Consumer Sentiment and its Power in Predicting Economic Success

Inge Siemes
ANR: 201668

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN COMMUNICATION AND INFORMATION SCIENCES,

MASTER TRACK DATA SCIENCE BUSINESS & GOVERNANCE, AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES OF TILBURG UNIVERSITY

Supervisor: dr. H. J. Brighton
Second reader: dr. M. Atzmüller

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands

May 2019

Preface

The basis for this research stems from my interest in the combination of human behavior and the use of data. By examining consumer sentiment and its effect on the state of the economy I believe to have proven my abilities necessary to complete my Master of Science degree in Data Science.

I sincerely want to thank my supervisor Henry Brighton for the excellent insights and straightforward advice in this brief period of time. Secondly, I want to thank my friends and family and in particular Adrian, who supported me during the final stage of this master.

Inge Siemes,
Tilburg, May 2019

Abstract

This thesis investigates how accurately consumer sentiment can predict various indicators of economic conditions within states of the United States. These economic conditions are business performance, coincident index, and economic recession. Both directions of the relationship between sentiment and the economic conditions are explored to account for causality issues. Based on data from the Michigan Consumer Sentiment Index from January 2005 to April 2018 the predictive power of consumer sentiment within the 50 states of the U.S. is assessed. While previous research has focused on examining the effects of consumer sentiment, the fact that state economies differ greatly received little attention in this regard. This research extends prior research by incorporating state-level information and controlling for economic recession. The temporal structure in the data is represented by the sliding window approach and the time windows are used to evaluate the predictive power of consumer sentiment. Several regression and classification models are built to predict future economic conditions. The results show that consumer sentiment has power in predicting business performance and coincident index. This predictive power does however not improve during economic recession and it does not differ over state clusters systematically. Consumer sentiment does not have predictive power for economic recession. Rather, it turns out that economic recession has power in predicting sentiment. While not fully consistent over the different models used in the analysis, the predictive power of sentiment regarding business performance and coincident index is higher than it is vice versa. Based on the results, it can be concluded that consumer sentiment has some predictive power for economic success.

Table of contents

1. Introduction

The creation of a strong consumer experience is a strategic priority for businesses (Zaki & Neely, 2019). The shift in marketing from an orientation on the short-term sales process to the long-term engagement point of view in which consumer activity towards the business is the focus marks the importance of the consumer in increasing business performance (Harrison-Walker, 2001; Calder et al., 2018). Economic success is furthermore dependent upon the economic situation in which consumers make purchases and the capacity to predict this product demand (Huth et al., 1994).

Consumer sentiment can be described as a statistical measure that indicates the overall health of the economy based on the experiences and opinions of consumers (Kenton, 2018). The Survey Research Center of the University of Michigan developed an index that tracks this sentiment in the United States. The index encompasses attitudes towards personal finances, general business conditions, and buying or market conditions and prices (Curtin et al., 2019). The index has proven to be an accurate indicator of the future course of the economy. It is consistent with the timing of business cycle peaks and troughs and conforms to business expansions and contractions. When many people change from an optimistic to a pessimistic view towards economic prospects, it has been found that expenditures are extensively postponed.

Consumers react differently to the same economic phenomena over time (Curtin et al., 2019). Understanding the rationale of consumers for their actions provides insight in why this is the case. Since the timing at which consumer purchases take place influences the entire course of the economy it is important to understand the direction of the relationship between consumer sentiment and economic conditions (Curtin et al., 2019). However, no unambiguous conclusion has been drawn in previous research. Researchers have hypothesized the direction of the relationship between the state of the economy and sentiment in different ways. Some results prove the dominant role of consumer sentiment on the state of the economy (Lahiri et al., 2016; Vuchelen, 2004; Lozza et al., 2016) whereas others state that changes in consumer sentiment are generally a reflection of the economic circumstances (Throop, 1992; Fuhrer, 1993) or question any relevance of sentiment (Barnes & Olivei, 2017; Carroll et al., 1994).

Furthermore, publications on consumer sentiment are mainly found in management journals and primarily focus on managerial actions rather than the underlying causes and consequences of consumer sentiment (Beyari et al., 2017). Besides, “the state of the economy” is mainly defined through consumption expenditures and does not cover the whole concept of economic health within a country. There are several indicators that play an important role in summarizing the health of the economy, but these have never been considered with respect to consumer sentiment. An example is coincident index. This index is created using a model developed by Stock and Watson (1989). The model is based on the belief that co-movements in macroeconomic time series can be captured using a single statistic representing the overall state of the economy (Megna & Xu, 2003). This coincident index is coincident with business cycles, which are "expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals…" (Burns & Mitchell, 1947, pp. 3). Another example is economic recession, basically interchangeable with the health of the economy. Previous studies have investigated the relationship between consumer sentiment and stock or oil crisis (Golinelli & Parigi, 2004; Zouaoui et al., 2011; Anderson et al., 2011), the mediated effect of news media on economic recession via consumer sentiment (Starr, 2012; Blood & Philips, 1995), and the effect of economic recession on consumer behaviour (Flatters & Willmott, 2009; Voinea & Filip, 2011), but relatively little attention has been paid to the direct effect of consumer sentiment on economic recession within a specific area.

To be able to prevent decreases in business performance, anticipate changes in coincident index, and diminish the effects of economic recession in an area it is important to investigate the effect of consumer sentiment on these various economic conditions. Understanding and predicting the impact of consumer sentiment will allow businesses and governmental agencies to react to a decrease in sentiment and prevent negative consequences (Nguyen et al., 2012). It would be even more valuable to be able to report any differences over economic phases in terms of the consequences of consumer sentiment. This might clarify co-existing circumstances which exacerbate or reduce the consequences of consumer sentiment, which in turn can be addressed too. Previous research only provides general statements and results without focusing on discrepancies between specific times or circumstances.

This thesis aims to contribute to prior research by investigating how accurately consumer sentiment can predict various indicators of economic conditions – business performance, coincident index, and economic recession – within states of the United States. In order to answer this question, the following research questions will be addressed:

RQ1: To what degree can consumer sentiment predict business performance, and can this prediction be done more accurately during economic recession?

RQ2: To what degree can consumer sentiment predict state recession, and can this prediction be done more accurately during economic recession?

RQ3: To what degree does the predictive power of consumer sentiment differ over state clusters with regard to business performance and coincident index?

Because there is not yet an unambiguous conclusion about the direction of the effect between consumer sentiment and the state of the economy this thesis will also investigate what predictive power is associated with a change of the causal direction in the last research question:

RQ4: To what degree would it be more insightful to change the causal direction between the indicators of economic conditions and consumer sentiment?

The main findings of this thesis are that consumer sentiment has predictive power for business performance and coincident index. This does however not improve during economic recession or differ over state clusters systematically. Furthermore, consumer sentiment does not predict economic recession. Rather, economic recession has predictive power for sentiment. While not fully consistent over the different models used in the analysis, the predictive power of consumer sentiment regarding business performance and coincident index is higher than it is when changing the causal direction of these relationships.

2. Related work

This section provides a theoretical background on the relationship between consumer sentiment and the indicators of economic conditions within states of the United States. First, a more detailed description of consumer sentiment and its importance is provided. The second section discusses the predictive power of consumer sentiment with respect to business performance, economic recession, and coincident index. Lastly, some previous findings with respect to the causality of the relationship are discussed.

2.1 Consumer sentiment

Consumer sentiment has an important role in the economic pricing market. Consumers form their perceptions based on available information, but they also behave according to their attitudes and experiences (Marcato & Nanda, 2016). Long term economic expectations are crucial in forming spending decisions and consumer confidence has been recognized as a key factor in shaping the direction in which the economy is heading (Curtin, 2011; Curtin et al., 2019).

A measure that is widely used to track consumer sentiment in the United States is the Michigan Consumer Sentiment Index. The fact that the index started as a project on an annual basis and is now regarded one of the leading indicators of consumer sentiment in the United States which is moreover published every month, marks its value (Cussen, 2019). The index is based on surveys conducted among a random sample of households or consumers (Curtin et al., 2019). Respondents are asked five questions, such as “Would you say that you are better off or worse off financially than you were a year ago?” and “Do you think that during the next 12 months, the country as a whole will have good times financially or bad times?”. To calculate the final index, a relative score for the five questions is computed (Curtin et al., 2019). The questions and formula are listed in Appendix A.

2.1.1 Importance of consumer sentiment

In order to understand how consumer sentiment became this important, the relation between economic events and consumer sentiment must be considered.

Several studies found consumer sentiment to be correlated with economic events. Richard Curtin (2003) for example found that sentiment can forecast changes in the unemployment rate. His results indicate that unemployment expectations of consumers are significantly correlated with future changes in unemployment. The expectations were even more closely correlated to future changes rather than past developments in the rate of unemployment. This indicates that unemployment expectations contain predictive information that is not contained in past trends nor captured by changes in other economic variables. The same conclusions can be found in the research of Thomas (1999). The sentiment of consumers was found to be very important in forecasting inflation. Thomas found that consumer expectations of the year-ahead inflation rate were strongly correlated with the actual inflation rate, and surprisingly also outperformed the forecasts of professional forecasters. The Survey Research Center of Michigan also investigated how accurately consumers anticipate or gauge future economic conditions. They found that consumers anticipate changes in the unemployment rate and the interest rate several months in advance of the actual change. Furthermore, changes in consumer price expectations preceded changes in the actual price index and consumers' assessments of the developments in the national economy show a close correspondence with the actual development (Curtin et al., 2019).

Curtin provided a more practical insight into the effect of sentiment on the economy at the Economic Outlook Conference of November 2013. Blanchard (1993) already stated that economic growth in the year preceding and the year following recession is anaemic too, without having an obvious cause. And indeed, Curtin (2013) also argued that after the Great Recession the economic growth was far from what it should have been. The sentiment index showed that the expectations and attitudes of consumers continued to decline. Three changes in this sentiment could be addressed, namely reduced income expectations, lower work motivation, and a loss of confidence in the economic policies of the government. These changed expectations had a great effect on the course of the economy. Government policies focused on repairing the economy by convincing consumers to spend more and take on more debt, whereas consumers tried to cut it back. As a result, consumer demand became inadequate while it accounts for two-thirds of the economy. Curtin (2013) remarked that consumer spending will continue to be the driving force behind the growth of the economy. Without taking sentiment into account, aimed growth of the economy will never be reached.

2.2 The predictive power of consumer sentiment

The fact that consumer sentiment is correlated with various economic conditions is thus acknowledged by several researchers. Consumer sentiment is moreover found to have predictive power with respect to the overall state of the economy.

2.2.1 Consumer sentiment and business performance

The economic optimism and confidence of households and consumers influence the course of the overall economy. Economic optimism makes individuals more willing to buy and to make debt commitments, whereas economic pessimism leads to a desire to cut expenditures and start saving (Curtin et al., 2019). A consumer suffering financial distress would decrease his demand and limit his purchases because he would prefer holding on to his liquid assets (Mishkin et al., 1978). Business performance is for a large part dependent upon the expenditures of consumers. In the literature there has been a prolonged interest in the power of consumer sentiment in predicting business performance fluctuations and consumption growth (Lahiri & Zhao, 2016). Measures of consumer sentiment are found to be statistically significant in relation to personal consumption expenditures in numerous studies (Juster & Wachtel, 1974; Bram & Ludvigson, 1998; Lahiri & Zhao, 2016). Past literature explained this finding mostly by income growth expectations (Lahiri & Zhao, 2016). Sentiment of consumers captures expectations of income growth and that explains why higher confidence levels lead to higher future consumption (Ludvigson, 2004). However, while it is known that states differ in fiscal and economic environment (Fiscal 50, 2019), the predictive power of consumer sentiment for business performance was never assessed at the state level. Furthermore, previous research never questioned whether this predictive power is stronger in times of economic recession. A business cycle turning point, which is reflected by economic recession, is associated with economic contraction and thus with a decline in personal income, retail sales and industrial production (Megna & Xu, 2003; The Balance, 2019). Because economic recession also affects business performance it is possible that the power of consumer sentiment in predicting business performance is stronger during economic recession.

2.2.2 Consumer sentiment and economic recession

Several studies found that consumer sentiment also has a dominant role in the prediction of the state of the economy, sometimes referred to as the “animal spirits” hypothesis. These animal spirits, e.g. sudden realizations of past overborrowing that lead to increasing prudence and panic, cause an impulse response towards consumer expenditures (Blanchard, 1993). Shifts in sentiment, which can also be positive shocks to expectations about future output or growth, can be self-fulfilling in a way that these expectations drive economic activity (Benhabib & Spiegel, 2018). It is thus believed that consumer sentiment can drive economic activity as fluctuations in sentiment cause shifts in economic cycles. Blanchard (1993) proposed in this regard that the cause of the 1990-1991 recession was a long-lasting negative consumption shock in combination with a shift in pessimism, accounting for the causal effect on the overall aggregate demand. More recent research also finds that recession is driven by extrinsic demand shocks. Angeletos and La’O (2013) show that economic outcomes co-move in response to these extrinsic shocks, which they call sentiment. Again, the predictive power of consumer sentiment for economic recession has never been regarded at the state level and neither was it assessed whether this could be stronger in recessionary periods.

2.2.3 Consumer sentiment and state clusters

Owyang et al. (2004) found that business cycles in the United States are often characterized as a sequence of distinct expansion and recession phases. Within these phases, states differ a lot in the levels of growth they experience. These differences are related to industry mix and education and age composition. The characterization of these business cycles is obtained by using state-level coincident indexes. These coincident indexes combine four indicators that summarize economic conditions at the state level, which are the unemployment rate, wage and salary disbursements, payroll employment, and the average hours worked in manufacturing (FRED, 2019). The indexes have value for the identification of business cycles, provide an indication of state GDP, and act as a signal for recession.

In their research, Owyang et al. (2004) thus find that state-level expansions and recessions differ greatly. When clustering states based on the moments at which they suffered from recession, four clusters of states can be identified, namely financial states, oil states, manufacturing states, and mixed economy states. People are said to be sensitive for day to day, personal economic experiences, which most likely will not be equal over the different state clusters as for the variety in recession timing (Linden, 1982; Blood & Philips, 1995). Because states differ in their industrial composition and in their business cycles in terms of coincident index, it is possible that the predictive power of consumer sentiment for both business performance and coincident index is higher in one state cluster than another. While it is known states experience very different economic phases, the effect of consumer sentiment in different state clusters was never assessed.

2.3 The predictive power of economic conditions

While many results prove the dominant role of consumer sentiment in predicting the economy, there is not yet a definite conclusion with regard to this relationship. In his research, Throop (1992) finds that changes in consumer sentiment are normally caused by purely economic factors and that sentiment is just a reflection of economic adversity or prosperity, reinforcing business cycles rather than initiating them. Barnes & Olivei (2017) also state that the role of sentiment in consumption is small and that the independent information from sentiment is limited when controlling for economic fundamentals. In order to draw the right conclusions, a change in the causal direction must be examined too.

This thesis addresses the beforementioned shortcomings of previous research by not only considering the power of consumer sentiment in predicting the state of the economy, but to examine this effect at the state level and, moreover, during economic recession. Furthermore, a change in the causal direction between consumer sentiment and indicators of economic conditions is examined to provide a definite conclusion about the direction of the relationship.

3. Experimental setup

This section describes the dataset and the experimental procedure used to address the research questions. The first subsection provides a description of the dataset. Subsection 3.2 describes what pre-processing is done. Subsection 3.3 presents an overview of the final dataset after pre-processing the data. Lastly, subsection 3.4 describes the experimental procedure. The software and packages used are listed in Appendix M.

3.1 Description of the raw dataset

In order to address how accurately consumer sentiment can predict various indicators of economic conditions, business performance, economic recession, and coincident index within a state will be analysed together with the Michigan Consumer Sentiment Index. The dataset that will be used in the analysis is retrieved from Kaggle (Kirsch, 2019).

This dataset was built using data collected from the Federal Reserve Bank of Philadelphia and St. Louis and the U.S. Bureau of Economic Analysis. The dataset includes monthly data for all 50 states of the United States of America from January 2005 to April 2018 and contains information about whether or not the state economy is in a recession at a specific month, the rate of the state coincident index, the personal consumption expenditures in different industries, and the monthly consumer sentiment index of the country as a whole. States are furthermore clustered based on the dates they were in recession. This clustering forms clusters of states that move together in terms of recession. The four state clusters are the financial cluster, oil cluster, manufacturing cluster, and mixed economy cluster.

The consumer sentiment index is based on surveys conducted by the Survey Research Center of the University of Michigan among random samples of households. The index summarizes attitudes and expectations of consumers towards personal finances, market conditions or prices, and general business conditions (Curtin et al., 2019). The index fluctuates between 55 and 101 between 2005 and 2018. Coincident index reflects four state-level indicators to summarize economic conditions in a state. These indicators are payroll employment, average hours worked in manufacturing by production workers, wage and salary disbursements deflated by the consumer price index, and the unemployment rate. The personal consumption expenditures per industry are used to provide an indication of the business performance in each state. The expenditures are the amount of goods and services in dollars purchased by U.S. residents (BEA, 2019). Lastly, economic recession is a binary variable where a “zero” indicates that the state was not in recession, while a “one” indicates that a state was in recession.

3.2 Data pre-processing

This section describes the pre-processing steps needed in order to prepare the data for analyses. The subsections describe which methods are used in order to handle the missing values, elaborate on the modification of the data, present an overview of the pre-processed data, and discuss the sliding window method used to create the final dataset.

