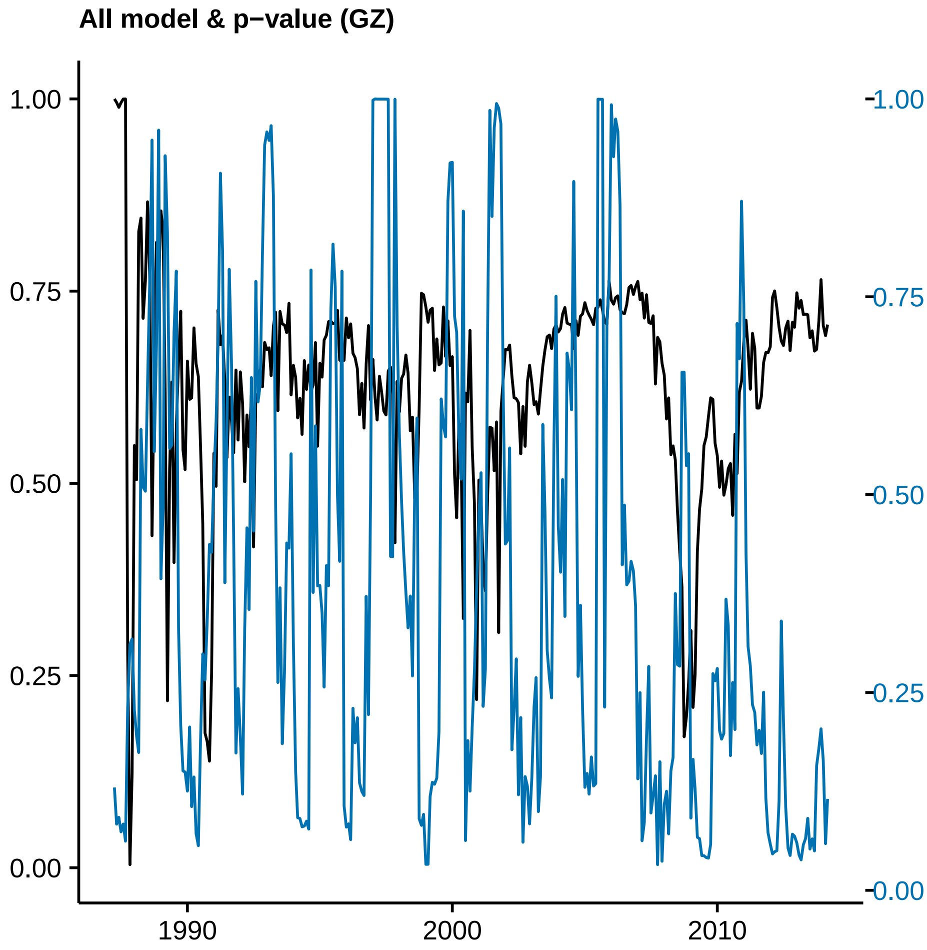

Figure 22: Probability of an UP Market (ALL EPU)

This figure illustrates between 01/1983 - 04/2014 the monthly probability of the combined model (ALL) forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. In addition, the colored line represents the statistical significance (p-value) of the monthly forecast of the (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index individually represented in the ALL model. Each probability and p-value is calculated on a 24-month rolling estimation window.

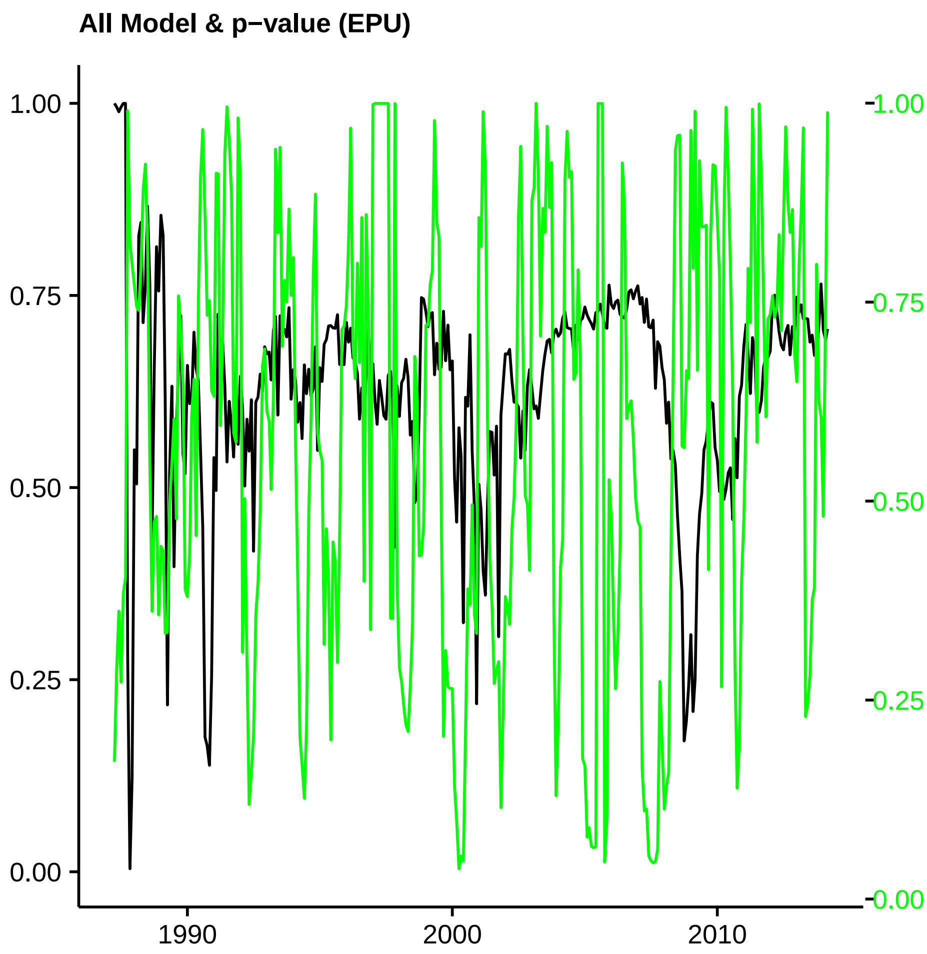

Figure 23: Probability of an UP Market (ALL FUI)

This figure illustrates between 01/1983 - 04/2014 the monthly probability of the combined model (ALL) forecast that the (SPX) S&P 500 Index will have positive returns in the next calendar month. In addition, the colored line represents the statistical significance (p-value) of the monthly forecast of the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index individually represented in the ALL model. Each probability and p-value is calculated on a 24-month rolling estimation window.

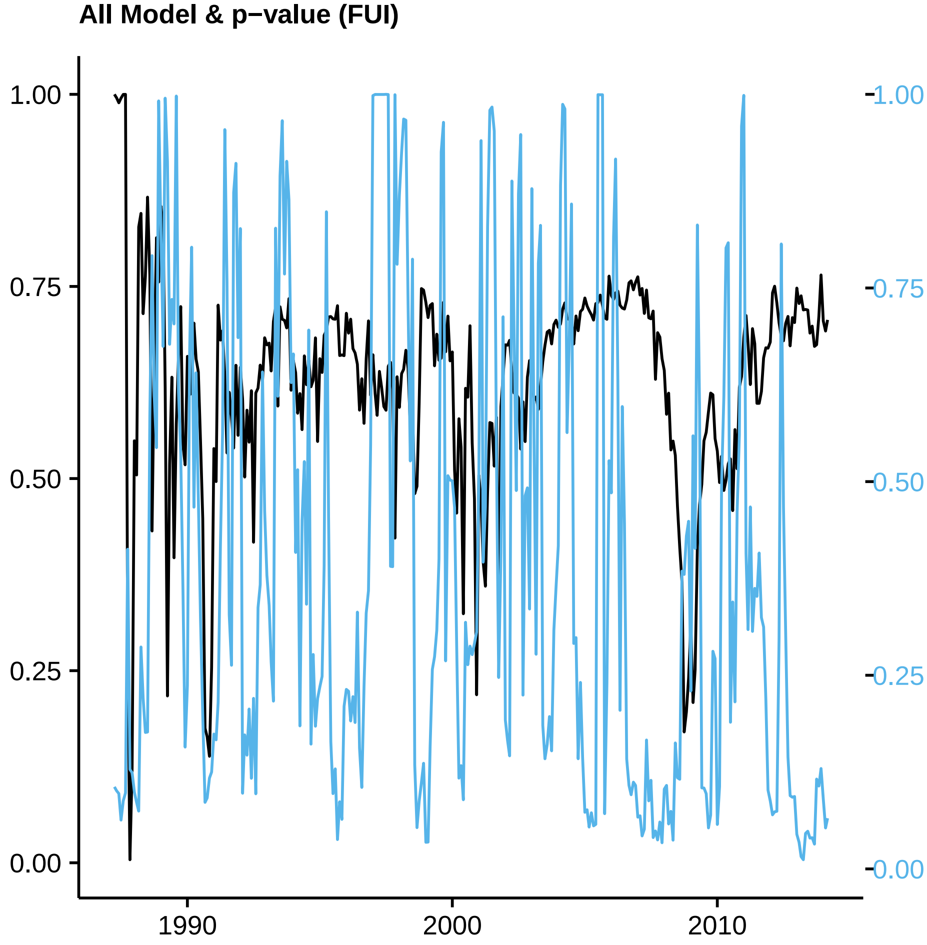

Figure 24: Persistence of Probability Forecast using Autocorrelation

The following six plots represent the persistence of probability forecasts among the five forecast model portfolios, and the benchmark model from 01/1983 - 04/2014. The (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (EPU) Baker, Bloom, Davis Economic Policy Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The "kitchen sink" (ALL) model, the Lasso model selection index. The blue horizontal dashes represent the 95% confidence interval. In addition, each plot illustrates the autocorrelation corresponding to each monthly lag based on the following AR(1) model: $X_k = \rho X_{k-1} + \epsilon$ where, k= [0, 1, 2,...,20]

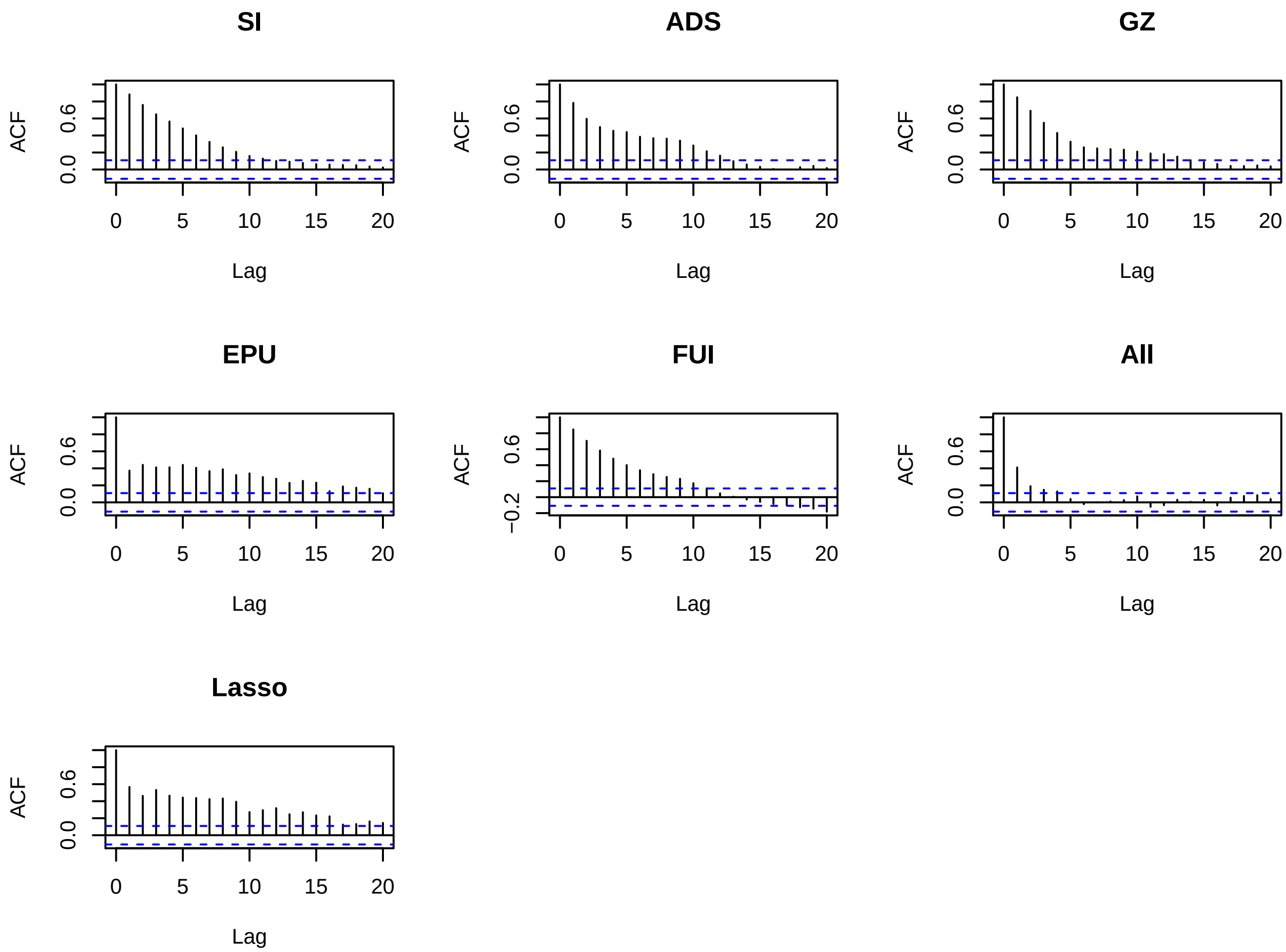

Figure 25: Wealth Index: S&P 500 vs. Forecasted Portfolios without EPU Index

This figure illustrates the cumulative wealth index of the buy and hold portfolio, the combination of four forecast model portfolios excluding the (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the model selection portfolio. The buy and hold portfolio is represented by the S&P 500 Index (SPX), and the model selection portfolio is the lasso model (LASSO). The combination of the remaining four forecasting models is represented as the all portfolio (ALL). The individual model portfolios are the following: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index.

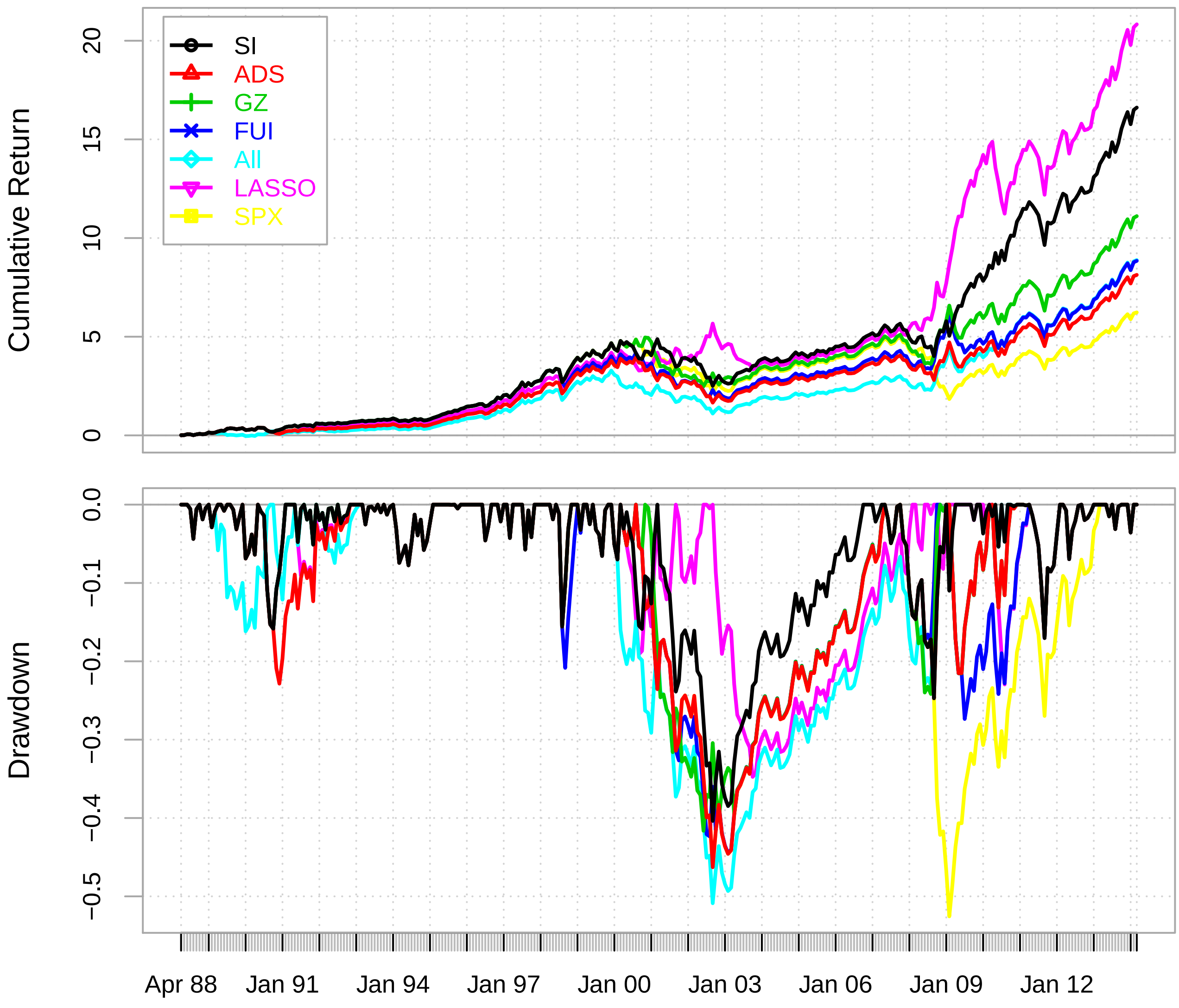

Table 1: Summary Statistics and Correlation Matrix

Panel A: Represents the summary statistics associated with each of the forecasting indices. The indices represented are for the sample period (01/1973 - 04/2014). All columns represent monthly returns. SPX1 represents the S&P 500 Index one-month-ahead returns. The forecasting indices represent the monthly raw data value. (SENT) Baker and Wurgler Sentiment Index, (SI) is the Huang, Jiang, and Tu Sentiment Index, (ADS) is the Arubold-Diebold-Scotti Business Conditions Index, (GZ) is the Gilchrist-Zakrajsek Credit Spread Index, and the (FUI) is the Jurado, Ludvigson, and Ng Financial Uncertainty Index.

Panel A: Summary Statistics

	SPX1	SENT	SI	ADS	GZ	FUI
mean	0.007	-0.037	-0.051	-0.113	1.778	0.922
sd	0.045	0.888	0.895	0.884	0.979	0.176
min	-0.218	-2.578	-1.701	-4.311	0.545	0.643
max	0.163	2.497	3.085	2.569	7.824	1.549
skew	-0.460	-0.134	1.335	-1.358	2.376	0.721
kurtosis	1.870	0.635	1.582	4.153	8.930	0.206

Panel B: Represents the correlation matrix between each of index introduced in Panel A above. All correlations are calculated from the monthly raw values of each index during the sample period (01/1973 - 04/2014).

Panel B: Correlation Matrix

	SPX1	SENT	SI	ADS	GZ	FUI
SPX1	1.000					
SENT	-0.015	1.000				
SI	-0.115	0.669	1.000			
ADS	0.002	-0.066	-0.029	1.000		
GZ	-0.077	0.114	0.137	-0.510	1.000	
FUI	-0.118	0.002	0.220	-0.410	0.503	1.000

Table 2: Correlation of Predictions

This table illustrates the correlations from (01/1973 - 04/2014) between the one-month-ahead S&P 500 Index (SPX1) return, the UP movement of the S&P 500 Index, the return predictions associated with each of the five forecasting models: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium.

Correlation Matrix of Monthly Return Predictions

	SPX1	UP	SENT	SI	ADS	GZ	FUI	VWDP
SPX1	1.000							
UP	0.758	1.000						
SENT	0.098	0.062	1.000					
SI	0.160	0.092	0.239	1.000				
ADS	0.084	0.068	0.185	0.429	1.000			
GZ	0.131	0.075	0.489	0.329	0.482	1.000		
FUI	0.060	0.047	0.052	0.370	0.590	0.347	1.000	
VWDP	0.047	0.112	0.069	0.414	-0.060	0.046	0.017	1.000

Table 3: Annual Holding Period Returns

This table illustrates annual holding period returns from (01/1983 - 04/2014) for the S&P 500 Index (SPX), the S&P 500 Index Perfectly Forecast Long/Short portfolio (SPXLS), the S&P 500 Index Perfectly Forecast Long/RF portfolio (SPXRF), and the five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All HPRs are estimated using a 120 month rolling estimation window.

	SENT	SI	ADS	GZ	FUI	VWDP	SPXRF	SPXLS	SPX
1983	0.104	-0.118	-0.103	-0.089	-0.103	0.104	0.258	0.309	0.125
1984	0.099	0.244	0.039	0.099	0.092	0.099	0.324	0.435	0.099
1985	0.179	0.179	0.179	0.179	-0.162	0.179	0.289	0.327	0.179
1986	0.294	0.294	0.294	0.294	0.121	0.294	0.600	0.881	0.294
1987	-0.062	-0.062	-0.062	-0.062	-0.062	-0.062	0.386	0.860	-0.062
1988	0.157	0.157	0.157	0.157	0.157	0.157	0.303	0.403	0.157
1989	0.106	0.106	0.106	0.106	0.106	0.106	0.316	0.457	0.106
1990	0.045	0.045	0.045	0.045	0.045	0.045	0.327	0.539	0.045
1991	0.189	0.189	0.189	0.189	0.189	0.189	0.382	0.544	0.189
1992	0.073	0.073	0.073	0.073	0.073	0.073	0.154	0.218	0.073
1993	0.098	0.098	0.098	0.098	0.098	0.098	0.170	0.222	0.098
1994	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	0.178	0.370	-0.023
1995	0.352	0.352	0.352	0.352	0.352	0.352	0.371	0.366	0.352
1996	0.236	0.236	0.236	0.236	0.236	0.236	0.335	0.414	0.236
1997	0.247	0.247	0.247	0.247	0.247	0.247	0.449	0.632	0.247
1998	0.305	0.305	0.305	0.305	0.316	0.305	0.595	0.861	0.305
1999	0.090	0.090	0.090	0.090	0.090	0.090	0.333	0.547	0.090
2000	0.064	-0.045	-0.020	0.075	-0.231	0.074	0.279	0.545	-0.020
2001	0.102	-0.039	-0.061	0.074	-0.191	-0.173	0.221	0.681	-0.173
2002	-0.243	-0.243	-0.243	-0.090	-0.079	-0.243	0.209	0.830	-0.243
2003	0.322	0.322	0.015	0.143	0.322	0.322	0.363	0.401	0.322
2004	0.044	0.044	0.044	0.044	0.044	0.044	0.152	0.257	0.044
2005	0.084	0.084	0.046	0.084	0.084	0.084	0.179	0.245	0.084
2006	0.124	0.124	0.124	0.124	0.124	0.124	0.164	0.195	0.124
2007	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042	0.184	0.389	-0.042
2008	-0.401	0.374	0.464	0.259	0.625	-0.401	0.091	0.835	-0.401
2009	0.300	0.621	0.017	0.017	-0.180	0.300	0.548	0.816	0.300
2010	0.198	0.117	0.198	0.198	0.198	0.198	0.451	0.737	0.198
2011	0.020	0.020	0.020	0.020	0.020	0.020	0.238	0.486	0.020
2012	0.141	0.141	0.141	0.141	0.141	0.141	0.252	0.367	0.141
2013	0.190	0.190	0.190	0.190	0.190	0.190	0.293	0.402	0.190
2014	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050

Table 4: Summary Statistics of Returns

This table shows the summary statistics of returns from (01/1983 - 04/2014) for the S&P 500 Index (SPX), the S&P 500 Index Perfectly Forecast Long/Short portfolio (SPXLS), the S&P 500 Index Perfectly Forecast Long/RF portfolio (SPXRF), and the five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All statistics are estimated using a 120 month rolling estimation window.

	SENT	SI	ADS	GZ	FUI	VWDP	SPXRF	SPXLS	SPX
Minimum	-0.2176	-0.2176	-0.2176	-0.2176	-0.2176	-0.2176	0.0000	0.0000	-0.2176
Quartile 1	-0.0165	-0.0156	-0.0174	-0.0165	-0.0189	-0.0164	0.0044	0.0133	-0.0189
Median	0.0113	0.0121	0.0107	0.0113	0.0099	0.0118	0.0108	0.0269	0.0096
Arithmetic Mean	0.0088	0.0105	0.0084	0.0097	0.0072	0.0080	0.0226	0.0347	0.0068
Geometric Mean	0.0078	0.0095	0.0074	0.0087	0.0062	0.0071	0.0222	0.0343	0.0057
Quartile 3	0.0361	0.0369	0.0360	0.0360	0.0350	0.0358	0.0360	0.0479	0.0360
Maximum	0.1318	0.1694	0.1694	0.1694	0.1694	0.1318	0.1630	0.2176	0.1630
Stdev	0.0436	0.0432	0.0437	0.0434	0.0439	0.0437	0.0257	0.0293	0.0449
Skewness	-0.7467	-0.4377	-0.4721	-0.3871	-0.2249	-0.8112	1.7729	1.7123	-0.4613
Kurtosis	2.4993	2.3918	2.2428	2.2753	1.9918	2.4686	3.6270	4.7048	1.8799

Table 5: Relationship between Probability Estimates (Dunn Test Ranking)

This table summarizes the Z-statistic and its relative p-values resulting from the Kruskal-Wallis rank sum test and the pairwise Dunn test. The time period is from (01/1983 - 04/2014) for the five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium.

	$SENT$	SI	ADS	GZ	FUI	$VWDP$
$SENT$	0	3.3028	0.9354	8.6666	16.754	5.4272
	0	(0.0072)*	(1.0000)	(0.0000)*	(0.0000)*	(0.0000)*
SI	3.3028	0	2.3674	5.3637	13.4511	2.1244
	(0.0072)*	0	(0.1343)	(0.0000)*	(0.0000)*	(0.2523)
ADS	0.9354	2.3674	0	7.7312	15.8186	4.4918
	(1.0000)	(0.1343)	0	(0.0000)*	(0.0000)*	(0.0001)*
GZ	8.6666	5.3637	7.7312	0	8.0873	3.2393
	(0.0000)*	(0.0000)*	(0.0000)*	0	(0.0000)*	(0.0090)*
FUI	16.754	13.4511	15.8186	8.0873	0	11.3267
	(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)*	0	(0.0000)*
$VWDP$	5.4272	2.1244	4.4918	3.2393	11.3267	0
	(0.0000)*	(0.2523)	(0.0001)*	(0.0090)*	(0.0000)*	0

$(pvalue < 0.05)*$

Table 6: Annual Forecast Errors of Predictions

This table shows the annual forecast errors of the predictive portfolios from (01/1983 - 04/2014). The S&P 500 Index (SPX) represents the number of UP vs DOWN months. The five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All forecasting errors are depicted from a 120 day rolling estimation window. ER1 represents forecasting errors in predicting UP months, and ER2 represents forecasting errors in DOWN months.

Year	SPX		SENT		SI		ADS		GZ		FUI		VWDP	
	UP	DOWN	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2
1983	6	4	0	4	6	1	6	0	2	3	6	0	0	4
1984	5	7	0	7	3	4	4	3	0	7	4	1	0	7
1985	7	5	0	5	0	5	0	5	0	5	7	0	0	5
1986	8	4	0	4	0	4	0	4	0	4	2	4	0	4
1987	8	4	0	4	0	4	0	4	0	4	0	4	0	4
1988	8	4	0	4	0	4	0	4	0	4	0	4	0	4
1989	8	4	0	4	0	4	0	4	0	4	0	4	0	4
1990	5	7	0	7	0	7	0	7	0	7	0	7	0	7
1991	9	3	0	3	0	3	0	3	0	3	0	3	0	3
1992	8	4	0	4	0	4	0	4	0	4	0	4	0	4
1993	8	4	0	4	0	4	0	4	0	4	0	4	0	4
1994	7	5	0	5	0	5	0	5	0	5	0	5	0	5
1995	10	2	0	2	0	2	0	2	0	2	0	2	0	2
1996	10	2	0	2	0	2	0	2	0	2	0	2	0	2
1997	9	3	0	3	0	3	0	3	0	3	0	3	0	3
1998	9	3	0	3	0	3	0	3	0	3	2	2	0	3
1999	7	5	0	5	0	5	0	5	0	5	0	5	0	5
2000	4	8	1	7	4	2	0	8	1	6	2	8	3	1
2001	6	6	4	1	3	4	2	3	6	1	2	5	0	6
2002	4	8	0	8	0	8	0	8	4	5	2	6	0	8
2003	9	3	0	3	0	3	2	3	2	1	0	3	0	3
2004	9	3	0	3	0	3	0	3	0	3	0	3	0	3
2005	5	7	0	7	0	7	1	6	0	7	0	7	0	7
2006	11	1	0	1	0	1	0	1	0	1	0	1	0	1
2007	7	5	0	5	0	5	0	5	0	5	0	5	0	5
2008	4	8	0	8	1	5	4	2	3	3	2	3	0	8
2009	9	3	0	3	0	1	4	1	4	1	6	1	0	3
2010	7	5	0	5	4	3	0	5	0	5	0	5	0	5
2011	5	7	0	7	0	7	0	7	0	7	0	7	0	7
2012	9	3	0	3	0	3	0	3	0	3	0	3	0	3
2013	10	2	0	2	0	2	0	2	0	2	0	2	0	2
2014	2	1	0	1	0	1	0	1	0	1	0	1	0	1
Total	233	140	5	134	21	119	23	120	22	120	35	114	3	133
Incorrect			139		140		143		142		149		136	
Correct			234		233		230		231		224		237	

Table 7: Annualized Performance Analysis

This table summarizes annualized return and risk information This table shows the summary statistics of returns from (01/1983 - 04/2014) for the S&P 500 Index (SPX), the S&P 500 Index Perfectly Forecast Long/Short portfolio (SPXLS), the S&P 500 Index Perfectly Forecast Long/RF portfolio (SPXRF), and the five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All statistics are estimated using a 120 month rolling estimation window.