3.2.1 Missing values

There are 746 observations that contain missing values for in total 8 of the 48 variables. Figure 1 in Appendix B visualizes the number of missing values per variable that is used in the analysis. Because deleting the observations with missing values is very wasteful and would lead to a loss of statistical power, multiple imputation is used to impute the missing values. This method produces unbiased parameters and correct standard deviations (Lang, 2018).

Different methods are used for the multiple imputation depending upon the distribution of the variable that contains missing values. These are listed in Appendix D. Bayesian linear regression is used for incomplete variables that have a normal distribution. Predictive mean matching (PMM) is used for the incomplete variables that are skewed. PMM maintains the observed support of an incomplete variable's distribution, it for example only imputes values that have actually been observed. For binary variables, logistic regression is used (Little et al., 2013).

The variables "date" and "state" are excluded from the multiple imputation analysis because they cannot be used as meaningful predictors in the imputation model. Furthermore, all lagged variables are excluded. The incorporation of the lagged values of state recession and coincident index causes autocorrelation in the multiple imputation model: the variables are correlated with previous copies of themselves (Dancho, 2017). The performance of the model decreases as these redundant columns and high correlations, called multicollinearity, lead to unreliable estimates (Lang, 2018). The variable "Consumer Sentiment Index" is included in the analysis because it is thought to be a key covariate. The quikpred algorithm used in the imputation process attempts to find good key covariates automatically, but this consumer sentiment index must be included in the imputation model regardless of what the quickpred algorithm suggests. This variable is believed to improve the imputation because this research expects consumer sentiment to have great predictive power for the (incomplete) variables reflecting economic conditions.

Since the actual values of the missing data are most likely dependent upon the observed data, it is most insightful to incorporate all data that may provide information about the missing values in the multiple imputation (Lang, 2018). Therefore, incomplete variables not used in the experimental analyses are also imputed. The data is only imputed once because of the sliding window method that is used in a later stage, described in section 3.2.4. As this approach generates a new dataset for each of the 50 states and furthermore incorporates many features, it is inconvenient to also have multiple imputed datasets.

3.2.2 Modification of the data

Additional variables. The "current" value of coincident index for each state in a specific month is not available in the Kaggle dataset. Each observation only contains lagged values of coincident index. A lag is a shift of a time series and looks back in time (Dancho, 2017). Because the research questions require the actual value of coincident index, this thesis retrieved this value from the lagged values for every observation.

Next to that, the Kaggle dataset contains personal consumption expenditures for 23 industries. Because this research is interested in predicting and assessing business performance as a single statistic, these personal consumption expenditures were summed into one variable for this thesis. This summation was done after multiple imputation as some industries contained missing values for the personal expenditure values. The Kaggle dataset contains a variable "all industries total" but those values do not correspond to the actual value of the total expenditures.

Scaling. Because the variable business performance is measured in dollars and reaches very large numbers the values cannot easily be compared to the other variables in the dataset. Moreover, the values cannot be meaningfully compared between states. To aid comparison the values for business performance are scaled and centered for this research (Diez et al., 2012). Centering is done by subtracting the column mean from every value. After that, scaling is performed by dividing the centered columns by their standard deviations (R Core Team, 2018). The values for consumer sentiment, coincident index, and economic recession are non-absolute since their values are either measured on an index or binary. Therefore, these variables can be meaningfully compared and were not scaled for analysis.

3.2.3 Pre-processed data

The remaining data after pre-processing consists of 8,000 observations and 48 variables. Table 1 below provides a short presentation of the data that is used in the analysis. Figures 2 to 5 in Appendix B provide a visualization of the data. The further organization of the dataset is presented in the next subsections.

Table 1: Description of the variables in the pre-processed dataset

Variable	Description
Date *d*	Each value for date *d* consists of the month and year of the observation
State *s*	State *s* is the name of the state an observation belongs to
Cluster *c*	The cluster *c* is the state cluster the state *s* belongs to in terms of the timing of economic recession *rs*
Consumer sentiment $cs_{s,d}$	Consumer sentiment *cs* is the value of the attitudes and expectations of consumers at each date *d* for the whole country, present for every state *s*
Coincident index $ci_{s,d}$	Coincident index *ci* reflects the value of coincident index at each date *d* for every state *s*
Business performance $bp_{s,d}$	Business performance *bp* is the scaled, summed value of all personal consumption expenditures in the industries of state *s*
Economic recession $rs_{s,d}$	Economic recession *rs* is a binary variable that indicates whether a state *s* was in recession at date *d* (1) or not (0)

3.2.4 Sliding window method

The data presents itself as a time series since the variables consumer sentiment, coincident index, business performance, and economic recession all involve sequences of observations at monthly intervals between 2005 and 2018. A time series is a sequence of events that occur during a certain period of time. Every event that occurs at a particular point within this time period has a value that is observed. The collection of all these values represents a time series (BenYahmed et al., 2015). A time series can typically be represented by T = (x_t, x_{t+1} …, x_{t+n}), where T is the time series and x_t is the observed value of variable x at time *t*. An important characteristic of such time series is the dependence of future values on current and past values. Modelling future values as a parametric function of current and past values enables one to use the results as a forecasting tool (Shumway & Stoffer, 2010).

To be able to use supervised learning algorithms whilst still considering the temporal patterns in the data, the data is reorganized in a new dataset using the sliding window method. This method maps the data into overlapping time windows. Given the time series T of length *n* and a subsequence length of *w*, all possible subsequences across T can be extracted by a sliding window of length *w* (Yu et al., 2014). The sliding window method implies that at each set point, say x_t to x_{t+n}, the value for variable $x = x_{t+1}, x_{t+2}, \ldots, x_{t+n}$ can be modelled using the history of variable x over a previous time sequence, i.e. $x = x_{t-1}, x_{t-2}, \ldots, x_{t-n}$ (Mozaffari et al., 2015).

For this research a dataset is created for each state, which results in 50 datasets. The sliding window method is applied to each of these datasets. Each sliding window covers three subsequent months starting at the value for variable x at the first month in the sliding window x_t, followed by x_{t+1} and x_{t+2}. With a slide of the sliding window the time sequence changes and covers the last two months of the previous window plus one month later. The months in each time window are presented in Appendix L. This ordering of the data enables using time sequences x_t, x_{t+1}, and x_{t+2}, as predictors. To illustrate this method, an example is given in figure 1. This figure is based on the data for one randomly picked state for the year 2009. The values of the consumer sentiment index are scaled for this visualization in order to be able to compare them with the scaled values of business performance, small values of coincident index, and the binary values of economic recession.

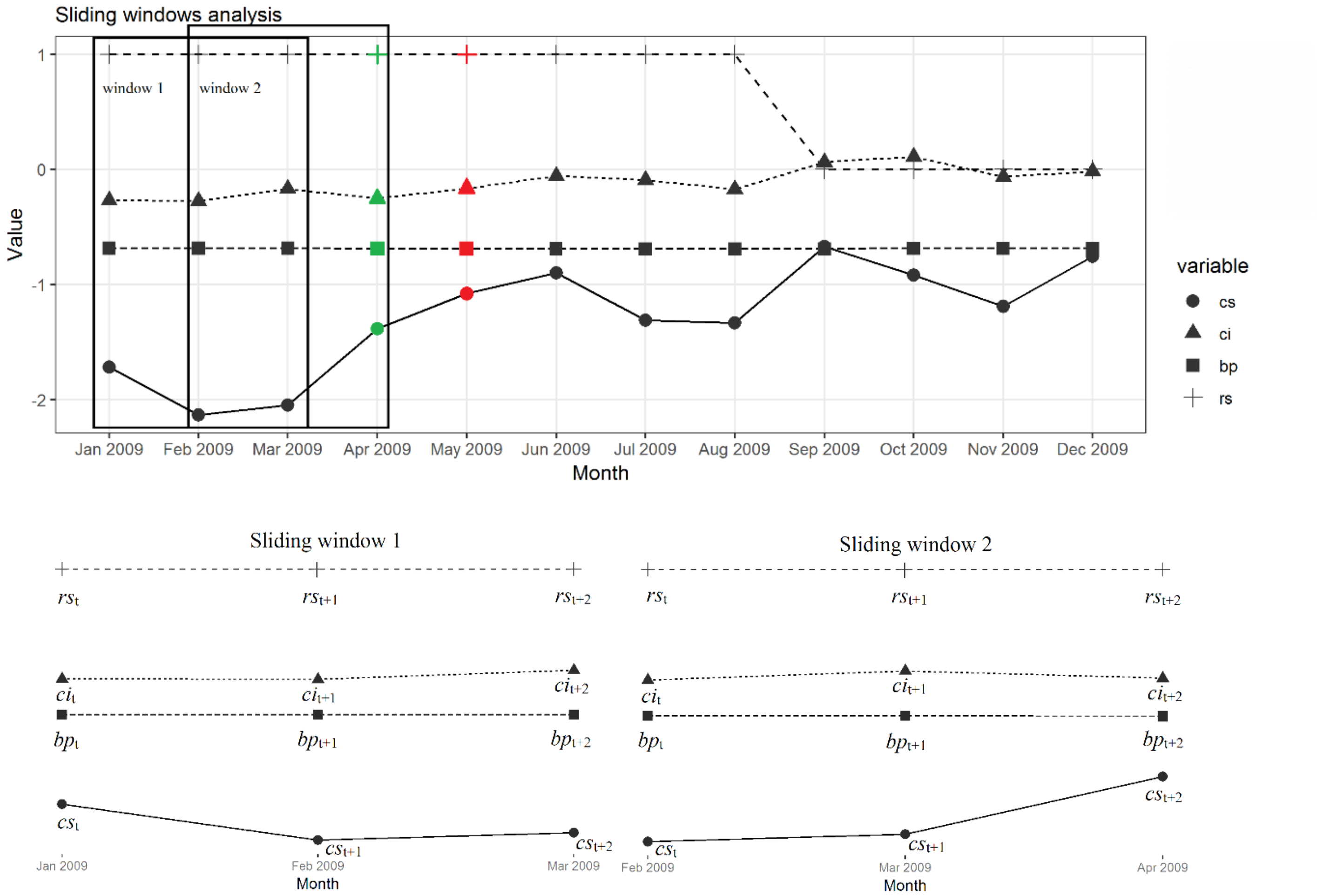

Figure 1: Sliding window method

3.2.5 Feature construction

The sliding window method creates windows that contain three data points for each variable. The method allows to code temporal relationships in the data by constructing additional features from these three values. This feature construction is performed because it discovers unknown information about the relationships between the three data points. Besides, the data is transformed and simplified as to make data mining techniques easier (Liu & Motoda, 1998). New representations from the original data are built to account for the temporal patterns in the data, improving the performance of various algorithms (Piramuthu et al., 1998).

For each sliding window of consumer sentiment, coincident index, business performance, and economic recession the mean, median, maximum, minimum, variance, the coefficient of range, and the coefficient of quartile deviation are constructed from the data points x_t, x_{t+1} and x_{t+2}. The two coefficients are relative measures of dispersion, which are measures of variance regardless of the unit of measure of the range of values (Master of Project Academy, 2017). This eliminates any differences that may still exist in the units of measure.

Moreover, the value of the change between the adjacent point outside each sliding window (x_{t+3}) relative to x_{t+2} is constructed for each sliding window. An example is given in figure 1. For window 1 x_{t+3} is marked green and for window 2 it is marked red. The change between x_{t+3} and x_{t+2} is the value that should be predicted for the regression problems. The value that should be predicted for the classification problem is the binary value of economic recession at x_{t+3} outside every sliding window. Therefore, this feature is included too.

Table 2: Formulas for feature construction

Feature	Formula
Mean	$\bar{x} = \frac{\sum x_t, x_{t+1,} x_{t+2}}{n}$
Median	$median = \left(\frac{n+1}{n}\right)^{th} term$
Maximum	$maximum = \max(x_t, x_{t+1,} x_{t+2})$
Minimum	$minimum = \min(x_t, x_{t+1,} x_{t+2})$
Variance	$variance = \sqrt{\frac{\sum x - \bar{x}}{n}}$
Coefficient of range	$CR = \frac{\max(x_t, x_{t+1,} x_{t+2}) - \min(x_t, x_{t+1,} x_{t+2})}{\max(x_t, x_{t+1,} x_{t+2}) + \min(x_t, x_{t+1,} x_{t+2})}$
Coefficient of quartile deviation	$CQD = \frac{Q3-Q1}{Q3+Q1}$
Change between x_{t+3} and x_{t+2}	$change\ x_{t+3;}\ x_{t+2} = x_{t+3} - x_{t+2}$
Economic recession at x_{t+3}	$binary\ value\ of\ rs_{t+3}\ (0;1)$

3.3 Final dataset

The final dataset after merging the datasets per state resulting from the implementation of the sliding window method contains 7,850 observations and 45 features. This reduction of 150 observations is the total of the reduction of 3 observations for every state. This reduction is first of all caused because of the sliding window width of three data points. The last two sliding windows would contain two and one data point respectively and are therefore disregarded. It is furthermore caused because the change between x_{t+3} and x_{t+2} is calculated through a lag, while the first sliding window does not have a previous value. Table 3 below provides a summary of the final dataset and the features it contains.

Table 3: Summary of the final dataset

Feature		Description
Consumer sentiment	cs_t, cs_{t+1}, cs_{t+2}	The values of the data points in the first, second and third sliding window (x_t, x_{t+1} and x_{t+2})
Coincident index	ci_t, ci_{t+1}, ci_{t+2}	
Business performance	bp_t, bp_{t+1}, bp_{t+2}	
Economic recession	rs_t, rs_{t+1}, rs_{t+2}	
Consumer sentiment	$cs_{\bar{x}}$, cs_{median}, cs_{max}, cs_{min}, cs_{σ^2}, cs_{CR}, cs_{CQD}, $change_{cst+3;\, cst+2}$	The values of the features derived from the data points x_t, x_{t+1} and x_{t+2}
Coincident index	$ci_{\bar{x}}$, ci_{median}, ci_{max}, ci_{min}, ci_{σ^2}, ci_{CR}, ci_{CQD}, $change_{cit+3;\, cit+2}$	
Business performance	$bp_{\bar{x}}$, bp_{median}, bp_{max}, bp_{min}, bp_{σ^2}, bp_{CR}, bp_{CQD}, $change_{bpt+3;\, bpt+2}$	
Economic recession	$rs_{\bar{x}}$, rs_{median}, rs_{max}, rs_{min}, rs_{σ^2}, rs_{CR}, rs_{CQD}, $change_{rst+3;\, rst+2}$, rs_{t+3}	

3.4 Experimental procedure

This section describes the experimental procedure conducted to address the research questions. First, the partitioning of the data into a train and test set is discussed. Subsection 3.4.2 describes in which way features were selected for each experiment. Subsection 3.4.3 provides a description of the experiments.

3.4.1 Data partitioning

The sliding window dataset is partitioned into a train and test set. To ensure that the models generalize to unseen data, the models are trained on data from states in the North East, Midwest, and South of the United States. To test the models, the states in the Western part of the United States are used. By partitioning the data in this way this thesis ensures that the results generalize to states in different geographical areas. This approach allows drawing conclusions at the state level, which is the aim of the research. Table 4 below presents the states within each set.

Table 4: Partitioning of the data

Dataset		States	
Train set	North East	Pennsylvania, New York, Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New Jersey	
	Midwest	North Dakota, Minnesota, Wisconsin, South Dakota, Nebraska, Iowa, Kansas, Missouri, Illinois, Indiana, Ohio, Michigan	38 (76%)
	South	Oklahoma, Texas, Arkansas, Louisiana, Kentucky, Tennessee, West Virginia, Virginia, Maryland, Delaware, North Carolina, South Carolina, Florida, Mississippi, Alabama, Georgia, Hawaii	
Test set	West	Washington, Oregon, California, Montana, Idaho, Wyoming, Colorado, New Mexico, Nevada, Arizona, Utah, Alaska	12 (24%)

3.4.2 Feature selection

The goal in the regression experiments is to predict the value of the change of target variable Y between point t+3 and t+2, thus change$_{Y_{t+3};Y_{t+2}}$. The goal when predicting whether the economy is in recession or not is to predict the binary value of economic recession at t+3. As this research expects consumer sentiment to have predictive power for target value Y, consumer sentiment at t, t+1, and t+2 and additional features of consumer sentiment constructed from the sliding window method are the features that can be used in the prediction task. Furthermore, as the sliding window method assumes dependence of future values on current and past values, the value of Y_t is considered a predictive feature too.

However, it is possible that predictions can be done more accurately using a subset of these features. Reducing the number of features by omitting irrelevant ones can lead to a reduced running time of algorithms and can avoid overfitting (Dash & Liu, 1997). Feature selection can be used to pick only the features relevant in predicting the target variable. Therefore, every experiment is conducted two times. Once using all features considered important for the research question it addresses and once using the features deemed to be important according to feature selection.

In order to select the most relevant features for each research question Monte Carlo simulation is performed. Appendix C provides a more detailed description of this simulation method. Some features appear more important than others during this process, but using all features results in more accurate predictions for each experiment. This indicates that all features contribute to the predictions. This is also true for the fourth experiment in which the causal direction between consumer sentiment and the economic conditions is changed. Therefore, no features are omitted for the analyses. The predictions using a subset of the features are also provided in Appendix C.