	SENT	SI	ADS	GZ	FUI	VWDP	SPXRF	SPXLS	SPX
Annualized Return	0.098	0.121	0.092	0.110	0.077	0.088	0.302	0.499	0.071
Annualized Std Dev	0.151	0.150	0.151	0.150	0.152	0.151	0.089	0.101	0.155
Annualized Sharpe (Rf=0%)	0.649	0.806	0.611	0.730	0.509	0.583	3.391	4.932	0.457
Sortino Ratio (MAR = 0%)	0.304	0.393	0.295	0.357	0.256	0.270	Inf	Inf	0.225
Omega (L = 0%)	1.696	1.889	1.651	1.792	1.535	1.619	Inf	Inf	1.484
Tracking Error	0.021	0.039	0.040	0.043	0.048	0.016	0.051	0.054	0.000
Annualised Tracking Error	0.074	0.134	0.137	0.150	0.167	0.057	0.177	0.186	0.000
Information Ratio	0.174	0.265	0.054	0.164	-0.046	0.057	1.306	2.307	
Semi Deviation	0.033	0.032	0.032	0.032	0.032	0.034	0.013	0.016	0.033
Gain Deviation	0.025	0.027	0.027	0.028	0.028	0.025	0.026	0.029	0.027
Loss Deviation	0.033	0.030	0.031	0.030	0.029	0.034			0.032
Maximum Drawdown	0.526	0.404	0.447	0.302	0.483	0.526	0.000	0.000	0.526
Up Capture	0.947	0.822	0.771	0.768	0.636	0.954	0.657	0.993	1.000
Down Capture	0.839	0.511	0.592	0.489	0.466	0.907	-0.643	-1.015	1.000

Table 8: Predictive Logistic Regression Estimation Results

Table 7 illustrates each coefficient (%), Newey-West t-statistic, and pseudo (McFadden) R^2 associated with the logistic regression model using a maximum likelihood approach:

$$\log \frac{p(x_{t+1}=1)}{1-p(x_{t+1}=1)} = \alpha_k + \beta_k F_{k,t}$$

where, $F_{k,t}$ is an index. Equivalently,

$$p(x_{t+1}=1) = \frac{1}{1+e^{-(\alpha_k+\beta_k F_{k,t})}}.$$

Each observation is the last month (December) of each year during the forecasting period (01/1983 - 04/2014). These results are based on the prediction of the five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All statistics are estimated using a 120 month rolling estimation window. (t-statistics above +/- 2.00, 2.50, and 3.00 are significant at the 10%, 5%, and 1% levels respectively)

	β (%)						$t-statistic$						$pseudo-R^2$					
	$SENT$	SI	ADS	GZ	FUI	$VWDP$	$SENT$	SI	ADS	GZ	FUI	$VWDP$	$SENT$	SI	ADS	GZ	FUI	$VWDP$
1983	0.559	0.488	0.447	0.510	0.459	0.581	1.016	-0.165	-1.134	-0.533	1.068	-1.711	0.431	0.298	0.212	0.342	0.237	0.468
1984	0.531	0.585	0.517	0.522	0.423	0.536	-0.053	-1.494	-1.681	0.866	3.136	-0.119	0.381	0.476	0.353	0.364	0.158	0.390
1985	0.532	0.556	0.520	0.533	0.459	0.531	0.387	-1.195	-0.574	0.113	2.663	-0.714	0.382	0.425	0.360	0.384	0.238	0.381
1986	0.550	0.612	0.544	0.568	0.696	0.547	-0.021	-1.557	-0.449	0.165	2.486	-0.349	0.415	0.519	0.404	0.446	0.645	0.410
1987	0.620	0.739	0.580	0.603	0.684	0.636	-1.285	-2.473	-0.114	0.390	1.020	1.253	0.532	0.704	0.467	0.505	0.629	0.557
1988	0.673	0.758	0.585	0.614	0.521	0.711	-1.837	-2.734	-0.271	0.583	0.976	1.689	0.612	0.729	0.475	0.523	0.362	0.666
1989	0.721	0.727	0.588	0.614	0.582	0.690	-1.822	-2.594	-0.588	0.970	0.616	1.785	0.680	0.689	0.481	0.523	0.471	0.636
1990	0.633	0.691	0.520	0.604	0.559	0.663	-0.942	-2.124	0.399	1.712	0.009	1.329	0.552	0.638	0.359	0.507	0.430	0.598
1991	0.645	0.584	0.610	0.627	0.600	0.600	-1.004	-1.897	-0.254	1.485	0.007	0.799	0.571	0.474	0.517	0.542	0.500	0.501
1992	0.632	0.627	0.582	0.614	0.656	0.617	-0.718	-1.093	-1.048	1.348	-0.667	0.623	0.551	0.543	0.470	0.522	0.587	0.528
1993	0.637	0.621	0.573	0.586	0.662	0.559	-0.903	-1.289	-0.715	1.368	-0.796	1.489	0.559	0.534	0.455	0.477	0.597	0.430
1994	0.631	0.641	0.619	0.609	0.666	0.641	0.273	-0.480	-0.207	0.785	-0.622	0.893	0.549	0.565	0.531	0.514	0.602	0.564
1995	0.666	0.652	0.656	0.655	0.711	0.668	0.787	-0.449	-0.469	0.239	-1.253	0.973	0.603	0.581	0.588	0.586	0.665	0.605
1996	0.762	0.661	0.649	0.643	0.690	0.656	1.061	-0.192	-0.733	0.814	-1.456	0.975	0.734	0.595	0.576	0.568	0.637	0.587
1997	0.738	0.683	0.664	0.674	0.647	0.737	1.145	-0.075	-0.541	0.600	-1.104	1.315	0.702	0.627	0.599	0.613	0.574	0.701
1998	0.707	0.694	0.689	0.852	0.619	0.743	1.433	-0.133	-0.131	1.449	-0.973	1.658	0.661	0.642	0.635	0.843	0.530	0.709
1999	0.782	0.705	0.689	0.791	0.581	0.582	1.442	0.112	0.130	1.233	-1.139	0.532	0.759	0.658	0.636	0.770	0.468	0.471
2000	0.425	0.318	0.680	0.357	0.533	0.625	-1.247	-2.387	-0.100	-1.590	-1.850	2.510	0.163	0.122	0.622	0.080	0.384	0.540
2001	0.586	0.664	0.535	0.520	0.560	0.725	-1.699	-2.436	1.490	-1.750	-2.204	2.312	0.476	0.599	0.389	0.359	0.432	0.685
2002	0.734	0.655	0.515	0.427	0.548	0.699	-1.960	-2.234	1.804	-2.084	-2.496	1.138	0.697	0.587	0.351	0.166	0.411	0.650
2003	0.742	0.680	0.628	0.643	0.731	0.656	-2.234	-2.463	1.554	-2.006	-2.708	1.224	0.708	0.623	0.545	0.568	0.693	0.587
2004	0.684	0.726	0.691	0.717	0.791	0.625	-2.356	-2.658	1.549	-2.351	-3.223	1.183	0.628	0.686	0.639	0.674	0.770	0.540
2005	0.639	0.679	0.686	0.598	0.732	0.578	-1.871	-1.873	1.508	-1.751	-2.098	1.255	0.561	0.621	0.632	0.497	0.694	0.464
2006	0.630	0.679	0.638	0.659	0.790	0.584	-2.026	-2.311	1.385	-1.524	-2.400	1.171	0.548	0.621	0.560	0.592	0.769	0.473
2007	0.621	0.612	0.566	0.547	0.607	0.587	-2.075	-2.272	1.455	-1.198	-2.191	1.108	0.534	0.520	0.443	0.410	0.512	0.478
2008	0.617	0.443	0.255	0.266	0.248	0.587	-2.112	-2.245	1.435	-1.319	-2.640	0.668	0.528	0.203	0.636	0.498	0.565	0.478
2009	0.607	0.514	0.580	0.582	0.540	0.579	-2.401	-2.602	0.691	-0.880	-2.123	1.017	0.511	0.348	0.467	0.470	0.396	0.465
2010	0.629	0.669	0.619	0.614	0.597	0.619	-2.094	-2.110	1.025	-0.981	-1.758	-0.654	0.545	0.606	0.531	0.523	0.496	0.530
2011	0.699	0.693	0.601	0.581	0.581	0.589	-2.180	-1.972	0.774	-0.796	-1.722	-0.965	0.650	0.642	0.501	0.468	0.469	0.483
2012	0.642	0.649	0.678	0.638	0.684	0.628	-1.268	-1.860	1.106	-0.813	-1.533	-0.296	0.567	0.577	0.620	0.560	0.629	0.545
2013	0.696	0.696	0.657	0.658	0.685	0.637	-1.228	-1.970	1.207	-0.832	-1.747	-0.295	0.646	0.646	0.589	0.590	0.630	0.558

Table 9: Certainty Equivalent Returns for Empirical Results

This Table illustrates the monthly CEQ performance for the S&P 500 portfolio Index (SPX), and the out-of-sample CEQ performance for the Forecast model strategies. We used the same CEQ calculation as DeMiguel et al (2009),

$$\widehat{CEQ} = \hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2,$$

where, $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are mean and variance of the out-of-sample excess returns for each model strategy k and risk-aversion γ.The time period is (01/1983 - 04/2014). These results are based on the out-of-sample prediction of the five forecasting model portfolios: (SENT) Baker and Wurgler Sentiment Index, (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the benchmark prediction variable: (VWDP) Baker and Wurgler Value-Weighted Dividend Premium. All statistics are estimated using a 120-month rolling estimation window. P-values are reported in parentheses to represent the difference between the S&P 500 Index (SPX) portfolio's Sharpe ratio compared to each forecast index. Any p-value below 0.050 would suggest the Sharpe ratios are statistically different.

	Logistic Regression Cut-off (0.50)					Logistic Regression Cut-off (0.75)				
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$
SPX	0.00350	0.00255	0.00159	0.00064	-0.00031	0.0035	0.00255	0.00159	0.00064	-0.00031
SENT	0.00446	0.00352	0.00258	0.00163	0.00069	-0.01107	-0.01204	-0.01301	-0.01398	-0.01494
	(0.00778)	(0.02279)	(0.05715)	(0.12325)	(0.23005)	(0.99999)	(1.00000)	(1.00000)	(1.00000)	(1.00000)
SI	0.00616	0.00522	0.00429	0.00335	0.00242	-0.01132	-0.01228	-0.01325	-0.01421	-0.01518
	(0.00075)	(0.00289)	(0.00958)	(0.02712)	(0.06593)	(0.99999)	(1.00000)	(1.00000)	(1.00000)	(1.00000)
ADS	0.00403	0.00307	0.00212	0.00116	0.00021	-0.01213	-0.01309	-0.01405	-0.01501	-0.01597
	(0.01312)	(0.55297)	(0.71057)	(0.1695)	(0.29666)	(1.00000)	(1.00000)	(1.00000)	(1.00000)	(1.00000)
GZ	0.00533	0.00439	0.00345	0.00251	0.00157	-0.00979	-0.01076	-0.01174	-0.01271	-0.01369
	(0.0025)	(0.00845)	(0.02442)	(0.06048)	(0.12897)	0.99991	0.99998	(1.00000)	(1.00000)	(1.00000)
FUI	0.00285	0.00188	0.00091	-0.00006	-0.00102	-0.01267	-0.01363	-0.01458	-0.01554	-0.0165
	(0.5915)	(0.74417)	(0.86007)	(0.93374)	(0.97299)	(1.00000)	(1.00000)	(1.00000)	(1.00000)	(1.00000)
VWDP	0.00372	0.00277	0.00182	0.00086	-0.00009	-0.01167	-0.01264	-0.0136	-0.01456	-0.01552
	(0.43907)	(0.60572)	(0.75475)	(0.86658)	(0.93704)	(1.00000)	(1.00000)	(1.00000)	(1.00000)	(1.00000)

Table 10: Summary Statistics and Correlation Matrix

Panel A: Represents the summary statistics associated with each of the forecasting indices. The indices represented are for the sample period (01/1985 - 04/2014). All columns represent monthly returns. SPX1 represents the S&P 500 Index one-month-ahead returns. The forecasting indices represent the monthly raw data value. (SI) is the Huang, Jiang, and Tu Sentiment Index, (ADS) is the Arubold-Diebold-Scotti Business Conditions Index, (GZ) is the Gilchrist-Zakrajsek Credit Spread Index, (EPU) is the Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) is the Jurado, Ludvigson, and Ng Financial Uncertainty Index.

Panel A: Summary Statistics

	SPX1	SI	ADS	GZ	EPU	FUI
mean	0.008	-0.173	-0.126	2.071	0.034	0.907
sd	0.044	0.732	0.705	1.002	0.293	0.184
min	-0.218	-1.440	-3.948	0.980	-0.601	0.643
max	0.132	3.085	1.872	7.824	1.934	1.549
skew	-0.820	1.926	-2.033	2.475	2.476	0.813
kurtosis	2.379	4.491	7.170	8.740	12.282	0.303

Panel B: Represents the correlation matrix between each of index introduced in Panel A above. All correlations are calculated from the monthly raw values of each index during the sample period (01/1985 - 04/2014).

Panel B: Correlation Matrix

	SPX1	SI	ADS	GZ	EPU	FUI
SPX1	1.000					
SI	-0.177	1.000				
ADS	0.129	-0.174	1.000			
GZ	-0.135	0.437	-0.713	1.000		
EPU	-0.151	0.064	-0.042	0.058	1.000	
FUI	-0.168	0.510	-0.443	0.686	0.119	1.000

Table 11: Correlation of Predictions

This table illustrates the correlations from (01/1985 - 04/2014) between the one-month-ahead S&P 500 Index (SPX1) return, the UP movement of the S&P 500 Index, the return predictions associated with each of the five forecasting models: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model.

Correlation Matrix of Monthly Return Predictions

	SPX1	UP	SI	ADS	GZ	EPU	FUI	All
SPX1	1.000							
UP	0.763	1.000						
SI	0.224	0.069	1.000					
ADS	0.074	0.011	0.251	1.000				
GZ	0.111	0.055	0.266	0.427	1.000			
EPU	0.050	-0.039	0.290	0.087	0.151	1.000		
FUI	0.062	-0.000	0.209	0.563	0.633	0.130	1.000	
All	0.151	0.036	0.520	0.469	0.381	0.418	0.498	1.000

Table 12: Annual Holding Period Returns for Each Forecasting Model

This table illustrates annual holding period returns from (01/1988 - 04/2014) for the S&P 500 Index (SPX), the LASSO selection criteria model, and the five forecasting model portfolios: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model.

Annual Return Series (01/1988 - 04/2014)

	SI	ADS	GZ	EPU	FUI	All	LASSO	SPX
1988	0.149	0.149	0.149	0.149	0.149	0.061	0.149	0.149
1989	0.106	0.106	0.106	0.106	0.106	-0.162	0.106	0.106
1990	0.045	-0.118	0.045	0.045	0.045	0.197	0.045	0.045
1991	0.189	0.189	0.189	0.039	0.189	0.039	0.086	0.189
1992	0.073	0.073	0.073	-0.008	0.073	-0.008	0.073	0.073
1993	0.098	0.098	0.098	0.098	0.098	0.098	0.098	0.098
1994	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023	-0.023
1995	0.352	0.352	0.352	0.352	0.352	0.352	0.352	0.352
1996	0.236	0.236	0.236	0.236	0.236	0.236	0.236	0.236
1997	0.247	0.247	0.247	0.140	0.247	0.140	0.247	0.247
1998	0.305	0.305	0.305	0.305	0.152	0.545	0.305	0.305
1999	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090
2000	-0.061	-0.020	0.064	-0.074	-0.020	-0.104	-0.020	-0.020
2001	-0.039	-0.173	-0.307	-0.202	-0.202	-0.073	-0.017	-0.173
2002	-0.243	-0.243	-0.020	-0.243	-0.214	-0.243	0.263	-0.243
2003	0.322	0.322	0.143	0.322	0.322	0.322	-0.057	0.322
2004	0.044	0.044	0.044	0.044	0.044	0.044	0.044	0.044
2005	0.084	0.084	0.084	0.084	0.084	0.084	0.084	0.084
2006	0.124	0.124	0.124	0.124	0.124	0.124	0.124	0.124
2007	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042	-0.042
2008	0.146	0.146	0.259	-0.156	0.368	0.368	0.102	-0.401
2009	0.300	0.018	0.017	0.300	-0.123	0.017	0.746	0.300
2010	0.370	0.198	0.198	0.198	0.198	0.423	-0.084	0.198
2011	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020
2012	0.141	0.141	0.141	0.141	0.141	0.141	0.135	0.141
2013	0.190	0.190	0.190	0.190	0.190	0.190	0.190	0.190
2014	0.050	0.050	0.050	0.050	0.050	0.050	-0.036	0.050

Table 13: Summary Statistics of Monthly Portfolio Returns

This table illustrates the summary statistics and return distributions using monthly holding period returns on an annual basis from (01/1988 - 04/2014). The buy and hold portfolio is represented by the S&P 500 Index (SPX), and the model selection portfolio is the lasso model (LASSO). The five forecasting model portfolios are represented as follows: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model.

	SI	ADS	GZ	EPU	FUI	All	LASSO	SPX
Minimum	-0.1458	-0.1458	-0.1458	-0.1458	-0.1458	-0.1100	-0.1458	-0.1694
Quartile 1	-0.0168	-0.0184	-0.0179	-0.0189	-0.0181	-0.0189	-0.0175	-0.0179
Median	0.0120	0.0111	0.0112	0.0103	0.0109	0.0101	0.0117	0.0112
Arithmetic Mean	0.0101	0.0080	0.0089	0.0071	0.0082	0.0086	0.0098	0.0073
Geometric Mean	0.0092	0.0071	0.0080	0.0062	0.0074	0.0077	0.0089	0.0064
Quartile 3	0.0360	0.0348	0.0351	0.0338	0.0348	0.0348	0.0351	0.0348
Maximum	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1694	0.1116
Stdev	0.0415	0.0420	0.0418	0.0421	0.0419	0.0419	0.0416	0.0421
Skewness	-0.1710	-0.1886	-0.1026	-0.2111	-0.1176	0.1255	-0.0875	-0.6024
Kurtosis	0.9803	0.9187	0.8773	0.9038	0.8710	0.6813	0.8901	1.1801

Table 14: Probability Estimates and Rank Sum Test

This table summarizes the Z-statistic and relative p-values resulting from the Kruskal-Wallis rank sum test and the pairwise Dunn test. The time period is from 01/1985 - 04/2014. (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices "kitchen sink" is represented as the (ALL) model, and the model selection index (LASSO).

	SI	ADS	GZ	EPU	FUI	ALL	$LASSO$
SI	0	1.7456	0.2231	0.2614	0.3895	0.3906	1.1156
	0	(0.8429)	1.0000	1.0000	1.0000	1.0000	1.0000
ADS	1.7456	0	1.5224	2.0069	1.3561	2.1362	2.8612
	(0.8429)	0	1.0000	(0.4699)	1.0000	(0.3429)	(0.0443)*
GZ	0.2231	1.5224	0	0.4845	0.1664	0.6138	1.3388
	1.0000	1.0000	0	1.0000	1.0000	1.0000	1.0000
EPU	0.2614	2.0069	0.4845	0	0.6509	0.1293	0.8542
	1.0000	(0.4699)	1.0000	0	1.0000	1.0000	1.0000
FUI	0.3895	1.3561	0.1664	0.6509	0	0.7801	1.5051
	1.0000	1.0000	1.0000	1.0000	0	1.0000	1.0000
ALL	0.3906	2.1362	0.6138	0.1293	0.7801	0	0.7250
	1.0000	(0.3429)	1.0000	1.0000	1.0000	0	1.0000
$LASSO$	1.1156	2.8612	1.3388	0.8542	1.5051	0.7250	0
	1.1156	(0.0443)*	1.0000	1.0000	1.0000	1.0000	0

$(pvalue < 0.05)*$

Table 15: Model Monthly Forecasting Errors

This table summarizes each model's monthly forecasting errors, on a annual basis. The buy and hold portfolio is represented by the S&P 500 Index (SPX), and the model selection portfolio is the lasso model (LASSO). The five forecasting model portfolios are represented as follows: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model. Each forecasting portfolio with the exception of the SPX, has two forecasting error columns: ER1 and ER2. ER1 refers to the error of predicting the market would go down, but the market went up: Market Up - Predicted Down. ER2 represents when the prediction was for an up market, but the market declined: Market Down - Predicted Up. The bottom two rows illustrate the total "Incorrect" and "Correct" forecasting errors.

Year	SPX		SI		ADS		GZ		EPU		FUI		All		LASSO	
	UP	DOWN	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2	ER1	ER2
1988	6	3	0	3	0	3	0	3	0	3	0	3	1	3	0	3
1989	8	4	0	4	0	4	0	4	0	4	0	4	2	4	0	4
1990	5	7	0	7	2	7	0	7	0	7	0	7	2	4	0	7
1991	9	3	0	3	0	3	0	3	1	3	0	3	1	3	1	3
1992	8	4	0	4	0	4	0	4	1	4	0	4	1	4	0	4
1993	8	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4
1994	7	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
1995	10	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2
1996	10	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2
1997	9	3	0	3	0	3	0	3	1	3	0	3	1	3	0	3
1998	9	3	0	3	0	3	0	3	0	3	1	3	1	2	0	3
1999	7	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
2000	4	8	3	6	0	8	1	7	2	8	0	8	3	7	0	8
2001	6	6	3	4	0	6	4	5	1	6	1	6	3	4	4	1
2002	4	8	0	8	0	8	3	5	0	8	2	7	0	8	4	0
2003	9	3	0	3	0	3	2	1	0	3	0	3	0	3	6	1
2004	9	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3
2005	5	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
2006	11	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
2007	7	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5
2008	4	8	1	6	1	6	3	3	0	7	2	4	2	4	4	2
2009	9	3	0	2	3	1	4	1	0	3	5	1	4	1	0	2
2010	7	5	1	4	0	5	0	5	0	5	0	5	2	3	3	4
2011	5	7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
2012	9	3	0	3	0	3	0	3	0	3	0	3	0	3	1	3
2013	10	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2
2014	2	1	0	1	0	1	0	1	0	1	0	1	0	1	1	1
Total	197	115	8	107	6	111	17	101	6	114	11	108	23	100	24	92
Incorrect			115		117		118		120		119		123		116	
Correct			161		159		158		156		157		153		160	

Table 16: Annualized Performance Analysis

This table summarizes annualized performance and risk statistics for the S&P 500 Index, the five forecasting models. The buy and hold portfolio is represented by the S&P 500 Index (SPX), and the model selection portfolio is the lasso model (LASSO). The five forecasting model portfolios are represented as follows: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting indices is represented as the ALL model.

	SI	ADS	GZ	EPU	FUI	All	LASSO	SPX
Annualized Return	0.117	0.089	0.101	0.077	0.092	0.097	0.113	0.079
Annualized Std Dev	0.144	0.145	0.145	0.146	0.145	0.145	0.144	0.146
Annualized Sharpe (Rf=0%)	0.811	0.611	0.695	0.527	0.633	0.667	0.783	0.542
Sortino Ratio (MAR = 0%)	0.403	0.300	0.347	0.260	0.314	0.345	0.393	0.255
Omega (L = 0%)	1.869	1.631	1.728	1.539	1.656	1.694	1.834	1.557
Tracking Error	0.032	0.031	0.044	0.022	0.038	0.049	0.044	0.000
Annualised Tracking Error	0.110	0.109	0.152	0.076	0.131	0.170	0.153	0.000
Information Ratio	0.341	0.089	0.142	-0.029	0.099	0.104	0.220	
Semi Deviation	0.030	0.031	0.030	0.031	0.030	0.029	0.030	0.032
Gain Deviation	0.027	0.026	0.027	0.026	0.027	0.028	0.027	0.024
Loss Deviation	0.027	0.028	0.026	0.028	0.027	0.024	0.026	0.031
Maximum Drawdown	0.404	0.463	0.416	0.504	0.445	0.432	0.230	0.526
Up Capture	0.928	0.897	0.764	0.938	0.843	0.685	0.732	1.000
Down Capture	0.670	0.784	0.507	0.917	0.680	0.408	0.387	1.000

Table 17: Predictive Logistic Regression Estimation Results

This table illustrates each coefficient (%), Newey-West t-statistic, and pseudo (McFadden) R^2 associated with the logistic regression model using a maximum likelihood approach following the same methodology of Mascio (2017). A kitchen sink variable, denoted ALL, is created by combining the five forecasting models using the following multivariate logistic regression,

$$\log \frac{p(x_{t+1}=1)}{1-p(x_{t+1}=1)} = a + \sum_{k=1}^{5} b_k F_{k,t}.$$

Each observation is the last month (December) of each year during the forecasting period (01/1985 - 04/2014). These results are based on the prediction of the five forecasting model portfolios: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the Lasso model is denoted *lasso*; its estimates are only shown under column 7 in the β (%) reporting section. All statistics are estimated using a 24-month rolling estimation window. (t-statistics above +/- 2.00, 2.50, and 3.00 are significant at the 10%, 5%, and 1% levels respectively)

	β (%)							$t-statistic$						$pseudo-R^2$					
	SI	ADS	GZ	EPU	FUI	ALL	lasso	SI	ADS	GZ	EPU	FUI	ALL	SI	ADS	GZ	EPU	FUI	ALL
1987	0.364	0.680	0.756	0.734	0.499	0.942	0.667	1.000	0.139	-1.130	-0.986	-1.471	1.945	0.618	0.381	0.552	0.505	0.316	0.605
1988	0.644	0.634	0.644	0.747	0.762	0.758	0.625	0.184	-0.839	0.650	-1.026	-0.765	1.693	0.295	0.269	0.293	0.531	0.565	0.556
1989	0.695	0.665	0.695	0.667	0.686	0.892	0.667	-0.897	-0.455	-0.444	0.682	0.539	1.603	0.416	0.347	0.416	0.350	0.395	0.817
1990	0.800	0.812	0.598	0.578	0.411	0.897	0.471	-1.265	-1.022	1.590	-0.488	-1.332	2.139	0.643	0.667	0.176	0.120	0.426	0.825
1991	0.252	0.683	0.938	0.543	0.611	0.974	0.688	-1.242	-1.722	2.228	-0.534	-0.747	2.295	0.665	0.389	0.697	0.020	0.210	0.857
1992	0.727	0.563	0.714	0.735	0.673	0.820	0.708	-1.504	-0.924	0.883	-0.361	0.227	1.780	0.488	0.080	0.459	0.507	0.365	0.681
1993	0.680	0.866	0.665	0.668	0.718	0.884	0.667	-0.305	1.143	0.013	-0.023	-0.660	1.429	0.381	0.770	0.347	0.353	0.468	0.802
1994	0.582	0.791	0.615	0.624	0.416	0.877	0.625	0.332	1.235	0.763	0.813	1.186	1.866	0.131	0.625	0.222	0.242	0.406	0.789
1995	0.662	0.728	0.980	0.731	0.665	0.983	0.708	-0.962	-0.350	1.570	-0.388	-1.152	1.885	0.338	0.492	0.668	0.498	0.346	0.873
1996	0.795	0.765	0.810	0.983	0.717	1.000	0.875	-0.399	-1.226	0.443	-1.961	-0.919	1.989	0.631	0.570	0.661	0.573	0.466	0.852
1997	0.814	0.823	0.862	0.710	0.677	0.844	0.792	-0.168	-1.672	0.916	-1.318	-1.047	2.024	0.670	0.688	0.762	0.451	0.374	0.728
1998	0.777	0.741	0.982	0.749	0.729	1.000	0.708	-0.195	0.243	0.994	0.035	-0.350	1.363	0.596	0.519	0.572	0.537	0.492	0.888
1999	0.858	0.999	0.721	0.655	0.680	1.000	0.667	0.452	2.158	1.056	-1.117	0.118	1.980	0.754	0.698	0.476	0.322	0.382	0.749
2000	0.363	0.377	0.370	0.455	0.418	0.320	0.302	-0.688	0.483	-0.372	0.053	-0.515	1.422	0.626	0.564	0.594	0.261	0.399	0.825
2001	0.546	0.465	0.381	0.332	0.344	0.193	0.417	-0.604	-0.945	1.362	1.289	0.436	1.927	0.031	0.228	0.546	0.767	0.710	0.755
2002	0.411	0.398	0.621	0.403	0.158	0.539	0.494	0.058	-0.757	2.094	0.861	1.762	2.106	0.424	0.476	0.236	0.458	0.663	0.789
2003	0.788	0.605	0.626	0.461	0.868	0.043	0.581	-2.624	0.494	-0.480	-0.845	-1.697	2.228	0.617	0.194	0.248	0.242	0.774	0.780
2004	0.970	0.872	0.863	0.706	0.784	0.994	0.708	-1.783	0.958	-1.166	0.264	-0.488	1.932	0.685	0.780	0.763	0.442	0.609	0.790
2005	0.467	0.588	0.556	0.647	0.300	0.335	0.625	1.171	0.022	-0.139	0.865	1.225	1.684	0.221	0.149	0.060	0.303	0.557	0.752
2006	0.570	0.644	0.537	0.664	0.971	0.996	0.667	-0.957	-0.365	0.879	-0.389	-2.046	1.927	0.097	0.294	0.003	0.343	0.552	0.779
2007	0.176	0.709	0.679	0.749	0.222	0.783	0.750	-1.900	0.614	-0.234	-1.090	-1.658	2.099	0.522	0.449	0.380	0.536	0.587	0.609
2008	0.231	0.200	0.314	0.478	0.248	0.903	0.199	-1.154	0.990	-0.545	-0.575	-1.132	1.879	0.457	0.664	0.556	0.669	0.594	0.836
2009	0.448	0.603	0.564	0.526	0.548	0.041	0.021	-1.606	0.481	-0.162	-0.964	-0.070	1.657	0.632	0.698	0.726	0.712	0.711	0.874
2010	0.940	0.659	0.676	0.734	0.702	0.961	0.667	-1.796	-0.121	-0.111	-0.893	-0.267	1.638	0.600	0.332	0.371	0.504	0.432	0.836
2011	0.455	0.497	0.676	0.449	0.470	0.990	0.839	0.341	-0.228	1.127	-0.685	0.394	1.555	0.262	0.121	0.373	0.285	0.209	0.883
2012	0.546	0.902	0.474	0.584	0.722	0.997	0.583	-0.501	1.340	1.558	-0.257	-0.922	1.916	0.030	0.835	0.543	0.137	0.478	0.795
2013	0.838	0.891	0.726	0.862	0.804	0.979	0.792	-0.718	1.478	0.380	0.882	-0.256	1.743	0.716	0.815	0.487	0.762	0.650	0.865

Table 18: Annualized Return Rankings among all Strategies

This table ranks each individual forecasting strategy based on its mean annual return between (01/1988 - 04/2014). These results are based on the prediction of the five forecasting models, the combination of the forecasting models, and the model selection strategy. The model selection portfolio is the lasso (LASSO), and the five forecasting model portfolios are represented as: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all five forecasting models is represented as the all portfolio (ALL). The buy and hold portfolio is the S&P 500 Index (SPX).

	Return	Std Dev	Sharpe	Drawdown
SI	0.117	0.144	0.811	0.404
LASSO	0.113	0.144	0.783	0.230
GZ	0.101	0.145	0.695	0.416
ADS	0.089	0.145	0.611	0.463
ALL	0.097	0.145	0.667	0.432
FUI	0.092	0.145	0.663	0.445
SPX	0.079	0.146	0.542	0.526
EPU	0.077	0.146	0.527	0.504

Table 19: Annualized Performance Analysis without EPU

This table summarizes annualized return and risk information for the S&P 500 Index, the four forecasting models excluding the (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, the ALL combination model, and the model selection strategy. These results are based on the prediction of each of the forecasting straties. The model selection portfolio is the lasso (LASSO), and the four forecasting model portfolios are represented as: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, and the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. The combination of all four forecasting models is represented as the all portfolio (ALL), and the buy and hold portfolio is the S&P 500 Index (SPX).