3.4.3 Description of the experiments

This thesis investigates whether consumer sentiment can predict various indicators of economic conditions within the United States. The experiments will therefore investigate whether the sentiment within a state can predict future business performance, coincident index, and economic recession, as well as what predictive power is concerned with changing the causal direction by looking at the degree to which these indicators can predict consumer sentiment. Different models are used to estimate the relationship between the sentiment and indicators of economic conditions. A seed of 1 is set to ensure reproducibility. The experiments are conducted using the dataset that was generated from the sliding window method. Each experiment investigates how accurately the target value can be predicted with the use of features constructed from this method.

3.4.3.1 Experiment 1

RQ1: To what degree can consumer sentiment predict business performance, and can this prediction be done more accurately during economic recession?

The first experiment addresses research question 1 by investigating whether it is possible to predict business performance within a state based on consumer sentiment. The goal is to predict the change of business performance for t+3, i.e. predict $change_{bpt+3;bpt+2.}$ To test whether the prediction can be done more accurately during recession, the models are applied to a subset of the data with a score of "1" on the variable economic recession. Table 5 below presents an overview of the features used to address research question 1.

Table 5: Features included in experiment 1

Features	Target variable
cs_t, cs_{t+1}, cs_{t+2}, $cs_{\bar{x}}$, cs_{median}, cs_{max}, cs_{min}, cs_{σ^2}, cs_{CR}, cs_{CQD}, bp_t	$change_{bpt+3;bpt+2}$

3.4.3.2 Experiment 2

RQ2: To what degree can consumer sentiment predict state recession, and can this prediction be done more accurately during economic recession?

The second experiment addresses research question 2. The goal is to predict economic recession three months ahead, i.e. rs_{t+3}. The features are listed in table 6. The second part of the research question examines whether the prediction can be done more accurately during recession. For this part of the research question the models are again applied to the subset of the data with a score of "1" on the variable economic recession. The models predict whether economic recession at $t+3$ is present (1) or absent (0).

Table 6: Features included in experiment 2

Features	Target variable
cs_t, cs_{t+1}, cs_{t+2}, $cs_{\bar{x}}$, cs_{median}, cs_{max}, cs_{min}, cs_{σ^2}, cs_{CR}, cs_{CQD}, rs_t	rs_{t+3}

3.4.3.3 Experiment 3

RQ3: To what degree does the predictive power of consumer sentiment differ over state clusters with regard to business performance and coincident index?

To address the third research question experiment 3 incorporates the state cluster for each state. Subsets are created for each of the four clusters so that every cluster is assigned the sliding window data for the states that belong to it. The state clusters in the train and test sets are listed in Appendix D. In order to address the question to what degree the predictive power of sentiment differs with respect to business performance the predictive performance between state clusters is assessed. The analysis is performed with the features listed in the top part of table 7. The goal is to predict the change of business performance for t+3, i.e. predict $\text{change}_{bpt+3;bpt+2.}$ In order to address the question to what degree this predictive power differs with respect to coincident index the predictive power of consumer sentiment regarding the change of ci_{t+3} relative to ci_{t+2} is assessed. This analysis is performed with the features listed in the bottom part of table 7.

Table 7: Features included in experiment 3

	Features	Target variable
Part 1	cs_t, cs_{t+1}, cs_{t+2}, $cs_{\bar{x}}$, cs_{median}, cs_{max}, cs_{min}, cs_{σ^2}, cs_{CR}, cs_{CQD}, bp_t	$\text{change}_{bpt+3;bpt+2}$
Part 2	cs_t, cs_{t+1}, cs_{t+2}, $cs_{\bar{x}}$, cs_{median}, cs_{max}, cs_{min}, cs_{σ^2}, cs_{CR}, cs_{CQD}, ci_t	$\text{change}_{cit+3;cit+2}$

3.4.3.4 Experiment 4

RQ4: To what degree would it be more insightful to change the causal direction between the indicators of economic conditions and consumer sentiment?

In order to address the last research question this experiment seeks to find the predictive power of the economic conditions regarding the change of consumer sentiment between t+3 and t+2, i.e. $\text{change}_{cst+3;cst+2}$. To assess whether the change in the causal direction is more insightful the performance of each of the models against the baseline is compared to the performance of the models in experiment 1, experiment 2, and experiment 3 respectively. The features are listed in table 8.

Table 8: Features included in experiment 4

	Features	Target variable
Part 1	bp_t, bp_{t+1}, bp_{t+2}, $bp_{\bar{x}}$, bp_{median}, bp_{max}, bp_{min}, bp_{σ^2}, bp_{CR}, bp_{CQD}, cs_t	$change_{cst+3;cst+2}$
Part 2	rs_t, rs_{t+1}, rs_{t+2}, $rs_{\bar{x}}$, rs_{median}, rs_{max}, rs_{min}, rs_{σ^2}, rs_{CR}, rs_{CQD}, cs_t	$change_{cst+3;cst+2}$
Part 3	ci_t, ci_{t+1}, ci_{t+2}, $ci_{\bar{x}}$, ci_{median}, ci_{max}, ci_{min}, ci_{σ^2}, ci_{CR}, ci_{CQD}, cs_t	$change_{cst+3;cst+2}$

3.4.4 Evaluation scheme

This section presents which evaluation criteria were used in order to assess the performance of the regression and classification models.

MSE/Accuracy. The performance of the regression models is derived by constructing Mean Squared Errors (MSE) that measure the average squared difference between the predicted and actual values in the data. A lower score indicates higher performance. For classification models, a confusion matrix is constructed that lists the number of instances that are correctly and incorrectly classified relative to the actual instances in the dataset. The accuracy is the amount of correctly classified records compared to the total amount of records. A higher score indicates higher performance.

Baseline. In order to evaluate the MSE and accuracy, a baseline model is constructed to interpret the model results. This baseline model assumes for the regression problems that the change between t+3 and t+2 for e.g. sliding window 2 is the same as it was for sliding window 1. For the classification problem the baseline assumes that the binary value of economic recession at t+3 for e.g. sliding window 2 is the same as it was for sliding window 1. It does not take any other attributes into consideration. This baseline is also called "naïve forecasting" as it takes the last sliding window's actual value as the forecast for the current sliding window without adjustment or establishing interactions or correlations (Hyndman, 2018).

Test set. In order to evaluate the generalization of the model to unseen data the MSE and classification accuracy on the test set are assessed. Because the models are trained on the training data to learn about the relationship between the features and the target value Y, it is possible that a model is fitted so closely to the training data that it fits noise instead of signal (Shmueli et al., 2017). Assessing the performance on the test set prevents this so-called overfitting.

4. Methods

This section describes the models used in this thesis. First, the models Elastic Net, Support-Vector Machine and Random Forest are discussed. After that, the classification models PART, Bagging, and k-Nearest Neighbour are presented.

This thesis investigates whether consumer sentiment can predict business performance and coincident index, and whether a change in the causal direction between consumer sentiment and economic conditions would be more insightful. The aim is to predict a continuous outcome; hence regression is used for these analyses. The aim when investigating whether consumer sentiment can predict economic recession, however, is to predict a binary outcome and therefore, classification is performed. Support-Vector Machine and Random Forest can handle regression and classification problems and are therefore suitable for both aims of this thesis.

Elastic Net is a regularization technique that combines two linear regression techniques Ridge and Lasso. Ridge is a shrinkage method that reduces the residual sum of squares of the regression coefficients using an L_2-norm penalty ($\alpha = 0$). This penalty biases the estimates but reduces the variance and therefore leads to better estimations than non-regularized regression models (Tutz & Binder, 2007). Lasso is a penalizing least squared technique that imposes an L_1-penalty on the coefficients ($\alpha = 1$), minimizing the sum of the regression coefficients (Tibshirani, 1997; 2011). Some of the coefficients are set to exactly zero, whereas others that are found important are used to train the model, effectively choosing a simpler model. The limitations of Lasso – it selects at most *n* variables and tends to pick only one variable from a group of highly correlated ones – are solved by Elastic Net (Zou & Hastie, 2005). The parameters α and λ are tuned for the best trade-off between Ridge and Lasso and the most effective penalty strength to achieve the best predictions (Zou & Hastie, 2005).

Support-Vector Machine (SVM) is a supervised learning algorithm that can be used for both regression and classification of data. The aim is to non-linearly map input vectors to a high-dimensional feature space in which a linear decision surface is constructed (Cortes & Vapnik, 1995). This kernel trick makes that the model can observe very complex relationships without having to transform the variables (Rüping, 2001). The algorithm builds a model that classifies or regresses the instances by finding an optimal hyperplane. This is the linear decision function with a maximal margin between the vectors of the different classes (Cortes & Vapnik, 1995). To determine this margin only a small number of the training instances has to be considered, the support vectors. The kernel-based parameters cost and gamma must be optimized in order to find the best results. The cost parameter is a penalty parameter which reflects the cost of violating constraints and gamma defines the influence of one training observation (Sarkar et al, 2016; R Core Team, 2018).

The Random Forest algorithm is a learning method which handles classification and regression problems too. The model generates many classifiers or regressions and aggregates the results, also called the ensemble method (Liaw & Wiener, 2002). The algorithm draws a user-defined number of bootstrap samples from the training data. For each sample it grows an unpruned decision tree by randomly sampling the predictors and choosing the best split among these variables to derive the values. It then predicts new data by aggregating its predictions over the trees (Liaw & Wiener, 2002). By doing so the algorithm compensates for the bias that is caused by the randomness and therefore increases generalizability and reduces variance.

PART is a rule-learning system that combines two rule learning paradigms, C4.5 and RIPPER. Both approaches iteratively perform global optimization on a set of rules that is initially induced. The PART algorithm combines these learning paradigms but adds simplicity by avoiding global optimization (Frank and Witten, 1998). The algorithm builds a rule, removes the instances that are covered by the rule, and continues creating rules for the remaining instances until none are left. These rules can then be used for classification. In order to make a rule, a pruned partial decision tree is built for the instances and the leaf with the highest coverage is transformed into a rule. The tree is then discarded (Frank and Witten, 1998).

Bootstrapped Aggregation (Bagging) is an ensemble learning method that creates multiple decision trees and aggregates the results to retrieve a predictor (Liaw & Wiener, 2002). It differs from Random Forest because it does not add randomness. A decision tree represents data as a collection of binary vectors in order to infer rules. It tries to maximize the information gain when splitting on attributes to derive the class of each instance (Emmery, 2018). The bagging algorithm constructs large classification decision trees which consist of multiple subtrees. The original train data is used to select the best subtrees by finding a sequence of the simplest trees with minimum classification error.

The k-Nearest Neighbour classifier finds a model that can predict the outcome of the class given the values of the attributes (Daniels, 2019). The algorithm computes the distance of a record to the other records according to some distance metric. It then identifies the user-defined *k* nearest neighbours and uses their class labels to determine the class of the unknown record with a majority vote. Small distances between the records imply that the discriminating attributes are equal, and the record is classified into the same class. Larger distances imply difference of the discriminating attributes and the record is not assigned to the majority class of those records (Daniels, 2019).

5. Results

This section presents the results of the experiments that were conducted. The first subsection presents the results of experiment 1. Subsection 5.2 provides results for experiment 2 and subsection 5.3 for experiment 3. The last subsection discusses the results of the change of the causal directions performed in experiment 4.

5.1 Experiment 1

The goal of this experiment was to address the first research question to what degree consumer sentiment could predict business performance and whether this prediction could be done more accurately during recession. The regression models were tuned and evaluated on the train data. For all regression problems in this thesis the tuned SVM makes use of the eps-regression method with a radial kernel and a cost of violating constraints of 1. The gamma is set to 0.09090909 which is the influence of one training observation on the model. Lastly, the epsilon is 0.1. For Elastic Net the parameters α and λ are tuned using repeated 5-fold cross-validation with 5 iterations. The optimal parameters are selected by picking the combination with the smallest Root Mean Squared Error. All results were reported in MSE on the train and test set. Tables 9a and 9b presents the results of the first experiment. The coefficients and feature importance of Elastic Net and Random Forest are listed in Appendix F.

5.1.1 Experiment 1.1

For the first part of the experiment the train data consists of 5966 observations and the test data of 1884 observations. All models outperformed the naïve baseline model which predicted that the change between point t+3 and t+2 would be the same as it was for the last sliding window. Random Forest performed best compared to the other models with the smallest MSE scores on both train and test data and a 67.88% improvement with respect to the baseline on the test set. The data is better modelled through this ensemble method than regularized linear regression using Elastic Net, which scores an MSE of 1.162. This model uses an alpha close to 0 and thus performs much like Ridge regression. SVM scores the highest MSE on both train and test data. For experiment 1 SVM makes use of 2523 support vectors, which is approximately 40% of the train data. The large differences between the MSE on the train and test data imply that all models overfit on the training set.

Table 9a: Results experiment 1.1 (10^{-4})

Model	MSE train		MSE test	
Baseline	2.721	-	8.675	-
Elastic Net	1.162	(57.30%)	3.476	(59.93%)
SVM	1.194	(56.12%)	3.726	(57.05%)
Random Forest	0.498	(81.70%)	2.786	(67.88%)

Results of experiment 1 part I. Parameters Elastic Net: alpha = 0.2121425; lambda = 0.001205047 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 2523 - Random Forest: ntree = 5000; importance = TRUE

Figure 2 below illustrates the predictions that the models made on the test set for 10 sliding windows, along with the true values of business performance. It again shows that Random Forest performs best by capturing the patterns in the data. SVM does not predict the true values that well. Elastic Net also predicts very small changes compared to the actual changes. Since the predicted changes are that small they cannot be visualized at the same scale as the actual changes and they appear almost as a straight line. This indicates that most coefficients are shrunk to around zero and are not considered important by the model.

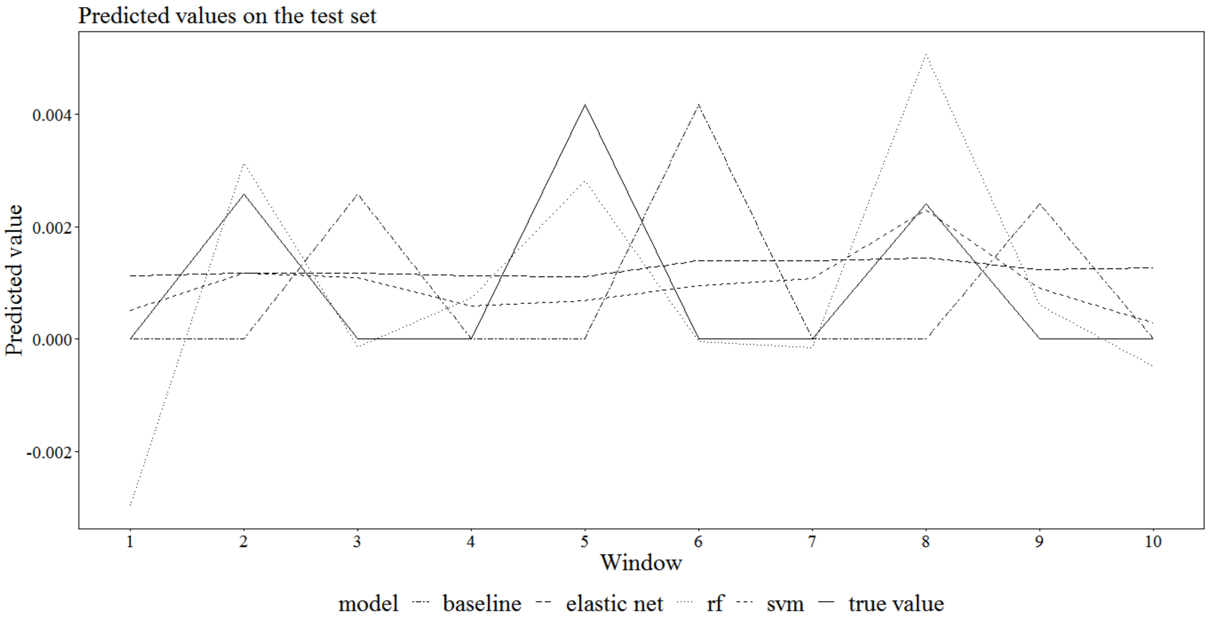

Figure 2: Actual and predicted values on the test set of experiment 1

5.1.2 Experiment 1.2

The second part of this experiment applied the models to a subset of the train and test data given economic recession at time *t*. This resulted in 846 observations for the train data and 284 observations for the test data. Table 9b presents the results of this analysis. The models outperform the baseline and Random Forest is again the best model with the lowest MSE on both train and test data. It outperforms the baseline with 75.50% and 55.29% respectively. Where SVM was the worst performing model in the first part, it is now second best. It uses approximately 37% of the train data in order to derive these results. Linear modelling with Elastic Net achieves the highest MSE. The models in this part of the experiment also overfit the training data since the MSE on the test set are higher than on the train set. The MSE for all models are lower than they were in the first part of this experiment, except for Random Forest on the train set, but the improvement against the baseline is also lower. This indicates that there is not more predictive power associated for consumer sentiment during economic recession.

Table 9b: Results experiment 1.2 (10^{-4})

Model	MSE train		MSE test	
Baseline	2.135	-	5.826	-
Elastic Net	1.025	(51.99%)	2.899	(50.24%)
SVM	0.965	(54.80%)	2.870	(50.74%)
Random Forest	0.523	(75.50%)	2.605	(55.29%)

Results of experiment 1 part II. Parameters Elastic Net: alpha = 0.5589081; lambda = 7.025305 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 316 – Random Forest: ntree = 5000; importance = TRUE

Figure 3 serves as an illustration of the predictions that the models made on the test set for the second part of this experiment. It shows no model predicts the patterns in the test data very well, which is also reflected through the relatively large MSE on the test set. Furthermore, Elastic Net now predicts the same value for every window. This is caused due to the fact that the relatively high lambda and alpha-parameter set all coefficients to exactly zero, meaning that the model could not find important coefficients when predicting business performance during economic recession. It is now an intercept-only model, also presented in Appendix F. This too indicates that consumer sentiment does not have more predictive power for business performance during recession.

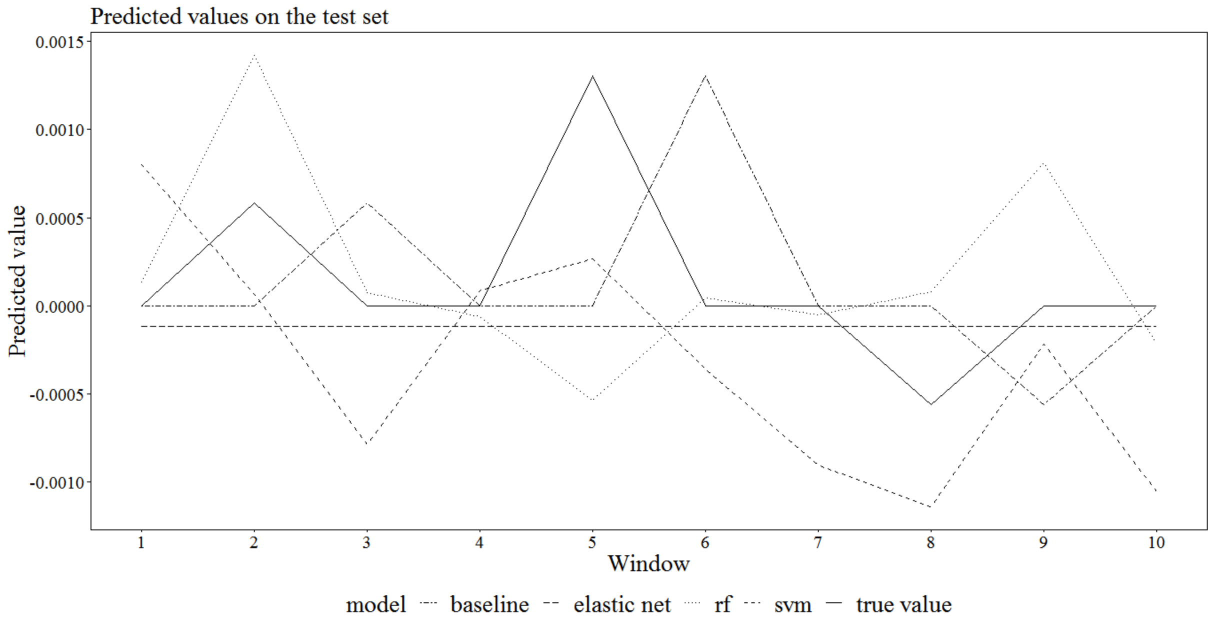

Figure 3: Actual and predicted values on the test set of experiment 1.1 (states in recession)

5.2 Experiment 2

The aim of experiment 2 was to address the question if consumer sentiment could predict state recession and whether the prediction could be done more accurately during economic recession. Classification models were used to derive the results. The models were tuned and evaluated on the train data. The tuned SVM makes use of C-classification with a radial kernel and a cost of violating constraints of 1. The gamma is set to 0.09090909. The k-NN model is fit using 5-fold cross-validation. The results are reported in classification accuracy and F1-scores. Accuracy indicates the amount of correctly classified instances with respect to the total amount of instances. Since accuracy could be very high while not all instances are equally well predicted, the F1-score is included. This is the harmonic mean of precision and recall and also provides information about observations not classified very well (Daniels, 2019). The best value of an F1 score is 1 and its worst is 0. Tables 10a and 10b present the results. The feature importance of PART, Bagging, Random Forest, and k-NN are listed in Appendix G.

5.2.1 Experiment 2.1

The train data for the first part of the experiment consists again of 5966 observations and the test data of 1884 observations. Table 10a shows that the models hardly outperform the baseline model, which has a very high accuracy on both the train and test data. Consumer sentiment has no greater predictive power than this naïve forecast. The results indicate that SVM, PART, Bagging, and Random Forest have approximately the same accuracy and F1-scores. This means there is some general pattern in the data that is recognized by all of these models. k-NN performs worst on both train and test data and achieves the lowest F1-scores. Since the results of the baseline are best, no informative conclusions can be drawn from the models. Adding the predictor features of consumer sentiment does not contribute to the prediction of economic recession.

Table 10a: Results experiment 2.1

Model	MSE train		MSE test		F1 train	F1 test
Baseline	0.985	-	0.985	-	0.991	0.991
SVM	0.985	-	0.985	-	0.991	0.991
PART	0.986	(0.1%)	0.985	-	0.992	0.991
Bagging	0.987	(0.2%)	0.985	(0.1%)	0.992	0.991
Random Forest	0.987	(0.2%)	0.985	-	0.992	0.991
k-NN	0.972	(-1.3%)	0.971	(-1.4%)	0.984	0.983

Results of experiment 2 part I. Parameters SVM: method = C-classification; kernel = radial; C = 1; gamma = 0.09090909; support vectors: 274 - Random Forest: ntree = 5000; importance = TRUE – k-NN: method = knn; trControl = cv (number: 5); k = 5

Figure 4 serves an illustrative purpose for the actual and the predicted classes in the test data. It is clear from this figure that there are no large differences between the models and the baseline.

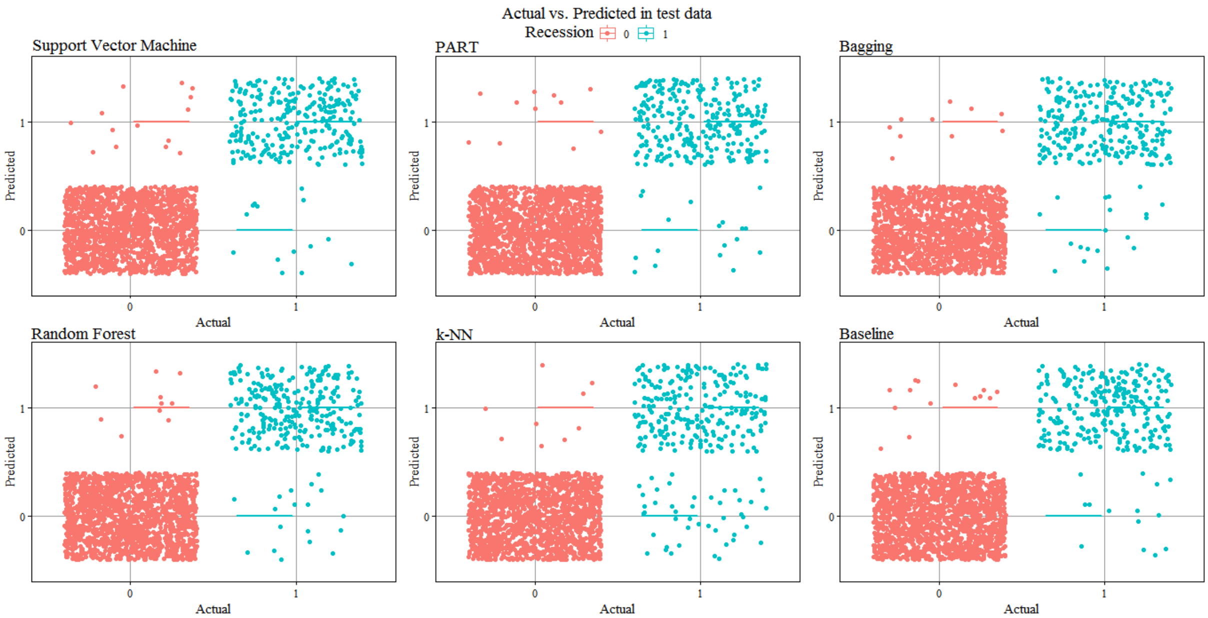

Figure 4: Actual and predicted classes in the test data of experiment 2.1

5.2.2 Experiment 2.2

Table 10b shows the results achieved by applying the models to the subset of the data with a value of "1" on recession at time *t*, resulting in 846 observations for the train data and 284 observations for the test data. In this second part of the experiment all models outperform the naïve baseline model by 5 to 7%. Bagging and Random Forest perform best on both train and test data as they have the highest accuracy. SVM, PART and k-NN perform slightly less. The classification accuracy is very large for every model, but when considering the F1 scores it shows that the models do not classify as well as the accuracy suggests. This is a result of unbalanced classes. In the test data for this part of the experiment 270 instances are labelled with class 1 and only 14 with class 0. Apparently, in case of recession at time *t*, only rarely recession is still apparent at time *t+3*. The models do not do a good job in classifying the instances for class 0 and the unbalanced classes cause a decreased performance. The accuracy is high because the models predict the majority class right, but F1 scores are low since the minority class is not captured.

Table 10b: Results experiment 2.2

Model	**MSE train**		**MSE test**		**F1 train**	**F1 test**
Baseline	0.893	-	0.901	-	0.022	-
SVM	0.956	(7.05%)	0.947	(5.11%)	0.373	0.286
PART	0.955	(6.94%)	0.947	(5.11%)	0.345	0.286
Bagging	0.957	(7.17%)	0.951	(5.55%)	0.400	0.364
Random Forest	0.957	(7.17%)	0.951	(5.55%)	0.400	0.364
k-NN	0.953	(6.72%)	0.947	(5.11%)	0.286	0.211

Results of experiment 2 part II. Parameters SVM: method = C-classification; kernel = radial; C = 1; gamma = 0.09090909; support vectors: 155 - Random Forest: ntree = 5000; importance = TRUE – k-NN: method = knn; trControl = cv (number: 5); k = 7

Figure 5 illustrates these findings. The figure shows that the baseline does not classify instances that are labelled with class 0 as such. The F1 scores for the baseline on the test set cannot be retrieved since it classifies all instances in class 1 and none in class 0, resulting in a precision and recall score of 0. The "harmonic mean" of those, the F1 score, can therefore not be calculated. The other models also classify instances of class 1 quite well, but not instances of class 0.

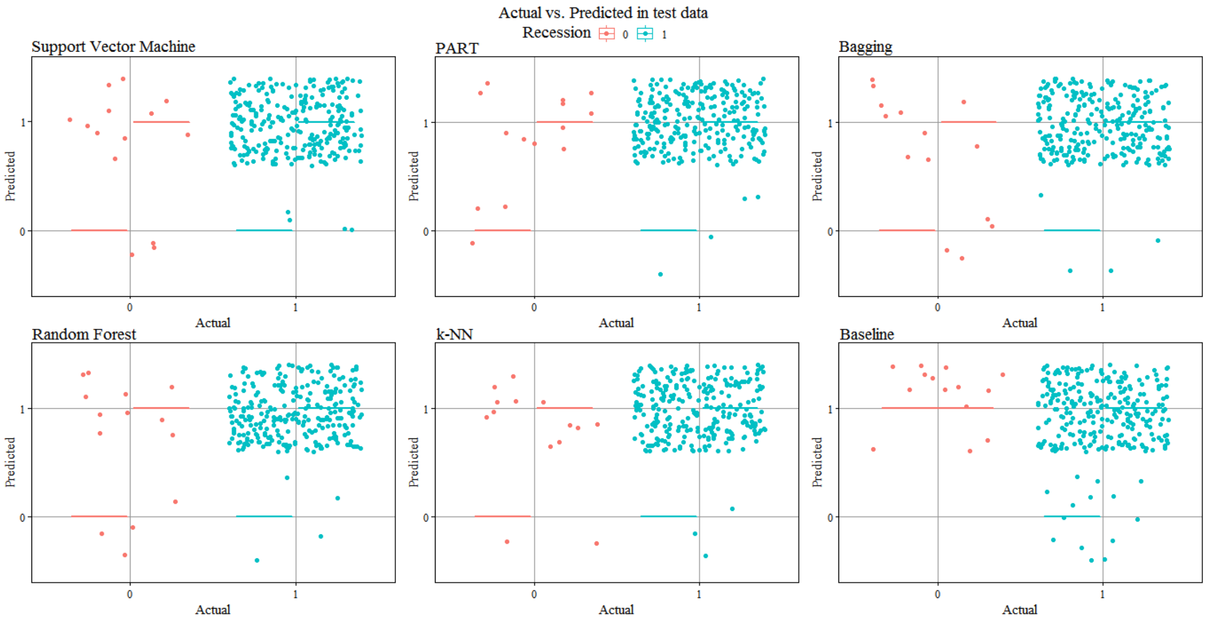

Figure 5: Actual and predicted classes in the test data of experiment 2.2 (states in recession)

To account for the imbalanced classes the minority class “1” is randomly oversampled to create a balanced train and test set. The models are again applied to this subset to assess whether the predictive power of consumer sentiment improves during recession. The results are listed in table 10c. It shows that the models now predict both classes quite well since the F1 scores increased from around 0.3 in the case with the imbalanced classes to around 0.8 in the case with balanced classes. However, the models do not outperform the baseline anymore. The improvements against the baseline are even worse compared to the first part of this experiment in which recession was not controlled for.

Table 10c: Results experiment 2.2

Model	MSE train		MSE test		F1 train	F1 test
Baseline	0.944	-	0.948	-	0.944	0.948
SVM	0.883	(-6.46%)	0.761	(-19.73%)	0.893	0.765
PART	0.881	(-6.67%)	0.807	(-14.87%)	0.892	0.819
Bagging	0.883	(-6.46%)	0.807	(-14.87%)	0.893	0.819
Random Forest	0.883	(-6.46%)	0.807	(-14.87%)	0.893	0.819
k-NN	0.868	(-8.05%)	0.828	(-12.66%)	0.882	0.844

Results of experiment 2 part II with balanced classes. Parameters SVM: method = C-classification; kernel = radial; C = 1; gamma = 0.09090909; support vectors: 476 - Random Forest: ntree = 5000; importance = TRUE - k-NN: method = knn; trControl = cv (number: 5); k = 5

Since accounting for the imbalanced classes by applying the models to a balanced subset of the data when controlling for recession leads to worse predictions in terms of improvement against the baseline, the conclusion that the overall prediction can be done more accurately during recession cannot be drawn.

5.3 Experiment 3

For experiment 3 the difference in the predictive power of consumer sentiment between state clusters is assessed. The experiment examines whether the predictive power of consumer sentiment differs in state clusters when predicting business performance and coincident index. The regression models were tuned and evaluated on the train data and the results were reported in MSE on the train and test set. The procedures for parameter optimization are the same as described in experiment 1. The number of instances in the train and test set for every state cluster are listed in table 11. Tables 12a to d present the results for the first part of this experiment and tables 13a to d for the second part. The coefficients and feature importance of Elastic Net and Random Forest are listed in Appendix H.

Table 11: Number of instances in train and test sets

Cluster	Train	Test
1. Financial cluster	1570	157
2. Oil cluster	628	471
3. Manufacturing cluster	2512	314
4. Mixed Economy cluster	1256	942

5.3.1 Experiment 3.1

The first part of this experiment examines whether the predictive power of consumer sentiment is different between state clusters regarding business performance.

For cluster 1, the MSE on the test set are much bigger than on the train set and also a lot bigger compared to all other experiments. Figures 6 and 7 also illustrate that the models overfit the training set. The values of the change between t+3 and t+2 are larger for the test set, whereas the models constantly predict small values since the data they were trained on did not contain these large changes. The business performance in California, which makes up the test set for cluster 1, has a scaled mean of 4.74 while the scaled mean for the train set is only 0.05. The difference in values explains the difference in performance. The figures also illustrate that Elastic Net does not predict the true values very accurately. The predictions are nearly equal over the different windows, indicating it could not find informative features to model the data. The data is not accurately represented using a linear technique.

Table 12a: Results experiment 3.1.1 - Cluster 1 (10^{-4})

Model	**MSE train**		**MSE test**	
Baseline	3.034	-	52.600	-
Elastic Net	1.378	(54.58%)	22.711	(56.82%)
SVM	1.436	(52.67%)	25.857	(50.84%)
Random Forest	0.511	(83.16%)	20.822	(60.41%)

Results of experiment 3 part I - Cluster I. Parameters Elastic Net: alpha = 0.05145107; lambda = 0.001965335 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 632 - Random Forest: ntree = 5000; importance = TRUE

The results for cluster 2 are listed in table 12b. The models overfit the training set since the MSE on the test set are somewhat higher. This is also illustrated by figures 6 and 7. The models predict the exact same patterns for the test set as they did for the train set. The models on the test set however do perform better compared to the baseline, they outperform the naïve forecast with around 60% whereas the models on the train set outperform the baseline with around 50%. Random Forest is an exception. The model performs best on the train set, but worst on the test set. Elastic Net performs best on the test set. Figures 6 and 7 show that cluster 2 is the only cluster in which Elastic Net models relatively large changes. This means the importance of the features of consumer sentiment is higher in this cluster and that its data can more accurately be modelled using this linear technique.

Table 12b: Results experiment 3.1.2 - Cluster 2 (10^{-4})

Model	**MSE train**		**MSE test**	
Baseline	7.705	-	15.539	-
Elastic Net	3.426	(55.54%)	5.792	(62.73%)
SVM	3.620	(53.02%)	5.798	(62.69%)
Random Forest	1.059	(86.26%)	5.886	(62.12%)

Results of experiment 3 part I - Cluster II. Parameters Elastic Net: alpha = 0.2121503; lambda = 0.002286553 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 259 - Random Forest: ntree = 5000; importance = TRUE

For cluster 3 the MSE on both train and test data are low. The models generalize well to unseen data. Random Forest outperforms the baseline with 89.14% on the train and 77.48% on the test set, which is the biggest increase of all. In order for SVM to map the data points it needs approximately 50% of the training instances. It has the lowest performance with regard to the baseline on both train and test data. Elastic Net uses an alpha close to the Ridge L_2-penalty. The nearly straight lines of Elastic Net in figures 6 and 7 indicate that features of consumer sentiment are not considered important in predicting business performance in cluster 3.