	SI	ADS	GZ	FUI	All	LASSO	SPX
Annualized Return	0.117	0.089	0.101	0.092	0.092	0.126	0.079
Annualized Std Dev	0.144	0.145	0.145	0.145	0.145	0.143	0.146
Annualized Sharpe (Rf=0%)	0.811	0.611	0.695	0.633	0.634	0.879	0.542
Sortino Ratio (MAR = 0%)	0.403	0.300	0.347	0.314	0.317	0.455	0.255
Omega (L = 0%)	1.869	1.631	1.728	1.656	1.657	1.956	1.557
Tracking Error	0.032	0.031	0.044	0.038	0.043	0.048	0.000
Annualised Tracking Error	0.110	0.109	0.152	0.131	0.150	0.167	0.000
Information Ratio	0.341	0.089	0.142	0.099	0.087	0.280	
Semi Deviation	0.030	0.031	0.030	0.030	0.030	0.029	0.032
Gain Deviation	0.027	0.026	0.027	0.027	0.027	0.028	0.024
Loss Deviation	0.027	0.028	0.026	0.027	0.027	0.024	0.031
Maximum Drawdown	0.404	0.463	0.416	0.445	0.509	0.347	0.526
Up Capture	0.928	0.897	0.764	0.843	0.754	0.699	1.000
Down Capture	0.670	0.784	0.507	0.680	0.542	0.262	1.000

Table 20: Certainty Equivalent Returns for Empirical Results

This Table illustrates the monthly certainty equivalent (CEQ) performance for the S&P 500 portfolio Index (SPX), and the out-of-sample CEQ performance for the five forecast model strategies, and the combined models. We used the same CEQ calculation as DeMiguel et al (2009),

$$\widehat{CEQ} = \hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2,$$

where, $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are mean and variance of the out-of-sample excess returns for each model strategy k and risk-aversion γ. The time period is (01/1983 - 04/2014). These results are based on the out-of-sample prediction of the five forecasting model portfolios: (SI) Huang, Jiang, and Tu Sentiment Index, (ADS) Arubold-Diebold-Scotti Business Conditions Index, (GZ) Gilchrist-Zakrajsek Credit Spread Index, (EPU) Baker, Bloom and Davis Economic Policy Uncertainty Index, the (FUI) Jurado, Ludvigson, and Ng Financial Uncertainty Index. Also, the combined portfolios: the "kitchen sink" (All), and the LASSO. All statistics are estimated using a 24-month rolling estimation window. P-values are reported in parentheses to represent the difference between the benchmark (SPX) portfolio's Sharpe ratio compared to each forecast index. Any p-value below 0.050 would suggest the Sharpe ratios are different.

	Logistic Regression Cut-off (0.50)					Logistic Regression Cut-off (0.75)				
	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$
SPX	0.0035	0.00255	0.00159	0.00064	-0.00031	0.00293	0.00196	0.001	0.00003	-0.00093
SI	0.00616	0.00522	0.00429	0.00335	0.00242	-0.00347	-0.00446	-0.00545	-0.00643	-0.00742
	(0.00075)	(0.00289)	(0.00958)	(0.02712)	(0.06593)	(0.78215)	(0.88016)	(0.94184)	(0.97518)	(0.99070)
ADS	0.00403	0.00307	0.00212	0.00116	0.00021	-0.00769	-0.00867	-0.00965	-0.01063	-0.01161
	(0.10461)	(0.03587)	(0.08403)	(0.16950)	(0.29666)	(0.99962)	(0.99790)	(0.99941)	(0.99986)	(0.99997)
GZ	0.00533	0.00439	0.00345	0.00251	0.00157	-0.00893	-0.0099	-0.01088	-0.01185	-0.01283
	(0.00250)	(0.00845)	(0.02442)	(0.06048)	(0.12897)	(0.99852)	(0.99960)	(0.99990)	(0.99998)	(1.00000)
EPU	0.00446	0.00352	0.00258	0.00163	0.00069	-0.00784	-0.00883	-0.00981	-0.01080	-0.01178
	(0.00778)	(0.02279)	(0.05715)	(0.12325)	(0.23005)	(0.99428)	(0.99822)	(0.99951)	(0.99988)	(0.99997)
FUI	0.00285	0.00188	0.00091	-0.00006	-0.00102	-0.00734	-0.00833	-0.00931	-0.01030	-0.01128
	(0.59150)	(0.74417)	(0.86007)	(0.93374)	(0.97299)	(0.99998)	(1.00000)	(1.00000)	(1.00000)	(1.00000)
All	0.00366	0.00243	0.00146	0.00049	-0.00049	-0.00083	-0.00181	-0.0028	-0.00379	-0.00477
	(0.45029)	(0.65944)	(0.79871)	(0.89661)	(0.95420)	(0.93508)	(0.97206)	(0.98949)	(0.99655)	(0.99901)
LASSO	0.00772	0.00602	0.00507	0.00411	0.00316	0.00414	0.00318	0.00223	0.00128	0.00032
	(0.00006)	(0.00096)	(0.00365)	(0.01184)	(0.03282)	(0.01050)	(0.02774)	(0.06405)	(0.12968)	(0.23140)

Sentiment Indices and their Forecasting Ability

1 Introduction

"Beta optimization" is defined as a market timing strategy designed to change the beta of a portfolio as equity market expectations change over time. The traditional market timing strategy attempts to generate excess returns by switching between common stocks and bonds or cash equivalents as the stock market fluctuates. If an investor can consistently identify the turning points in the market, the timing strategy can substantially enhance investment performance. Originally, the market timing objective was to be long common stocks (100%) during bull markets and long cash equivalents (100%) during bear markets. Timers make their portfolios more or less sensitive to the market by switching from stocks (beta approximately 1.0) to bonds or cash (beta approximately 0.0) and back as their outlook for the market changes.

Portfolio betas can be increased to 1.0 through the use of leverage or can be made negative by shorting stocks or index futures. The beta optimization market timing objective is to create a portfolio with (1) a high beta when the market is expected to go up, and (2) a beta of 0.0 when the market is expected to be neutral. When a bear market is expected a conservative investor will use cash or bonds to create (3) a zero beta portfolio while a more aggressive manager will use short positions to create (4) a negative beta portfolio. Successful application of this strategy generates returns that are consistent with a long call option position when using cash or bonds; a look-back option straddle payoff is observed when short equity positions are used to create a negative beta. In either case, positive returns are generated in both up and down markets.

Research into market timing has a long history. For example, Treynor and Mazuy (1966) point out that successful market timing managers will shift the composition of their portfolios between more and less volatile securities as their outlook for the markets change over time. Merton (1981) points out that a market timer or "macro-forecaster" tries to predict whether stocks (bonds) will outperform bonds (stocks), and that the returns generated by successful market timers' returns are "virtually indistinguishable" from successful option strategies. The resultant payoffs for successful market timers resemble that of a call option; funds are in stocks during market increases and in bonds or cash during stock market declines. If the manager has the ability to use short selling, a look-back option payoff can be generated.

The success of any timing strategy depends on successful market forecasting and market participants are continually experimenting with modeling changes in financial markets. While a brief overview of forecasting is presented below, in this research effort five recently developed indices are used to forecast the direction of the market (the S&P 500) in the upcoming month. The S&P 500 Index (SPX) with a beta of 1.0 is considered the market portfolio.

In the results presented below, if a bull market is forecast, a long position in the SPX results in a beta of 1.0. Obviously, in practice the beta of the portfolio can be increased through the use of leverage or margin. If a bear market is expected, a zero beta is achieved by moving from the SPX into cash equivalents, or a beta of -1.0 is generated by going short the SPX. Both strategies are examined and compared.

This paper examines the efficacy of the beta optimization strategy using three types of market forecasting models, and the market timing results of these models are compared to a benchmark index by Baker and Wurgler (2004)[1]. An investor sentiment index developed by Baker and Wurgler (2007) represent the first type of model. This model combine six proxies for sentiment to generate a "sentiment" index.

The second two models are the ADS index from the Philadelphia Federal Reserve Bank developed by Aruoba, Diebold and Scotti (2009) and the improved investor sentiment index developed by Huang et al (2015). The ADS model combines quarterly, monthly and

[1] See Appendix: A for a detailed description of each model

weekly macroeconomic information to develop an outlook for the overall economy, while the Huang et al (2015) model was created to forecast aggregate stock market performance, and act as a proxy for macroeconomic variables.

The final two models are the Gilchrist and Zakrajsek (2012) GZ Spread Index, and the Jurado, Ludvigson, and Ng (2015) Financial Uncertainty Index. The GZ Spread Index is a corporate fixed income credit spread model, and the financial uncertainty index is an independent time-varying econometric model. Both are created to predict the overall financial market risk.

The stock market forecasting ability of these five empirical forecasting models are demonstrated with the following criteria. First, how many monthly market movements, both up and down, are correctly forecast? Second, how effective are the models in predicting the big down-market movements, the "Black Swan" events that can devastate portfolios? Avoiding the worst months can enhance portfolio returns substantially; some models may be superior in forecasting severely down markets. Third, most importantly, what is the overall significance of the each model's predictions. Finally, all forecasts are compared to two perfect monthly timing strategies (or perfect forecasting strategies).

While the SPX represents a one-asset portfolio, the procedures outlined here can be useful for both individual investors and portfolio managers. Individual investors can switch between equity mutual funds and bond or money market funds while active portfolio managers can adjust the betas of their portfolios by increasing (decreasing) portfolio betas through stock selection and switching to bonds or money market instruments or to short positions as their outlook for the market changes.

The paper proceeds as follows. Section 2 reviews the theory behind positive and negative alphas and Section 3 presents the issues associated with market timing, stock returns, investor sentiment, and macroeconomic forecasting models. Section 4 reviews the value-weighted dividend premium benchmark. Data and methodology are presented in Section 5. Empirical results are shown in Section 6 while concluding remarks are in Section 7.

2 Origin of Positive and Negative Alphas

Since the development of the capital asset pricing model, CAPM, in the mid-1960s by Sharpe (1964), Lintner (1965), and Mossin (1966), beta, the measure of systematic risk, has been a major focus of both academic and practitioner research. The CAPM is an ex-ante model that relates an asset's required return to its expected risk. The model divides total risk, the asset's return variance, into systematic and unsystematic risk. Systematic risk is measured by beta and relates individual asset returns to the returns of a market index. Unsystematic or idiosyncratic risk is the variance of the residual terms in the regression. Theoretically, this risk may be diversified away and, hence, is not priced in efficient markets. Because expected returns are unavailable, empirical testing relies on the assumption that ex-post return distributions reflect ex-ante return distributions. In addition, early empirical tests generally assume that the probability distributions generating the ex-post returns are stationary over time. The original CAPM assumes all market participants share identical return distribution expectations.

The CAPM posits returns on individual assets as well as portfolios are explained by a risk-free rate and a risk premium, where the risk premium is a function of the asset's systematic risk, beta. This relation between returns and beta is known as the Security Market Line (SML). Very risk-averse investors can invest 100% in the risk-free rate, generally proxied by T-Bills; moderately risk-averse investors can vary their investment between the risk-free rate and the market portfolio. More aggressive or optimistic investors can invest 100% in the market portfolio. But to get returns greater than the market portfolio, according to the CAPM, the investor can lever the portfolio by borrowing at the risk-free rate and investing in the market index.

Early tests by Black, Jensen and Scholes (1972), Miller and Scholes (1972), and Fama and McBeth (1974) provide evidence that, consistent with theory, a linear relation exists between asset returns and beta. However, these studies also report that the empirical security market line is flatter than the theoretical security market line, with low-beta (high-beta) stocks generating positive (negative) alphas. Alpha is the difference between an asset's actual return and the return expected based on its risk. A positive (negative) alpha indicates the asset's return is greater (less) than the risk-adjusted expected return. A number of different reasons have been put forward to explain these results. For example, Black (1972, 1993) shows that a CAPM with restricted borrowing and short-selling impacts the efficient frontier and can result in low beta assets generating higher returns than high beta assets. With these restrictions, the theory suggests investors wanting returns greater than the market index must invest in higher beta stocks and, hence, bid up the prices of high beta stocks. This causes the subsequent empirical returns on high beta stocks to be less than their theoretical returns, flattening the empirical security market line. As a result, the returns of low beta stocks are above the SML and are considered to generate "positive alpha" while high beta stocks generate "negative alpha".

This "beta anomaly" has been the focus of much recent research. Baker, Bradley, and Wurgler (2011) examine separately months when returns are above and below their median. Consistent with extant theory, higher (lower) returns were earned by higher beta stocks in up (down) markets. However, on a risk-adjusted basis, the low beta anomaly was evident, and in fact, their empirical results indicate the risk-return relation over the long term becomes inverted. Finally, they suggest mutual fund managers' generally forego the use of leverage that would allow them to take advantage of the low beta anomaly.

Eisele (2012) and Frazzini and Pedersen (2014) create a "Betting Against Beta" (BAB) factor by leveraging a long position in low beta assets and taking a short position in high beta assets. Eisele reports the importance of the BAB factor is second only to the market factor and is more important than the size and value factors of Fama-French (1993) and the momentum factor of Carhart (1997). Frazzini and Pedersen (2014) report abnormal monthly returns and a Sharpe ratios higher than the momentum and value factors. Even with the Carhart (1997) momentum and Pastor and Stambaugh (2003) liquidity factors, the BAB factor still generates positive abnormal monthly returns. They argue that Berkshire

Hathaway, Warren Buffett's firm, successfully uses a BAB strategy by applying leverage in the purchase of low beta assets.

The present paper uses portfolio betas in a market timing framework based on the expected future equity market risk environment. Rather than attempting to arbitrage the differences in beta, the portfolio's beta is adjusted over time as risk expectations change. If the market is expected to increase (decrease), the portfolio is fully invested in the market (short the market). Thus, the beta optimization process is a finer tuning of the traditional market timing model.

3 Market Timing and Forecasting Strategies

Generally, there are two ways to "beat the market", market timing and superior stock selection. As mentioned above, the original market timing strategy moves monies between the stock market and the bond market or cash equivalents as expectations change over time. Empirical studies of market timing strategies report mixed results, especially when the frequency of switching is considered.

An early study by Sharpe (1975) uses annual returns and concludes a market timing strategy will not outperform the market unless the timer can be correct approximately 70% of the time. Merton (1981) points out that a market timer or "macro-forecaster" tries to predict whether stocks (bonds) will outperform bonds (stocks), and that the returns generated by successful market timers' returns are "virtually indistinguishable" from successful option strategies. Droms (1989) shows that with perfect monthly timing the manager only has to be correct 51% of the time to outperform the market. Interestingly, Phillips and Lee (1989) suggest that market timers may view risk differently from many money managers. They state "A market timer is not concerned with risk in a portfolio sense; risk to a market timer is being in or out of the market at the wrong time." (p 15).

Bollen and Busse (2001) suggest that fund managers may "...execute timing strategies dynamically" and conclude that managers utilize intra-month portfolio adjustments in their investment strategies. Similarly, Fung and Hsieh (2001) consider timing strategies and distinguishes between "market timers" and "trend followers". They use options to create a look-back straddle that models a "...primitive trend following strategy..." Their model exhibits positive skewness and delivers positive returns in extreme bull and bear markets, especially in the currency, commodity and interest rate markets. A "truly perfect timing strategy" is developed by Lam and Li (2004) using daily S&P 500 returns. For timers with low transactions costs, daily switching can generate excess annual returns and the correct prediction probability needed to outperform the market is about 60%.

Jiang, Yao, and Yu (2007) examine the market-timing ability of mutual funds by comparing return-based timing measures with holdings-based timing measures. They conclude that the average performance attributable to timing skills is positive and that the excess returns appear to be generated by managers' response to public information such as aggregate earnings-to-price ratios and aggregate dividend yields. Because Avramov and Wermers (2006) report that over the business cycle fund managers vary their industry composition substantially, Jiang, Yao, and Yu (2007) also examine industry shifts in mutual fund portfolios. Their findings with those of Avramov and Wermers; "...fund managers respond to macroeconomic information..." and tend to switch from high (low) beta industries to low (high) beta industries as market conditions change over time. Jiang, Yao and Yu (2007) conclude that "...on average, actively managed U.S. domestic equity funds possess positive timing ability."

Cremers and Petajisto (2009) evaluate mutual fund performance using an active share and timing measure, while Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014) consider both market timing and stock picking in their evaluation of fund managers. They define skill as a "...general cognitive ability to pick stocks or time the market." They report that a subset of active managers generate abnormal returns by successfully performing both tasks. Throughout the economic cycle, these managers hold more cash and rotate into defensive industries during recessions and, overall, their portfolios tend to have lower market betas. During economic expansions these successful managers invest in cyclical industries. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014) also suggest that managers with the necessary cognitive skills will focus more on macroeconomic information during recessionary periods and microeconomics factors during economic expansions. That is, their focus shifts between the aggregate economy and firm-specific information over the economic cycle.

Successful market timers generate non-linear payoffs. Hubner (2010) points out that many hedge fund strategies are intended to provide convex payoffs while traditional performance measures are designed to evaluate performance in a stationary mean-variance environment. Using linear measurement tools to appraise convex payoffs may cause positive (negative) timers to incorrectly exhibit negative (positive) performance. Furthermore, Jagannathan and Korajczyk (2015) point out that the CAPM divides total variation into systematic and idiosyncratic components and that the idiosyncratic risk is assumed to be unpredictable. However, it is this component that generates the abnormal return or the "alpha" for a portfolio. They state "…the ability to forecast the idiosyncratic returns of assets… represents skill." Because certain asset classes have non-linear, option-like payoffs, they observe that dynamic hedge fund strategies and strategies that use derivative instruments can provide nonlinear payoffs. Traditional performance measures are inadequate and require measures that can identify stock picking and market timing abilities.

Fisher and Statman (2003) succinctly outline the issues associated with consumer confidence, investor sentiment, and stock returns. "We know that consumer confidence predicts economic activity. Consumer confidence is a component of the Index of Leading Economic Indicator, and it predicts household expenditures. But does consumer confidence also predict stock returns? Do stock returns affect consumer confidence? And what is the relationship between confidence and investor sentiment?" (p.115)

Fisher and Statman (2003) report a statistically significant positive relation between changes in consumer confidence and contemporaneous stock returns. However, they observe that consumer confidence may be a contrarian indicator. Confidence is reflected in the bidding up of stock prices today; prices are bid too high and a negative relation is observed between consumer confidence and future stock returns. They also report a statistically significant positive relation between changes in consumer confidence and changes in investor sentiment for individual investors, but no relation with changes in institutional investor sentiment. Using surveys of both individual and professional investors, Brown and Cliff (2004) examine the relations between survey-type sentiment measures and technical sentiment indicators.[2] Their results show that market returns predict future individual and institutional investor sentiment, but find little evidence that the sentiment measures can predict future stock returns. In this paper we show evidence that is contrary their findings. We demonstrate through our beta optimization that investor sentiment does in fact predict future stock returns.

4 Benchmark Index

In order to determine if the five forecast indices are superior predictors of the one-month ahead returns on the S&P 500 Index, we compare their forecast ability with a well-established benchmark index developed by Baker and Wurgler (2004) called the value weighted dividend premium (VWDP). The index is calculated by taking the difference between in the logs of the average market-to-book ratio of firms that pay dividends (dividend payers), and firms that do not pay dividends (non-dividend payers). These ratios are log-normally distributed, and each firm is value-weighted in the model[3].

This simple model was initially developed to determine the level at which managers would pay dividends to its shareholders[4]. Baker and Wurgler (2004) discover that when

[2]They distinguish between individual and professional investors by using survey results from the American Association of Individual Investors (AAII) and Investors Intelligence (II). The technical indicators are measures of trading volume, type of trade, derivatives, and "other".

[3]The market-to-book value of each firm is calculated using the Fama and French (2001) methodology

[4]According to Baker and Wurger (2004) the lagged dividend premium variable explains nearly 60% of

the dividend premium proxy is low(high), the inflation rate increase(decrease). This indirectly results in equity mispricing associated with dividend payments. Their results also suggest, "dividends are highly relevant to a company's share price, but in different directions at different times." Moreover, Shleifer and Vishny (2003) conclude that dividend premium variables are possibly driven by investor sentiment resulting in market timing strategies designed to exploit inefficiencies within the capital markets. Therefore, instead of benchmarking the investor sentiment models used in this paper to a single stock dividend variable, we use a simple non-estimated proxy for dividends to compare effectiveness.

5 Data and Methodology

5.1 Data

This research effort uses five models to forecast the movement of the stock market one month ahead. The models are the Baker and Wurgler (2007) sentiment index (SENT), the Huang, Jiang and Tu (2015) Improved Sentiment Index (SI), the Aruoba, Diebold, and Scotti (2009) Market Conditions Index (ADS), the the Gilchrist and Zakrajsek (2012) Credit Spread Index (GZ), and the Jurado, Ludvigson, and Ng (2015) Financial Uncertainty Index (FUI). The investment results for each forecasting model are compared on both a cross-sectional and time-series basis.

The common date for all the forecast indices, the benchmark index, and the S&P 500 index in this study are from January 1, 1973 through April 30, 2014. The beginning and ending dates result in 592 months of data. Data comes from a number of sources; the month ending price of the S&P 500 index is taken directly from Bloomberg, and T-Bill rates are from CRSP. Data for the Baker and Wurgler (2007) sentiment index, SENT, was taken from Jeffrey Wurgler's website.[5] The Huang, Jiang and Tu (2015) improved sentiment index

the annual change in the inflation rate

[5] The website address is: http://people.stern.nyu.edu/jwurgler/.

data, SI, was received directly from the authors. The Philadelphia Federal Reserve Bank is the source of the data for the Aruoba, Diebold and Scotti (2009) ADS Business Conditions Index.[6] The GZ Credit Spread Index data comes directly from Gilchrist, Zakrajsek, and Favara (2016).[7], and the Financial Uncertainty Index (FUI) data are available from the Board of Governors of the Federal Reserve System.[8]

5.2 Methodology: Logistic Regression

The five forecasting indices form the basis for beta optimization. Each index uses information that can be employed to predict future market direction. The objective is to generate an "up" or "down" market prediction for the next month. Logistic regression procedures can be used when the dependent variable takes on one of two mutually exclusive values. In this study, when the market is predicted to be up (down) the subsequent month, the variable is assigned a value of 1 (0). Each of the five indices described above is used as a forecasting factor, $F_{k,t}$, where k refers to the indices, $k = 1, \dots 5$, and t is the month.

To begin the forecasting process we first categorize the market movement as "Up" if the market return in the next month was positive, and "down" otherwise – by doing this, a binomial (binary) variable x_{t+1} is created for each month t. Second, we line up the current month indices $F_{k,t}$ with the next month market movement x_{t+1}. Third, at the end of each month, we use the previous 120 months $(t - 120, t - 1)$ of observations to estimate the models shown below[9]

Defining the probability of an upmarket $x = 1$ as $p(x)$; the logistic regression model is

$$\log \frac{p(x_{t+1} = 1)}{1 - p(x_{t+1} = 1)} = \alpha_k + \beta_k F_{k,t}, \tag{1}$$

[6]The address for this index is: http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/

[7]The Fed Notes Web address link is: https:www.federalreserve.goveconresdata/notes/feds-notes

[8]https://www.federalreserve.gov/econresdata/workingpapers.htm

[9]We tested 24, 48, 72, 96, 120 as well as the entire sample period of 592 months to determine most the optimal estimation window. The criteria was based on accuracy of the prediction, mean monthly forecasted return and statistical significance (p-value). All results are available upon request.

where, $F_{k,t}$ is an index. Equivalently,

$$p(x_{t+1} = 1) = \frac{1}{1 + e^{-\left(\alpha_k + \beta_k F_{k,t}\right)}}. \tag{2}$$

The parameters are estimated using a maximum likelihood approach. We use the value of the index at the end of the month and the estimated logit model parameters using the previous 120 months to forecast the probability of the market up movement for the coming month (month 26th for the first case). An up market (down market) is predicted when $p(x = 1) \geq 0.5$ ($p(x = 1) < 0.5$.). That is, if $p(x) \geq 0.5$, we predict $x = 1$; if $p(x) < 0.5$, $x = 0$ is predicted.

Once the logistic predictions are generated for each forecasting model, assets are allocated to the S&P 500 Index (SPX) or allocated to a short SPX position. Portfolio return results are generated for each model.

5.3 Calculating Certainty Equivalents

To determine the risk appetite of an individual investor and their willingness to invest in equities versus the risk-free asset, we use DeMiguel et al (2009) to calculate the certainty equivalent (CEQ). We assess the economic significance of the expected returns of the five forecast models and the benchmark index compared to the S&P 500 Index. By comparing the average CEQ for a mean-variance investor with risk aversion of [$\gamma = 1, 2, 3, 4, 5$], a portfolio allocation of 100% in stocks can be determined when equal emphasis is placed on expected return and volatility. The average utility of the this type of model is based on the overall utility of each predictor during the forecast sample period. We calculate the CEQ performance for each k model in the following equation,

$$\widehat{CEQ} = \hat{\mu}_k - \frac{\gamma}{2}\hat{\sigma}_k^2, \tag{3}$$

where, $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the out-of-sample mean excess returns and variance for each model k, while γ is the measurement of an investor's risk aversion. In order to determine if the results of the CEQ are statistically different, we use the paired test for equality of Sharpe ratio based on Lo (2002)[10]. This approach compares the difference of each strategies' Sharpe ratio to that of the S&P 500 Index (SPX). If the resulting p-value of the difference is above 0.050, then we would accept the null-hypothesis that their is no difference between each model and the SPX.

6 Empirical Results

6.1 Descriptive Statistics

The first step in the process is to determine the sequence of returns that could be generated if the portfolio manager had perfect foresight for both (1) the traditional market timing strategy and (2) a long/short beta optimization strategy. First, using the S&P 500 Index (SPX), the returns from the traditional market timing approach is determined. If the S&P 500 Index is up for the month, the portfolio is 100% in the SPX; if the Index is down, the portfolio is 100% in T-bills. This is the SPX Risk-free (SPXRF) model. Trading costs are ignored for all timing models.[11] Next, using the beta optimization framework, in months when the market is up, the strategy invests 100% in the SPX, resulting in a beta of 1.0. Conversely, when the S&P 500 Index is down, a 100% short position in the SPX is taken; the portfolio now has a beta of -1.0.

Figure 1 shows graphically the results of *perfect* foresight. The horizontal axis shows the returns (red/purple line) on the S&P 500 Index (SPX) ranged from a low of -40.1% in 2008 to a high of 35.2% in 1995. The blue/purple line reflects the traditional market timing

[10]This test statistic performs a hypothesis test of equality of Sharpe ratios of p assets given paired observations. Steven E. Pav formally developed the statistic in R as the SharpeR package

[11]Full service discount broker, Charles Schwab, offers zero trading cost brokerage accounts see http://www.schwab.com/onesource

strategy, which is 100% in T-Bills when S&P 500 Index returns are negative and 100% in the SPX purple line when the index is positive. This strategy is designated as SPXRF. The more aggressive market timing model (purple line) generates positive returns by taking a short position in the SPX when the S&P 500 Index is negative, and a long position in the SPX when the S&P 500 Index is positive. This strategy is the SPXLS model. All four models overlap to the right of the zero midpoint.[12]

The monthly data for the initial analysis starts in January, 1973 and goes through April, 2014. For example, in 2008, the lowest performance of the the S&P 500 Index (SPX) during the sample period returned -40.1%. Conversely, the SPXRF was up 9.1%, and the SPYLS was up 83.5%, representing a short SPX position. On the positive end, the best performance of the S&P 500 Index during the sample period was in 1995. The SPX was up 35.2%, the SPXRF was up 36.6%, and the SPXLS was up 37.1%. It can be seen in Figure 1 that the SPXRF strategy has the payoff of a long call position as suggested by Treynor and Mazuy (1966) and the SPXLS strategy generates a payoff consistent with the option straddle position described by Fung and Hsieh (2001).

While perfect market timing is not possible in the real world, the efficacy of different forecasting models is compared with a simple buy-and-hold strategy. Table 1 shows summary statistics and correlations for the five forecasting models. Panel A shows a mean value of 7 basis points for the returns of the one-month ahead S&P 500 Index, SPX1, for the 592 months of return data. The SPX1 correlation with the five forecasting models is less than 0.002, as shown in Panel B.

The last five columns in Panel A present the statistics of the five indices that are used for forecasting the direction of the S&P 500 Index in the next month. Care must be exercised when examining these results because of the different characteristics of the indices. The summary statistics for the Baker and Wurgler (2007) sentiment index (SENT) and the Huang, Jiang, and Tu Sentiment Index (SI) are very similar; both indices are standardized

[12]The portfolio beta could be increased to be greater than 1.00 by using leverage. This would increase portfolio returns when the S&P 500 returns are positive.

to have a mean of zero and a standard deviation of one. Over the period examined, their sentiment indices mean values were slightly negative. For the other models, the magnitudes of the range of values and the standard deviations show substantial variation, primarily due to the financial crisis. The ADS index is also designed to have a mean value of zero; business economic conditions are better (worse) the greater the positive (negative) values. The ADS index exhibits a small negative mean value and a negative skewness that is driven by the large negative values or "black swan events" that occurred during the financial crisis. The GZ spread (GZ) and the financial uncertainty (FUI) models generally move in an opposite direction to business conditions. Both their mean values and skewness are positive. The greater the value of GZ and FUI, the more negative the market direction. Also, GZ and FUI are always positive and peak around significant economic and political events. Because of their construction, its values are several orders of magnitude greater than the other models. In addition, high values represent greater uncertainty. Skewness is positive but larger relative to the standard deviation.

Panel B is discussed in three sections; first is an examination of the correlations between the S&P 500 Index and the timing strategies, second is a discussion of the correlations among the forecasting indices, and finally is a review of the relations between the timing strategies and the forecasting indices. The discussion will focus on the one-month ahead return of the SPX.

When the forecasting models are examined, SENT and SI are positively correlated, but, somewhat surprisingly, almost independent of the ADS, GZ, and FUI. The ADS is a measure of business conditions, and firms operate best when the markets are functioning well and when managers are not faced with uncertainty. Hence, ADS is negatively correlated with all 4 models; GZ spread (-0.510), FUI (-0.410), SENT, (-0.066), and SI (-0.029).

The signs for the correlations between SPX1 returns and the ADS are as expected. Favorable business conditions this month is positively related to next month's SPX1 return, but the ADS is negatively correlated to the other four indices. The intuition is that during periods when market conditions of uncertainty is high, prices are bid down and this is the basis for greater returns in the following period. When the FUI is down, this signals the timer to move to cash and/or go short the SPX. The ADS is just the opposite; when the ADS is up (down), a stronger (weaker) economy is foreseen, and the S&P 500 Index SPX moves up (down), on positive (negative) information.

6.2 Logistic Regression Results

To quantitatively examine the relationship among the five models and the benchmark index Table 2 presents the correlations among the timing models and their relationship with the SPX's one-month-ahead return (SPX1) and the average forecast for an "up" or "down" market in the following month. First, it is observed that the correlation between the one-month-ahead portfolio return, SPX1, and the UP measure of market direction is 0.758. However, the correlations of the forecasting models' predictions related to SPX1 are substantially lower. SI and GZ are the highest with correlations of 0.160 and 0.131 respectively with next month's return, while FUI is lowest at 0.060. Surprising, the benchmark index, VWDP, has the lowest correlation of the all models at 0.047. It should be noted that these correlations are based on the [0,1] classifications generated by the logit model. Interestingly, none of the models have correlations greater than 0.450 with respect to each other with the exception of the FUI and the ADS correlating at 0.590.

Table 3 illustrates annual holding periods of the forecast models and the benchmark index from 1983 - 2014. The SPX recorded seven years of negative returns.[13] In 1987, 1994, and 2007 all five forecast model portfolios and the benchmark index identical negative returns as the SPX -0.062, -0.023 and -0.042 respectively. The most noticeable difference in the model returns are in 2000, 2001, 2002 and 2008, all being recessionary periods. The SENT and GZ models had positive returns in both 2000 and 2001 versus negative returns

[13]The following years were negative returns for the S&P 500 Index: 1987 (-0.062), 1994 (-0.023), 2000 (-0.020), 2001 (-0.173), 2002 (-0.243), 2007 (-0.042), 2008 (-0.401)

for the SPX. In 2002, GZ was -0.079, where all other models including the SPX where down significantly at -24%. The most notable difference in returns is in 2008 when the SPX and VWDP recorded the worst returns in the sample period of -40.1%. Conversely, in the same year SI, ADS, GZ, and FUI all had positive returns.

The summary statistics of monthly returns associated with the predictions of each model are shown in Table 4. The distribution of the SPX represents a buy-and-hold strategy. The maximum drawdown return for the SPX, SENT and the VWDP is 52.6%, which occurred from late 2008 to early 2009. During that same time period, the SI, ADS, GZ, and FUI had a maximum drawdown of 40.4%, 44.7%, 30.2%, and 48.3% respectively[14]. The annualized returns of the SENT and ADS are considerably higher than the SPX (7.1%) at 9.8% and 9.2% respectively; the FUI (7.7%) is slightly higher than the SPX, but lower than VWDP (8.8%). The SI index generates the best returns at 12.1%, followed by GZ at 11.0%. The standard deviation of returns of the five forecast models, the benchmark, and the market are virtually the same at 15%. The Sharpe ratio of all the forecast models is greater than the SPX at 0.457. While the perfect timing models, SPXRF and SPXLS, exhibit Sharpe ratios seven to ten times greater than the SPX due to both substantially greater returns and smaller standard deviations.

In order to better understand the statistical relationship between the five predictive models and the benchmark index, we perform non-parametric tests to determine if the probability estimates are the same. Initially, referring to Figure 2 we can visually conclude that the forecast indices do not have the same structure. Each dot represents outlier estimates, the horizontal lines are observations at the median, the grey boxes are probability estimates that are within the 50th percentile, and the vertical lines represent observations above or below the previous threshold. The box-plot suggests that our probability estimates are more than likely non-Gaussian estimators. Therefore, the following statistical tests will primarily evaluate stochastic dominance between each estimator.

[14]As constructed, SPXRF and SPXLS had zero drawdown as they are either in US T-Bills or are short the SPX when the market is down.

Figure 3 represents each models' distribution of probability estimates. The range of each indices bandwidth is between 0.01574 (VWDP) and 0.02956 (FUI), while the highest(lowest) density of each model is 7.0(4.0) with 374 monthly observations. SENT has a range of probability estimates between 0.300 and 0.900 with a peak density near 6.5. SI has a lower left-tail than SENT at 0.240 with a similar right-tail at 0.880, but displays higher kurtosis and a peak density at above 6.5. ADS has a high amount of right skew and kurtosis with the lowest probability estimate (0.086), and the lowest maximum probability estimate (0.760) of all predictors. GZ and the benchmark, VWDP, display similar characteristics with density readings above 6.25 and a range of probability estimates between 0.150 and 0.912. FUI seems to be the most evenly distributed with the highest range of probability estimates (0.100 - 0.950) and lowest density (4.05) compared to the other forecast models. All the models seem to have a right skew, and high kurtosis.

To confirm that the probability estimates of the five forecast indices and the benchmark model are non-Gaussian estimates, we perform the Shapiro-Wilks normality test[15]. Figure 4 represents the results of each normality test in the form of a Q-Q plot, with the theoretical quantiles on the horizontal axis and sample quantiles on the vertical axis. Based on the visual representation of each figure, we determine that each model's probability estimates are non-Gaussian. Moreover, the SENT, SI, ADS, and FUI all display observations below the line of normality on both the upper and lowers tails. Conversely, GZ and the benchmark model, VWDP, have estimates above the upper line of normality. All models display probability estimates near the lower theoretical (-3) and sample (0.200) quantiles.

To determine the homogeneity of the variance among the entire population of the prediction variable estimates, we run a Bartlett test[16]. This test is determines if the population variances are equal. If the p-value is higher than 0.05, then the null hypothesis is rejected, thereby accepting that the variances are unequal. The results of the test give a k-squared

[15] Shapiro-Wilkes (1965) is a normality test statistic to determine if a given set of data is from a normally distributed population. Given as $W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, where is x is the sample mean and a_i is the constants

[16] Snedecor and Cochran (1989), improved the x^2 distribution test if there are k samples, S_i^2, sample variances

value of 110 with 5 degrees of freedom, and a p-value of zero. Therefore, we conclude that our forecast variable estimates' variances are not equal.

We also run a Kruskal-Wallis test[17] by ranks of the probability estimates for each of the predictors. This non-parametric test is used for comparing if two or more variables samples sizes belong to the same distribution. Moreover, the test statistic will indicate if one sample stochastically dominates over the others. Once again, if the p-value is below the 0.05 threshold, we can reject the null hypothesis that our probability estimates come from the same distribution. The test results of our sample return a X^2 value of 380 with 5 degrees of freedom and p-value of zero, allowing us to conclude that the estimates do not come from the same distribution.

Table 5 displays the results of the pairwise Dunn-Test[18] between the five forecast models and the benchmark model. The null hypothesis is that all the indices probability estimates are the same when comparing each variable in a pairwise framework. Each model is represented both in the first column and row of the table, with the Z-statistic score on the top and its relative p-value in parentheses. In order to reject the null hypothesis a p-value needs to be lower than 0.05. The highest resulting Z-statistic is SENT-FUI (16.754), and the lowest is ADS-SENT (0.9354). The mid-range is between 2.1244 (SI-VWDP) and 15.8186 (FUI-ADS). Each pairwise test results in p-values below the 0.050 threshold with the exception of the ADS-SENT (1.000). Therefore, in this instance the null hypothesis would be accepted, and it can be determined that the probability estimates between the ADS and the SENT indices are statistically the same. All other pairwise tests results reject the null hypothesis.

[17] Kruskal-Wallis (1952), ranks all group members values from 1 to N, and returns an X^2 value. This test does not determine a relationship between each group pair individually.

[18] Dunn (1961)

6.3 Forecasting Errors and Portfolio Performance

Table 6 shows the monthly forecasting errors on a year-by-year basis for all five forecasting models and the benchmark index. The objective of the forecasting models is to correctly predict the direction of the S&P 500 Index in the upcoming month. The "Total" row at the bottom of the table shows a summary of the "up" and "down" months for the S&P 500 Index and the numbers of each type of errors for each model. The two rows along the bottom of the table show the number of incorrect and correct predictions. Over the 373 months from 1983-2014, the S&P 500 Index was "up" in 233 months and "down" in 140 months. Forecasts of the Baker and Wurgler (2007) sentiment index (SENT) was the best overall predictor of the market direction, being incorrect (correct) in 139 (234). In five of the incorrect months, SENT predicted the market would go "down" but it went "up", and in 134 of the incorrect months the prediction was that the market would go "up" but it went "down". Thus, the portfolio would be short in 5 months when the market went up but long in 134 months when the market went down. In comparison, the benchmark index (VWDP) was incorrect (correct) in 136 (237).

While the number of incorrect predictions is important, the market movement can exacerbate incorrect predictions. Consider the financial crisis year of 2008; the market was up in four of the months and down in eight. SENT and the benchmark index, VWDP, correctly predicted all the "up" market months, but missed on all eight of the "down" market months. Thus, the portfolio was long the SPX in the eight months the market was down.

By contrast, ADS was particularly bearish in 2008, incorrectly predicting the "up" months, but correctly predicting six of the eight down months and, hence, was short the SPX during those six months. As shown later, this strategy generates an impressive return of 32.6% for that 2008 Black Swan year; the SI model correctly predicted three of the eight down months, while the GZ and the FUI got five of the eight down months correct. Overall, four of the five forecasting models accurately predicted the market drawdown in October 2008. During the recovery period, 2009 - 2014, all five indices and the benchmark index accurately predicted all market up months, with a few exception in 2009.

Table 7 presents the annualized performance analysis of monthly returns for the perfect timing strategies, the five individual forecasting models and the benchmark model. SPX provides the naïve investment strategy of simply buying and holding the S&P 500 Index while SPXLS, the perfect timing model, provides an upper bound for portfolio returns. Successful strategies avoid down markets and capture up markets. The performance measures of the SI index are substantially better than the other forecasting models. It exhibits a greater mean return and with a standard deviation similar to those of the other models. SI generates substantially greater Sharpe (0.806) and Sortino (0.393) ratios.[19] Additionally, the SI Omega ratio (1.889) is greater than the naive strategy of buying-and-holding the SPX (1.484).[20] Importantly, in addition to the better performance measures, the GZ was the best in mitigating the large market declines with the least drawdowns of 0.302 compared to 0.404 for the next best, SI. The gain and loss deviation is basically the same for all indices. Over the entire sample period, if you had a perfect market timing your annualized returns in the SPXRF would have been 30.2% and the SPXLS would be 49.9%. All five forecast models have higher annualized returns than the SPX, and four of the five models outperform the benchmark index, VWDP. Figure 5 illustrates the Wealth Index of all five forecast indices, the benchmark model and the SPX. The best performing forecast model is SI, and the worst performing index is the FUI. All other forecast models significantly outperform the SPX and the benchmark, VWDP, during the sample period.

6.4 Statistical Significance of Model Estimates

Table 8 illustrates the results of logistic regressions from equations (1, 2). Each coefficient reported, β (%), is the model estimate resulting in a binary prediction of the direction

[19] It should be noted that the Sortino ratio cannot be calculated for the SPXRF and SPXLS because there are no negative returns to be used in the computation.

[20] The Omega ratio is estimated as the returns above a threshold level divided by returns below the threshold level. We use a threshold level, L, of a zero return.

of the S&P 500 index (SPX) as either UP or Down in the immediate next month. An UP market is forecasted when the $p(x = 1) \geq 0.5$, and Down market is forecasted when ($p(x = 1) < 0.5$.). So, if $p(x) \geq 0.5$, we predict $x = 1$; if $p(x) < 0.5$, $x = 0$ is predicted. The β coefficients represent the probability predictions for each of the forecast models. The middle columns represent the Newey-West t-statistics associated with the logistic regression results. The columns on the far right of the table represent the McFadden's $pseudo-R^2$.[21] Each fitted model's results is reported for the month of December each year starting in 1983 through 2013. All coefficients are the result of a 120-month rolling estimation period. For most of the reporting period each predictive model forecasts an UP market, but there are a few notable exceptions.

In December 2000, the SENT, SI, and the GZ incorrectly forecast a down month with estimates of 0.425, 0.318, and 0.357 respectively. Conversely, the ADS, FUI and the benchmark model, VWDP, record accurate estimates of 0.680, 0.533, and 0.625 respectively. Each of the incorrect model forecasts have an absolute value for the t-statistic that is less than 2.00 with the exception of the SI (2.387), as well as rather low $pseudo - R^2$ of SENT (0.163), SI (0.122), and GZ (0.080). Of the correctly forecasted indexes, only the benchmark model, VWDP, has a t-statistic above 2.00, but each index has a higher $pseudo - R^2$ of ADS (0.622), FUI (0.384), and VWDP (0.540).

During the height of the financial crisis in 2008, the forecast models once again differed in their predictions. The SENT (0.617) and the VWDP (0.587) coefficients both predicted an UP market, while SI (0.443), ADS (0.255), GZ (0.266), and FUI (0.248) forecasted correctly a Down market. The t-statistics of the models which forecast an UP market all had an absolute value below 2.00, with the exception of SENT (2.112); the models that correctly predicted a down market recorded t-statistics of SI (2.245), ADS (1.435), GZ (1.319), and FUI (2.640). All $pseudo - R^2$ for this time period were above 0.450, with the

[21] McFadden and Domencich (1974), In McFadden's model the log likelihood approach is applied to the full model divided by the intercept model, which is the total sum of the squares with no predictors: $pseudo - R^2 = 1 - \frac{ln(\hat{L}(M_{full}))}{ln(\hat{L}(M_{intercept}))}$

exception of the SI (0.203).

Considering specifically the β (%) coefficient, which represent probabilities, only 13 of the 186 sample model-months had values below 0.50. Three of the models forecast down months in two of the years, (1983 and 2000), and four of the models forecast down months in the crisis year of 2008. The three lowest coefficients are observed in that year; ADS (0.255), GZ (0.266), and FUI (0.248). Interestingly, FUI shows four coefficients below 0.500, with three occurring in the first three years of the investigation period. Focusing on the higher probabilities, there are 24 coefficients greater than 0.700; SENT appears to be the most optimistic model with seven coefficients greater than 0.700, while ADS is the least optimistic with none greater than 0.700. Interestingly, FUI had values greater than 0.700 for four consecutive years (2003 through 2006).

Thirty-nine of the t-statistics observed have absolute values greater than 2.00; only five of the values are positive, FUI with three in the years 1984-1986, and VWDP in 2000 and 2001. Overall, 72, nearly half of the t-statistics are positive. FUI has a run of nine consecutive years of significant t-statistics from 2001 through 2009, while SI has the most observations exceeding 2 (14 in total). There were 147 monthly regressions that had no statistically significant coefficients.

Finally, a perusal of the aggregate $pseudo - R^2$ observations reveal that a fairly large proportion of the variation is explained by the regressions. There are 65 regressions with $pseudo - R^2$ less than 0.500, with ADS (14), GZ (13), FUI (13) and VWDP (12) with similar values. At the upper end of the $pseudo - R^2$ values, SENT has four values greater than 0.700 while GZ has three. ADS has no values greater than 0.70.

6.5 Prediction Accuracy

Logistic regressions are used to develop monthly predictions. Figure 3 - Figure 7 illustrates each models probability of the S&P 500 Index being up or down in the next calendar month. Each figure also shows the prediction's statistical significance (p-value). It is observed that the ADS (Figure 8), the GZ (Figure 9), and the FUI (Figure 10) models, are in general agreement on market direction while SENT (Figure 6) and SI (Figure 7) are most closely related. The benchmark, VWDP (Figure 11), model seems to act independent relative to the other models. Practically, the portfolio will be invested in the SPX when the probability of an UP market is greater than 0.50, and will be short the SPX when the probability is less than 0.50.[22] Larger probability values will not change the investment decision, though the confidence of the prediction is higher. These figures show the magnitude of the forecast, the direction, its statistical significance. Overall, the models are in substantial agreement and are robust as the probability is generally much higher or much lower than the 0.50 probability threshold. Each model has discriminatory power with high level of significance in identifying substantial market movements, particularly during the mid to late 1990's, the bursting of the tech bubble in early 2000's, and the 2007-2009 financial crisis though the recovery period.

Figure 6 illustrates the probability of an UP market for the SENT model and its statistical significance. Between January 1988 though May 1990, the mean monthly p-value of the SENT index is 0.0724 and the monthly prediction is correct 62.1% of the time. During the recession period from 2000-2003 the mean monthly p-value was 0.0552, correctly forecasting 55.4% of the time. The best forecasting period of the SENT is from January 2008 through July 2011 with a mean p-value of 0.0416, and a forecast accuracy of 64.2%, but during the financial crisis (2008-2009) the mean p-value is 0.0384 with a prediction accuracy of only 55.0%.

The SI index shown in Figure 7 has the best forecast accuracy and lowest mean monthly p-value of all the models tested. From November 1986 through January 1992 the mean monthly p-value is 0.0224 and the model correctly forecasts 62.1% of the time. During the recession period between July 2000 through December 2003 the mean monthly p-value is 0.0225 and has a forecast accuracy of 55.7%. The most impressive forecast period is

[22] If the portfolio uses the traditional stock/bond timing approach, the shift would be from stocks to bonds/cash in down markets.

from March 2008 through November 2009, with a monthly mean p-value of 0.0216 and prediction accuracy of 81.1%. During the recovery period, from November 2011 through March 2014 the forecast accuracy is 79.3% with a monthly mean p-value of 0.0564.

Illustrated in Figure 8, the ADS model throughout the mid 1980's and the entire decade of the 1990's has a high mean monthly p-value (0.6629), but a good prediction accuracy of 66.7%. During the financial crisis, from March 2008 to November 2009 the ADS index has a forecast accuracy of 57.1%, and a monthly mean p-value of 0.2421. Throughout the recovery period the model has a high prediction accuracy (79.3%), but the statistical significance is rather weak with a monthly mean p-value of 0.2973.

The GZ model, Figure 9, is similar to the SI in its forecast ability, but has a lower level of statistical significance. The model does particularly well during the 1990-91 recession period. From October 1990 to July 1991, the index has a 90.1% forecast accuracy rate with mean monthly p-value of 0.0850. The model also performs well from October 2000 to March 2006, which includes the recession period, recording a mean monthly p-value of 0.0537 and correctly forecast the market direction 55.5% of the time. During the financial crisis, the GZ accurately predicted five of the six largest draw-down months between August 2008 through January 2009. During this time period, the index was correct 83.3% of the time with a respectable monthly p-value of 0.1214. The model has a high prediction rate during the recovery period (November 2011 - March 2014) of 79.3%, but a rather low statistical significance (0.4353 mean monthly p-value).

The FUI, Figure 10, has the lowest mean monthly p-value of all the models tested with a value of 0.19976 during the entire sample period. Though, referring back to Figure 5, the FUI is the worst performing index with an average annual rate of return of 7.7%. The most notable forecast period for the FUI was during the recovery from November 2011 through March 2014. The forecast accuracy rate was 79.3% with a monthly mean p-value of 0.1097.

The benchmark model, VWDP (Figure 11), only predicts a down market during the financial crisis in 2008. The index's statistical significance is generally below 0.2515 before the recovery period, but from November 2011 to March 2014 its predication accuracy and p-value underperform all five forecast models. Its only period of meaningful prediction is during the recessionary period between 2001 to 2002.

6.6 Persistency of the Probability Estimates

In order to determine the overall persistency of the probability estimates we follow similar autoregressive time series (AR) models as Ding, Granger, and Engle (1993), Ferguson, Sarkassian, and Simin (2003), Campbell and Thompson (2008), and most recently Lo and Remorov (2017). Each report statistically significant autocorrelations between 1 and 20 month lagged periods. We use a straightforward AR model to track persistency of the probability estimates from the five predictive indices and the benchmark portfolio. We use a first-order autocorrelation process represented by an AR(1) model as a standard linear difference equation,

$$X_k = \rho X_{k-1} + \epsilon, \tag{4}$$

where $k = 0, 1, 2, ...$ and the ϵ_k are the innovation (error terms) of the variability of the time series. This difference equation relates X_k to its own value at some previous time k as a lagged variable expressed at X_{k-1}.

Referring to Figure 12, there are six computational estimates of the autocorrelation function (acf) for each of the five predictive models and the benchmark index. Each plot illustrates the number of lag periods in months along the horizontal axis and the acf on the vertical axis. Each plot is a result of an AR(1) model. Each of the predictive models demonstrate meaningful persistency for the probability estimates at least 18 monthly lag periods with a 95% confidence interval with the exception of the benchmark index, VWDP.

The SENT model at a 4-month lag period has an ACF value over 0.80. At the 5-month lag persistence levels, the acf remains over 0.70. The acf each month gradually decreases below 0.40 at the 12-month lag period and then drops below the 0.10 threshold at the 20-month lag. The SI, ADS, GZ, and FUI indices all have an acf value that exceeds 0.90 at the 1-month lag. All four models demonstrate persistency above 0.50 through the 10-month lag period, and continue above the 0.10 level through the 19-month lag. However, the GZ model drops below the 0.10 value at the 18-month lag. The only model to remain above the 0.10 level through the 20-month lag period is the ADS index. Each model seems to have a high level of persistency with only the downward slope of the acf plot differing as each monthly lag is increased. The VWDP benchmark model differs in that its persistency drops at a higher rate than the other five predicted models.

6.7 Certainty Equivalent Results

Table 9 provides a comparison of the performance based on the CEQ analysis. As expected, the overall returns and significance results are consistent with each forecast portfolio's annual geometric performance. The first five columns of the table display CEQ returns for different values of risk aversion $\gamma = 1, 2, 3, 4, 5$ as a result of the logistic regression probability cut-off of 0.50. The first row shows the CEQ returns for the S&P 500 Index (SPX), and the next 13 rows are the CEQ returns and p-values associated with the forecast model strategies.

At a risk aversion level of $\gamma = 1$, which is the same used in the study by DeMiguel et al (2009), we observe that all forecast models with the exception of the FUI have higher returns than the SPX. In addition, the p-values comparing the difference between the SPX and each forecasting model portfolio's Sharpe ratio are below 0.050 with the exception of the FUI and the VWDP. This suggests that each of the former mentioned strategies Sharpe ratios are statistically different than the SPX. The best performing monthly CEQ returns are the SI (0.00616), GZ (0.0053), SENT (0.00446), and ADS (0.00403), which are all higher than the SPX (0.00350).

As the risk aversion increases from $\gamma = 2$ to $\gamma = 5$, the CEQ returns decrease and the p-values increase for all strategies. At $\gamma = 2$, the SENT [0.00352, (0.02279)], SI [0.00522, (0.00289)], and GZ [0.00439, (0.00845)] are all higher than the SPX (0.00255), and their p-values are below the 0.050 threshold. The remaining strategies have lower returns than the SPX, and p-values higher than 0.050. At $\gamma = 3$, the SPX's monthly return is 0.00159 compared to the SI [0.00429, (0.00958)] and the GZ [0.00345, (0.02442)], which are the only two strategies that have returns higher than the SPX, and p-values below 0.050. Once the risk aversion level reaches $\gamma = 4$, the SI and the GZ strategies both have higher returns than the SPX, but the SI is the only model with a p-value below 0.050. At $\gamma = 5$, none of the forecast strategies have p-values below the 0.050 threshold.

As a final robustness check, we refer to Table 9 to compare the CEQ returns and p-values of the forecast strategies to the SPX. Instead of running the logistic regression at the 0.50 probability cut-off, we change it to 0.75. The last five columns of the Table 9 illustrate these results. At every level of risk aversion $\gamma = 1, 2, 3, 4, 5$, none of the forecast strategies have higher returns than the SPX or p-values below 0.050.

7 Concluding Remarks

This paper uses five indices to forecast the one-month-ahead S&P 500 Index returns. The focus is on market timing and the changing of the portfolio's beta as market expectations change. The changing of the beta is referred to as "beta optimization". Operationally, when the market is expected to be up (down) the portfolio is adjusted to have a long (short) position in the S&P 500 Index (SPX). A number of indices have been created with a view to explain or predict market direction. Baker and Wurgler (2007) develop the sentiment index (SENT); Huang et al (2015) build upon the the SENT model to develop an improved sentiment index (SI); Aruoba, Diebold and Scotti (2009) create the ADS that is designed to reflect current business conditions; Gilchrist and Zakrajsek (2012) focus on economic activity through fixed income credit spreads (GZ); Jurado, Ludvigson, and Ng (2015) create an independent time-varying macro-economic uncertainty index (FUI). These five indices are used in a logistic regression procedure to develop *predictions* of market movement. These predictions are used to create portfolios composed of a long or short SPX position.

By construction, perfect market timing generates the greatest returns with the lowest risk while the naïve buy-and-hold strategy provides a comparison of the strategies associated with the forecasting models. During the 592 months of the study, two serious downturns occurred; the bursting of the technology bubble in early 2000 and the financial crisis of 2008. Over this period, there were 233 months of up-markets and 140 months of down-markets. Hence, the models were tested during volatile market conditions. Since the market was up approximately two-thirds of the months, the buy and hold strategy was moderately successful.

Four of the five models, SENT, SI, ADS, and GZ outperformed the buy-and-hold strategy (SPX) as well as the Baker and Wugler (2004) benchmark index, VWDP, with periods of high relative predictions accuracy and low mean monthly p-values. But the merit of the timing strategy becomes particularly apparent during the financial crisis of 2008-2009 when both the SI and GZ models correctly forecast more than 80% of the time, and are statistically significant 90% of the time. In volatile markets investors are keen to avoid downside risk; the SI and GZ indices appear to be the best at addressing this concern. Moreover, we find that the CEQ returns are significant for the SENT, SI, and GZ models at risk aversion levels of $\gamma = 1, 2, 3$. The SI is also significant at $\gamma = 4$, and the its mean CEQ returns are higher than the SPX at all levels of risk aversion.

There is generally concurrence among the models, disagreements exist, particularly around significant economic and political events. Having a battery of models allows the investor to make better informed decisions. Consistent with prior research, more can be done by combining the five forecasts and/or applying model selection criteria to improve prediction accuracy, significance and overall forecasted returns.

As noted the S&P 500 Index (SPX) represents a one-asset portfolio. However, individual investors could use the procedures outlined above to adjust the betas of their portfolios by switching from equity funds to bond or money market funds as their forecasts or expectations change. Active asset managers could modify the procedures by shifting to/from high beta stocks and adjusting the portfolio's beta by using derivatives. This is an area for future research as well.

A Caveat: Making successful predictions also requires an appropriate metric to evaluate the managers' decisions. Research has shown that successful market timing strategies yield option-like payoffs and the risk of such a strategy cannot be satisfactorily quantified. Even though the beta optimization strategy has worked well over the examined time period, caution must be exercised because as John Maynard Keynes observed, "the market can stay irrational longer than you can remain solvent."

References

[1] Avramov, Doron, and Russ Wermers, 2006. Investing in mutual funds when returns are predictable. Journal of Financial Economics 81, 339–377.

[2] Aruoba, S. Borağan, Francis X. Diebold and Chiara Scotti 2009 Real-Time Measurement of Business Conditions, Journal of Business & Economic Statistics 27, 417-427.

[3] Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2013. Measuring economic policy uncertainty, Working Paper, Stanford University.

[4] Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler, 2011, Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly, Financial Analysts Journal 67, 40-54.

[5] Baker, Malcolm, and Jeffrey Wurgler, 2004, A Catering Theory of Dividends, Journal of Finance 59, 1125-1165.

[6] Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor Sentiment and the Cross-Section of Stock Returns, Journal of Finance 61, 1645-1680.

[7] Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor Sentiment in the Stock Market Journal of Economic Perspectives 21, 129-151

[8] Baker, Malcolm, Jeffrey Wurgler and Yu Yuan, 2012, Global, local, and contagious investor sentiment, Journal of Financial Economics 104, 272–287.

[9] Baker, Scott R., Nicholas Bloom, and Steven J. Davis. (2015). Measuring economic policy uncertainty. Working Paper, Stanford University.

[10] Bates, J.M., and C.W.J. Granger, (1969), "The combination of forecasts", Operational Research Quarterly, 20,451-468.

[11] Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, Journal of Business 444-455.

[12] Black, Fischer, 1993, Estimating expected return, Financial Analysts Journal 20, 36-38.

[13] Black, Fischer, Michael C. Jensen and Myron S. Scholes, 1972, The Capital Asset Pricing Model: Some Empirical Tests, Studies in the Theory of Capital Markets, Praeger Publishers Inc. Available at SSRN: http://ssrn.com/abstract=908569

[14] Bollen, Nicolas P. B., and Jeffrey A. Busse, 2001. On the Timing Ability of Mutual Fund Managers. Journal of Finance 56: 1075 -1094.

[15] Brown, Gregory W., and Michael T. Cliff, 2004, Investor sentiment and the near-term stock market, Journal of Empirical Finance 11, 1 –27.

[16] Campbell, John and Thompson, Samuel, 2008 Predicting Excess Stock Returns out of Sample, The Review of Financial Studies 21, 1509-1531

[17] Carhart, Mark M., 1997, On persistence in mutual fund performance, Journal of Finance 52, 57–82.

[18] Chauvet, Marcelle, and Jeremy Piger, 2008, A comparison of the real-time performance of business cycle dating methods, Journal of Business and Economic Statistics 26, 42–49.

[19] Clemen, Robert T., (1989), Combining forecasts: A review and annotated bibliography, International Journal of Forecasting 5, 559-583.

[20] Cremers, K. J. Martijn and Antti Petajisto, 2009, How Active Is Your Fund Manager?A New Measure That Predicts Performance, The Review of Financial Studies 22, 3329-3365.

[21] Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The Sum of All FEARS Investor Sentiment and Asset Prices, The Review of Financial Studies 28 1-32.

[22] Daniel, K. and Titman, S. 1997, Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns, The Journal of Finance 52, 1-33.

[23] Del Guercio, Diane, and Jonathan Reuter, 2014, Mutual Fund Performance and the Incentive to Generate Alpha, Journal of Finance 69, 1673-1704.

[24] Ding, Z, Granger, C and Engle, R.F, 1993, Along memory property of stock market returns and a new model, Journal of Empirical Finance 1, 83-106.

[25] Droms, William G., 1989, Market Timing as an Investment Policy, Financial Analysts Journal 45, 73-77.

[26] Dunn, O. J. 1961. Multiple comparisons among means. Journal of the American Statistical Association 56: 52–64.

[27] Durbin, J. and Koopman, S., 2001, Time Series Analysis by State Space Methods, Oxford Statistical Series 20, 139-141.

[28] Eisele, Alexander, 2012, Beta Arbitrage and Hedge Fund Returns. Available at SSRN: http://ssrn.com/abstract=2026601 or http://dx.doi.org/10.2139/ssrn.2026601.

[29] Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3–56.

[30] Fama, Eugene F., and James D. McBeth, 1974, Tests of the Multiperiod Two-Parameter Model. Journal of Financial Economics 1, 43-66.

[31] Federal Reserve Bank of Philadelphia, (2013) Updates on ADS Index Calculation, www.phil.frb.org.

[32] Fisher, Kenneth L., and Meir Statman (2003) Consumer Confidence and Stock Returns, Journal of Portfolio Management 30, 115-127.

[33] Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta. Journal of Financial Economics 111, 1–25.

[34] Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious Regressions in Financial Economics, The Journal of Finance 58, 1393-1413.

[35] Fung, William, and David A. Hsieh, 2001, "The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers." Review of Financial Studies 14, 313–341.

[36] Genrea, Veronique, Geoff Kennya, Aidan Meyler, and Allan Timmermann, 2013, Combining expert forecasts: Can anything beat the simple average?, International Journal of Forecasting 29, 108–121.

[37] Gilchrist, Simon and Zakrajsek, Egon 2012, Credit Spreads and Business Cycles, American Economic Review 102, 1692–1720.

[38] Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou, 2015, Investor Sentiment Aligned: A Powerful Predictor of Stock Returns, Review of Financial Studies 28, 791-837.

[39] Hubner, Georges, 2010, The Alpha of a Market Timer, HEC Management School — University of Liege, Belgium School of Business and Economics, Maastricht University, the Netherlands, Gambit Financial Solutions Ltd, Belgium.

[40] Jagannathan, Ravi, and Robert A. Korajczyk, 2015, Market Timing, (February 9, 2015) http://ssrn.com/abstract=2516550 or http://dx.doi.org/10.2139/ssrn.2516550

[41] Jiang, G. J., Yao, T., and Yu, T., 2007. Do mutual funds time the market? Evidence from portfolio holdings. Journal of Financial Economics 86: 724-758.

[42] Jurado, Kyle, Ludvigson, Sydney, and Ng, Serena, 2015, Measuring Uncertainty, American Economic Review 105, 1177–1216.

[43] Kacperczyk, Marcin, Stijn Van Nieuwerburgh and Laura Veldkamp, 2014, Time-Varying Fund Manager Skill, Journal of Finance 69, 1455-1484.

[44] Kelly B. and Pruitt S., 2013, Market Expectations in the Cross-Section of Present Values, Journal of Finance 66, 1721-1756.

[45] Kruskal, Wallis (1952), Use of ranks in one-criterion variance analysis, Journal of the American Statistical Association. 47 (260): 583?621.

[46] Lam, K., and Wei Li, 2004, Is the 'Perfect' Timing Strategy Truly Perfect? Review of Quantitative Finance and Accounting 22, 39–51.

[47] Lintner, John, 1965, Security Prices, Risk and Maximal Gains from Diversification, Journal of Finance 20, 587-615.

[48] Lo, Andrew W. "The statistics of Sharpe ratios." Financial Analysts Journal 58, no. 4 (2002): 36-52.

[49] Lo, Andrew W., and Remorov, Alexander, 2017. Stop-loss strategies with serial correlation, regime switching, and transaction costs, Journal of Financial Markets 34, 1?15.

[50] Makridakis, S. and M. Hibon, 1979, "Accuracy of forecasting: An empirical investigation (with discussion), Journal of the Royal Statistical Society, Series A, 142, 97-145.

[51] McFadden, Daniel L. and Tom Domencich, 1974, Urban Travel Demand: A Behavioral Analysis, North Holland Publishing, Chapter 15.