Table 12c: Results experiment 3.1.3 - Cluster 3 (10^{-4})

Model	MSE train		MSE test	
Baseline	0.967	-	0.977	-
Elastic Net	0.429	(55.64%)	0.409	(58.14%)
SVM	0.444	(54.08%)	0.444	(54.55%)
Random Forest	0.105	(89.14%)	0.220	(77.48%)

Results of experiment 3 part I - Cluster III. Parameters Elastic Net: alpha = 0.08075406; lambda = 0.00160513 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1312 - Random Forest: ntree = 5000; importance = TRUE

Lastly, table 12d presents the results for cluster 4. An interesting result is that the MSE on the test set are lower than on the train set, meaning the models capture patterns in unseen data better. The performance against the baseline is however higher on the train set. The improvement against the baseline for Elastic Net on the train set is highest compared to the other clusters. On the test set it is somewhat lower than for cluster 2. Figure 7 illustrates this finding. Whereas Elastic Net predicts nearly equal values over the different windows in cluster 1 and 3, it now captures patterns in the test data better. This indicates it can find more informative features to predict the true changes of business performance in cluster 4.

Table 12d: Results experiment 3.1.4 - Cluster 4 (10^{-4})

Model	MSE train		MSE test	
Baseline	3.354	-	0.345	-
Elastic Net	1.200	(64.22%)	0.142	(58.84%)
SVM	1.215	(63.77%)	0.147	(57.39%)
Random Forest	0.617	(81.60%)	0.199	(42.32%)

Results of experiment 3 part I - Cluster IIII. Parameters Elastic Net: alpha = 0.08436307; lambda = 0.001500391 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 538 - Random Forest: ntree = 5000; importance = TRUE

Compared to the other clusters, Elastic Net and SVM yield the highest overall improvements with respect to the baseline of the train set in cluster 4, but Random Forest performs worst. Furthermore, the improvements on the test set are not the highest of all. Except for Random Forest, the models in cluster 2 yield the highest improvements on the test set, whereas the results on the train set are second lowest. The improvements against the baseline of cluster 3 for Elastic Net and SVM are second best and for Random Forest best over the clusters, but the improvements on the test set are among the lowest. The results indicate that none of the state clusters scores the highest overall performance on both the train and test set.

Figure 6 below illustrates the predicted and actual values of the models on the train set for the 4 clusters. The figure displays the first 10 sliding windows of all train sets.

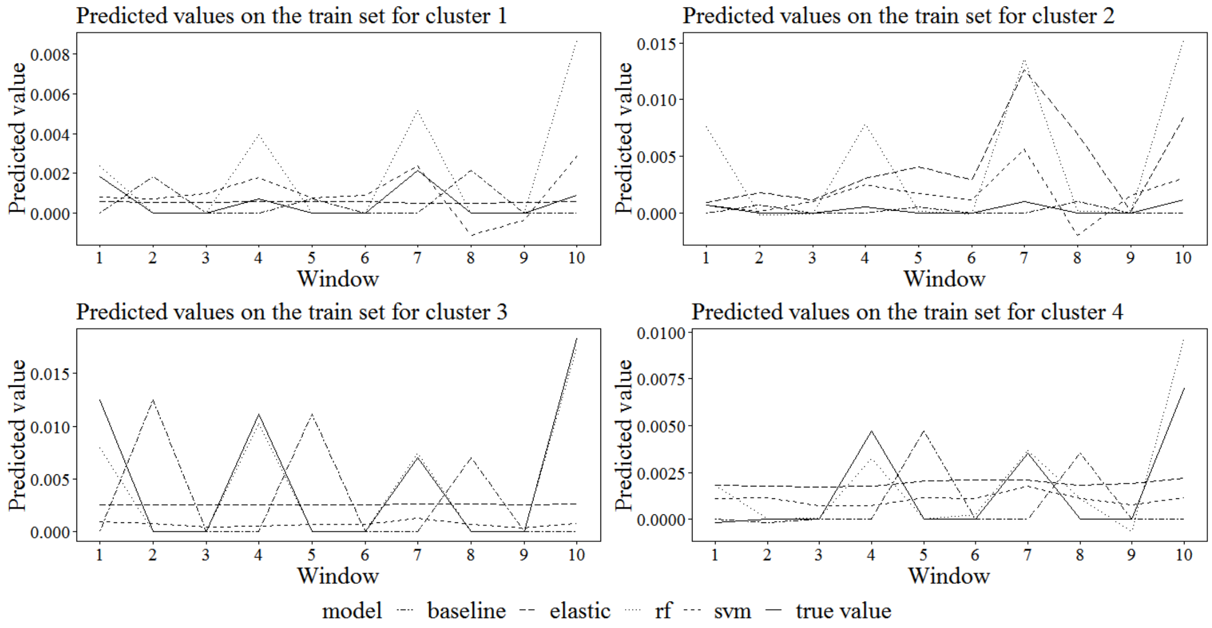

Figure 6: Actual and predicted values on the train set of experiment 3.1

Figure 7 below illustrates the predicted and actual values of the models on the test set for the 4 clusters. The figure displays the first 10 sliding windows of all test sets.

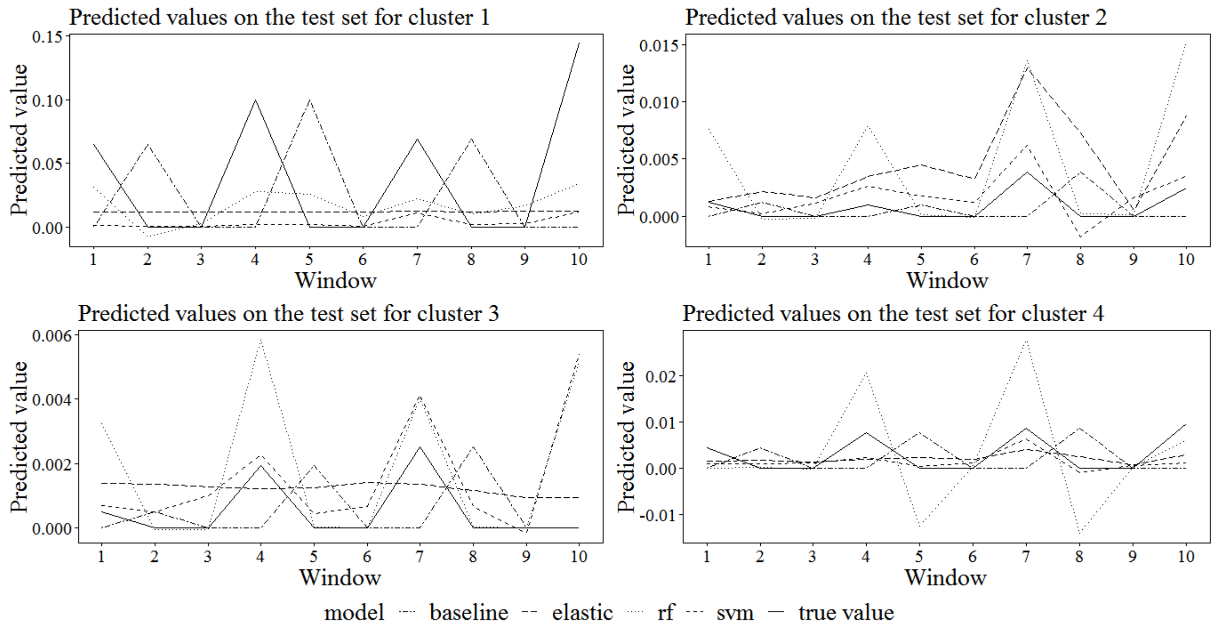

Figure 7: Actual and predicted values on the test set of experiment 3.1

The results indicate that the test sets are very different from the train sets for every cluster since the results of the test data mostly do not correspond to the results of the train data. Therefore, this part of the experiment was conducted a second time using a different partitioning of the train and test data, namely a 70/30 split into a training and test set with the values of the change in business performance between t+3 and t+2 equally distributed across both sets. In this way, the sets are more equal in terms of the values the models try to predict. The results are listed in Appendix I. Still, no state cluster includes models with the highest overall performance on both the train and test set. To conclude, the predictive power of consumer sentiment does differ over state clusters but not systematically.

5.3.2 Experiment 3.2

The second part of this experiment examines whether the predictive power of consumer sentiment is different between state clusters regarding coincident index.

For cluster 1 the results are listed in table 13a. The MSE on the test set are lower than on the train data, which means the models fit the unseen data better. Random Forest has the smallest MSE on the train data and performs equal to SVM on the test data. SVM however uses 1220 support vectors, which is 86% of the train observations. This indicates it could not find a clear pattern in the data. Elastic Net performs slightly less on both the train and test set. Figures 8 and 9 illustrate the predictions. It shows that the data is not captured very well using a linear technique since Elastic Net predicts almost no fluctuation in the data. This indicates the model shrunk many coefficients to zero when predicting business performance in cluster 3.

Table 13a: Results experiment 3.2.1 - Cluster 1 (10^{-4})

Model	**MSE train**		**MSE test**	
Baseline	0.134	-	0.070	-
Elastic Net	0.045	(66.42%)	0.024	(65.71%)
SVM	0.044	(67.16%)	0.023	(67.14%)
Random Forest	0.024	(82.09%)	0.023	(67.14%)

Results of experiment 3 part II - Cluster I. Parameters Elastic Net: alpha = 0.5492338; lambda = 0.009094559 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1220 - Random Forest: ntree = 5000; importance = TRUE

The test data of cluster 2 are also better modelled than the train data since the MSE are lower, except for Random Forest. Random Forest fits the train data very well and yields an improvement of 82.64% against the baseline, but the MSE on the test set is twice as big and yields the smallest improvement against the baseline. This means the model overfits the training data. Elastic Net performs worse on both train and test data. SVM has the second lowest MSE on the train data and the lowest on the test data. For this cluster it also used a lot of support vectors, corresponding to 86% of the train data.

Table 13b: Results experiment 3.2.2 - Cluster 2 (10^{-4})				
Model	**MSE train**		**MSE test**	
Baseline	0.144	-	0.134	-
Elastic Net	0.056	(61.11%)	0.046	(65.67%)
SVM	0.054	(62.50%)	0.045	(66.42%)
Random Forest	0.025	(82.64%)	0.052	(61.19%)

Results of experiment 3 part II - Cluster II. Parameters Elastic Net: alpha = 0.5617253; lambda = 0.03044812 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 538 - Random Forest: ntree = 5000; importance = TRUE

Table 13c presents the results for cluster 3. For this cluster the MSE on the test set are higher than those of the train set, which means the models overfit the training data. Figures 8 and 9 illustrate this too, the models capture the general patterns in the train set better than they do for the test set. Random Forest performs best on the train set with an increase in performance of 80.65%. The performance on the test set is however lowest. Elastic Net and SVM yield an MSE of 0.070 whereas the error of Random Forest is 0.074. SVM again uses 87% of the training instances to model the data. The MSE of all models are quite comparable, except for the low error of Random Forest on the train set. This indicates that all models generally predict the same patterns in the data.

Table 13c: Results experiment 3.2.3 - Cluster 3 (10^{-4})				
Model	**MSE train**		**MSE test**	
Baseline	0.124	-	0.187	-
Elastic Net	0.045	(63.71%)	0.070	(62.57%)
SVM	0.043	(65.32%)	0.070	(62.57%)
Random Forest	0.024	(80.65%)	0.074	(60.43%)

Results of experiment 3 part II - Cluster III. Parameters Elastic Net: alpha = 0.5887024; lambda = 0.001491015 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 2184 - Random Forest: ntree = 5000; importance = TRUE

Lastly, cluster 4 yields the smallest MSE on the train data and the largest improvements with regard to the baseline compared to the other clusters. For the test set, the MSE and improvements are not best. The models overfit the training data since the MSE on the test set are somewhat higher. Random Forest is the best model on the train data, but has the highest MSE on the test set, implying it heavily overfits the train data. For SVM it can again be concluded that it uses almost 87% of the train instances to model the data. It is second best on the train set and performs equally well as Elastic Net on the test set. Elastic Net uses an alpha close to the L_1-penalty of Lasso regression. This results in small coefficients and therefore nearly equal predictions over the windows. Figures 8 and 9 illustrate this.

Table 13d: Results experiment 3.2.4 - Cluster 4 (10^{-4})

Model	MSE train		MSE test	
Baseline	0.088	-	0.122	-
Elastic Net	0.029	(67.05%)	0.042	(65.57%)
SVM	0.028	(68.18%)	0.042	(65.57%)
Random Forest	0.014	(84.09%)	0.045	(63.11%)

Results of experiment 3 part II - Cluster IIII. Parameters Elastic Net: alpha = 0.8793684; lambda = 0.005933738 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1091 - Random Forest: ntree = 5000; importance = TRUE

The MSE and improvements against the baseline do not differ heavily over the four clusters and the results on the test data mostly do not correspond to the results on the train data. The predictive power of consumer sentiment for coincident index is highest for the train set of cluster 4 since it has the biggest improvements with respect to the baseline. It however overfits the training data and the MSE and improvements of the test data are not the highest, cluster 1 and 2 perform better in this regard. Cluster 1 and 2 also generalize better to unseen data since the MSE on the test set are lower than the train set, whereas both cluster 3 and 4 overfit the training data. The MSE of the train set of cluster 2 are among the largest and the improvements against the baseline are the smallest, with the exception of Random Forest that yields a very high improvement. However, the improvements on the test set are relatively high, which is not in line with the results on the train set. No cluster shows clear results indicating a high or low performance on both train and test sets.

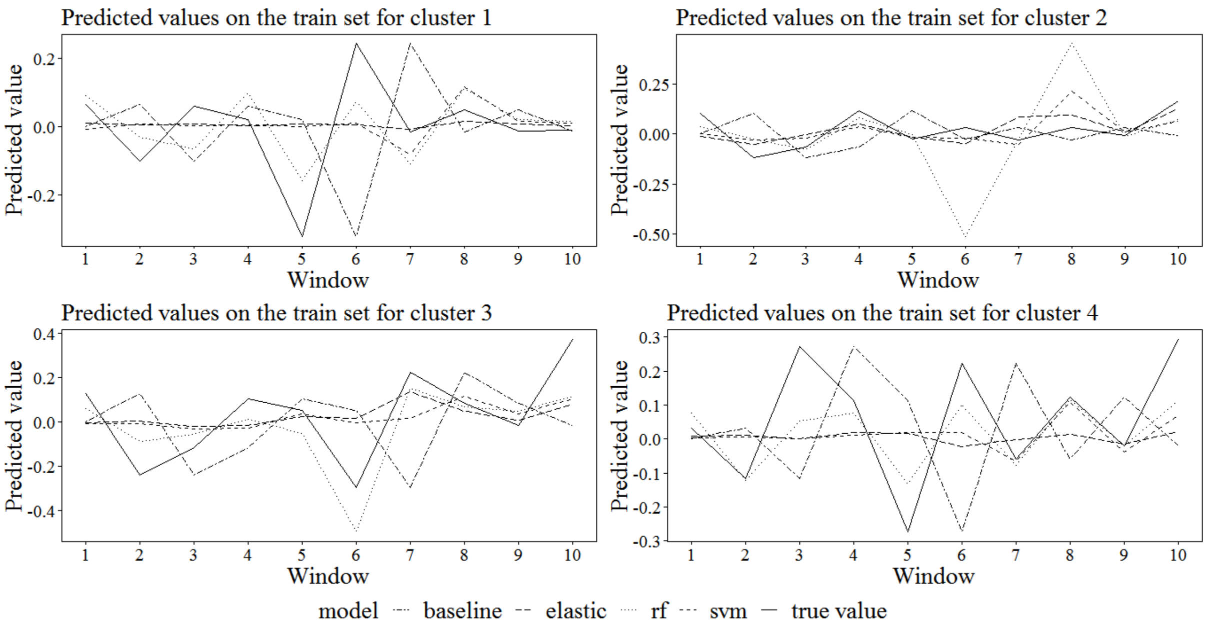

Figure 8: Actual and predicted values on the train set of experiment 3.2

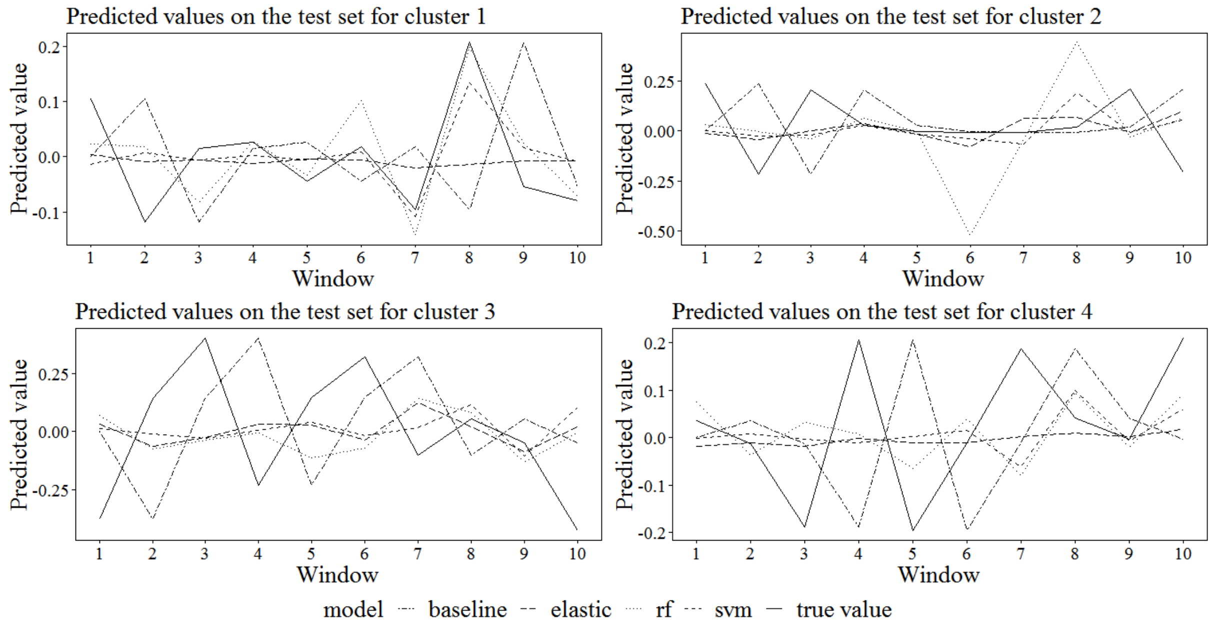

Figure 9: Actual and predicted values on the test set of experiment 3.2

Since these results indicate that the test sets are very different from the train sets for every cluster, this part of the experiment was also conducted a second time using a different partition of the train and test data, namely a 70/30 split into a training and test set with the values of the change in coincident index between t+3 and t+2 equally distributed across both sets. This causes the sets to be more equal in terms of the values the models try to predict. The results are listed in Appendix J. It shows that, still, no state cluster includes models with the highest overall performance on both the train and test set. The predictive power of consumer sentiment for coincident index does differ over clusters, but not systematically.