[52] Miller, Merton H., and Myron Scholes, 1972, Rates of return in relation to risk: A reexamination of some recent findings. Studies in the theory of capital markets 23.

[53] Merton, Robert C., 1981, On Market Timing and Investment Performance. I. An Equilibrium Theory of Value for Market Forecasts, Journal of Business 54, 363-406.

[54] Monteforte, Libero and Gianluca Moretti, (2013), Real-Time Forecasts of Inflation: The Role of Financial Variables, Journal of Forecasting, 32, 51–61.

[55] Mossin, Jan, (1966) "Equilibrium in a Capital Asset Market," Econometrica 34, 768-783.

[56] Newbold, Paul, and Clive W.J. Granger, 1974, Spurious regressions in econometrics, Journal of Econometrics 2, 111-120.

[57] Pasquariello, Paolo, 2014, Financial Market Dislocations, The Review of Financial Studies 27, 1868-1914.

[58] Pastor, Lubos, and Robert F. Stamaugh, 2003, Liquidity risk and expected stock returns, Journal of Political Economy 111, 642-685.

[59] Phillips, D., and J. Lee, 1989, Differentiating Tactical Asset Allocation from Market Timing, Financial Analysts Journal 45, 14-16.

[60] Sharpe, William F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," Journal of Finance 19, 425-442.

[61] Sharpe, William F., 1975, Likely gains from market timing, Financial Analysts Journal 31, 60-69.

[62] Shapiro, S. S., Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)", Biometrika. 52 (3?4): 591-611.

[63] Snedecor, George W. and Cochran, William G. (1989), Statistical Methods, Eighth Edition, Iowa State University Press.

[64] Treynor, Jack L., and Kay K. Mazuy, 1966, Can Mutual Funds Outguess the Market? Harvard Business Review 44, 131-136.

[65] Wallis, Kenneth F., 2011, Combining forecasts – forty years later, Applied Financial Economics 21, 33–41.

[66] Wold, Herman, 1966, Estimation of Principal Components and Related Models by Iterative Least Squares, Multivariate Analysis, 1391-1420.

[67] Wold, Herman, 1975, Soft Modeling by Latent Variables: The Nonlinear Iterative Partial Least Squares (NIPALS) Approach, Perspectives in Probability and Statistics, 117-142.

A Computation of Forecasting Models

A.1 The Investor Sentiment Index

Baker and Wurgler (2006) posit that investor sentiment is different for different types of companies and use six proxies for sentiment in the development of two investor sentiment indexes.[23] The initial investor sentiment index (ST) was developed in 2006, with the assumption firms that are most sensitive to investor sentiment are also more difficult to value and arbitrage. The characteristics of these stocks include low capitalization, younger, unprofitable, high-volatility, non-dividend paying, growth companies or firms in financial distress. Their analysis is centered around the following form,

$$E_{t-1}[R_{it}] = a + a_1 T_{t-1} + \mathbf{b}_1^{'} \mathbf{x}_{it-1} + \mathbf{b}_2^{'} T_{t-1} \mathbf{x}_{it-1}, \tag{5}$$

where i denotes companies, t is time, and $\mathbf{x}$ is a vector of characteristics, T is the proxy for investor sentiment, a_1 is the *generic* effect of sentiment, and $\mathbf{b}_1$ is the vector of *generic* effect of characteristics. The null assumption is that $\mathbf{b}_2 = 0$ or any non-zero effect is a compensation for systematic risk. Or, the alternative is that $\mathbf{b}_2$ is nonzero, which reveals the cross-sectional movement in sentiment-driven mispricing. This equation is called a *conditional characteristics model* adding conditional terms to the Daniel and Titman (1997) model.

The final form of the initial Baker and Wurgler (2006) sentiment index ST is the first principal component of the correlation matrix of the final six economic variables illustrated as a lead or a lag. The coefficients are rescaled based on each proxy's level of correlation.

$$\begin{aligned} ST_t = & -0.24CEFD_t + 0.24TURN_{t-1} + 0.253NIPO_t \\ & + 0.257RIPO_{t-1} + 0.112S_t - 0.283P_{t-1}^{D-ND} \end{aligned} \tag{6}$$

where $CEFD$ is the closed-end fund discount, $TURN$ is the natural log of the raw turnover ratio, $NIPO$ is the number of initial public offerings (IPOs), $RIPO$ is the average first day returns of the IPOs, S is the equity share in new issues, and P^{D-ND} is the dividend premium. They use this model to relate sentiment to future stock returns. In addition, they find a contrarian market view. When investor sentiment is below (above) its historical average by more than one standard deviation, subsequent monthly returns are positive (negative).

The follow-up investor sentiment index developed by Baker and Wurgler (2007), (SENT) uses the residuals from the same variables orthogonalized against four macroeconomic variables and a recessionary dummy variable. The second index differs from the first index by using the first principal components of the *changes*, Δ in the six original proxies.

Therefore, the model takes the following form,

$$\begin{aligned} \Delta SENT_t = & -0.17\Delta CEFD_t + 0.32\Delta TURN_{t-1} + 0.17\Delta NIPO_t \\ & + 0.41\Delta RIPO_{t-1} - 0.28\Delta S_t - 0.49\Delta P_{t-1}^{D-ND}, \end{aligned} \tag{7}$$

[23] They develop the sentiment indices using both levels and the changes in levels for the explanatory variables. In addition orthoganlized residuals are used in a second set of models.

where, the change in sentiment is denoted by $\Delta SENT_t$, $\Delta CEFD$ is the change in the closed-end fund discount, $\Delta TURN$ is change in the natural log of the raw turnover ratio, $\Delta NIPO$ is the change in number of initial public offerings (IPOs), $\Delta RIPO$ is the change in the average first day returns of the IPOs, ΔS is the change in the equity share in new issues, and ΔP^{D-ND} is the change in the dividend premium. Each variable is standardized; the turnover, first-day return on IPOs, and the dividend premium are lagged 12-months with a mean zero and a variance of one over the 40-year period.

More recently, Baker, Wurgler, and Yuan (2012) extend the investor sentiment models to five international markets using four indicators of sentiment; a volatility premium, two initial public offering indicators and a market turnover proxy. The resulting explanatory power (R^2) of the models range from 37% for Japan to 48% for Germany. They predict one-month-ahead market returns; global sentiment is the major driver behind country-level results and is a statistically significant contrarian indicator. Baker, Wurgler, and Yuan (2012) conclude that "...investor sentiment affects the time-series of international market-level returns..."For the purpose of our study, we use the Baker and Wurgler (2007) sentiment indices rather than the (2006), because of the improved explanatory power of the one-month-ahead market returns. Data for the indices are from Wurgler's website.

A.2 The Improved Sentiment Index

A new Sentiment Index by Huang et al (2015)[24], (SI) was created to forecast aggregate stock market returns. It also acts as a better proxy than most macroeconomic variables predicting returns sorted by industry, size, value and momentum. The model is primarily based on the Baker and Wurgler's (2007) index, but it greatly reduces the approximation errors that are not relevant in the first principal component for forecasting returns. From an economic perspective, the new index separates out the errors from the stock returns by using a partial least squared (PLS) method first introduced by Wold (1966, 1975), and then improved by Kelly and Pruitt (2013).

The first part of the process involves running N OLS regressions on each individual constant x_i and realized stock returns. R_t,

$$x_i = \pi_{i,0} + \pi_i R_t + u_{i,t-1}, \tag{8}$$

where, $t = 1, ..., T$, and the loadings from π_i are designed to capture the sensitivities of each proxy x_i. The second part the T cross-sectional regressions are run for each time period t for the $\hat{\pi}_i$ estimated in equation (4).

$$x_{i,t} = c_t + S_t^{PLS} \hat{\pi}_i + v_{i,t} \tag{9}$$

where, $i = 1, ..., N$, and S_t^{PLS} is the coefficient of the estimated regression in equation (4). For each i regression the first stage loadings are the independent variables and their

[24]Huang, Jiang, Tu, and Zhou (2015) use a partial least squares methodology and the six sentiment proxies of Baker and Wurgler (2006, 2007) to create a "new investor sentiment index" designed to predict future market returns. Their "aligned investor sentiment index" outperforms the Baker-Wurgler sentiment index with an R^2 value five to six times greater than that of the BW model.

respective sentiment S_t^{PLS} coefficients are estimated. The results become a joint system. Implementing the PLS methodology, the following model is the improvement on the six individual sentiment proxies introduced by Baker and Wurgler (2006, 2007),

$$\begin{aligned} SI^{PLS} &= -0.22CEFD_t + 0.16TURN_{t-1} - 0.04NIPO_t \\ &\quad + 0.63RIPO_{t-1} + 0.07S_t + 0.53P_{t-1}^{D-ND}, \end{aligned} \tag{10}$$

where, SI^{PLS} is investor sentiment using the partial least squared methodology. The new index eliminates the noise in sentiment proxies and is more statistically successful than the Baker and Wurgler (2007) sentiment index in explaining the cross-sectional stock returns.

A.3 The ADS Business Conditions Index

Aruoba, Diebold, and Scotti (2009) observe that "real world" businesses need accurate and timely information about the status of economic conditions. They develop a six-variable dynamic factor model to generate a business conditions index. The index uses variables observed at different frequencies; GDP is quarterly; employment, industrial production, personal income less transfer payments, and manufacturing and trade sales are all monthly; initial jobless claims are reported weekly.

The ADS Index was constructed to track economic activity at different frequencies including real time. The model[25] uses a dynamic factoring technique using precise filtering, and has four different ingredients boosting its ability to track latent macroeconomic variables in a real-time structure. The following is a list of the main features the model:

1. Since business conditions are unobservable variables, the constructs of the index use a dynamic factor modeling approach to bridge these variables into observed indicators.

2. The model integrates business condition variables at different intervals.

3. High frequency indicators are explicitly observed.

4. All business conditions indicators are linear and involve some approximations[26]

The modeling framework takes into account daily, weekly, monthly and quarterly frequencies with data that is not observable. The original model, published in 2009, included four variables: (1) yield curve term premium[27], (2) initial claims for U.S. unemployment, (3) employees on non-agricultural payrolls, and (4) real GDP. Since the release of the original index there are now six variables instead four, all of which are measured at different time intervals. The basic framework of the model is centered on business conditions, x_t, that evolve daily at t, and follow an $AR(\rho)$ process:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + ... + \rho_p x_{t-p} + e_t, \tag{11}$$

[25] As of the date of this paper, the Philadelphia Fed uses the model for macro economic forecasting

[26] In their original index Aruoba, Diebold, and Scotti (2009) incorporated into the model *no approximations*. In 2013, an updated white paper is published citing several changes to the construction of the model.

[27] Yield curve term premium is the difference between the 10-year and 3-month U.S. Treasury yields

where, e_t is white noise with a unit variance and x_t is a scalar of the single-factor model. Then, let y_t^i, represent the ith daily macroeconomic or financial variable at time t with varying lags based on exogenous inputs.

$$y_t^i = c_i + \beta_i x_t + \delta_{i1} w_t^1 + ... + \delta_{ik} w_t^k + \gamma_{i1} y_{t-D_i}^i + ... + \gamma_{in} y_{t-nD_i}^i + u_t^i, \tag{12}$$

where, w_t represents the exogenous variables, u_i are the uncorrelated innovations, and D_i is the series of lags of y_t^i. Since the value of x_t is not observable, a Kalman filter by Durbin and Koopman (2001) is used to smooth and extract the latent state of business activity.

Once this process is complete, the estimation technique, is a *Gaussian pseudo log-likelihood function*,

$$logL = -\frac{1}{2}\sum_{t=1}^{T}[Nlog2\pi + (log|\mathbf{F}_t| + \mathbf{v}_t^{'}\mathbf{F}^{-1}\mathbf{v}_t]. \tag{13}$$

If the components of y_t are not observable, then the likelihood is zero at time t. Hence, if elements of y_t are observed, then N^* is the number variables that will determine $\mathbf{F}_t^*$ and $\mathbf{v}_t^*$ through the filtering process of y_t^*[28].

The results from the model are consistent with the peaks and troughs identified by the NBER, although, the indicator generally reaches its peaks and troughs earlier than the NBER. Data frequency is also considered by Monteforte and Moretti (2013) who combine monthly and daily data to forecast inflation in Europe, and report forecast accuracy is improved substantially.

A.4 The GZ Spread Index

Gilchrist and Zakrajsek (2012) develop a corporate fixed income credit spread model[29] , *GZ Spread*, that measures economic activity comprised primarily of secondary market bond prices issued by U.S non-financial firms. These firms are covered by the S&P Compustat and the Center for Research in Security Prices (CRSP) databases. The month-end prices for the actively-traded fixed income securities in the secondary market are obtained from the databases of Barclay's/Warga and Bank of America's Merrill Lynch[30]. The sample is comprised of senior unsecured debt issued with a fixed coupon schedule. These bonds are selected to make sure the borrowing costs of each firm are on the same time schedule in their capital structure.

Take into consideration a series of cash flows represented by $C(s) : s = 1, 2, 3, ..., S$, comprised of normal coupon payments and return of principle at maturity. The fixed in-

[28]For a detailed explanation of the contemporaneous Kalman filter and the derivation of $\mathbf{F}_t^*$, $\mathbf{v}_t^*$, and y_t^* see section 3.2 of Aruoba, Diebold, and Scotti (2009).

[29]Gilcshrist and Zakrajsek (2012) also use the corporate bond market to calculate an investor's risk appetite by measuring what they call the *excess bond premium* (EBP). This indicator is used to predict the probability of a (NBER) National Bureau of Economic Research dated recession in the next 12 months.

[30]These data sources have been updated from Gilchrist et. el (2009) paper. In addition, they include secondary market prices from dollar-denominated credit securities publicly issued in the United States corporate cash-equivalent market.

come price is represented by,

$$P_{it}[k] = \Sigma_{s=1}^{S} C(s) D(t_s), \tag{14}$$

where, $C(s)$ is the discount cash flow series of the continuously compounded zero-coupon Treasury yields estimated daily by Gurkaynak et al. (2007). The discount function in period t is $D(t) = e^{-r_t t}$, and the price of the risk-free asset is denoted by $P_t^f[k]$. The risk-free asset is used to calculate the yield, $y_t^f[k]$, of the theoretical Treasury rate with the same cash-flows as the primary corporate credit instrument. The yield on the corporate bond k is represented by $y_{it}[k]$, which is represented by the spread between the corporate bond rate, and the estimated yield of the treasury assset with same maturity.

Therefore, the credit spread is represented by,

$$S_{it} = y_{it}[k] - y_t^f[k]. \tag{15}$$

The initial sample period of the Gilchrist and Zakrajsek (2012) model is January 1973 - September 2010[31] with an initial sample size of 5,982 fixed income securities. To make sure their results are not biased by a small number of extreme observations they omit bonds that have credit spreads below 5 basis points and above 3,500 basis points. They also eliminate securities from the sample with a par value below $1 million, and with a maturity date of less than one year or more than 30 years. The last screen is to match the remaining corporate bonds with its issuer's balance sheet and quarterly income data from Compustat and the equity valuations from CRSP. The remaining number of securities in the sample becomes 1,112.

The overall distribution of the sample exhibits some interesting characteristics.[32]

1. There is a positive skew, because many companies have several issues traded on the secondary market

2. There is a small number of senior unsecured debt obligations trading during the sample period.

3. The market values of the outstanding bonds range from $1.2 million to as high as $5.6 billion.

4. A mean maturity date in the sample is 11.3 years, but the average duration is much shorter (6.47 years), because the cash flows are much higher with corporate securities.

5. The average coupon rate is 7.34%.

The final model, uses the yield spreads for each debt security to replicate an artificial risk-free credit instrument that simulates the cash flows of each corporate bond issue. The

[31] The model is updated 2012 and 2014. In 2016, the index is adopted by the FOMC, and is publicly introduced in the FRB FEDS notes: Recession Risk and the Excessive Bond Premium, April 6, 2016

[32] Data taken directly from Gilchrist and Zakrajsek (2012), Table 1: Summary Statistics of Corporate Bond Characteristics

basic framework of the GZ Spread index, S_t^{GZ}, is the mean of the cross-section of credit spreads in the current month t,

$$S_t^{GZ} = \frac{1}{N_t}\Sigma_i\Sigma_k S_{it}[k], \tag{16}$$

where, $S_{i,t}[k]$ is the spread of bond k and N_t is the number of fixed income securities in month t.

A.5 The Financial Uncertainty Index

Jurado, Ludvigson and Ng (2015) JNL develop an independent time-varying macroeconomic uncertainty index. The authors' primary goal was to estimate a superior econometric model without having dependencies on a small number of economic inputs. They begin the process of determining if the overall economy is more or less predictable than in the past. This premise is formalized based on an assumption of uncertainty. Consider h-periods of forward uncertainty, $y_{jt} \in Y_t = (y_{1t}, ..., y_{Nyt})'$ represented by, $U_{jt}^y(h)$, which is the conditional variance of non-forecastable future value,

$$U_{jt}^y(h) = \sqrt{E[(y_{jt+h} - E[y_{jt+h}|I_t])^2|I_t]}, \tag{17}$$

where, $E(.|I_t)$ is the economic variable related to information, I_t, available at a certain periods of time. Assuming, t, is conditioned on information available today, forecast uncertainty, y_{jt+h}, will increase, if today's expectation of the squared error rises.

The index construction assumes two important distinctions of uncertainty. First, the authors differentiate uncertainty in y_{jt} and its conditional variability by removing the forecasting variable $E[y_{jt+h}|I_t]$, prior to calculating volatility. If this step is not performed, estimates may be misspecified, and should be removed from the entire forecast. While this may seem obvious, JNL point out that a good majority of the literature calculating both the cross-section or equity market volatility ignores this step.[33] Second, y_{jt} represent an individual uncertainty but may not be equivalent to a macroeconomic uncertainty spread among multiple series of uncertainty. Most econometric models assume macroeconomic uncertainty is a result of exogenous shocks related to technology advancements, agent preferences, changes in monetary/fiscal policy or firm level volatility and growth.[34] The importance of understanding the volatility of business cycle shocks demonstrates the influence of macroeconomic inputs among different industries and geographic regions.

The overall objective of the (FUI) is to calculate estimates of future uncertainty, $U_{jt}^y(h)$. In order to do this there are three key factors.

1. Calculate an estimate of the forecast $E[y_{jt+h}|I_t$ from an extensive set of forecast vari-

[33]Gilchrist, Sim, and Zakrajsek (2010), implement the Fama and French (1992) financial factors, and Bachmann, Elstner, and Sims (2013) are two notable exceptions.

[34]See Bloom (2009), Sims (2013), who use subjective forecasts of analysts. 5 See, e.g., Bloom (2009), Arellano, Bai, and Kehoe (2012), Bloom, Floetotto, and Jaimovich (2010), Gilchrist, Sim, and Zakrajsek (2010), Schaal (2011); Bachmann and Bayer (2011), and Herskovic et al. (2014) for evidence concerning volatility and firm-level returns.

ables $\{X_{it}\}, i = 1, 2, ..., N$, that range similar to I_t, then approximate the expectation by a diffusion index.

2. Construct a step ahead, h, forecast error, $V^y_{jt+h} \equiv y_{jt+h} - E[y_{jt+h}|I_t]$. The authors require an estimate of conditional volatility of the error, in which they model both the one-step ahead, h, forecast errors and the prediction errors in the inputs.

3. Estimate the macroeconomic uncertainty, $U^y_t(h)$ derived from the individual uncertainty observations, $U^y_{jt}(h)$, which is the equally-weighted estimate of $U^y_t(h)$ of each uncertainty.

The FUI model is created using two different datasets. The first dataset is from Ludvigson and Ng (2010) consisting of 132 macroeconomic variables. Each input is a time series consisting of income, employment, retail, manufacturing real output, sales, unfilled and filled orders, compensation, labor costs, price indices, stock and bond market indices, foreign exchange, and inventories. The second dataset is comprised of 147 financial times series variables from Ludvigson and Ng (2007). Unlike the first dataset, this series includes traditional valuation ratios like earnings-price ratios, dividend-price ratios, growth ratios of dividends, default spreads, corporate bonds yields, bond yields of different ratings, yields on Treasuries, and the Fama and French (1992) cross-section of industry, size, book-market, and momentum portfolio equity returns.

By combining the financial ratio data and the macroeconomic time series data into one aggregate database, the total number of estimators is equal to 279. The authors' emphasize not to *over-represent* the financial ratio series, because the data set has a higher variability than the macroeconomic series. So, they estimate the macroeconomic uncertainty, $U^y_t(h)$, from the 132 macro-series exclusively.

The Markov chain Monte Carlo (MCMC) estimation method is used to calculate the volatility parameters from the least squared residuals[35]. The base model is the mean of the model parameters used in the MCMC estimation to calculate the measure of uncertainty. To obtain the final estimate of macroeconomic uncertainty, $U^y_t(h)$, for h period ahead by averaging,

$$\overline{U}^y_t(h) = \frac{1}{N^y} \sum_{j=1}^{N^y} \hat{U}^y_t(h) \tag{18}$$

where, the mean of uncertainty does not assume individual uncertainties more than the latent variability structure.

[35]They use the stochvol package in R from Kastner and Frothwirth-Schnatter (2013)

Market Timing using Forecasting Indices and Lasso

B Introduction

Active money managers are continually seeking ways to "beat the market" and, generally, market timing and/or superior stock selection strategies are used by money managers. To use market timing to beat the market, the portfolio manager must successfully forecast future market directions and switch monies between common stocks and bonds or cash equivalents as the markets fluctuate over time. Originally, the market timing objective was to be long common stocks (100%) during bull markets and long cash equivalents (100%) during bear markets; risk-adjusted returns were yet to be developed. Interestingly, Phillips and Lee (1989)[50] point out market timers may view risk differently from many money managers and suggest that *A market timer is not concerned with risk in a portfolio sense; risk to a market timer is being in or out of the market at the wrong time.* (p 15). Treynor and Mazuy (1966)[56] point out that the payoffs from successful market timing will resemble that of a call option as managers shift funds between more and less volatile securities as their market forecasts change over time; funds will be invested in stocks (bonds or cash) when the market is forecast to increase (decrease).

The most crucial facet of market timing is the accuracy of market forecasts. Fama, Fisher, Jensen and Roll (1969)[19] posits that because equity market prices "fully reflect" available information, markets are efficient and stock prices follow a "random" walk. This "efficient market hypothesis " implies investors cannot outperform the market over time and that trying to time the market is a futile effort. Early empirical studies of market timing report mixed results, especially when the frequency of switching is considered. However, as discussed below, more recent studies provide evidence that some fund managers outperform the market on a risk-adjusted basis and generate positive alphas over time.

To use market timing to generate positive alphas, money managers must be successful in forecasting future market directions. Because the issue is such a hot-button topic, forecasting research efforts are numerous in both academic and practitioner journals. The focus of this research effort is to assess the efficacy of five forecasting indices and the combination of these indices to forecast or *time* the market.

As pointed out above, successful market timing depends on successful forecasts of market direction. As to be expected, numerous research efforts focusing on different forecasting objectives and different procedures have been published; the efficacy of these procedures are also well researched. There have been substantial changes in the forecasting environment over time. More timely and precise data are available at both the macro- and microeconomic levels. Also, more powerful statistical and econometric techniques are available and the speed of analysis has increased significantly[36]. Because of these technological innovations, it should not be surprising that more current studies may generate results different from earlier studies.

The capital asset pricing model (CAPM) divides an asset's total risk, the return variance, into systematic and unsystematic risk. The model was developed by Sharpe (1964)[53], Lintner (1965)[39], and Mossin (1966)[47]and they posit that an asset's required or expected returns are directly related to its expected risk[37]. Expected returns are unavailable and empirical tests make the assumption that ex-post return distributions can serve as adequate proxies for ex-ante return distributions. The systematic risk measure, beta, is estimated by regressing individual asset returns on a market index. The regression residuals are idiosyncratic and represent unsystematic risk, and in efficient markets, can be diversified away. Fama, Fisher, Jensen and Roll (1969)[19] posits that because equity market prices *fully reflect* available information, markets are efficient and stock prices follow a *random walk*. This *efficient market hypothesis* implies investors cannot outperform the market over time.

[36]An article in March of 2016 by Sprothen and Vera in the The Wall Street Journal reports high-frequency traders can now trade at a speed of 86 nanoseconds. A nanosecond is one-billionth of a second.

[37]In equilibrium, required and expected returns are equivalent. Practically, required returns are generally a function of systematic risk(s) while expected returns are often based on expected cash flows.

Numerous research efforts have examined the CAPM as well as the efficient market hypothesis. Early evidence of a linear relation between beta and asset returns is provided by Black, Jensen, and Scholes (1972)[9], and Fama and McBeth (1974)[20]. They report, however, that the empirical security market line (SML) is flatter than that offered by theory. This results in high-beta (low-beta) assets returns being lower (higher) than the returns predicted by the CAPM[38]. An asset's risk-adjusted return is difference between its actual return and it's expected return; this difference is known as alpha. A positive (negative) alpha indicates the asset's return is greater (less) than the risk-adjusted expected return. Investors seek is to generate positive alphas through market timing and/or security selection. The "flatter" empirical security market line causes the returns of low beta stocks to be above the SML and generate "positive alpha" while high beta stocks generate "negative alpha"Ṫhis "beta anomaly" has evolved into a "Betting Against Beta" (BAB) market factor[39].

The focus of this paper is market timing. Specifically, this paper examines the efficacy of the market timing strategy using five indexes: (1) The improved sentiment index (SI) of Huang et al (2015)[30], (2) The business conditions index (ADS) developed by Aruoba, Diebold, and Scotti (2009)[2] for the Philadelphia Federal Reserve Bank, (3) The Gilchrist and Zakrajsek (2012)[26] credit spread index (GZ), (4) The Economic Policy Uncertainty index (EPU) developed by Baker, Bloom, and Davis (2016)[7], and (5) The Financial Uncertainty Index (FUI) from Jurado, Ludvigson, and Ng (2015)[34].

The stock market forecasting abilities of these five empirical forecasting models are compared and are then combined to generate a better performing model[40]. Finally, the Tibshirani's (1996)[55] "least absolute shrinkage and selection operator" or "lasso" procedures are used to develop a forecasting model that generates an optimal combination of the five models. The models will be compared in at least two ways. First, how many market move-

[38]A number of different reasons have been put forward to explain these results. One such explanation is that some investors face borrowing restrictions and others may be reluctant to borrow. See, for instance, Black (1972, 1993).

[39]See, for instance, Baker, Bradley, and Wurgler (2011), Eisele (2012), and Frazzini and Pedersen (2014).

[40]The forecasting literature indicates combining forecasts can provide a model that outperforms individual forecasts , which is discussed in Section 3.3

ments, both up and down, are correctly forecast? Second, how effective are the models at predicting the big down market movements? For example, two models may exhibit similar forecast accuracy in the number of months correctly forecast but one may be superior in forecasting the down markets. Avoiding the worst months can enhance portfolio returns substantially[41].

The paper proceeds as follows. Section 2 reviews the issues associated with market timing. Section 3 outlines the relation between stock returns and (1) investor sentiment, (2) macroeconomic models and (3) their combinations. Section 4 describes the five forecast indexes and the lasso model selection criteria. Section 5 presents the data and methodology while Section 6 summarizes the empirical results. Concluding remarks are in Section 7.

C Market Timing

An early study by Sharpe (1975)[54] uses an annual investment period to examine market timing and concludes the timer must be correct approximately 70% of the time to outperform the market. Merton (1981)[45] uses the term *macroforecaster* to denote a market timer that predicts whether stocks (bonds) will outperform bonds (stocks). He also observes that successful market timers' will generate portfolio returns that are "virtually indistinguishable" from successful option strategies. Henriksson and Merton (1981)[29] develop a model to identify and specify gains from both market timing and stock selection. They observe that, generally, a researcher will not have access to fund managers' market timing forecasts but may use portfolio returns to infer the forecasts[42]. In an interesting contrast to Sharpe (1975)[54], Droms (1989)[17] shows that a market timer only has to be correct 51% of the time with perfect monthly timing[43].

[41] See Seyhun (2004) for examples associated with avoiding the worst months, and Savor and Wilson (2014) on average announcement excess daily return

[42] They point out that the inferences will "provide noisy estimates of the forecasts." They did not test the model empirically.

[43] More frequent switching generates greater transactions costs; the market timer must be successful enough in predicting the market to overcome these costs. Droms (1989) also points out that the periods under consid-

More recently, Bollen, and Busse (2001)[10] suggest that fund managers may utilize intra-month portfolio adjustments in their timing strategies and use daily returns in their analysis. Interestingly, 34.2 (33.3) percent of the funds outperform (underperform) their benchmark, suggesting a fairly uniform distribution. Fung and Hsieh (2001)[22] examine *market timers* and *trend followers*; a market timer forecasts the future price direction while a trend follower looks for specific price patterns. They observe, like Merton (1981), that the risk for these types of investors cannot be captured by standard linear-factor models because their payoffs are non-linear. Lam and Li (2004)[37] find daily switching can generate excess annual returns of approximately 80% for timers with low transactions costs. In addition, the manager needs to outperform the market only 60% of the time.

Bollen and Busse (2005) extend the Henriksson and Merton (1981) model using the Carhart (1997)[12] factors as explanatory variables. They observe relative portfolio performance seems to persist over time even though many earlier studies report negative excess returns. Their top decile of funds generate statistically significant excess returns of 25-39 basis points per quarter and they attribute this outperformance to momentum and manager skill.

Jiang, Yao, and Yu (2007)[33] examine the market-timing ability of mutual funds and conclude that the average performance attributable to timing skills is positive, and could amount to up to 0.6% annually. Managers response to public information such as aggregate earnings-to-price ratios and aggregate dividend yields appear to be the source of the excess returns[44]. Following Avramov and Wermers (2006)[3], Jiang et al. (2007) also examine industry shifts in mutual fund portfolios over the business cycle. They report fund managers shift their industry composition substantially as they respond to macroeconomic informa-

eration can greatly affect the outcomes. He indicates that during the 1970s and 1980s market timers benefited substantially from high interest rates.

[44]Jiang, Yao and Yu (2007) examine five macroeconomic variables and suggest further research should consider a wider range of variables. The variables they consider are (1) short-term interest rate (one-month T-bill yield), (2) term spread premium (10-year T-bond yield minus the one-month T-bill yield), (3), credit premium (Moody's Baa-rated yield minus Aaa-rated corporate bonds), (4) aggregate dividend yield, and (5) aggregate earnings-to-price ratio of the S&P 500 index.

tion. When the aggregate dividend yield is low (high), the aggregate earnings-to-price ratio is low (high), or when the short rate is high (low), managers switch from high (low) beta industries to low (high) beta industries. Interestingly, Jiang et al. (2007) conclude that "on average, actively managed U.S. domestic equity funds possess positive timing ability."

Cremers and Petajisto (2009)[15] develop an *Active Share* measure to evaluate mutual fund performance. Active Share is the "fraction of the portfolio that is different from the benchmark index" and shows how individual securities are over/underweighted relative to the benchmark. To outperform a benchmark, a fund must differ from the benchmark through factor timing or security selection. When they examine the returns of the funds, they find three features are related to superior performance; high active shares, smaller asset funds, and the best prior-year performance. The return above the benchmark is 6.5% per year net of fees and expenses.

Hubner (2010)[31] observes that many hedge fund strategies are intended to provide convex payoffs while traditional measures are designed to evaluate performance in a stationary mean-variance environment. Like the earlier observations by Treynor and Mazuy (1966) and Merton (1981), Hubner (2010) notes that using linear measurement tools to appraise convex payoffs may cause positive (negative) timers to incorrectly exhibit negative (positive) performance.

Both market timing and stock picking skills are considered by Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014)[35] (KVV) in their evaluation of fund managers. Skill is defined as a "...general cognitive ability to pick stocks or time the market..." They report that a subset of smaller, more active funds generate abnormal returns by successfully performing both tasks. During recessions these successful managers hold more cash and rotate into defensive industries; portfolio betas tend to be lower. During economic expansions, successful managers invest in cyclical industries. Managers with the necessary cognitive skills focus more on microeconomic factors during economic expansions and on macroeconomic information during recessionary periods.

Jagannathan and Korajczyk (2014)[32] (JK) point out that the CAPM divides total variation into systematic and idiosyncratic components. While idiosyncratic risk is assumed to be unpredictable, this is the component that generates the abnormal return or the *alpha* for a portfolio. They define skill as the "the ability to forecast the idiosyncratic returns of assets (which are unconditionally unforecastable)." [45] Also, JK concur with Treynor and Mazuy (1966), Merton (1981), and Hubner (2010) that non-linear payoffs can be generated by dynamic trading strategies and/or the use of derivative securities. For these types of payoffs, traditional performance measures are inadequate and measures that can identify stock picking and market timing abilities are required.

Cullen et al (2015)[23] test their contention that fund managers employ market timing to enhance portfolio performance[46]. The stocks held by a mutual fund are ranked by beta for each fund-quarter and are assigned to twenty equal-value buckets[47]. A weighted average beta of each bucket is established and the value of the stocks in each bucket that are traded during the quarter are regressed on the bucket betas. The estimated coefficients are denoted *timing trade betas* and those that are statistically significant coefficients provide evidence of market timing[48]. They also report a strong association between success in market timing and successful stock selection. This focus provides greater insights into the results of Cremers and Petajisto (2009)[15] and Kacperczyk, Van Nieuwerburgh, and Veldkamp (2014).

[45] They point out that the skill component is also known as "the risk-adjusted return, the abnormal return, Jensen measure, or alpha of the portfolio."

[46] The methodology is outlined in Cullen, Gasbarro, and Monroe (2010).

[47] Because the holdings of the funds are only reported quarterly, inter-quarter trades cannot be considered.

[48] A statistically positive (negative) timing trade beta identifies trades designed to increase (reduce) the portfolio beta by tilting the fund's portfolio towards higher (lower) beta stocks.

D Stock Returns

D.1 Investor Sentiment and Stock Returns

Fisher and Statman (2003)[21] examine consumer confidence, investor sentiment, and stock returns. They say "...consumer confidence predicts economic activity..." and ask "... does consumer confidence also predict stock returns...?" and "...what is the relationship between confidence and investor sentiment...?" (p.115) Their analysis reveals a significantly positive relation between changes in consumer confidence and contemporaneous stock returns. However, consumer confidence appears be a contrarian indicator as an increase in confidence causes the bidding up of stock prices; prices are bid too high causing negative future stock returns. Also, a positive relation is observed between changes in consumer confidence and changes investor sentiment for individual investors, but not with changes in institutional investor sentiment.

Brown and Cliff (2004)[11] examine the relations between several technical sentiment indicators and survey-type investor sentiment measures. Kalman filters, principal component analysis and vector autoregression procedures are used in the analysis. Surveys from the American Association of Individual Investors and Investors Intelligence are used to distinguish between individual and professional investors. Trading volume, type of trade, derivatives, and *other* are the technical indicators. They find that although market returns predict future individual and institutional investor sentiment, but find little evidence that the sentiment measures can predict future stock returns.

Amenc, Curtis and Martellini (2004)[1] use an "almost exhaustive set" of models to examine the risk-adjusted performance of hedge fund managers[49]. Using a normal distribution, positive alphas are generated for all the models; but when the entire return distribution

[49] The models include the CAPM and four CAPM-adjusted models, a payoff distribution pricing model, CAPM with multiple rewarded-risk factors, an Implicit Factor Model using Principal Component Analysis, and a Multi-Index Model. Control variables include: yield on 3-month T-Bills, dividend yield, term spread of 10-year yield minus 3-month yield on treasuries, and a credit spread of AAA bond yield minus Baa bond yield.

is used the average hedge fund alpha is not significantly positive. The alpha rankings of the models are reasonably consistent even though there is substantial disagreement concerning the alphas generated by the different models. Interestingly, a subset of the funds exhibit significantly positive alphas across all models; large funds outperform small funds, and newer funds outperform older funds. As expected, high incentive funds outperform low incentive funds, and market neutral funds outperform the average of other funds. Additionally, negative betas are associated with short-selling funds and administrative fees appear to be irrelevant.

Investor sentiment is the focus of Baker and Wurgler (2006[4], 2007[5]). They develop two investor sentiment indexes, one for sentiment levels and one for changes in sentiment based on their belief that investor sentiment is different for different types of companies. They conjecture that stocks that are more difficult to value and arbitrage are more sensitive to investor sentiment. Hard to value and arbitrage stocks are "...low capitalization, younger, unprofitable, high-volatility, non-dividend paying, growth companies or firms in financial distress..." Six proxies for sentiment are used[50] and a principal components methodology is used to construct the two sentiment indices. The first principal component for both the levels and the changes in levels are the designated sentiment indexes. Finally, the residuals from variables orthogonalized against four macroeconomic variables and a recessionary dummy variable create an orthogonalized index. Baker and Wurgler (2007) relate sentiment to future stock returns and find their sentiment index expresses a contrarian sentiment. When investor sentiment is below (above) its historical average by more than one standard deviation, subsequent monthly returns average 2.75 (-0.41)% for equal weighted indices and 1.18 (-0.34) percent for value-weighted indexes. For the difficult-to-value and arbitrage firms, future returns are especially low (high) when sentiment is high (low). The more stable, value stocks are less affected by sentiment.

[50]In addition to the U.S., the markets of Canada, France, Germany, Japan, and the U.K. are included in the study. The six sentiment proxies are: Closed-end fund discount rate (CEFD), Share turnover (TURN), Number of IPOs (NIPO), First-day returns of IPOs (RIPO), Dividend premium (PDND), and Equity share in new issues (S)

Baker, Wurgler, and Yuan (2012)[6] use four sentiment indicators to investigate sentiment in five international markets[51]. The indicators are an idiosyncratic volatility premium, the number of initial public offerings, the first day of IPO returns, and a market turnover proxy. The models explain a range of the variations from a low 37% for the Japanese market to a high of 48% for Germany. They conclude that international market-level returns are affected by investor sentiment and global sentiment and is a statistically significant *contrarian* indicator.[52]

A *new investor sentiment index* designed to predict future market returns is developed by Huang et al (2015)[30]. They use a partial least squares methodology to create Sentiment Partial Least Squares (SPLS) and report the explanatory power of their model is five to six times larger than the Baker-Wurgler (2006, 2007) sentiment index. Their results are both statistically and economically significant; an increase of one standard deviation in the sentiment index is related to an economically significant negative expected excess market return (-0.58%) in the following month.

D.2 Macroeconomic Models and Stock Returns

As was pointed out earlier, the objectives of different statistical and econometric models can be quite diverse. In addition, is the model designed to explain or to predict the information under consideration? If a model can successfully explain the information of interest, can it be successful in predicting future values? The following section provides a review of several alternative research efforts and the next section reviews the models that will be used in the empirical modeling. A review of several relevant research efforts is presented below in a roughly chronological sequence. It will be observed that the models represent different objectives, data sets and methodologies.

[51] The markets are Canada, France, Germany, Japan, the United Kingdom as well as the U.S.

[52] "A one-standard-deviation increase in the global sentiment index is associated with (45 basis points per month) lower value-weighted market returns and (47 basis points per month) lower equal-weighted market returns." Baker, Wurgler, and Yuan (2012)(p. 282).

Lettau and Ludvigson (2001)[38] examine the relation between the macroeconomic environment and the financial markets. They develop an aggregate consumption-wealth ratio, cay^t, based on consumption, assets and income, and provide evidence deviations from the shared trend is useful in predicting stock returns[53]. They report deviations from the trend can explain a significant proportion of next quarter's returns; positive (negative) trend deviations precede large positive (negative) excess returns. The economic intuition is that if returns are expected to decline in the future, investors will reduce consumption below its long-term relation with income and wealth in order to insulate future consumption from lower returns expected in the future. They conclude that expected returns vary over time and that fluctuations in cay^t, the aggregate consumption-wealth ratio, are useful in forecasting both real and excess stock returns.

Ludvigson and Ng (2007)[41] observe that much of the earlier empirical work use a limited number of conditioning variables in estimating the conditional values related to excess stock market returns. To address this issue, they use 209 quarterly macroeconomic activity variables and 172 financial variables in a dynamic factor analysis framework to examine excess stock market returns[54]. They report the first factor is strongly related to variation in stock market returns and identify it as a *volatility factor.* The second factor is a *risk premium factor* because it is related to the market return and two factors of the Fama-French (1993) three-factor model, the high-minus-low (HML) and the small-minus-big (SMB). When they combine these two factors with a consumption-wealth variable developed by Lettau and Ludvigson (2001), they find that the model predicts 16% of the one-quarter-ahead excess market returns. They conclude that the stability and statistical significance of the factors represents an improvement over the earlier "...low-dimensional forecasting regressions..."

[53] The data definitions and sources are found in the appendix. The macroeconomic data are quarterly values. "The consumption data are for nondurables and services excluding shoes and clothing asset holdings data is the household net worth series (from) the Federal Reserve (and) labor income is defined as wages and salaries plus transfer payments plus other labor income minus personal contributions for social insurance minus taxes."

[54] A list of the variables is provided in an appendix to the paper.

Gilchrist and Zakrajsek (2007)[25] use corporate bond prices, based on secondary market trades, to examine the relation between interest rates and corporate investment spending. Using firm-specific data[55], they construct a user cost of capital and relate this cost to a rate of investment. They report a strong negative relation between corporate bond yields and the rate of investment growth; a 1.0 percentage increase in firms' cost of capital is related to a 50 to 75 basis point reduction in the rate of investment. When default risk increases and yield spreads widen, investment spending is reduced, and overall economic growth slows.

Philippon (2009)[49] uses the bond market to predict corporate investment growth. He posits that a Tobin's q based on bond prices will perform better than a q based on equity prices because "...the bond market is less susceptible to bubbles than the equity market..." His q^{bond} "is a very significant predictor of corporate investment growth" and outperforms the traditional q in predicting capital investment[56]. He suggests the cash flow characteristics of bonds and the specialized knowledge of institutional investors that dominate the corporate bond market cause that market to be less volatile than the equity markets.

Ludvigson and Ng (2009)[42] use dynamic factor analysis and 132 measures of economic activity to determine whether excess bond returns are predictable[57]. The first factor is identified as a *real factor* as it is heavily loaded with measures of production and employment. Interest rate spreads are loaded in the second factor. The third and fourth factors are considered inflation factors as they are correlated with consumer and commodity prices as well as nominal interest rate levels. Interestingly, a stock market factor is the eighth factor and is related to the S&P Index and S&P dividend yield. Their models that include macroeconomic information in addition to financial information increases the predictabil-

[55]The data set is comprised of monthly bond prices of nonfinancial firms with long-term debt actively traded in the secondary market.

[56]Institutional investors with specialized knowledge dominate the corporate bond market.

[57]The data represented "broad categories of macroeconomic time series: real output and income, employment and hours, real retail, manufacturing and sales data, international trade, consumer spending, housing starts, inventories and inventory sales ratios, orders and unfilled orders, compensation and labor costs, capacity utilization measures, price indexes, interest rates and interest rate spreads, stock market indicators, and foreign exchange measures." Their data run from 1964:1 through 2003:12.

ity of excess bond returns. Their models are both statistically and economically significant and the factors explain some 21% to 26% of excess bond returns.

Businesses in the *real world* need timely and accurate information about the status of economic conditions and the Philadelphia Federal Reserve Bank provides data on the Business Conditions Index developed by Aruoba, Diebold and Scotti (2009)[2]. The ADS index uses four variables observed at different frequencies in a dynamic factor model to create an index of macroeconomic conditions. The variables and their frequencies are GDP (quarterly), Employment (monthly), Initial Jobless Claims (weekly) and the Slope of the yield curve term premium (daily)[58]. To accommodate the different frequencies of the observations among the variables, ADS use a Kalman filter process. The model generates an index with the peaks and troughs consistent with those identified by the NBER, although the ADS index tends to reach its peaks and troughs earlier than the NBER[59].

Gilchrist and Zakrajsek (2012)[26] use the prices of outstanding corporate bonds to create the GZ Credit Spread Index and use the resulting index to examine the relation between credit spreads and economic activity. The GZ credit spread index is used to forecast (1) the growth of private (nonfarm) payroll employment, (2) the change in the civilian unemployment rate, and (3) the growth of manufacturing industrial production. Also, they split the index into two components; (1) a component that captures expected default risk of individual firms and (2) a residual component, the excess bond premium, (EBP), that captures credit market sentiment. Their empirical results show that "...both the excess bond premium and the predicted GZ credit spread contain significant independent explanatory power for all three economic indicators, at both the 3- and 12-month forecast horizons..."

[58] The slope of the yield curve is measured as the 10-year yield minus the 3-month yield on U.S. Treasuries. The GEIS index is now known as the Aruoba-Diebold-Scotti Business Conditions Index (the ADS Index) and is available from the Philadelphia Federal Reserve Bank at: http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/

[59] The current ADS index contains six macroeconomic indicators: real GDP (quarterly), payroll employment, industrial production, real personal income less transfers, real manufacturing and trade sales (monthly) and initial jobless claims (weekly). The daily slope of the yield curve has been eliminated. A chronology of extensions may be found at: https://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/extensions.pdf

Also, they observe that during the 2007 to 2009 financial crisis, the creditworthiness of cash-market financial intermediaries weakened substantially resulting in an increase in the credit spread. The decline in the risk-bearing capability of this sector substantially reduced the supply of credit and caused serious adverse consequences for the macroeconomy.[60]

Monteforte and Moretti (2012)[46] use a MIDAS[61] regression approach to combine a monthly core inflation index with daily financial markets data to forecast euro area inflation. The daily variables include commodity prices, exchange rates, short- and long-term interest rates and interest rate spreads; the change in oil price and lagged inflation and are the monthly variables. The inflation futures contracts traded on the Chicago Mercantile Exchange (CME) are used to compare the euro inflation predictions of the MIDAS model. As expected, the models that combine monthly and daily variables outperform models that only use monthly variables. Forecast accuracy is improved about 30% on current inflation and, interestingly, the greatest predictive power is associated with a five-month lag in the core inflation index.

Daily information is used by Da, Engelberg, and Gao (2015)[16] to develop a sentiment index called FEARS, and acronym for "Financial and Economic Attitudes Revealed by Search." The index is based on information found by screening the internet for terms such as *recession, unemployment and bankruptcy* that may cause anxiety for investors. They relate changes in asset prices, volatility and fund flows to the FEARS index[62]. Stock returns are contemporaneously negatively correlated with the index, but over the following two days a reversal occurs. The results provide support for Baker and Wurgler (2006, 2007) in that the strongest reactions are for stocks that are more difficult to value and arbitrage. When the CBOE's volatility index futures are related to the FEARS index a similar reaction is observed; high contemporaneous correlation and a reversal in VIX futures returns over the next two trading days. Additionally, equity (bond) mutual funds experience an outflow

[60] Gilchrist and Zakrajsek (2011) do not attempt to forecast movements in the stock market.
[61] MIDAS is the acronym for MIxed DAta Sampling; the models combine mixed-frequency data.
[62] Google Trends in the Search Volume Index (SVI) is used: http://www.google.com/trends/

(inflow) on the day following the FEARS increase, indicating a *flight to safety* when the FEARS index spikes. Overall, Da, Engelberg, and Gao (2015) conclude aggregate market returns are somewhat predicted by the FEARS index.

Baker, Bloom, and Davis (2016)[7] investigate how newspaper coverage, tax law expirations and economic forecaster disagreement affect micro- and macroeconomic factors. Similar to the FEARS index, Baker, Bloom, and Davis (2016) create an Economic Policy Uncertainty (EPU) index. They screen newspaper coverage for words that suggest economic uncertainty[63]. The second uncertainty variable is the dollar amount of scheduled tax expirations obtained from the Congressional Budget Office. The Philadelphia Federal Reserve Bank's measure of economic forecaster disagreement is the model's third component. Weights are assigned to each factor and the results are summed to complete the EPU Index[64]. They report that stock price volatility is positively related to the EPU while investment and employment in government-dependent sectors are negatively related to the EPU. Three measures of equity uncertainty, the CBOE VIX, an equity market uncertainty index, and a Stock-Jump measure are positively related to the EPU[65].

Jurado, Ludvigson, and Ng (2015)[34] (JLN) observe that "...uncertainty is typically defined as the conditional volatility of a disturbance that is unforecastable..." However, they point out that a highly volatile time series can be divided into two components, a *forecastable* component and an *unforecastable* component. Hence, they define uncertainty as "...the conditional volatility of the purely unforecastable component..." JLN examine macroeconomic uncertainty using both macroeconomic and financial time series data[66]. They point out that the volatility of the stock market is a popular proxy for uncertainty,

[63]Their screens use words such as *economic, economy, uncertain, uncertainty, congress, deficit, Federal Reserve, legislation, regulation or White House*. Their data are available at www.policyuncertainty.com/ . Bloomberg, FRED and Reuters also provide the information.

[64]Their data are available at www.policyuncertainty.com/ . Bloomberg, FRED and Reuters also provide the information.

[65]The correlations between the EPU and the VIX and the Stock-Jump measure are 0.578 and 0.575 respectively. The correlation between the EPU and the equity market uncertainty index is 0.743. The equity market uncertainty index is created in a similar fashion as the EPU index using stock market related words in the screens. Stock-Jumps are defined as +/- 2.5% stock market jumps.

[66]The macroeconomic time series data include 25 financial indicators.

but they conclude that "much of the variation is not driven by uncertainty"and that periods of macroeconomic uncertainty occur less frequently than suggested by *popular* uncertainty proxies. JLN examine the relation between macroeconomic uncertainty and business cycles using principal components analysis. They focus on industrial production, employment and hours as reflecting the business cycle[67]. Twelve factors explain approximately 54% of the total variation; factor one is strongly related to stock market information and explains 37%, factors two and three are related to real economic activity, and risk and bond market term spreads and explain 8% and 3%, respectively. Three episodes of macroeconomic uncertainty are identified, all occurring during recessionary periods[68]. Using autoregression procedures, they find that periods of macroeconomic uncertainty have a half-life of 53 months compared to a half-life of four months for stock market volatility.

Favara et al. (2016)[27] follow-up the study by Gilchrist and Zakrajsek (2012) and use the GZ spread, the Treasury term spread[69], and the real federal funds rate in a probit regression model to forecast the probability of a recession occurring within the next 12 months. They report all three variables are statistically significant[70]. They partition the GZ spread into its two components; the excess bond premium, EBP, and the predicted GZ credit spread and rerun the regressions. The excess bond premium, EBP, is highly statistically significant while the predicted GZ credit spread exhibits no statistical significance. The model "...fits the data quite well..." as few false positives or negatives were generated. Their conclusion is "...that the EBP provides a timely and useful leading indicator of economic downturns..."

Five of the models described above are used to create a one-month-ahead market forecast, and the forecasts are then combined as discussed below. The models are (1) the

[67]The authors provide data for the Financial Uncertainty Index and the Macro Uncertainty Index at their websites. The data are monthly and forecast the indexes 1, 3, and 12 months ahead.

[68]The period examined runs from July, 1960 through December, 2012. The three recessionary periods are 1973-1974, 1981-1982, and 2007-2009.

[69]The term spread is the difference between the 10-year Treasury note yield and the 3-month T-Bill rate.

[70]The GZ credit spread is significant at the 0.01 level; the term spread and the real federal funds rate are significant at the 0.05 level, and the pseudo R^2 is 0.426.

Sentiment Index (SI) of Huang et al (2015), (2) the Business Conditions Index (ADS) of Aruoba, Diebold, and Scotti (2009), (3) the Gilchrist and Zakrajsek (2012) Credit Spread Index (GZ), (4) the Economic Policy Uncertainty Index (EPU) created by Baker, Bloom, and Davis (2015), and (5) the Financial Uncertainty Index (FIU) developed by Jurado, Ludvigson, and Ng (2015).

The one-month ahead forecast for each index is F_{kt}, where k refers to the particular forecasting model, $k \in (1:5)$ and t is the month. The forecasts are rolled forward one month at a time, and the return in the following month, R_{t+1}, is regressed on the forecast as shown:

$$R^m_{t+1} = \alpha + \beta F^k_t + \epsilon_{t+1} \tag{19}$$

The investment results for each forecasting model are compared on both a cross-sectional and time-series basis. The forecasts are then combined as explained in the next section.

D.3 Combining Forecasts

Accurate forecasting can be very difficult and the combining of forecasts is even more complex. Because of the vast forecasting literature, only a few relevant articles are presented here. Two excellent summaries of the forecasting literature are Clemen (1989)[14] who reviews the first 20 years and Wallis (2011)[57] who reviews the combining of forecast literature *forty years later*. Clemen (1989) observes there is an impressive set of forecasting literature and that many disciplines such as "...forecasting, psychology, statistics, and management science..." have made significant contributions to the literature. Both Clemen (1989) and Wallis (2011) consider Bates and Granger's (1969)[8] publication to be a seminal article in the combining of forecasts; they are the first to observe that linear combinations of forecasts can outperform individual forecasts. Wallis (2011) also observes that better forecasts are generated when the forecasts being combined are based on different sets of information. In addition, Rapach, Strauss, and Zhou (2009)[52] also recommend combining individual forecast forecast, because it delivers "...statistically and economically significant out-of-sample gains...".

Granger and Newbold (1974)[28] examine how the correlations among forecasts affect the combined forecast; they report that the correlations should be ignored. Makridakis, Hibon, and Moser (1979)[43] advocate using a simple average of forecasts as simple forecasting models generally outperform the more sophisticated models.

Investment analysts frequently attempt to forecast corporate earnings and these forecasts can be found in popular investment services such as the Value Line Investment Survey and I/B/E/S and Bloomberg. Because of this interest, numerous research efforts with many different methodologies have been used to test forecast accuracy[71]. For example, an early study by Newbold, Zumwalt, and Kannan (1987)[48] compares the forecasting ability of an ARIMA time series model and with the forecasts of Value Line analysts. While the Value Line analysts outperform the time series model on an individual basis, a combination of the two forecasts outperformed the individual forecasts. Lobo (1991)[40] compares corporate earnings forecasts from Value Line and I/B/E/S with statistical forecasts. His results show the combining of statistical model forecasts with analysts' forecasts can improve forecast accuracy. Kim, Lim, and Shaw (2001)[36] use two types of analysts' information, common and private, to examine the efficacy of combining forecasts. Interestingly, the common information is over-weighted relative to the private information when analysts' forecasts are combined. Additionally, forecast errors are reduced by using forecast momentum to adjust the forecast.

Business cycle forecasting has also been the focus of forecast combination. For example, Chauvet and Piger (2008)[13] use a dynamic-factor Markov-switching model and a nonparametric algorithm to examine forecast accuracy associated with NBER economic turning points[72]. Both models identify business cycle troughs better than the NBER, but

[71] Other early corporate earnings forecasts cited by Clemen (1989) include Cragg and Malkiel (1968), Elton, Gruber and Gultekin (1981), Conroy and Harris (1987), Guerard (1987) and Guerard and Beidleman (1987).

[72] The four variables used in their forecasts are (1) nonfarm payroll employment, (2) industrial production,

neither was helpful in forecasting peaks[73]. The forecasting of stock market volatility also has an extensive literature and Poon and Granger (2003)[51] provide an excellent summary in *Forecasting Volatility in Financial Markets: A Review.*[74] They observe that volatility is not the same as risk and concentrate on two issues: (1) *is volatility forecastable?* and (2) *which method will provide the best forecasts*. Their conclusion is that the volatility is forecastable and suggest future issues for research include the relation between forecast accuracy and how far ahead to forecast, and how forecastable is volatility changes.

Genrea et al (2013)[24] ask if the simple average of expert forecasts is the optimal way to combine forecasts. They compare the average of expert forecasts with a number of time series, recursive and principal components models for one- and two-year-ahead forecasts of GDP growth, Harmonized Index of Consumer Prices (HICP) inflation, and the unemployment rate from the European Central Bank's Survey of Professional Forecasters. While a number of combination strategies improve the forecast relative to the benchmark, the gains are modest and they could not identify one particular strategy that dominated across different horizons or variables.

Tibshirani (1996)[55] introduces *lasso* regression procedures[75] to increase prediction accuracy and to ease interpretation issues. The procedures are designed to identify the most important predictor variables from a larger set. Lasso performs covariate selection related to stepwise regression and shrinks large coefficients associated with ridge regression.

The lasso selection criteria model begins with the ordinary least squares (OLS) estimation technique, which minimizes the residual squared error term. Many data scientists are not satisfied with the results of this regression method, according to Tibshirani (1996)

(3) real manufacturing and trade sales, and (4) real personal income excluding transfer payments.

[73]For other articles focusing on business conditions or macroeconomic forecasting see Lettau and Ludvigson (2001), Gilchrist and Zakrajsek. (2007), Aruoba, Diebold and Scotti (2009), Gilchrist and Zakrajsek (2012), Monteforte and Moretti (2013), and Baker, Bloom and Davis (2015).

[74]Poon and Granger (2003) review 93 studies. While recent articles on forecasting volatility are too numerous to cite here, recent papers include Becker and Clements (2008), Xing, Zhang, and Zhao (XZZ) (2010), Cremers and Weinbaum (2010) Bekaert and Hoerova (2014), Fernandes, Medeiros, and Scharth (2014), Engle, Ghysels, and Sohn. (2013), and Suhas (2015).

[75]*lasso* is the acronym for "least absolute shrinkage and selection operator."

for two reason: First, the *prediction accuracy* of the estimates in many cases have a low bias, but have a high variance. This can be improved by *shrinking the coefficients* or setting them to 0, which will slightly increase the bias, but may improve the prediction results and lower the variance. Second, with a large group of predictors often times the *interpretation results* will not exhibit a smaller set of predictors with robust outcomes. In order to improve the overall forecast ability of OLS, Tibshirani (1996) introduces the least absolute shrinkage and selection operator, which "...represents least absolute shrinkage and selection operator..." [76] Lasso shrinks the sum of the coefficients, and reduces other less significant coefficients to 0.

In a recent paper, Elliott and Timmermann (2016)[18] provide an excellent review of the issues faced in forecasting and review the evolution of the processes designed to improve forecast accuracy. Emphasis is on the presumption that economic forecasting is a decision problem and the loss function should be linked to the economic costs of prediction errors. Decision makers must understand the underlying loss functions associated with different methods of forecasting. They point out that "...different forecasting methods often can be combined to produce improved forecasts..." Elliott and Timmermann (2016) point out that combining lasso forecasts with dynamic factor models have the potential to improve forecasts. They also observe that it is *surprisingly difficult* to forecast combinations that outperform an equally-weighted average. They point out that a post-lasso method utilizes lasso to identify the important predictors, other variables are dropped out and the important variables are re-estimated using OLS.

E Forecasting Models

Four of the seven forecasting models used in this paper are presented in Mascio (2017); first, is the Investor Sentiment Index introduced by Huang et al (2015), second, the ADS

[76] Prior to Tibshirani (1996) there were two techniques available for improving OLS: subset selection, and ridge regression.

Business Conditions Index created by Aruoba, Diebold, and Scotti (2009), third, the GZ spread index by Gilchrist and Zakrajsek (2012), and fourth is the Financial Uncertainty Index develop by Jurado, Ludvison, and Ng (2015). The fifth index is the Economic Policy Uncertainty Index, which was developed by Baker, Bloom, and Davis (2015).[77][78] The sixth model combines the five indices into a "kitchen sink" index, and the seventh model is *least absolute shrinkage and selection operator* (lasso), which is a technique used to choose the best forecasting index each period.

Elliot and Timmermann (2016) determine that α will vary the strength of the forecast variables based on understanding the input terms of the initial hypothesis. They create a series of Monte Carlo simulations using a linear forecast methodology and a mean square error (MSE) loss setting. Their experiment uses 10 potential predictor variables that are mutually correlated, and are joint normally distributed random variables. The results are determined by setting $k = 3$ or $k = 6$, so $10 - k$ predictors become irrelevant when $\hat{\beta} = 0$. The forecasts are determines four different way; (1) OLS applying all parameters, (2) complete subset regression (CSR) combining $k = 3$, (3) Akaike Information Criterion (AIC), a weighted average of forecasts across all estimators, and (4) the lasso. They conclude that the *lasso is the preferred model selection method, because it takes a more parsimonious approach using the loss function, when predictor variables are not statistically obvious to include*.

[77]See Appendix A for a detailed description of each forecast index.

[78]This model replaces the Investor Sentiment Index by Baker and Wurgler (2007) from Mascio's (2017) paper The primary reason for the substitution is to have a more diverse group of forecasting models to create a lower correlation amongst the models.

F Data and Methodology

F.1 Data

Five forecasting indexes used to predict the direction of the S&P 500 Large Cap Stock Index (SPX) one month ahead. The models are; the Huang, Jiang and Tu (2015) Improved Sentiment Index (SI), the Arubold, Diebold, and Scotti (2009) Business Conditions Index (ADS), the Gilchrist and Zakrajsek (2012) Credit Spread Index (GZ), the Baker, Bloom, and Davis (2016) Economic Policy Uncertainty Index (EPU), and the Jurado, Ludvigson, and Ng (2015) Financial Uncertainty Index (FUI). The performance results for all of the predictive models are compared on a cross-sectional and time series basis.

In order to match the same date range for all the forecast indexes, the first month of the study is January 1985, and the last month is April 2014, resulting in 396 monthly observations in the sample period. This period differs from Mascio (2017) in order to determine how well the models forecast in a low inflationary environment[79]. The actual data came from several difference sources. The S&P 500 Index monthly data came from Bloomberg. Data for the improved Sentiment Index (SI) came directly from Juang, Jiang, and Tu (2015)[80], and data for the ADS Business Conditions Index came directly from the Philadelphia Federal Reserve Bank[81]. The GZ Credit Spread Index data comes directly from Gilchrist, Zakrajsek, and Favara (2016)[82]. The Economic Policy Uncertainty (EPU) Index is available online[83], and the Financial Uncertainty Index (FUI) data are available from the Board of Governors of the Federal Reserve System.[84]

[79] Mascio (2017) examined the time period from 1/1973 to 3/2014, in which the yield on the US 10-year Treasury note during the 1970's average nearly 12.5%

[80] We sincerely appreciate the responsiveness from Dashan Huang in providing the entire dataset from their 2015 paper.

[81] The web address link is: http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/

[82] The Fed Notes Web address link is: https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/files/ebp.csv

[83] The data are available from: www.policyuncertainty.com/

[84] https://www.federalreserve.gov/econresdata/workingpapers.htm

F.2 Methodology: Logistic Regression

The five predictive indexes from above create the foundation for *beta optimization* introduced by Mascio (2017)[44]. Each model's information can be used to forecast the future direction of the S&P 500 Stock Index's (SPX) direction. The goal is to calculate econometrically an "up" or "down" market forecast for the next month. The dependent variable (SPX1) will have two mutual exclusive outcomes based on a logistic regression procedure.

Specific to this experiment, when the market is forecasted to be up (down) in the subsequent month, the parameter is assigned a value of 1 (0). The five models introduced above are the forecasting factor, $F_{k,t}$, where k refers to the models, $k = 1, \dots 5$, and t is the month.