5.4 Experiment 4

For experiment 4 the predictive power associated with a change in the causal direction of the relationship between consumer sentiment and economic conditions is assessed. The experiment investigates if it is more insightful to predict consumer sentiment by business performance, economic recession, and coincident index. In order to meaningfully assess predictive power, the improvements against the baseline of the first, second, and third part of this experiment are compared to experiment 1, 2, and 3 respectively. Appendix K lists the coefficients and feature importance of Elastic Net and Random Forest.

5.4.1 Experiment 4.1

The first part of this experiment examines if it is more insightful to explain consumer sentiment by business performance. The predictive performance with respect to the baseline is better for Random

Forest on both test and train data as compared to experiment 1. For experiment 1 the predictive performance for Random Forest on the train and test set was 81.70% and 67.88% respectively, and it now scores 92.96% and 70.77% respectively. However, both Elastic Net and SVM yield higher improvements in experiment 1. Furthermore, SVM uses 90% of the train data to regress the data points, which means it could not find a clear pattern in the data. In experiment 1, it used only 40% of the data.

Table 14a: Results experiment 4.1

Model	**MSE train**		**MSE test**	
Baseline	35.284	-	35.291	-
Elastic Net	17.575	(50.19%)	17.580	(50.19%)
SVM	17.406	(50.67%)	17.563	(50.23%)
Random Forest	2.485	(92.96%)	10.315	(70.77%)

Results of experiment 4 part I (business performance). Parameters Elastic Net: alpha = 0.6246774; lambda = 0.07549938 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 5365 – Random Forest: ntree = 5000; importance = TRUE

Figure 10 illustrates that Random Forest best models the data since it fits the true values almost perfectly. The other models do not do a great job in predicting consumer sentiment. The Elastic Net uses an alpha close to one which means it performs much like Lasso regression. Most coefficients are shrunk close to zero and this explains the rather straight line in the figure, which was also the case in experiment 1. This indicates that the model does not consider features of business performance more informative in predicting consumer sentiment than it did vice versa.

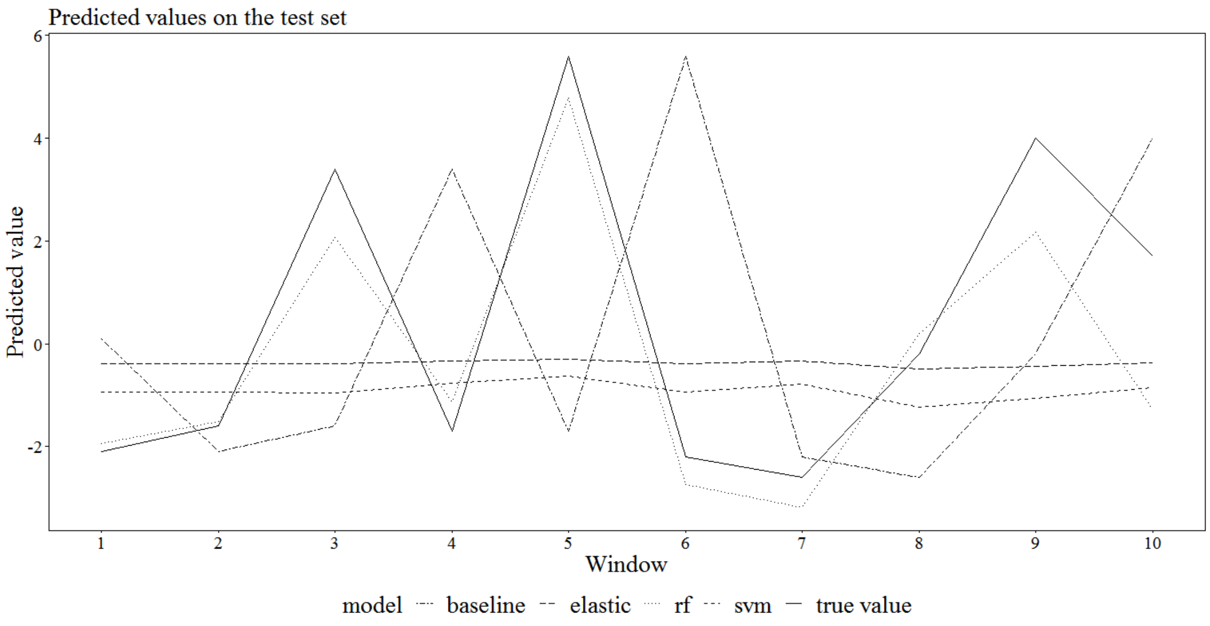

Figure 10: Actual and predicted values on the test set of experiment 4.1

5.4.2 Experiment 4.2

The second part of the experiment examines whether consumer sentiment is better predicted by recession by comparing the results of the first part of experiment 2 with table 14b. While in experiment 2 the classification models hardly outperformed the baseline all regression models in this experiment outperform the baseline by around 50%. Random Forest performs best. The MSE are however relatively high.

Table 14b: Results experiment 4.2

Model	**MSE train**		**MSE test**	
Baseline	35.284	-	35.291	-
Elastic Net	17.454	(50.53%)	17.553	(50.26%)
SVM	17.392	(50.71%)	17.496	(50.42%)
Random Forest	16.042	(54.53%)	16.177	(54.16%)

Results of experiment 4 part II (economic recession). Parameters Elastic Net: alpha = 0.9419739; lambda = 0.007084022 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 5246 - Random Forest: ntree = 5000; importance = TRUE

Figure 11 below illustrates the results for 10 sliding windows. No model predicts the true values very well. A change in causality between sentiment and recession appears to be more insightful since the models now outperform the baseline, but they yield very high MSE and according to the figure do not capture patterns in the data.

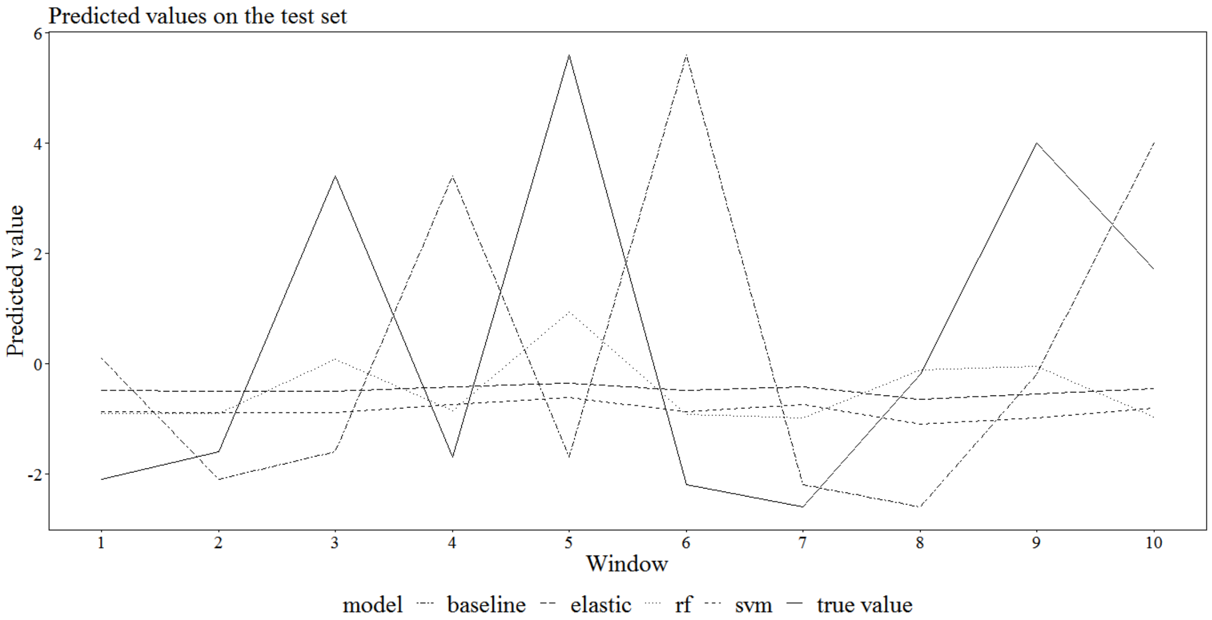

Figure 11: Actual and predicted values on the test set of experiment 4.2

5.4.3 Experiment 4.3

Lastly, the predictive power of coincident index regarding consumer sentiment is assessed. Since the effect of consumer sentiment on coincident index was only examined per state cluster, the performances of the models against the baseline are averaged over the clusters. This results in the percentage improvement listed in table 14c.

Table 14c: Average results experiment 3.2

Model	**Train set**	**Test set**
Baseline	-	-
Elastic Net	64.64%	64.88%
SVM	65.87%	65.43%
Random Forest	82.28%	62.97%

From the results it can be concluded that only Random Forest performs better. It almost yields a hundred percent improvement against the baseline. It however overfits the training data since the test set has a relatively high MSE compared to the train set. The other models perform worse. Elastic Net scores the highest MSE. It uses an alpha of 0.957, close to Lasso regression.

Table 14d: Results experiment 4.3

Model	**MSE train**		**MSE test**	
Baseline	35.284	-	35.291	-
Elastic Net	17.277	(51.03%)	17.389	(50.73%)
SVM	15.452	(56.21%)	16.105	(54.37%)
Random Forest	0.203	(99.42%)	1.099	(96.89%)

Results of experiment 4 part III (coincident index). Parameters Elastic Net: alpha = 0.9570266; lambda = 0.001278133 - SVM: method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 5254 - Random Forest: ntree = 5000; importance = TRUE

The results are illustrated in figure 12. Random Forest models the data very well, whereas the other models do not capture the real patterns in the test data. Elastic Net does not capture actual fluctuations in its predictions. The large alpha increases the likelihood that many coefficients are shrunk to zero. This was also the case in most clusters in experiment 3.2, so the model considers features of coincident index not more important in predicting consumer sentiment than it did vice versa.

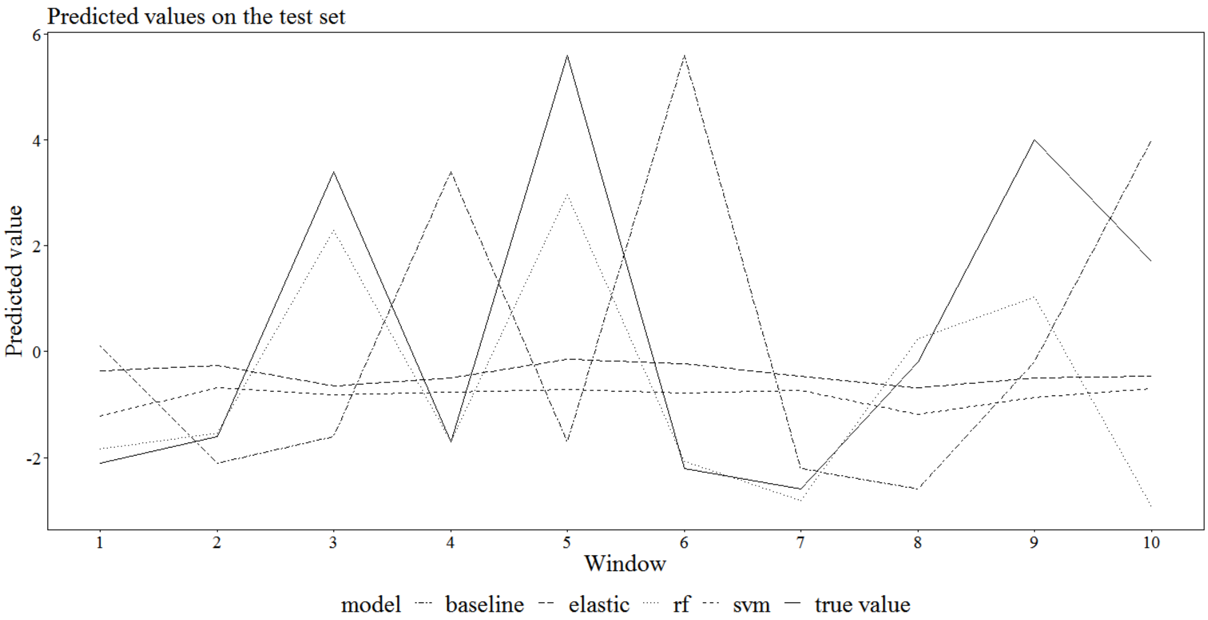

Figure 12: Actual and predicted values on the test set of experiment 4.3

From these results it can be concluded that it is more insightful to predict consumer sentiment by economic recession rather than the other way around. When predicting recession by sentiment the models hardly outperform the baseline, whereas the models perform around 50% better than the baseline when changing the causal direction. The models however yield very high MSE. For business performance only Random Forest resulted in higher improvements against the baseline when changing causality. Business performance is better explained by consumer sentiment when using SVM and Elastic Net. Lastly, when changing the relationship between consumer sentiment and coincident index again only Random Forest results in higher improvements with respect to the baseline. It furthermore overfits the training data. Coincident index is thus better explained by consumer sentiment using SVM and Elastic Net.

6. Discussion

The aim of this research was to examine how accurately consumer sentiment can predict various indicators of economic conditions within states of the United States. These indicators are business performance, coincident index, and economic recession. The Michigan Consumer Sentiment Index over the period of January 2005 to April 2018 was used to retrieve monthly values of consumer sentiment, which are the attitudes and expectations of consumers towards their personal finances, general business conditions, and market conditions. To address the overall research question, four experiments were conducted in which the relationship between sentiment and economic conditions was assessed. The main goal of these experiments was to build predictive models which not only examine the predictive power associated with consumer sentiment, but also investigate whether it is more insightful to predict consumer sentiment by those indicators of the economy. A sliding window approach was used to extract temporal patterns from the data. This method mapped the observations into overlapping time windows of three months and constructed features from this data to code temporal relationships between variables. The main objective of each experiment was to predict the future state of the economy based upon features of consumer sentiment.

The first experiment addressed the question to what degree consumer sentiment can predict business performance and whether this prediction can be done more accurately during economic recession. The aim was to predict the value of the change in business performance between the last observation in the time window and the first observation directly after it, corresponding to the value of business performance three months ahead. This thesis found consumer sentiment can accurately predict business performance, since the models Elastic Net, SVM, and Random Forest outperformed the baseline which predicted the value of the change to be the same as it was for the previous observation, also called the naïve forecast. This finding is in accordance with previous findings of Juster & Wachtel (1974), Bram and Ludvigson (1998), and Lahiri and Zhao (2016) that sentiment is related to business performance. An interesting finding was that the predictive power of sentiment is not higher during economic recession. The improvements against the baseline were lower when applying the models to the subset of sliding windows with a positive score on recession. Since prior research has not investigated the effect of the relationship during recession it is not directly clear what factors are likely to contribute to this effect, but it could be insightful for future research to use these findings as a potential basis for analysis.

The second experiment examined to what degree consumer sentiment can predict economic recession and whether this predictive power is higher during economic recession. The aim was to predict economic recession three months ahead. Whereas some previous research has found a relation between consumer sentiment and recession (Blanchard, 1993; Angeletos & La'O, 2013) the models SVM, PART, Bagging, Random Forest, and k-NN in this research hardly outperformed the baseline which predicted economic recession to be the same as the previous observation. An explanation could be that sentiment is indeed just a reflection of economic circumstances, a finding supported by Throop (1992) and Barnes and Olivei (2017). Economic recession could be apparent before a change in sentiment and as a result, sentiment does not predict recession. When applying the models to the subset of time windows suffering from recession it appeared that the models accurately predicted the majority class – no recession three months ahead – but did not capture the minority one – recession three months ahead. After accounting for this imbalance by applying the models to a balanced subset of the data, the predictions in terms of improvement against the baseline deteriorated when controlling for recession. Therefore, the conclusion that the overall prediction will be more accurately during recession cannot be drawn. As stated before, it is possible that consumer sentiment does not cause recession. In that case it is not surprising that the prediction cannot be done more accurately during recession. Again, previous research has not investigated the effect between sentiment and recession during recession. There might be factors not considered in this research that contribute to this relationship. Future research could build upon these findings by identifying possible mediating factors.