The prediction procedure begins by first categorizing the market direction as "Up" if the SPX performance in the next month was positive, and "down" if the SPX performance is negative. As a result a binary variable x_{t+1} is calculated for each month t. Next, we organize the current month indexes (factor) $F_{k,t}$ with the next month SPX direction x_{t+1}. Lastly, we use the previous 24 months $(t-25, t-1)$ of observations to generate a rolling estimation window[85]. The probability of an upmarket is denoted by $x = 1$ as $p(x)$. The following is the logistic regression model,

$$\log \frac{p(x_{t+1} = 1)}{1 - p(x_{t+1} = 1)} = \alpha_k + \beta_k F_{k,t}, \tag{20}$$

where, $F_{k,t}$ is an index. Equivalently,

$$p(x_{t+1} = 1) = \frac{1}{1 + e^{-\left(\alpha_k + \beta_k F_{k,t}\right)}}. \tag{21}$$

The maximum likelihood approach is used to estimate the parameters, and the index values at the end of the month are the estimated logit model parameters. The probability

[85] Similar to Mascio (2017) we tested the following estimation window: 24, 48, 72, 96, 120 during the sample period of 351 months. The criteria was based on accuracy of the predication, mean monthly forecasted return and statistical significance (p-value). All results are available upon request.

of the SPX having a positive return for the next month is determined by the previous 24 months of observations. An upmarket (down market) is forecasted when $p(x = 1) \geq 0.5$ ($p(x = 1) < 0.5.$). That is, if $p(x) \geq 0.5$, we predict $x = 1$; if $p(x) < 0.5$, $x = 0$ is predicted. Once the logit forecasts are calculated for each predictive model, either a long (short) position is taken in the SPX[86], depending on the model forecast being Up (Down) for the next month. Then each portfolio performance is calculated based on each models results.

F.3 The "Kitchen Sink" Model

Based on previous forecast literature, a combination of the above predictive models, should outperform the individual forecasts. A kitchen sink variable, denoted ALL, is created by combining the five forecasting models using the following multivariate logistic regression,

$$\log \frac{p(x_{t+1} = 1)}{1 - p(x_{t+1} = 1)} = a + \sum_{k=1}^{5} b_k F_{k,t}. \tag{22}$$

Once these results are calculated the same long (short) positions are taken in the SPX and performance is recorded.

F.4 The LASSO Model

The OLS regression assumes the observations are independent or conditionally independent given a set of predictors. Tibshirani (1996) builds the model based on a given a set of data $(\mathbf{x^i}, y_i), i = 1, 2, ..., N$, where $\mathbf{x^i} = (x_n, ..., x_{ip})^T$represent the forecast variables and y_i are the responses. Assume x_{ij} are standardized where $\frac{\Sigma_i x_{ij}}{N} = 0$, and $\frac{\Sigma_i x_{ij}^2}{N} = 1$. The coefficients are represented by $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_p)^T$. The solution to the following model is a

[86] Trading costs are ignored for all timing models. Discount broker, Charles Schwab, offers trading accounts with no transaction fees see http://www.schwab.com/onesource

quadratic programming problem that demonstrates linear inequality constraints.

$$(\hat{\alpha}, \hat{\beta}) = \arg\min \left\{ \sum_{i=1}^{N} \left(y - i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right\} + \lambda_i \sum_{j=1}^{N} \hat{w}_{i,j} |b_{i,j}|, \tag{23}$$

where, $\lambda_i \sum_{j=1}^{N} \hat{w}_{i,j}|b_{i,j}|$ is the penalty function associated with the lasso parameter λ_i[87], $t \geq 0$ is the tuning parameter subject to $\sum_j |\beta_j| \leq t$. Generality is not lost if $\bar{y} = 0$ when α is deleted, and t is solved when α is $\hat{\alpha} = \bar{y}$. The overall solution is determined when $t \geq 0$ controls the degree of shrinkage applied to the estimates. If $\hat{\beta}_j^o$ is the least squared estimate and $t_0 = \Sigma|\hat{\beta}_j^o|$, when $t < t_0$ the solutions will decrease towards 0, and in some instances the coefficients will be equal to 0.

As a final estimation, the *lasso model* is calculated from section 4.6, where equation 8 is substituted for equation 11. The same methodology is followed to generate the long (short) portfolios in the SPX, and returns are recorded.

F.5 Certainty Equivalents from Logistic Regression

To calculate the certainty equivalent (CEQ), we use the same methodology as DeMiguel et al (2009). This method determines the trade-off between an investor risk tolerance towards investing in stocks or the risk-free asset. We use the S&P 500 Index (SPX) as a comparison portfolio to the five predictive models, the kitchen sink model, and the Lasso to determine the economic significance of the expected returns. A mean-variance investor that is assigned a risk aversion of $\gamma = 1, 2, 3, 4, 5$ can choose between a portfolio allocation that is entirely in stocks based on their return and variance expectations.

An investors risk aversion will determine their peak utility of of each forecast variable in the out-of-sample period. We compute the CEQ return for each k strategy,

$$\widehat{CEQ} = \hat{\mu}_k - \frac{\gamma}{2} \hat{\sigma}_k^2, \tag{24}$$

[87] lasso parameter is selected based on a cross-validation methodology.

where, $\hat{\mu}_k$ and $\hat{\sigma}_k^2$ are the out-of-sample mean excess performance and volatility for each strategy k. And γ is an investor's risk aversion. To conclude if the CEQ returns are statistically different from the SPX, we use the Sharpe ratio test for equality initially developed by Lo (2002)[88]. This methodology calculates the difference between each model's Sharpe ratio and the S&P 500 Index (SPX). When the p-value is above 0.050, we accept the null-hypothesis that the Sharpe ratios are not statistically different.

G Empirical Results

G.1 Descriptive Statistics and Performance of Predictors

Table 10 presents the summary statistics and correlations among the five forecasting models and the S&P 500 index (SPX). Panel A shows the one-month-ahead returns of the SPX has a mean value of eight basis points and a substantially greater standard deviation of 44 basis points. The minimum value of a minus 21.8% that occurred in October 2008, while the maximum of 13.2% occurred in September 1999. Skewness is negative, implying a large left tail, and kurtosis is reasonably close to a normal distribution for the SPX. The other five columns show the same statistics for the individual indices. It should be noted that the values here are *not* comparable with the SPX the returns or with each other as the indices all have different characteristics. However, it is observed that over the period the sentiment index and the business conditions index exhibit negative values while the GZ spread, the economic policy uncertainty, and the financial uncertainty index are all positive. Interestingly, the standard deviations of the GZ index and the FUI index are substantially lower than their mean value. Of the five indices only the ADS exhibits a negative skewness. Four of the five indices indicate substantial kurtosis, with the FUI index being relatively flat.

Panel B presents the correlations of the indices with the one month ahead SPX return as

[88] The statistical test calculates a hypothesis test between Sharpe ratio pairs of given assets. Steven E. Pav developed the statistical test in an R package named "SharpeR".

well as the correlations among themselves. Only the ADS, the business conditions index, is positively related to the SPX1 one-month-ahead return providing support for Arouba Diebold and Scotti (2009). It should be noted that all of the correlations relative to the SPX are modest and have absolute values less than 0.200. Correlations among the indexes, however, range from a low of -0.713 between ADS and GZ, and a high of 0.686 between the financial uncertainty index, FUI, and the GZ spread index. The ADS index is negatively correlated with the four other indices.

As noted in the methodology section, each of the five indices are used to predict the direction of the market one-month-ahead. In addition, an equally weighted model that combine the five individual forecasts, ALL and a lasso model are used in the forecasting. Table 11 presents the correlations between the SPX, the UP movement, and all of the models. First, the correlation between the Up movement and the one-month-ahead returns SPX1 is 0.774, suggesting the SPX has positive returns 77% of the time. Among the forecasting models, the sentiment index has the highest relation with the SPX1 at 0.193, and the lasso model has the second highest correlation at 0.157. The EPU index shows the least correlation at 0.012. When the relations among the monthly return predictions of the models are examined, a substantial range of correlations is observed. The highest correlation is between the GZ credit spread index and the financial uncertainty index, FUI, at 0.653. The returns based on the forecasts of GZ index and the FUI index indicate their forecasts are the most similar among all the models. The smallest correlation is between the economic policy uncertainty index, EPU, and the lasso index at 0.057. The modest correlations among these measures suggest their combinations could be useful in explaining or forecasting SPX returns. It should be noted the correlations between the individual forecasting models and the equally weighted ALL index model are ≥ 0.40. The relationship of the five indexes with the lasso methodology is much more variable, with a range from 0.057 to 0.519, though most are below 0.200.

Annual holding period returns for all of the models and the SPX are presented in Table

12 and some interesting results are observed. For example, in a number of years the returns of each index as well as the combinations are identical to the SPX. For example, in 1993 the SPX return was 9.8%; all of the models generated similar returns. On the other hand, there are instances when the models substantially outperformed or underperformed the index. In 1989, for example, all of the indices and the SPX generate returns of 10.6% except for the ALL model which shows a loss of 16.2%. In years where the SPX exhibits substantial losses, some of the indices perform better and some worse than the SPX. For example, in 2001 and 2002, the SPX had losses of 17.3% and 24.3%, respectively; the lasso index showed a loss of 1.7% in 2001 but a substantial gain of 26.3% in 2002. The following year, 2003, the SPX gained 32.2% while the lasso model generated a loss of 5.7%. Interestingly, four of the individual models and the ALL model generated returns equal to the SPX. In 2008 and 2009, the SPX generated a loss of 40.1% and then a gain of 30%. In these years lasso generated 10.2% and an impressive 74.6% return.

G.2 Logistic Regression Results and Statistical Tests

The summary statistics for monthly portfolio returns for each individual forecasting model, the ALL model, and the lasso model, are presented in Table 13. It can be seen that the minimum value for each of the individual indices and the lasso was a -14.58%; the SPX had a loss of 16.94% and the best performer was the equally-weighted ALL model with a loss of 11.0%. The sentiment index, SI, records the largest median, arithmetic mean and geometric mean among the models and the SPX. The SI index outperforms the SPX on all three measures. The lasso model is very similar with only a few basis points difference for those three measures. The standard deviations are all very similar at approximately 4.2%. The skewness measure for all of the models are negative except for the ALL model. However, the skewness value for the SPX is -0.6024 while the skewness for the models are range from -0.0875 for the lasso model to a high of -0.2111 for the EPU model. The kurtosis values are all approximately 1.0, except for the ALL model with a value of 0.6813.

In order to determine if the probability estimates have the same statistical structure, we performed various non-parametric test. First we examine Figure 1, which is a box-plot of the logistic regression probability estimates. The probability estimate is on the vertical axis and the name of the forecast index is on the horizontal axis. All indices seem to have a similar median, variance, and distribution. The SI, GZ, EPU, FUI, ALL, and LASSO display median probability estimates at or near 0.65, but the ADS is 0.72. The range of estimates between all forecast variables are between 0.05 and 1.000, signaling similar variances. Also, estimates in the 50th percentile (grey box) of each forecast model, lie between 0.50 and 0.75 with the exception of the ADS (0.35, 0.85) and the LASSO (0.425, 0.770). Overall, the box-plot suggests the probability estimates may have similar statistical structure.

Figure 2 illustrates the distribution of 325 monthly probability estimates of each forecast index. The density of each forecast model is on the y-axis and the range of the monthly estimates is on the x-axis. The bandwidths of the SI, ADS, GZ, and the FUI are between 0.0580 and 0.5122, but both the EPU (0.0374) and the LASSO (0.0338) have a narrower range. Not surprising, the ALL model has the largest bandwidth at 0.0968. The peak density of the SI, ADS, GZ, and FUI are all near 1.600, and the ALL model is at 1.5. The EPU has a density just over 3.5, and the LASSO has the highest density of all models at 4.0. The shape of the distribution curve for the SI, ADS and the GZ are similar with a with a flattening of the curve to the left of each model's peak density. Both the EPU and the FUI have distribution curves that start off flat from the origin moving left to right, then gradually increase to a peak probability of 0.600. The ALL model's curve is flat at a density of 0.500 through probabilities 0.012 to 0.750, and peaks at a density of 1.5 and a probability of 1.000. The LASSO model has the most interesting distribution with two different peaks. The first occurs at density 1.25 and probability 0.400; the second is at the peak of the curve at density 4.0. All the models display a right skew and a high amount of kurtosis.

To confirm if the forecast model's probability estimates are normally distributed, we run a Shapiro-Wilks test for normality[89]. Figure 3 illustrates a Q-Q plot of each forecast indices' probability estimates. The vertical axis is the sample quantiles and the horizontal axis is the theoretical quantiles. Each model seems to have a non-Gaussian distribution. The SI, ADS, GZ, EPU, and FUI show upper tail probability estimates that lie below the line of normality; while the combined models (ALL and LASSO) display upper tail estimates above the line of normality. The EPU, FUI, and the LASSO have estimates that lie below the line of normality in the lower tail. All models illustrate properties that lie in both the upper (lower) theoretical and sample quantiles [3, 0.80 (-3, 0.01)].

To determine if the sample variances of the probability estimates are equal, we run a Bartlett test[90]. This test ultimately determines the homogeneity of the probability estimates. A p-value that is higher than 0.050, would suggest that the sample variances are equal, and the null-hypothesis would be rejected. Any p-value below this threshold would mean the variances are unequal. Our test results in a p-value of almost zero and a k-squared value of 320 with 6 degrees of freedom. Therefore, we must determine that the variance of our probability estimates are not equal.

The next statistical test we performed was a Kruskal-Wallis test[91] of the all the forecast models' estimates. We use this non-parametric test to determine stochastic dominance and distributional relationship between the models' probability estimates. If the p-value of the test is below 0.050, we would reject the null-hypothesis that our probability estimates come from the same distributional structure. This test results in a p-value of 0.2250, and a X^2 value of 9.3 with 6 degrees of freedom. In this instance, we would accept the null-hypothesis that our probability estimates are from the same distribution.

In order to compare each forecast models' probability estimates in a pairwise order, we computed a non-parametric Dunn-Test[92], which is illustrated in Table 14. Once again,

[89] Shapiro-Wilkes (1965) test statistic is given as $W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, where is x is the sample mean and a_i is the constants

[90] Snedecor and Cochran (1989), updated the original Bartlett test in their book *Statistical Methods*

[91] Kruskal-Wallis (1952), is a rank sum test of all estimates from $1 to N$; The test will result in a X^2 value, and does not determine a relationship between each estimate in a pairwise order.

[92] Dunn (1961)

if a p-value is higher than 0.050, the null-hypothesis would be accepted, based on the premise that the probability estimates are the same. Each test comparison will result in a Z-statistic value, and its related p-value is parentheses. In every pairwise test the Z-statistics is below 2.850, and all p-values exceed the 0.050 threshold. In addition, nearly all the pairwise tests result in a p-value of 1.000. The SI-ADS (0.8429), ADS-EPU (0.4666), ADS-ALL (0.3429), and the ADS-LASSO (0.0443) are the only model comparisons that have p-values below 1.000. The ADS-LASSO test is the only test statistic that we would reject null-hypothesis that the probability are the same with a Z-value 2.861 and a p-value of 0.0443.

G.3 Forecasting Errors and Model Performance

Monthly forecasting errors are shown in Table 15. The objective of the forecasting models is to correctly predict the direction of the one-month-ahead SPX1; an UP (DOWN) forecast indicates a positive (negative) return is expected. First, consider the two rows at the bottom of the table. These two rows show the number of incorrect and correct forecasts for the 276 months of this study. For example, it can be seen that the sentiment index correctly predicted 161 of the months and was incorrect 115 months. (the market was up 197 months and down 115 months.) It is observed that the range of incorrect forecasts was from 115 to 123. The values in the columns represent the errors associated with the type of error. ER 1 represents an error of predicting the market will go down but the market went up. ER 2 is the reverse, that is, predicting the market will go up but the market actually goes down. For example, consider the year 2000. There are four months when the market went up and eight months when the market went down. The sentiment index SI was incorrect nine of the months, predicting three times down when the market went up and six times up when the market went down. The ADS index was correct on all of the up markets with zero errors but incorrect for all eight of the down markets. This indicates the ADS index predicts an up month every month during the year. Reading across the table, it can be seen that the FUI index and the lasso index predicted an up market every month. The ALL model miss-predicts three of the four up months and seven of the eight down months.

The total row at the bottom of the tables shows the number incorrect predictions overall, the sentiment index SI missed eight of the up markets and 107 of the down markets; the GZ spread index, the FUI index, and the lasso index miss-predicted fewer of the down markets. The seriousness of miss-prediction depends on the return for the month. For example, a prediction that the market will go up when it actually goes down slightly is not nearly as serious as if the market goes down substantially. That is, being wrong when the market is down 20 basis points is not nearly as serious as being wrong when there is a 10% market correction.

Table 16 presents the annualized performance analysis for all models and the SPX, For example, over the period considered, the SPX generated an annual return of 7.9%. It is seen that all models except EPU generated higher returns; the sentiment index SI and lasso are the better performers. Figure **??** illustrates the SI index and the lasso index generate the highest amount of cumulative wealth during the sample period. Though, the lasso index (Figure **??**) has a much higher cumulative return than the ALL model.

The standard deviations for all models and the SPX are similar at approximately 14.5%. The sentiment index, SI, and the lasso are the two best performers when considering the annualized Sharpe ratio. They are also the best performers when considering the Sortino ratio (downside variation is used in a Sharpe-type ratio) and when considering the cumulative positive and negative returns as shown by the omega ratio. The economic policy uncertainty, EPU, model shows the least tracking error; the semi-deviation and gain deviation are reportedly similar across the models and with the SPX. All the models demonstrate slightly lower loss deviations compared to the SPX. The maximum drawdown measure shows that all models outperform the SPX. In this case, the lasso model reduces drawdown to less than one-half of the drawdown of the SPX. The EPU index has the best up capture ratio at 0.938, closely followed by the sentiment index SI at 0.928. Lasso at 0.387 and the

ALL index of 0.408 are best at down capture. The probability of a of an Up movement illustrated in Figure **??**, shows the ALL model and the lasso model have a much larger variance in their probability forecasts than the individual models.

G.4 Statistical Significance of Model Probability Estimates

Table 17 illustrates the results of monthly logistic regressions using the maximum likelihood approach. Each observation is recorded at the end of each year for the month of December for the period between 01/1985 - 04/2014 for the five individual forecasting models, the All model, and the Lasso model. All reported statistics are on a rolling 24-month estimation window. The far right seven columns display each coefficient (β %), which results in a binary forecast of the S&P 500 index (SPX) as UP or Down in the following next month. An UP month is predicted when the $p(x = 1) \geq 0.5$, and Down month is predicted when ($p(x = 1) < 0.5$). If $p(x) \geq 0.5$, we forecast $x = 1$; if $p(x) < 0.5$, $x = 0$ is predicted. The middle six columns show the Newey-West t-statistics, and the far right six columns are the McFadden $pseudo - R^2$.

Each β coefficient is a forecast of the next month's market (SPX) direction. Of the 27 years in the sample period there are eleven years in which the forecast indices differ in their prediction. For example, in 1990, the FUI (0.411) and the lasso (0.471) model both correctly forecast a Down market for the month, while the other three forecast an UP market. All the models predict an Up month with coefficients between 0.578 and 0.897. Overall, ignoring statistical significance here, the models are predicting an upward trend for the market with probabilities greater than 0.50 for 144 of the possible 189 forecast-months. Two of the models, FUI and SI, predict the most down months at 10 and 9 months, respectively. The GZ model is the most optimistic, only predicting four down months while the All and Lasso models predict 5 and 6 down months, respectively. All seven models correctly anticipate the bursting of the tech bubble in 2000 with probabilities below 0.50. The five individual models and the Lasso correctly predicted the 2008 down month, but, unexpectedly, the All model predicted an up month for 2008. The All model exhibits the greatest average probability over the period at nearly 78% while FUI average probability is the least at about 59%. The Lasso model average probability is 61%.

When statistical significance is considered, the All model is clearly the most significant. For this model, there are six coefficients which are statistically different from zero with t-statistics greater than 2.00. The average t-statistic (1.84) of the All model is the greatest compared to the other models. The EPU model has no significant coefficients. The GZ model has two that are significant and each of the remaining models exhibit only one statistically significant coefficient. The All model appears to be more effective than the individual models from a significance evaluation.

The $pseudo - R^2$ values provide similar results. The average $pseudo - R^2$ for the All model (0.78) is substantially greater than for the other models. In addition, it has the greatest maximum value at 0.89 and the greatest minimum value at 0.56. None of the individual models have average $pseudo - R^2$ exceeding 0.50 and four of the five models have minimum $pseudo - R^2$ values less than 0.10.

G.5 Prediction Accuracy and Significance of the Combined Model

The five individual forecast indices display similar statistical significance (p-values) as discussed in Mascio (2017). The primary focus of this section is to discuss each model's forecast ability and significance within the combined ALL model, as well as the lasso model[93].

In this section, we will focus on four different time periods to evaluate the statistical significance of how well the "kitchen sink" (ALL) and the lasso model does at predicting

[93] According to Hastie, Tibershirani, and Wainwright's (2016) book, typical significance tests assume the null is a random variable with little relationship on the final outputs. Therefore, applying a traditional significance test to the lasso model would result in biased outcomes, because lasso already selects the best variables from the penalty function. However, there are some recent papers that have developed methodologies to extract significance of the variables within the model selection frame work. (For example: see Dezeure, Brenhlmann, Meier and Meinshausen (2015))

the one month ahead returns of the S&P 500 Index. The first is the three month period before and after the 1987 stock market crash. The second period is the recession period (2000-2003) following the dot.com bubble and the 9/11 terrorist attacks on New York City. Third, is the financial crisis and recession between 2008-09, and lastly, the recovery period from 2011-14.

In our sample period, the worst month of returns with respect to the S&P 500 Index was a -21.73% in October 1987. The ALL model forecasted correctly this large market drawdown with a probability of an UP market at 0.00505. Of the five forecasting indices that make up the ALL model, the FUI is the main contributor to the accurate prediction for the October 1987 crash. Referring to Figure 11, the FUI variable has a mean monthly p-value of 0.0865, and a monthly prediction accuracy of 80% during the three month period before and after the stock market crash of October 1987. Conversely, the lasso model during the same period forecasts a probability an UP market at 75%, resulting in an inaccurate prediction.

Both the ALL and lasso models do especially well at forecasting the returns of the S&P 500 Index between the period of March 2000 and February 2003. During this recessionary period, the ALL and lasso model (Figure **??**) have a mean monthly forecast of an UP market at 41.15% and 41.87% respectively. Accurately forecasting the negative monthly returns of the SPX nearly 75% of the time.

The individual variables of the ALL model have meaningful p-values during the period. For example, the first three months of the recession, the SPX had a total return of -5.27%. The SI (Figure 7), ADS (Figure 8), and the EPU (Figure 10) recorded mean monthly p-values of 0.0817, 0.0723 and 0.0460 respectively. Moreover, the ALL model had a prediction accuracy of 66% during this period.

During this thirty-five month period, the SPX had ten months of negative returns greater than 5.00%.[94] The lasso model predicted all ten negative months accurately, with the ex-

[94]Top ten negative SPX return months from March 2000 - February 2003; 10/2000 (-8.00%), 2/2001 (-9.22%), 7/2001 (-6.41%), 8/2001 (-8.17%), 3/2002 (-6.14%), 5/2002 (-7.24%), 6/2002 (-7.90%), 8/2002

ception of August 2002. The ALL model correctly forecasted six of the ten months, but was successful in predicting the -11.00% SPX return in August of 2002, which was the largest monthly drawdown during the period. The main contributor to the accuracy of the ALL model forecasts during these months was the GZ index. Referring to Figure 9, the GZ variable had a p-value of 0.1207 in 6/2002, 0.0607 in 8/2002, and 0.0839 in 11/2002.

From May 2008 through February 2009 represents the worst ten months of consecutive returns for the SPX during our sample period. The highest(lowest) monthly returns for the period was realized in July 2008(September 2008) of 1.21%(-16.9%). The total drawdown of the SPX was a -60.6%. Both the ALL and the lasso model accurately predicted each negative monthly return during the period, with the exception of 9/2008 the ALL model forecasted an Up month. The lasso(ALL) portfolios had a mean monthly probability of an Up market of 41.19%(19.95%), resulting in a 85% prediction accuracy for both models. The most interesting aspect of this period was the ALL model's significance among its individual variables. The lowest(highest) monthly p-values recorded was 0.1176(.9659) by the FUI(EPU) variables. Overall, the mean monthly p-value of the ALL model's five forecasted variables was 0.4119. Even though the ALL model's prediction accuracy was excellent, the individual ALL model inputs significance was poor.

The SPX total return between 9/2011 through 3/2014 was 52.89%. The lasso(ALL) model correctly forecasted the monthly market return 76.01%(61.29%) of the time. From November 2011 to February of 2012, both the lasso and the ALL model accurately predicted the 12.41% return of the SPX each month during the period. The primary variables of the ALL model responsible for the accurate prediction during this time period was the SI, GZ, and FUI with a mean monthly p-value of 0.0760, 0.0564, and 0.0832 respectively.

Another noteworthy time period is between March 2012 through August 2013, when the SPX was up 15.94%. Once again the ALL and the lasso models correctly predicted the monthly returns on the market more 65% of the time. The main contributors to the

(-11.00%)

ALL model's forecast accuracy was the GZ and FUI indices recording monthly p-values of 0.0589 and 0.0495 respectively. The EPU index in the ALL model, reference Figure 10, had a mean monthly p-value of 0.6925 during this period. This is the only index within the combined ALL model that does not contribute to the forecast of accuracy during this time period.

G.6 Persistency of the Estimates using Autocorrelation

Determining the persistency of the probability estimates follows a similar autoregressive time series (AR) process as Mascio (2017). The reported statistics have autocorrelation monthly lags between 1 and 20. We use an AR(1) model to compare the probability estimates of the five predictive models, the kitchen sink (ALL) model, and the Lasso model. The first-order autocorrelation model is represented as a standard linear difference equation,

$$X_k = \rho X_{k-1} + \epsilon, \tag{25}$$

where $k = 0, 1, 2, ...$ and ϵ_k is the error terms computed from the time series. The difference equation relates X_k, which is the original value at some previous time k, and a lagged parameter at X_{k-1}.

Figure 12 represents seven probability estimates of the autocorrelation function (acf) for the five predictive models (SI, ADS, GZ, EPU, and FUI) and the combined models (ALL and Lasso). The plots display the lag periods in months on the horizontal axis and the acf on the vertical axis. Each model is a result of an AR(1) process. Each of the forecast models show significant persistency of the probability estimates, but differ from Mascio (2017). The primary difference is the longevity of the lag period of persistency. For example, Mascio (2017) has at least an 18-month lag period among all the variables with a 95% confidence interval. Our results have significant lag periods that last in some cases only five months.

The SI, ADS, GZ, and FUI models drop below the 0.10 acf value by the 12-month lag period. Each model does have a value over 0.80 in the first 2 lagged months, but by the 5th lag the value drops below the 0.60 threshold. The SI at lag month 11, drops below the 0.10 level, and continues to fall through the 20-month lag period. The ADS model drops below the 0.50 acf value at lag 5, but levels out at roughly the same value until lag 10 (when it then falls to 0.05 by lag 20). The GZ and the FUI index both have similar characteristics, with the steepest slope from lag 1(0.75) to lag 5(0.40). The GZ, however, drops below the 0.10 value at lagged month 15, while the FUI falls below the same threshold at lag 11. At the 5-month lag, persistence levels of all the models remains above 0.40, with the exception of the ALL model. Only the Lasso has persistency beyond the 15th lag period. The SI, ADS, GZ, and FUI indices seem to all have similar behavior for all lagged periods.

The EPU model is the only forecast model that has a low persistency for all lagged periods. In the first lagged month the acf value is below 0.50, and then slightly increases in lags 2 and 3, but peaks at lag 5 (0.55). The model continues to decrease in persistency from lag 6 to 20, recording the lowest values of all the individual predictors.

The ALL model has the lowest persistency of all the models, with a 1-month lag value at 0.50. The model then drops below 0.10 at lag month 5, remaining at that level through the 20th lagged period. Conversely, the Lasso is the only model that remains above the statistically significant level of 0.10 for all 20 lagged months.

G.7 Forecast Portfolio Rankings

Table 18 ranks each individual forecasting strategy based on its mean annual return over the entire sample period (1988-2014). Referring to figure **??**, the sentiment index (SI), generates the highest annual return at 11.7%, closely followed by lasso at 11.3%. These two models have identical standard deviations; the sentiment index has a higher Sharpe ratio. The drawdown associated with the lasso is approximately one-half the drawdown of the other models and less than one-half that of the SPX. The economic policy uncertainty

(EPU) index is the only index that underperforms the SPX.

Based on the fact that the EPU index annual returns underperforms the SPX, and it has the highest monthly mean p-value of all the indices, we eliminate it from the dataset, and re-estimate the ALL and lasso models. These results are shown in Table 19. Meaningful improvements are shown with the lasso model but not the ALL model. The annualized lasso return is increased from 11.3% to 12.6%, which is an annual gain of 1.3%. Figure **??** shows the lasso model without the EPU, illustrating the highest cumulative wealth of all strategies. Table 19 also represents the annualized Sharpe and Sortino ratios improve for the lasso model, along with the omega ratio. While the semi-deviation, the gain deviation, loss deviation, and the up capture for the lasso are all similar to the previous results. The maximum drawdown (34.7%) and the down capture (26.2%) are the only statistics the lasso model did not improve upon.

G.8 Certainty Equivalent Performance Results

The results of the certainty equivalent returns are reported in Table 20. Similar to Mascio (2017), the CEQ performance results are consisted with the geometric returns of each forecasting portfolio. Each column represent an investor's risk aversion $\gamma = 1, 2, 3, 4, 5$. The first row are the CEQ returns of the S&P 500 Index (SPX), and below the line displays the CEQ monthly performance of each forecast model. Under each model's CEQ returns in parentheses is the p-value associated with the Sharpe ratio equality test with the SPX. The base CEQ model is calculated as a result of the logistic regression with a probability threshold of 0.50.

When an investor's risk aversion is $\gamma = 1$, which is the base CEQ returns for DeMiguel et al (2009), we find similar result to Mascio (2017). The CEQ return for the SPX (0.00350) is lower than all forecast strategies with the exception of the FUI (0.00285). The ALL [0.00366, (0.45029)] only slightly outperforms the SPX. In addition, The SI [0.00616, (0.00075)], GZ [0.00533, 0.00250], EPU [0.00446, (0.00778], and LASSO [0.00772, (0.00006)] have higher CEQ returns than the SPX, and its p-value are below 0.050 suggesting that the Sharpe ratios are different than the SPX.

At $\gamma = 2$, The CEQ returns of the SPX are 0.00255 compared to the highest returns, LASSO [0.00602, (0.00096)], and the lowest, FUI [0.00091, (0.86007)]. Once again the SI [0.00522, (0.00289)], GZ [0.00439, (0.00845)], EPU [0.00352, (0.02279)], and LASSO all have higher returns than the SPX as well as p-values below 0.050. At $\gamma = 3$, SI [0.00429, (0.00958)], GZ [0.00345, (0.02442)], and LASSO [0.00507, (0.00365)] display returns higher than SPX (0.00159), and p-values below the 0.050. All other models may have higher CEQ returns than the SPX, but their p-values are above 0.050.

When risk aversion is $\gamma = 4$, the SI [0.00335, (0.02712)], and LASSO [0.00411, (0.01184)] are the only models to outperform the SPX (0.00064), and record a p-value that are statistically different than the SPX's Sharpe ratio in a pairwise test. At $\gamma = 5$, LASSO [0.00316, (0.03282)] is the only model that has returns higher than the SPX (-0.00031), and a p-value below 0.050.

Lastly, we perform a robustness test to determine if the logistic regression probability cut-off is optimal. Similar to the results of Mascio (2017), we changed the probability from 0.50 to 0.75 and rerun the logistic regression model. We then calculate the CEQ returns, and record them in Table 20. At all levels of risk aversion, $\gamma = 1, 2, 3, 4, 5$, we find that each forecast index have lower CEQ values than the SPX, and p-values above 0.050. Interestingly, the LASSO has higher returns than the SPX, and p-values below 0.050 for $\gamma = 1, 2$.

H Conclusion

The focus of this paper is to determine the efficacy of the five forecasting indexes, a combined *kitchen sink* forecast model, and a lasso model for predicting the one-month-ahead S&P 500 Index (SPX) returns. A market timing strategy called "beta optimization" introduced by Mascio (2016) is used to change the beta of a portfolio if future market expectations change. Functionally, when the SPX is forecasted to be up (down) the portfolio is directed to have a long (short) position.

Since the mid-2000's many indexes have been developed to explain or forecast the overall stock market's direction. Huang, Jiang, and Tu (2015) create the improved sentiment index (SI); Aruoba, Diebold, and Scotti (2009) develop the (ADS), which is designed to measure current business conditions; Baker, Bloom, and Davis (2015) construct the Economic Policy Uncertainty (EPU); Gilchrist and Zakrajsek (2012) create the GZ Spread model (GZ), and Jurado and Ludvigson (2015) construct a time-varying Financial Uncertainty Index (FUI). These five models are used in an econometric framework to construct *forecasts* of the SPX monthly direction. These forecasts are used to build either long or short positions in the S&P 500 Index.

A buy-and-hold strategy is used as a comparison benchmark to illustrate the effectiveness of the five forecast model portfolios, a kitchen sink portfolio, and lasso model portfolio at predicting the direction of the market. Over the sample period the SPX was Up 197 months and Down 115 months. The ADS and the EPU had the lowest forecast errors, when considering Up market movement. Both have six forecast errors, but, interestingly, the errors occur in different months. Conversely, all five forecast model including the kitchen sink model (ALL) are inaccurate over 100 months, when predicting a down month for the SPX. The lasso model was the best at forecasting down months for the SPX, being correct 23 times. In addition, the lasso model was most successful at predicting the major market drawdowns during the recession of the early 2000's, and the financial crisis of 2008-2009. Moreover, in volatile markets investors would benefit the most from mimicking the forecasts of the lasso model to avoid large monthly losses. Also, consistent with prior research, combining the five forecasts improved prediction accuracy, but the timing of the predictions did not produce favorable returns. Conversely, the LASSO model produced the highest average annual returns, the highest CEQ returns, the most statistically significant returns and lowest monthly drawdown by nearly 50% over the individual five forecast portfolios. In addition, after dropping the EPU index from the estimation sample, the LASSO model has the highest annualized performance of all the models tested.

Active investment managers have the ability to modify their strategies by shifting between high and low beta stocks. Though, in order to measure a managers ability to make successful predictions requires an appropriate methodology to evaluate investment managers'; decision making ability. Past research has shown that successful market timing strategies can yield option-like payoffs, but overall volatility risk of these strategies has not accurately quantified. Even though the beta optimization strategy using the lasso model has worked well over the examined time period, investors should always be cognizant of any forecasting strategy, because as John Maynard Keynes observed, "the market can stay irrational longer than you can remain solvent."

References

[1] AMENC, N., CURTIS, S., AND MARTELLINI, L. The alpha and omega of hedge fund performance measurement. *Lille: EDHEC Risk and Asset Management Research Centre* (2004).

[2] ARUOBA, S. B., DIEBOLD, F. X., AND SCOTTI, C. Real-Time Measurement of Business Conditions. *Journal of Business & Economic Statistics 27*, 4 (2009), 417–427.

[3] AVRAMOV, D., AND WERMERS, R. Investing in mutual funds when returns are predictable. *Journal of Financial Economics 81*, 2 (2006), 339–377.

[4] BAKER, M., AND WURGLER, J. Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance LXI*, 4 (2006), 1645–1680.

[5] BAKER, M., AND WURGLER, J. Investor sentiment in the stock market. *Journal of Economic Perspectives 21*, 2 (2007), 129–151.

[6] BAKER, M., WURGLER, J., AND YUAN, Y. Global, local, and contagious investor sentiment. *Journal of Financial Economics 104*, 2 (2012), 272–287.

[7] BAKER, S., BLOOM, N., AND DAVIS, S. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics 131*, 4 (2016), 1–69.

[8] BATES, J. M., AND GRANGER, C. W. J. The Combination of Forecasts. *Journal of the Operational Research Society 20*, 4 (1969), 451–468.

[9] BLACK, F., JENSEN, M. C., AND SCHOLES, M. *The Capital Asset Pricing Model: Some Empirical Tests*, vol. 81. Praeger Publishers Inc., 1972.

[10] BOLLEN, N. P., AND BUSSE, J. A. On the Timing of Mutual Fund Managers. *Journal of Finance 56*, 3 (2001), 1075–1094.

[11] BROWN, G. W., AND CLIFF, M. T. Investor sentiment and the near-term stock market. *Journal of Empirical Finance 11*, 1 (2004), 1–27.

[12] CARHART, M. On persistence in mutual fund performance. *Journal of Finance 52*, 1 (1997), 57–82.

[13] CHAUVET, M., PIGER, J. M., AND FEDERAL RESERVE BANK OF ST. LOUIS. A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics 1*, 1 (2008), 42–49.

[14] CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*, 4 (1989), 559–583.

[15] CREMERS, K. J. M., AND PETAJISTO, A. How Active Is Your Fund Manager A New Measure That Predicts Performance. *Review of Financial Studies 22*, 9 (2009), 3329–3365.

[16] DA, Z., ENGELBERG, J., AND GAO, P. The Sum of All FEARS: Investor Sentiment and Asset Prices. *Review of Financial Studies 28*, 1 (2015), 1–40.

[17] DROMS, W. G. Market Timing as an Investment Policy. *Financial Analysts Journal 45*, 1 (1989), 73–77.

[18] ELLIOT, G., AND TIMMERMANN, A. G. Forecasting in Economics and Finance, 2016.

[19] FAMA, E. F., FISHER, L., JENSEN, M. C., AND ROLL, R. The Adjustment of Stock Prices To New Information. *International Economic Review 10*, 1 (1969), 1.

[20] FAMA, E. F., AND MACBETH, J. D. Long Term Growth In A Short Term Market. *The Journal of Finance 29*, 3 (1974), 857–885.

[21] FISHER, K. L., AND STATMAN, M. Consumer Confidence and Stock Returns. *The Journal of Portfolio Management 30*, 1 (2003), 115–127.

[22] FUNG, W., AND HSIEH, D. A. The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies 14*, 2 (2001), 313–341.

[23] GASBARRO, D., CULLEN, G., AND LEMONROE, G. Mutual Fund Trades: Timing Sentiment and Managing Tracking Error Variance. *SSRN Electronic Journal* (2015).

[24] GENRE, V., KENNY, G., MEYLER, A., AND TIMMERMANN, A. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting 29*, 1 (2013), 108–121.

[25] GILCHRIST, S., AND ZAKRAJSEK, E. Investment and the Cost of Capital: New Evidence from the Corporate Bond Market. *National Bureau of Economic Research Working Paper Series 14863* (2007).

[26] GILCHRIST, S., AND ZAKRAJSEK, E. Credit Spreads and Business Cycle Fluctuations. *American Economic Review 102*, 4 (2012), 1692–1720.

[27] GILCHRIST, S., ZAKRAJSEK, E., FAVARA, G., AND LEWIS, K. Recession Risk and the Excess Bond Premium. *Fed Notes April*, 8 (2016), 1–3.

[28] GRANGER, C., AND NEWBOLD, P. Spurious Regression in Econometrics. *Journal Econometrics 2* (1974), 111–120.

[29] HENRIKSSON, R. D., AND MERTON, R. C. On market timing and investment performance. II. statistical procedures for evaluating forecasting skills. *The Journal of Business 54*, 4 (1981), 513.

[30] HUANG, D., JIANG, F., TU, J., AND ZHOU, G. Investor Sentiment Aligned: A Powerful Predictor of Stock Returns. *Review of Financial Studies 28*, 3 (2015), 791–837.

[31] HÜBNER, G. The Alpha of a Market Timer. *SSRN Electronic Journal* (2010), 1–44.

[32] Jagannathan, R., and Korajczyk, R. A. Market Timing. *SSRN Electronic Journal* (2014).

[33] Jiang, G. J., Yao, T., and Yu, T. Do mutual funds time the market? Evidence from portfolio holdings. *Journal of Financial Economics 86*, 3 (2007), 724–758.

[34] Jurado, K., Ludvigson, S. C., and Ng, S. Measuring Uncertainty. *American Economic Review 105*, 3 (2015), 1177–1216.

[35] Kacperczyk, M., Nieuwerburgh, S. V., and Veldkamp, L. Time-varying fund manager skill. *Journal of Finance 69*, 4 (2014), 1455–1484.

[36] Kim, O., Lim, S., and Shaw, K. The Inefficiency of the Mean Analyst Forecast as a Summary Forecast of Earnings. *Journal of Accounting Research 39*, 2 (2001), 329–335.

[37] Lam, K., and Li, W. Is the Perfect Timing Strategy Truly Perfect? *Review of Quantitative Finance and Accounting 22*, 1 (2004), 39–51.

[38] Lettau, M., and Ludvigson, S. Consumption, aggregate wealth, and expected stock returns. *Journal of Finance 56*, 3 (2001), 815–849.

[39] Lintner, J. Security Prices, Risk, and Maximal Gains From Diversification. *The Journal of Finance 20*, 4 (1965), 587–615.

[40] Lobo, G. J. Alternative methods of combining security analysts' and statistical forecasts of annual corporate earnings. *International Journal of Forecasting 7*, 1 (1991), 57–63.

[41] Ludvigson, S. C., and Ng, S. The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics 83*, 1 (2007), 171–222.

[42] Ludvigson, S. C., and Ng, S. Macro factors in bond risk premia. *Review of Financial Studies 22*, 12 (2009), 5027–5067.

[43] MAKRIDAKIS, S., HIBON, M., AND MOSER, C. Accuracy of Forecasting: An Empirical Investigation. *Journal of the Royal Statistical Society. 142*, 2 (1979), pp. 97–145.

[44] MASCIO, D. A. *Sentiment Indices and their Forecasting Ability*. PhD thesis, EDHEC Risk Insitute, 2016.

[45] MERTON, R. C. On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business 54*, 3 (1981), 363–406.

[46] MONTFORTE, L., AND MORETTI, G. Real-Time Forecasts of Inflation: The Role of Financial Variables. *The Journal of Forecasting 32*, 1 (2012), 51–61.

[47] MOSSIN, J. Equilibrium in a capital asset market. *Econometrica 34*, 4 (1966), 768–783.

[48] NEWBOLD, P., ZUMWALT, J. K., AND KANNAN, S. Combining forecasts to improve earnings per share prediction. An examination of electric utilities. *International Journal of Forecasting 3*, 2 (1987), 229–238.

[49] PHILIPPON, T. The Bond Market's q. *Quarterly Journal of Economics 124*, 3 (2009), 1011–1056.

[50] PHILLIPS, D., AND LEE, J. Differentiating Tactical Asset Allocation from Market Timing. *Financial Analysts Journal 2*, March-April (1989), 14–16.

[51] POON, S. H., AND GRANGER, C. Forecasting financial market volatility: A review. *Journal of Economic Literature 41*, June (2003), 478–539.

[52] RAPACH, D. E., STRAUSS, J. K., AND ZHOU, G. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies 23*, 2 (2009), 821–862.

[53] SHARPE, W. F. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance 19*, 3 (1964), 425–442.

[54] SHARPE, W. F. Likely Gains from Market Timing. *Financial Analysts Journal 31*, 2 (1975), 60–69.

[55] TIBSHIRANI, R. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society B 58* (1996), 267–288.

[56] TREYNOR, J., AND MAZUY, K. Can mutual funds outguess the market? *Harvard Business Review 44*, 4 (1966), 131–136.

[57] WALLIS, K. F. Combining forecasts: Forty years later. *Applied Financial Economics 21*, No.1-2 (2011), 33–41.

A Forecasting Models Construction

A.1 The Modern Investor Sentiment Index

The initial series of modern investor sentiment indices were primarily introduced by Baker and Wurgler (2006, 2007), which were primarily based on the behavioral decision making ability of institutional investors. The two indices are comprised of six different proxies to measure investor sentiment. The explanatory variables in the first index (2006) were from *levels*, and the second index (2007) were from *changes in levels where the residuals were orthogonalized*. Baker and Wurgler (2006) made the assumption that individual companies are most sensitive to investor sentiment when the firm is hard to value and arbitrage. These characteristics are prevalent in stocks with low capitalization, less mature, unprofitable, high volatility, non-dividend paying, growth companies or firms in financial distress. The initial model was centered around the following equation.

$$E_{t-1}[R_{it}] = a + a_1 T_{t-1} + \mathbf{b}_1^{'} \mathbf{x}_{it-1} + \mathbf{b}_2^{'} T_{t-1} \mathbf{x}_{it-1}, \tag{26}$$

where i denotes firms, t is the time variable, and $\mathbf{x}$ is a vector of characteristics, T represents investor sentiment, a_1 is the *generic* effect of sentiment, and $\mathbf{b}_1$ is the vector of *generic* effect of characteristics. The null is $\mathbf{b}_2 = 0$ as a compensation for systematic risk; the alternative is $\mathbf{b}_2 \neq 0$ which reveals the cross-sectional movement in sentiment characterized by mispricing. This equation is called a conditional characteristics model adding similar terms to the Daniel and Titman (1997) model.

The Baker and Wurgler (2006) sentiment index is made of six economic variables from the first principal component of the correlation matrix in a lead or lag form. Each coefficient is re-measured based on each variables level of correlation. The following is the final model of the first index.

$$\begin{aligned} SENT_t &= -0.24CEFD_t + 0.24TURN_{t-1} + 0.253NIPO_t && (27)\\ &+ 0.257RIPO_{t-1} + 0.112S_t - 0.283P_{t-1}^{D-ND}, && (28) \end{aligned}$$

where $CEFD$ is the closed-end fund discount, $TURN$ is the natural log of the raw turnover ratio, $NIPO$ is the number of initial public offerings (IPOs), $RIPO$ is the average first day returns of the IPOs, S is the equity share in new issues, and P^{D-ND} is the dividend premium. This equation directly relates sentiment to future stock returns, which can be construed as a contrarian indicator[95]. If sentiment is below its historical average by more than one standard deviation, subsequent monthly returns are positive. The converse of these assumptions also holds true.

The following year, Baker and Wurgler (2007), use the same residuals from the six proxies orthogonalized against four economic variables in the 2006 index, but they add a recessionary dummy variable. They also measured investor sentiment based on the first

[95] Baker, Wurgler and Yuan (2012) extend the investor sentiment models to five international markets using four indicators of sentiment; a volatility premium, two initial public offering indicators and a market turnover proxy. The resulting explanatory power (R^2) of the models range from 37% for Japan to 48% for Germany. They predict one-month-ahead market returns; global sentiment is the major driver behind country-level results and is a statistically significant contrarian indicator

principal components of the changes in the six economic proxies. The final model takes the following form,

$$\begin{aligned} \Delta SENTO_t &= -0.17\Delta CEFD_t + 0.32\Delta TURN_{t-1} + 0.17\Delta NIPO_t \quad (29) \\ &+ 0.41\Delta RIPO_{t-1} - 0.28\Delta S_t - 0.49\Delta P_{t-1}^{D-ND}, \quad (30) \end{aligned}$$

where, the Δ in sentiment is denoted by $\Delta SENTO_t$, the change in the closed-end fund discount is represented by $\Delta CEFD$, $\Delta TURN$ is change in the natural log of the raw turnover ratio, $\Delta NIPO$ is the change in number of initial public offerings (IPOs), $\Delta RIPO$ is the change in the average first day returns of the IPOs, ΔS is the *change* in the equity share in new issues, and the change in the dividend premium ΔP^{D-ND}. Each variables macroeconomic state is omitted and then standardized. The model's levels, turnover, first-day return on IPOs, and the dividend premium have a twelve month lag period with mean of zero and the variance spans a forty year period.

The Huang et al (2015) model was developed to predict aggregate equity market performance, and act as a proxy for macroeconomic variables. The economic perspective of the improved model removes the errors from the equity returns by applying a partial least squared (PLS) methodology introduced by Wold (1966, 1975) and later improved by Kelly and Pruitt (2013). The foundation of the model is the Baker and Wurgler (2007) index, though it significantly reduces the *approximation errors* that are irrelevant in the first principal component for predicting future performance.

Initially, by calculating OLS regressions on N individual constants $x_{i,t-1}$ and past stock returns. R_t, the individual constants take on the following value,

$$x_{i,t-1} = \pi_{i,0} + \pi_i R_t + u_{i,t-1}, \quad (31)$$

where, $t = 1, ..., T$, and the loadings from π_i are meant to capture the sensitivities of each variable $x_{i,t-1}$.

After calculating the OLS regressions, the final step is to estimate the T cross-sectional regressions for each time period t for the $\hat{\pi}_i$ estimate from the sentiment index (SI) developed by Baker and Wurgler (2007) sentiment index.

$$x_{i,t} = c_t + S_t^{PLS}\hat{\pi}_i + v_{i,t} \quad (32)$$

where, $i = 1, ..., N$, and S_t^{PLS} is the estimated coefficient from the (2007) model. The i regressions are the independent variables and each respective sentiment S_t^{PLS} coefficients are estimated, resulting in a joint system. Applying this new PLS method improves the model first introduced by Baker and Wurgler (2006, 2007), as follows,

$$\begin{aligned} SI^{PLS} = &-0.22CEFD_t + 0.16TURN_{t-1} - 0.04NIPO_t \\ &+ 0.63RIPO_{t-1} + 0.07S_t + 0.53P_{t-1}^{D-ND}, \end{aligned} \quad (33)$$

where, SI^{PLS} is the partial least squared investor sentiment. This new model completely reduces the noise in six economic proxies, and is better at explaining the cross-sectional stock returns than the Baker and Wurgler (2007) sentiment index.

A.2 The ADS Business Conditions Index

Aruoba, Diebold, and Scotti (2009) created a dynamic business conditions index based on six different proxies in order to measure the risk associated with economic conditions. The model boasts in its ability to combine variables at different time frequency. For example, gross domestic product is measured quarterly; personal income and industrial production is monthly; initial jobless claims are weekly. The ultimately goal of the index is to measure economic activity in real time.

The Aruoba, Diebold, and Scotti (ADS) model[96] implements a dynamic factoring procedure with four different ingredients having the ability to track latent macroeconomic variables in a real-time structure. There are four main aspects of the index. First, since business conditions are unobservable variables, the constructs of the index use a dynamic factor modeling approach to bridge these variables into observed indicators. Second, the model integrates business condition variables at different intervals. Third, high-frequency indicators are explicitly observed. Fourth, all business conditions indicators are linear and involve some approximations[97]

The framework within the index takes into account quarterly, monthly, weekly and daily non-observable data points. In 2009, when the initial model was developed, the model included four proxies: initial claims for U.S. unemployment, yield curve term premium[98], employees on non-agricultural payrolls, and real GDP. The model was updated in 2013, increasing the number of variables from four to six calculated at different frequencies.

The current index follows an $AR(\rho)$ process centered on business conditions, x_t, measured daily at t:

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + ... + \rho_p x_{t-p} + e_t, \tag{34}$$

where, e_t is white noise with a unit variance and x_t is a scalar of the single-factor model.

The ith daily macroeconomic variable at time t is illustrated by y_t^i at different lags influenced by exogenous variables.

$$y_t^i = c_i + \beta_i x_t + \delta_{i1} w_t^1 + ... + \delta_{ik} w_t^k + \gamma_{i1} y_{t-D_i}^i + ... + \gamma_{in} y_{t-nD_i}^i + u_t^i, \tag{35}$$

where, w_t represents the exogenous inputs, and u_i are the non-correlated innovations. D_i is the series of lags of y_t^i, and because the value of x_t is non-observable,

The final estimation process is a Gaussian pseudo log-likelihood, which applies a Kalman filter introduced by Durbin and Koopman (2001), used to level and remove the latent state of business activity.

$$logL = -\frac{1}{2}\sum_{t=1}^{T}[Nlog2\pi + (log|\mathbf{F}_t| + \mathbf{v}_t^{'}\mathbf{F}^{-1}\mathbf{v}_t]. \tag{36}$$

Since the elements of y_t are non-observable, the likelihood is zero at time t. If the components of y_t are observed, then N^* is the number variables which will determine $\mathbf{F}_t^*$ and $\mathbf{v}_t^*$

[96]The Philadelphia Fed uses this index for macro economic forecasting

[97]In their original index, Aruoba, Diebold, and Scotti (2009) do not use approximations. A white paper was published in 2013 explaining several changes to the construction of the model.

[98]Yield curve term premium is the difference between the 10-year and 3-month U.S. Treasury yields

via the Kalman filtering process of y_t^{*}[99]. The index results are consistent with the economic expansions and contractions identified by NBER, even though the signals reach their peaks and troughs sooner than NBER.[100]

A.3 The Economic Policy Uncertainty Index

The basic premise of the economic policy uncertainty index (EPU) developed by Baker, Bloom and Davis (2015) is to measure how different news events affects the behavior of investors within the financial markets. The EPU tracks newspaper coverage, tax law expirations and economic forecaster disagreement to construct the index. The main focus of the index is how newspaper articles effect the overall economic environment. Search words such as "economic", "economy", "uncertain", "uncertainty", "congress", "deficit", "Federal Reserve", "legislation", "regulation" or "White House" are used in their screens. The model makes use of the Congressional Budget Office's dollar value of scheduled expirations to develop the second variable, which is based on the uncertainty surrounding whether congress will renew various different tax code provisions.

A similar index is created by Da, Engelberg, and Gao (2015). They search the internet for items causing anxiety to investors, such as *recession, unemployment or bankruptcy*. Their Financial and Economic Attitudes Revealed by Search (FEARS) model confirms Baker and Wurgler's (2007) premise that stocks are hard to value and arbitrage when they exhibit strong reaction to sentiment behavior.

The authors try to alleviate some of the difficulties related to the newspaper articles reliability, accuracy, bias and consistency in three ways. Initially, they calculate the uncertainty of implied and realized stock market volatility to measure uncertainty. Then, they evaluate the EPU against other sentiment models[101]. Lastly, they perform explicit human audits of over 10,000 US newspaper articles from data providers such as Bloomberg, Reuters, Haver, and FRED.[102]

A.4 The GZ Spread Index

Gilchrist and Zakrajsek (2012) develop a corporate fixed income credit spread model[103], *GZ Spread*, which measures economic activity and overall financial market risk. This index is primarily comprised of secondary market bond prices issued by U.S non-financial firms. These firms are covered by the S&P Compustat and the Center for Research in Security Prices (CRSP) databases. The month-end prices for the actively-traded fixed income

[99] A descriptive explanation of the contemporaneous Kalman filter and the derivation of $\mathbf{F}_t^*$, $\mathbf{v}_t^*$, and y_t^* reference section 3.2 of Aruoba, Diebold, and Scotti (2009).

[100] Monteforte and Moretti (2013) combine monthly and daily data to forecast inflation in Europe, and report forecast accuracy is substantially improved using varying degrees data frequency.

[101] These models include US Federal Reserve Beige book, and forecaster disagreement for US federal expenditures

[102] This suggests an information uniformity through a group of University of Chicago students whom read each article based on a 64 page training guide.

[103] Gilcshrist and Zakrajsek (2012) also use the corporate bond market to calculate an investor's risk appetite by measuring what they call the *excess bond premium* (EBP). This indicator is used to predict the probability of a (NBER) National Bureau of Economic Research dated recession in the next 12 months.

securities in the secondary market are obtained from the databases of Barclay's/Warga and Bank of America's Merrill Lynch[104]. The sample is comprised of senior unsecured debt issued with a fixed coupon schedule. These bonds are selected to make sure the borrowing costs of each firm are on the same time schedule in their capital structure.

Take into consideration a series of cash flows represented by $C(s) : s = 1, 2, 3, ..., S$, comprised of normal coupon payments and return of principle at maturity. The fixed income price is represented by,

$$P_{it}[k] = \Sigma_{s=1}^{S} C(s) D(t_s), \tag{37}$$

where, $C(s)$ is the discount cash flow series of the continuously compounded zero-coupon Treasury yields estimated daily by Gurkaynak et al. (2007). The discount function in period t is $D(t) = e^{-r_t t}$, and the price of the risk-free asset is denoted by $P_t^f[k]$. The risk-free asset is used to calculate the yield, $y_t^f[k]$, of the theoretical Treasury rate with the same cash-flows as the primary corporate credit instrument. The yield on the corporate bond k is represented by $y_{it}[k]$, which is not prejudice to the spreads represented by paring the corporate bond rate with the estimated yield of the Treasury asset with same maturity.

Therefore, the credit spread is represented by,

$$S_{it} = y_{it}[k] - y_t^f[k]. \tag{38}$$

The initial sample period of the Gilchrist and Zakrajsek (2012) model is January 1973 - September 2010[105] with an initial sample size of 5,982 fixed income securities. To make sure their results are not biased by a small number of extreme observations they omit bonds that have credit spreads below 5 basis points and above 3,500 basis points. They also eliminate securities from the sample with a par value below $1 million, and with a maturity date of less than one year or more than 30 years. The last screen is to match the remaining corporate bonds with its issuer's balance sheet and quarterly income data from Compustat and the equity valuations from CRSP. The remaining number of securities in the sample becomes 1,112.

The overall distribution of the sample exhibits some interesting characteristics.[106]

1. There is a positive skew, because many companies have several issues traded on the secondary market

2. There is a small number of senior unsecured debt obligations trading during the sample period.

3. The market values of the outstanding bonds range from $1.2 million to as high as $5.6 billion.

[104]These data sources have been updated from Gilchrist et. el (2009) paper. In addition, they include secondary market prices from dollar-denominated credit securities publicly issued in the United States corporate cash-equivalent market.

[105]The model is updated 2012 and 2014. In 2016, the index is adopted by the FOMC, and is publicly introduced in the FRB FEDS notes: *Recession Risk and the Excessive Bond Premium*, April 6, 2016

[106]Data taken directly from Gilchrist and Zakrajsek (2012), Table 1: Summary Statistics of Corporate Bond Characteristics

4. A mean maturity date in the sample is 11.3 years, but the average duration is much shorter (6.47 years), because the cash flows are much higher with corporate securities.

5. The average coupon rate is 7.34%.

The final model, uses the yield spreads for each debt security to replicate an artificial risk-free credit instrument that simulates the cash flows of each corporate bond issue. The basic framework of the GZ Spread index, S_t^{GZ}, is the mean of the cross-section of credit spreads in the current month t,

$$S_t^{GZ} = \frac{1}{N_t} \Sigma_i \Sigma_k S_{it}[k], \tag{39}$$

where, $S_{i,t}[k]$ is the spread of bond k and N_t is the number of fixed income securities in month t.

A.5 The Financial Uncertainty Index

Jurado, Ludvigson, and Ng (2015) develop an independent time-varying macroeconomic uncertainty index. The authors' primary goal was to estimate a superior econometric model to measure the market risk environment without having dependencies on a small number of economic inputs. They begin the process of determining if the overall economy is more or less predictable than in the past. This premise is formalized based on an assumption of uncertainty.

Consider you have h-periods of forward uncertainty, $y_{jt} \in Y_t = (y_{1t}, ..., y_{Nyt})'$ represented by, $U_{jt}^y(h)$, which is the conditional variance of non-forecastable future value,

$$U_{jt}^y(h) = \sqrt{E[(y_{jt+h} - E[y_{jt+h}|I_t])^2|I_t]}, \tag{40}$$

where, $E(.|I_t)$ is the economic variable related to information, I_t, available at a certain periods of time. Assuming, t, is conditioned on information available today, forecast uncertainty, y_{jt+h}, will increase, if today's expectation of the squared error rises.

The index construction assumes two important distinctions of uncertainty. First, the authors differentiate uncertainty in y_{jt} and its conditional variability by removing the forecasting variable $E[y_{jt+h}|I_t]$, prior to calculating volatility. If this step is not performed, estimates may be misspecified, and should be removed from the entire forecast. While this may seem obvious, JNL point out that a good majority of the literature calculating both the cross-section or equity market volatility ignores this step.[107] Second, y_{jt} represent an individual uncertainty but may not be equivalent to a macroeconomic uncertainty spread among multiple series of uncertainty. Most econometric models assume macroeconomic uncertainty is a result of exogenous shocks related to technology advancements,

[107] Gilchrist, Sim, and Zakrajsek (2010) and Bachmann, Elstner, and Sims (2013), implement the Fama and French (1992) financial factors within their estimates

agent preferences, changes in monetary/fiscal policy or firm level volatility and growth.[108] The importance of understanding the volatility of business cycle shocks demonstrates the influence of macroeconomic inputs among different industries and geographic regions.

The overall objective of the Financial Uncertainty Index (FUI) is to calculate estimates of future uncertainty, $U^y_{jt}(h)$. In order to do this there are three key factors.

1. Calculate an estimate of the forecast $E[y_{jt+h}|I_t$ from an extensive set of forecast variables $\{X_{it}\}, i = 1, 2, ..., N$, that range similar to I_t, then approximate the expectation by a diffusion index.

2. Construct a step ahead, h, forecast error, $V^y_{jt+h} \equiv y_{jt+h} - E[y_{jt+h}|I_t]$. The authors require an estimate of conditional volatility of the error, in which they model both the one-step ahead, h, forecast errors and the prediction errors in the inputs.

3. Estimate the macroeconomic uncertainty, $U^y_t(h)$ derived from the individual uncertainty observations, $U^y_{jt}(h)$, which is the equally-weighted estimate of $U^y_t(h)$ of each uncertainty.

The FUI model is created using two different datasets. The first dataset is from Ludvigson and Ng (2010) consisting of 132 macroeconomic variables. Each input is a time series consisting of income, employment, retail, manufacturing real output, sales, unfilled and filled orders, compensation, labor costs, price indices, stock and bond market indices, foreign exchange, and inventories. The second dataset is comprised of 147 financial times series variables from Ludvigson and Ng (2007). Unlike the first dataset, this series includes traditional valuation ratios like earnings-price ratios, dividend-price ratios, growth ratios of dividends, default spreads, corporate bonds yields, bond yields of different ratings, yields on Treasuries, and the Fama and French (1992) cross-section of industry, size, book-market, and momentum portfolio equity returns.

By combining the financial ratio data and the macroeconomic time series data into one aggregate database, the total number of estimators is equal to 279. The authors' emphasize it is not to *over-represent* the financial ratio series, because the data set has a higher variability than the macroeconomic series. So, they estimate the macroeconomic uncertainty, $U^y_t(h)$, from the 132 macro- series exclusively. The base model is the mean of the model parameters used in the Monte Carlo Markov Chain estimation to calculate the measure of uncertainty. To obtain the final estimate of macroeconomic uncertainty, $U^y_t(h)$, for h period ahead by averaging,

$$\overline{U}^y_t(h) = \frac{1}{N^y} \sum_{j=1}^{N^y} \hat{U}^y_t(h) \tag{41}$$

where, the mean of uncertainty does not assume individual uncertainties more than the latent variability structure.

[108] See Bloom (2009), Sims (2013), who use subjective forecasts of analysts. See, e.g., Bloom (2009), Arellano, Bai, and Kehoe (2012), Bloom, Floetotto, and Jaimovich (2010), Gilchrist, Sim, and Zakrajsek (2010), Schaal (2011); Bachmann and Bayer (2011), and Herskovic et al. (2014) for evidence concerning volatility and firm-level returns.