The third experiment assessed the predictive power of consumer sentiment over state clusters regarding both business performance and coincident index. The coincident index summarizes state-level economic conditions in a single statistic by combining employment measures and wage and salary disbursements. The state clusters are the financial, oil, manufacturing, and mixed economy clusters and are constructed based on the periods in which states suffered from recession. The goal was to predict the value of the change between the last observation in the time window and the subsequent point outside it, which is the change in business performance and coincident index three months ahead. To compare the predictive power of Elastic Net, SVM, and Random Forest over state clusters, the improvements of the models against the naïve forecast baseline were examined. The expectation that the effect of sentiment differs over state clusters, based on the findings of Owyang et al. (2004), was confirmed, but not systematically. First, for business performance none of the state clusters yielded the highest overall performance on both train and test data. The results indicated differences when comparing the performance on the train sets, but this did not generalize to the test sets and vice versa. The results indicated that the clusters in the test set contained patterns not captured by the models trained on the train set. Second, for coincident index the results also did not indicate a high or low performance on both train and test set over the clusters. The errors and improvements against the baseline did not differ heavily over the four clusters and the results on the test data did not correspond to the results on the train data. Again, the test set was too different from the train set. Therefore, both experiments were conducted a second time using a different partitioning of the train and test set that equally distributed the values the models tried to predict over the sets. These results, listed in Appendix I and J, also indicate that there are differences in predictive power over the state clusters, but still not systematically. A limitation that could have caused these findings is that the Michigan Consumer Sentiment Index is only measured over the United States as a whole and not per state. Future research could investigate if state-level sentiment might make a difference.

The last experiment accounted for the possibility of a causality issue by examining whether or not more predictive power was associated with predicting consumer sentiment by business performance, economic recession, and coincident index. It was clear from the results that it is more insightful to predict sentiment by economic recession rather than the other way around, since the change in causality led to more informative predictions in terms of improvement against the naïve forecast baseline. This finding is also supported in previous research (Throop, 1992). The models used in this thesis did however score very high MSE on the train and test set. For both business performance and coincident index, the results were not that obvious. Only Random Forest resulted in higher improvements against the naïve baseline when changing causality. Business performance and coincident index were better explained by consumer sentiment when using Elastic Net and SVM. These results partly refute the findings of Barnes & Olivei (2017) that the independent information from sentiment is limited with regard to economic conditions.

6.1 Limitations

There are several limitations with regard to the data used in this thesis. First of all, the personal consumption expenditures over all industries in a state are regarded as business performance, but business performance is not only reflected through financial performance. Future research could develop a more sophisticated measure for this. Next, the Michigan Consumer Sentiment Index does not measure sentiment per state but produces one aggregate monthly score over the United States as a whole. Future research can build on this research by using an index that measures state-level sentiment. Besides, the amount of data was very limited with a total of 7,850 observations. Conducting the analyses with a larger amount of data can possibly result in more informative outcomes. Using data from different countries instead of just considering the United States as done in this thesis could probably also result in more informative outcomes. Lastly, the sliding window approach used time windows with a width of three months. Increasing the width of the windows increases the data available for making predictions and this could enhance the predictive power of the models.

7. Conclusion

This thesis aimed to answer how accurately consumer sentiment can predict various indicators of economic conditions within states of the United States by formulating four research questions:

RQ1: *To what degree can consumer sentiment predict business performance, and can this prediction be done more accurately during economic recession?*

RQ2: *To what degree can consumer sentiment predict state recession, and can this prediction be done more accurately during economic recession?*

RQ3: *To what degree does the predictive power of consumer sentiment differ over state clusters with regard to business performance and coincident index?*

RQ4: *To what degree would it be more insightful to change the causal direction between the indicators of economic conditions and consumer sentiment?*

By introducing the predictive classification and regression models Elastic Net, Support Vector Machine, Random Forest, PART, Bagging, and k-NN, that incorporate both state-level information and control for economic recession, this thesis provided more insight into the predictive power of consumer sentiment at the state level. State-level generalization was accounted for by training the models on North East, Midwest, and South America and testing them on the Western part of America.

The results indicate that consumer sentiment has power in predicting business performance and coincident index but that this predictive power does not improve during economic recession or differ over state clusters systematically. Furthermore, consumer sentiment does not predict economic recession. Rather, it showed that economic recession has power in predicting sentiment. While not fully consistent over the different models used in the analysis, the predictive power of consumer sentiment regarding business performance and coincident index was higher than it was when changing the causal direction of these relationships.

Sources

Anderson, S. T., Kellogg, R., Sallee, J. M., & Curtin, R. T. (2011). Forecasting gasoline prices using consumer surveys. *American Economic Review, 101*(3), 110-14.

Angeletos, G. M., & La'O, J. (2013). Sentiments. *Econometrica, 81*(2), 739-779.

Barnes, M. L., & Olivei, G. P. (2017). Consumer Attitudes and Their Forecasting Power for Consumer Spending. Journal of Money, *Credit and Banking, 49*(5), 1031-1058.

Benhabib, J., & Spiegel, M. M. (2018). Sentiments and economic activity: Evidence from US states. *The Economic Journal*, *129*(618), 715-733.

BenYahmed, Y., Abu Bakar, A., Hamdan, A. R., Ahmed, A., & Syed Abdullah, S. M. (2015). Adaptive sliding window algorithm for weather data segmentation. *Journal of Theoretical and Applied Information Technology, 80*(2), 322-333.

Beyari, H., Abareshi, A., & Elferjani, I. (2017). An exploration of the factors that impact customer experience in social commerce websites. *ITS 2017*, 51-58.

Blanchard, O.J. (1993). "Consumption and the Recession of 1990-1991." *American Economic Review, 83*(2), 270-274.

Blood, D. J., & Phillips, P. C. (1995). Recession headline news, consumer sentiment, the state of the economy and presidential popularity: A time series analysis 1989–1993. *International Journal of Public Opinion Research*, *7*(1), 2-22.

Bram, J., & Ludvigson, S. C. (1998). Does consumer confidence forecast household expenditure? A sentiment index horse race. *Economic Policy Review*, *4*(2).

Burns, A. F., & Mitchell, W. C. (1947). Measuring business cycles.

Calder, B. J., Hollebeek, L. D., & Malthouse, E. C. (2018). Creating stronger brands through consumer experience and engagement. *Customer engagement marketing*, 221-242.

Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why?. *The American Economic Review, 84*(5), 1397-1408.

Curtin, R. T. (2003). Unemployment expectations: The impact of private information on income uncertainty. *Review of Income and Wealth*, *49*(4), 539-554.

Curtin, R. (2011). The Stagnation Two-Step: Consumer Pessimism and Consumer Spending. *Economic Outlook Conference*, University of Michigan.

Curtin, R. (2013). Consumer Behavior Adapts to Fundamental Changes in Expectations. *Economic Outlook Conference*, University of Michigan.

Curtin, R., Tuba Suzer-Gurtekin, Z., Ellcey, E., Fu, Y., Burgett, V., & Zajkowski, J. (2019). Survey Research Center of the University of Michigan: Consumer Sentiment. Retrieved from Surveys of Consumers; https://data.sca.isr.umich.edu/, April 2, 2019.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Dancho, M. (2017, August 30). Tidy Time Series Analysis, Part 4: Lags and Autocorrelation. Retrieved from www.business-science.io/timeseries-analysis/2017/08/30/tidy-timeseries-analysis-pt-4, April 2, 2019.

Daniels, H. (2018). *Classification Methods* [Powerpoint slides]. Retrieved from the Business Intelligence and Data Management Blackboard, April 15, 2019.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis, 1*(1-4), 131-156.

Diez, D. M., Barr, C. D., & Cetinkaya-Rundel, M. (2012). *OpenIntro statistics*, 174-175.

Emmery, C. (2018). Working with text data – part I [Powerpoint slides]. Retrieved from the Data Mining Blackboard, April 15, 2019.

Federal Reserve Bank of Philadelpha (FRED). (2019). State Coincident Indexes. Retrieved from FRED; https://www.philadelphiafed.org/research-and-data/regional-economy/indexes/coincident/, April 25, 2019.

Fiscal 50. (2019). Fiscal 50: State Trends and Analysis. Retrieved from PEW; https://www.pewtrusts.org/en/research-and-analysis/data-visualizations/2014/fiscal-50#ind0, April 18, 2019.

Flatters, P., & Willmott, M. (2009). Understanding the post-recession consumer. *Harvard Business Review, 87*(7/8), 106-112.

Fonti, V., & Belitser, E. (2017). Feature selection using lasso. *VU Amsterdam Research Paper Business Analytics.*

Frank, E., & Witten, I. H. (1998). Generating accurate rule sets without global optimization. *Working paper 98*(2), 1-16.

Fuhrer, J. C. (1993). What role does consumer sentiment play in the US macroeconomy?. *New England Economic Review*, 32-44.

Gelper, S., Lemmens, A., & Croux, C. (2007). Consumer sentiment and consumer spending: decomposing the Granger causal relationship in the time domain. *Applied Economics, 39*(1), 1-11.

Golinelli, R., & Parigi, G. (2004). Consumer sentiment and economic activity. *Journal of Business Cycle Measurement and Analysis, 2004*(2), 147-170.

Harrison-Walker, L. J. (2001). The measurement of a market orientation and its impact on business performance. *Journal of Quality management, 6*(2), 139-172.

Huth, W. L., Eppright, D. R., & Taube, P. M. (1994). The indexes of consumer sentiment and confidence: Leading or misleading guides to future buyer behavior. *Journal of Business Research, 29*(3), 199-206.

Hyndman, R. J., & Athanasopoulos, G. (2018). F*orecasting: principles and practice.* OTexts.

Jolliffe, I. (2011). Principal component analysis. *Springer Berlin Heidelberg*, 1094-1096

Kenton, W. (2018). Consumer Sentiment. Retrieved from https://www.investopedia.com/terms/c/consumer-sentiment.asp, April 9, 2019.

Kirsch, I. (2019). State Recessions, version 2 (industries_recessions). Retrieved from https://www.kaggle.com/kirschil/state-recessions#industries_recessions.csv, February 1, 2019.

Komorowski, J. (2015). Decision trees and Monte Carlo feature selection [Powerpoint slides]. Retrieved from http://bioinf.icm.uu.se/rsb15/5_DTreesAndMCFS-Gliwice.pdf, April 8, 2019.

Lahiri, K., Monokroussos, G., & Zhao, Y. (2016). Forecasting consumption: The role of consumer confidence in real time. Journal of Applied Econometrics, 31(7), 1254-1275.

Lang, K.M. (2018). Data Cleaning: Missing Data & Outliers [Powerpoint slides]. Retrieved from the Statistics and Methodology Blackboard, April 2, 2019.

Linden, F. (1982). The consumer as forecaster. *Public Opinion Quarterly*, *46*(3), 353-360.

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2013). On the joys of missing data. *Journal of pediatric psychology*, *39*(2), 151-162.

Liu, H., & Motoda, H. (Eds.). (1998). Feature extraction, construction and selection: A data mining perspective. *Springer Science & Business Media*, 453.

Lozza, E., Bonanomi, A., Castiglioni, C., & Bosio, A. C. (2016). Consumer sentiment after the global financial crisis. *International Journal of Market Research, 58*(5), 671-691.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic perspectives*, *18*(2), 29-50.

Marcato, G., & Nanda, A. (2016). Information content and forecasting ability of sentiment indicators: case of real estate market. *Journal of Real Estate Research, 38*(2), 165-203.

Master of Project Academy. (2017). Relative Measure of Dispersion. Retrieved from MPA; https://blog.masterofproject.com/relative-measures-of-dispersion/, April 7, 2019.

Megna, R., & Xu, Q. (2003). Forecasting the New York State economy: The coincident and leading indicators approach. *International Journal of Forecasting, 19*(4), 701-713.

Mishkin, F. S., Hall, R., Shoven, J., Juster, T., & Lovell, M. (1978). Consumer sentiment and spending on durable goods. *Brookings Papers on Economic Activity*, *1978*(1), 217-232.

Mozaffari, L., Mozaffari, A., & Azad, N. L. (2015). Vehicle speed prediction via a sliding-window time series analysis and an evolutionary least learning machine: A case study on San Francisco urban roads. *Engineering science and technology, an international journal, 18*(2), 150-162.

Nguyen, L. T., Wu, P., Chan, W., Peng, W., & Zhang, Y. (2012, August). Predicting collective sentiment dynamics from time-series social media. *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining* (p. 6).

Piramuthu, S., Ragavan, H., & Shaw, M. J. (1998). Using feature construction to improve the performance of neural networks. *Management Science, 44*(3), 416-430.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rüping, S. (2001). SVM kernels for time series analysis. *Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, 2001*(43), 1-13.

Sarkar, S., Vinay, S., Pateshwari, V., & Maiti, J. (2016, December). Study of optimized SVM for incident prediction of a steel plant in India. *2016 IEEE Annual India Conference*, 1-6.

Scholkopf, B., Sung, K. K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T., & Vapnik, V. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing, 45*(11), 2758-2765.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.

Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications. *Springer.*

Starr, M. A. (2012). Consumption, sentiment, and economic news. *Economic Inquiry*, *50*(4), 1097-1111.

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, 4, 351-394.

The Balance, The Worst Economic Contractions in U.S. History. Retrieved from The Balance; https://www.thebalance.com/economic-contraction-4067683, February 14, 2019.

Thomas, L. B. (1999). Survey measures of expected US inflation. *Journal of Economic perspectives*, *13*(4), 125-144.

Throop, A. W. (1992). Consumer sentiment: Its causes and effects. *Federal Reserve Bank of San Francisco Economic Review*, 1, 35-59.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in medicine, 16*(4), 385-395.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(3), 273-282.

Tutz, G., & Binder, H. (2007). Boosting ridge regression. *Computational Statistics & Data Analysis, 51(*12), 6044-6059.

U.S. Bureau of Economic Analysis (BEA), Consumer Spending. Retrieved from BEA.gov, Bureau of Economic Analysis; https://www.bea.gov/data/consumer-spending/main, April 2, 2019.

Voinea, L., & Filip, A. (2011). Analyzing the main changes in new consumer buying behavior during economic crisis. *International Journal of Economic Practices and Theories, 1*(1), 14-19.

Vuchelen, J. (2004). Consumer sentiment and macroeconomic forecasts. *Journal of economic psychology*, *25*(4), 493-506.

Yu, Y., Zhu, Y., Li, S., & Wan, D. (2014). Time series outlier detection based on sliding window prediction. *Mathematical problems in Engineering*.

Zaki, M., & Neely, A. (2019). Customer Experience Analytics: Dynamic Customer-Centric Model. *Handbook of Service Science Volume II*, 207-233.

Zouaoui, M., Nouyrigat, G., & Beer, F. (2011). How does investor sentiment affect stock market crises? Evidence from panel data. *Financial Review, 46*(4), 723-747.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320.

Appendix A: Index questions and calculation

x1. "We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?"

x2. "Now looking ahead – do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?"

x3. "Now turning to business conditions in the country as a whole – do you think that during the next twelve months we'll have good times financially, or bad times, or what?"

x4. "Looking ahead, which would you say is more likely – that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?"

x5. "About the big things people buy for their homes – such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?"

To calculate the final index, the relative score plus 100 for the five questions (x1 to x5) is computed. This score is the percentage giving favourable replies minus the percentage giving unfavourable replies. The sum of the relative scores is divided by the 1966 base period total and a constant of 2 is added to correct for changes in the sample design. Until 1972 no constant was added. From 19724 until 1981 the constant was 2.7, and from 1981 to present the constant is 2.0. The final formula is:

$$Index = \frac{x1+x2+x3+x4+x5}{6.7558} + 2.0$$

Appendix B: visualization of the data

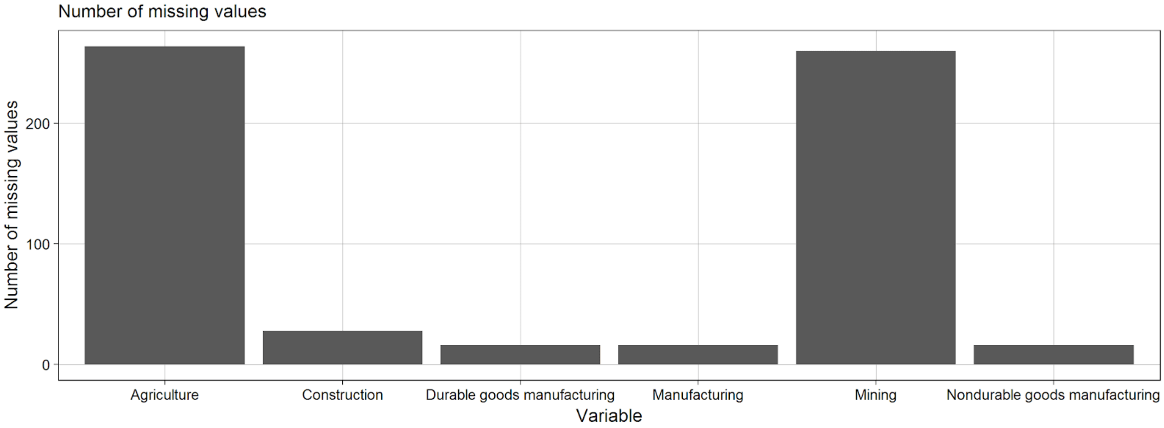

Figure 1 Appendix B: Number of missing values in the variables used for analyses

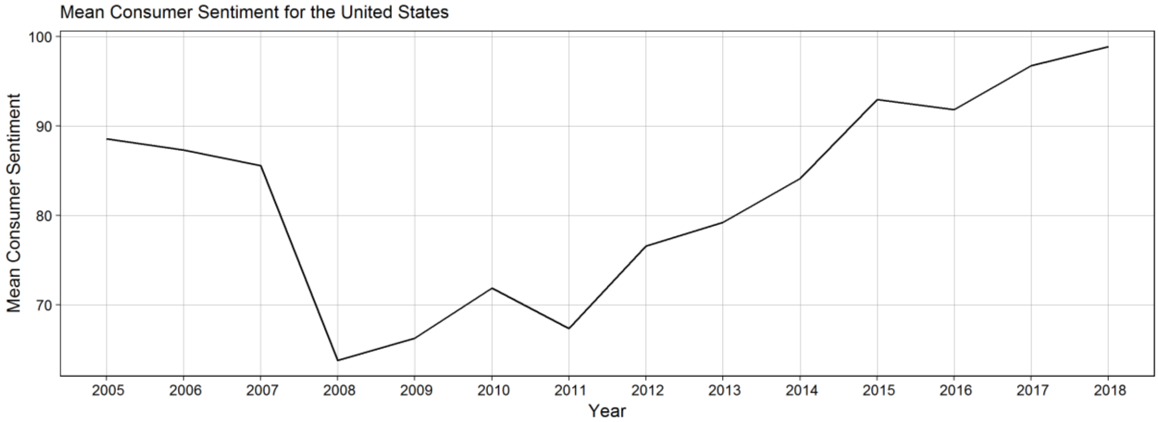

Figure 2 Appendix B: Mean consumer sentiment for the United States

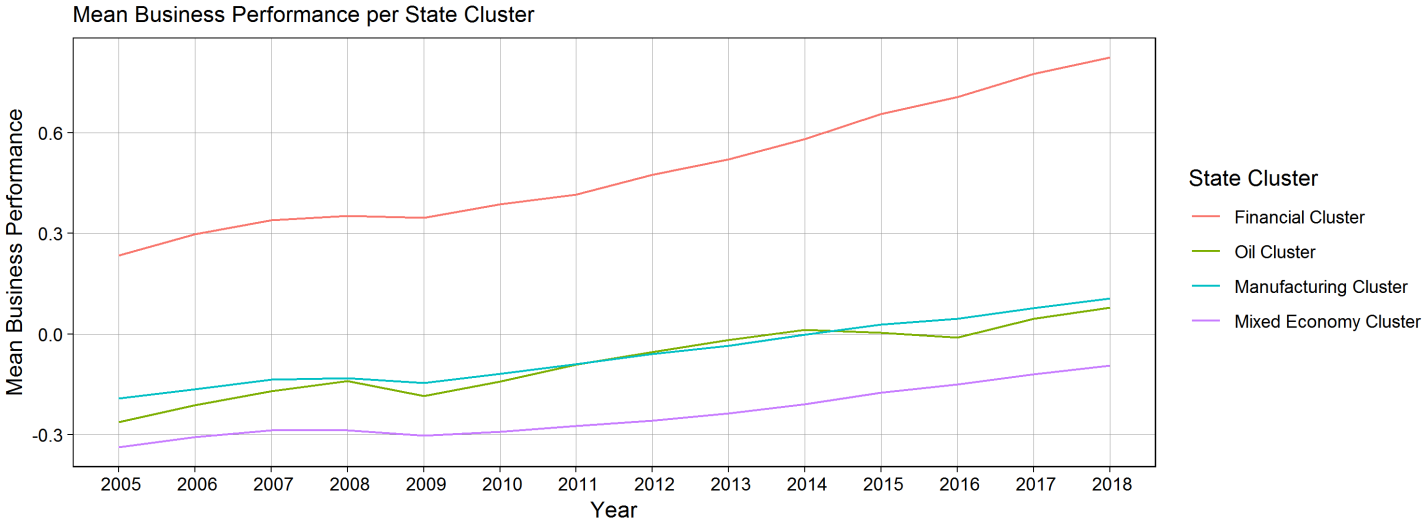

Figure 3 Appendix B: Mean business performance per state cluster

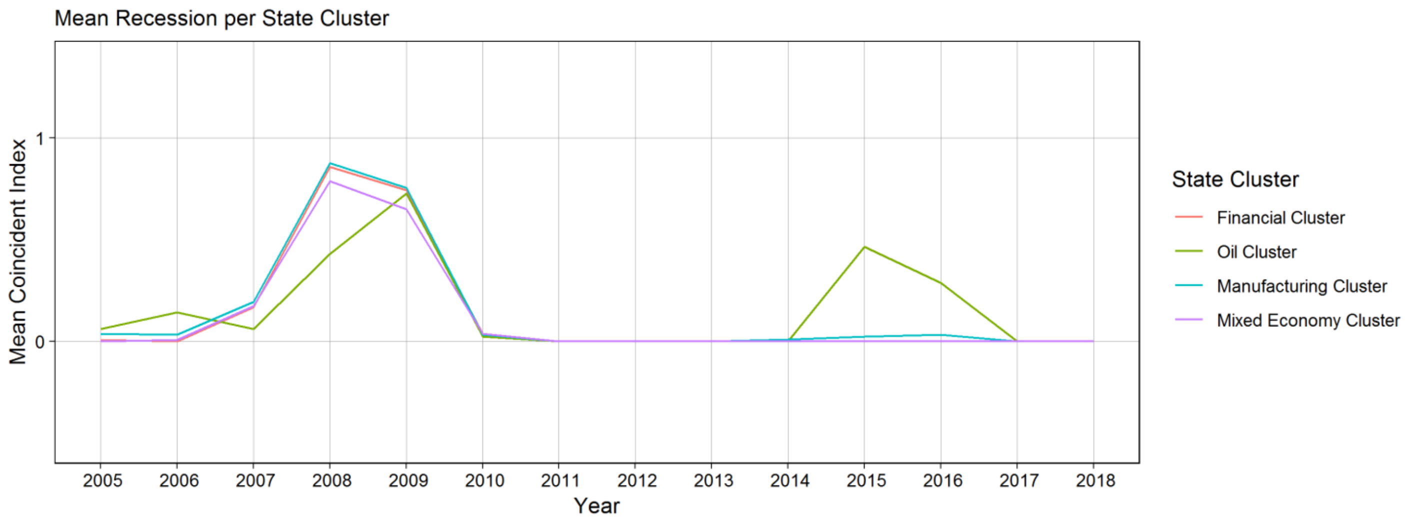

Figure 4 Appendix B: Mean recession per state cluster

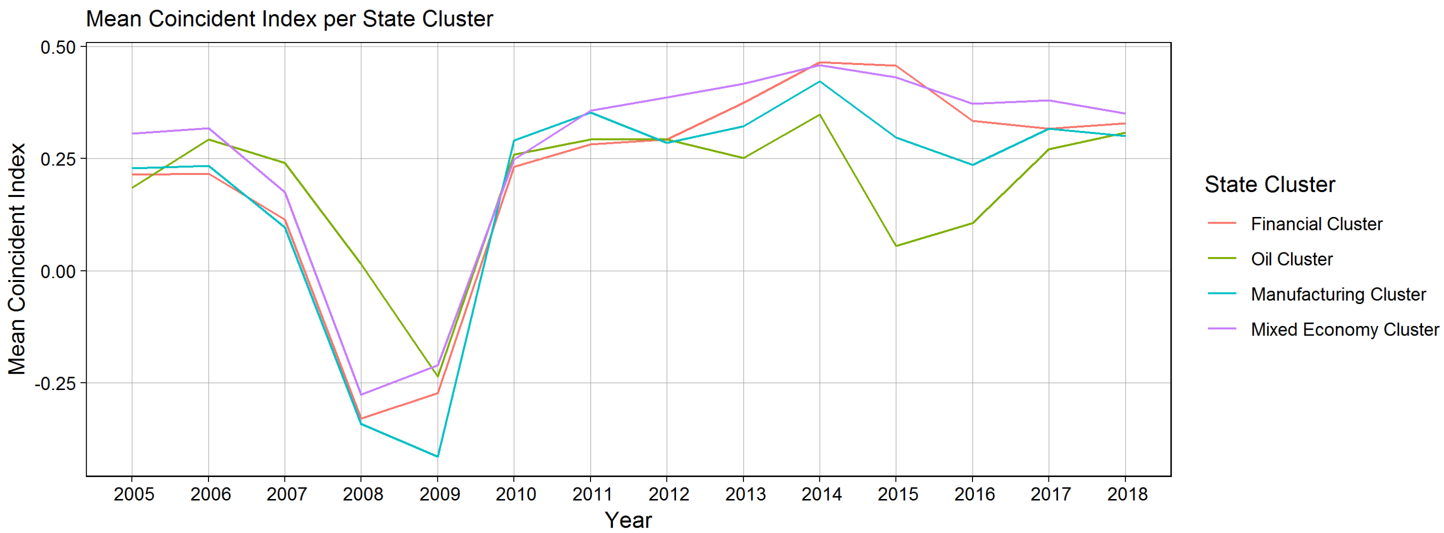

Figure 5 Appendix B: Mean coincident index per state cluster

Appendix C: Monte Carlo simulation for feature selection

To find the most relevant features for each research question composite scores were constructed using principal component analysis (PCA). PCA tries to identify the space in which the data points approximately lie (Jolliffe, 2011). It computes new variables called principle components which are obtained from linear combinations of the original features. By doing so, the goal of the PCA is to extract the most important features (Abdi & Williams, 2010). To find how many principal components should be computed a subset of the training data, which included the sliding windows and additional features for consumer sentiment and the value of Y_t, was used to compute the proportion of variance explained.

The principal components with the highest weights were then used in the Monte Carlo simulation. Monte Carlo is a simulation method that relies on repeated random sampling. The algorithm creates subsets of randomly chosen features and divides the objects in each subset in train and test sets (Komorowski, 2015). For each combination of features 10-fold cross validation was performed and the Mean Squared Errors (MSE) were computed on the test set. The combination of features with the lowest test MSE score were the features considered most relevant to each research question.

C1. MSE per feature combination

Experiment	**Target value**	**Features**	**MSE**
1 & 3.1	change$_{bpt+3;bpt+2}$	cs_{max}, cs_{min}, $cs_{median,}$ cs_{CQD}	1.228 E-4
2	rs_{t+3}	cs_{max}, cs_{min}, cs_{CQD}, cs_{t+1}	8.811 E-2
3.2	change$_{cit+3;cit+2}$	cs_{max}, cs_{min}, $cs_{CR,}$ ci_t	4.255 E-2
4.1	change$_{cst+3;cst+2}$	bp_{min}, bp_{CR}, bp_{CQD}, cs_t	17.577
4.2	change$_{cst+3;cst+2}$	rs_{max}, $rs_{min,}$ rs_t	17.637
4.3	change$_{cst+3;cst+2}$	$ci_{\bar{x}}$,, ci_{σ^2}, , ci_{t+2}, cs_t	17.456

C2. Predictions with feature selection

Experiment 1

Part	Model	MSE train	MSE test	Parameters
I	Baseline	2.721	8.675	-
	Elastic Net	1.230	3.754	*alpha = 0.2121425; lambda = 0.001205047*
	SVM	1.227	3.770	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 2591*
	Random Forest	0.993	3.440	*ntree = 5000; importance = TRUE*
II	Baseline	2.135	5.826	-
	Elastic Net	1.021	2.892	*alpha = 0.006356115; lambda = 0.006691759*
	SVM	0.983	2.830	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 316*
	Random Forest	0.809	2.680	*ntree = 5000; importance = TRUE*

Experiment 2

Part	Model	Acc. train	Acc. test	F1 train	F1 test	Parameters
I	Baseline	0.985	0.985	0.991	0.991	-
	SVM	0.902	0.895	0.944	0.940	*method = C-classification; kernel = radial; C = 1; gamma = 0.25; support vectors: 1302*
	PART	0.940	0.935	0.966	0.963	-
	Bagging	0.941	0.937	0.966	0.964	-
	RandomForest	0.941	0.938	0.966	0.964	*ntree = 5000; importance = TRUE*
	k-NN	0.940	0.938	0.965	0.964	*method = knn; trControl = cv (number: 5); k = 5*
II	Baseline	0.894	0.901	0.0217	-	-
	SVM	0.946	0.951	-	-	*method = C-classification; kernel = radial; C = 1; gamma = 0.25; support vectors: 105*
	PART	0.955	0.947	0.345	0.286	-
	Bagging	0.957	0.951	0.400	0.364	-
	RandomForest	0.957	0.951	0.400	0.364	*ntree = 5000; importance = TRUE*
	k-NN	0.953	0.947	0.286	0.211	*method = knn; trControl = cv (number: 5); k = 7*

Experiment 3.1

Cl.	Model	MSE train	MSE test	Parameters
1	Baseline	3.034	52.600	-
	Elastic Net	1.469	25.782	*alpha = 0.06905604; lambda = 0.00229073*
	SVM	1.476	26.220	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 616*
	Random Forest	1.202	23.097	*ntree = 5000; importance = TRUE*
2	Baseline	7.705	15.539	-
	Elastic Net	3.722	5.899	*alpha = 0.2426044; lambda = 6.790757*
	SVM	3.688	5.794	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 253*
	Random Forest	2.391	6.544	*ntree = 5000; importance = TRUE*
3	Baseline	0.967	0.977	-
	Elastic Net	0.444	0.437	*alpha = 0.1274343; lambda = 0.001274972*
	SVM	0.453	0.455	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 1285*
	Random Forest	0.226	0.307	*ntree = 5000; importance = TRUE*
4	Baseline	3.354	0.345	-
	Elastic Net	1.251	0.152	*alpha = 0.2827089; lambda = 0.002055782*
	SVM	1.236	0.149	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 532*
	Random Forest	0.953	0.185	*ntree = 5000; importance = TRUE*

Experiment 3.2

Cl.	Model	MSE train	MSE test	Parameters
1	Baseline	0.126	0.070	-
	Elastic Net	0.042	0.024	*alpha = 0.3073374; lambda = 0.00926526*
	SVM	0.041	0.023	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 1357*
	Random Forest	0.023	0.023	*ntree = 5000; importance = TRUE*
2	Baseline	0.144	0.134	-
	Elastic Net	0.057	0.046	*alpha = 0.8686318; lambda = 0.008117249*
	SVM	0.055	0.045	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 541*
	Random Forest	0.023	0.049	*ntree = 5000; importance = TRUE*
3	Baseline	0.124	0.187	-
	Elastic Net	0.046	0.070	*alpha = 0.4247356; lambda = 0.002565703*
	SVM	0.043	0.070	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 2188*
	Random Forest	0.025	0.072	*ntree = 5000; importance = TRUE*
4	Baseline	0.088	0.122	-
	Elastic Net	0.029	0.042	*alpha = 0.5303174; lambda = 0.007506053*
	SVM	0.029	0.042	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 1098*
	Random Forest	0.014	0.044	*ntree = 5000; importance = TRUE*

Experiment 4

Prt.	Model	MSE train	MSE test	Parameters
1	Baseline	35.284	35.291	-
	Elastic Net	17.567	17.614	*alpha = 0.1561852; lambda = 0.1073471*
	SVM	17.620	17.919	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 5185*
	Random Forest	4.578	10.578	*ntree = 5000; importance = TRUE*
2	Baseline	35.284	35.291	-
	Elastic Net	17.623	17.743	*alpha = 0.09788889; lambda = 0.005984826*
	SVM	17.660	17.795	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 5329*
	Random Forest	17.661	17.711	*ntree = 5000; importance = TRUE*
3	Baseline	35.284	35.291	-
	Elastic Net	17.432	17.526	*alpha = 0.2203477; lambda = 0.003436149*
	SVM	16.919	17.433	*method = eps-regression; kernel = radial; C = 1; gamma = 0.25; epsilon = 0.1; support vectors: 5235*
	Random Forest	2.742	11.486	*ntree = 5000; importance = TRUE*

Appendix D: States in state clusters

Cluster	Train set	Test set
1. Financial cluster	Connecticut, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, Vermont, Virginia	California
2. Oil cluster	Louisiana, North Dakota, Oklahoma, Texas	Alaska, Wyoming, New Mexico
3. Manufacturing cluster	Alabama, Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Mississippi, Missouri, Ohio, Pennsylvania, South Carolina, Tennessee, West Virginia, Wisconsin	Washington, Montana
4. Mixed economy cluster	Arkansas, Delaware, Florida, Georgia, Hawaii, Nebraska, North Carolina, South Dakota	Oregon, Idaho, Colorado, Nevada, Arizona, Utah

Appendix E: Multiple Imputation

Variable	Imputation method
sixmonthsout	Logistic regression
Oil price	Bayesian linear regression
Oil state	PMM
Agriculture	PMM
Mining	PMM
Construction	PMM
Manifacutring	PMM
Durable goods	PMM
Nondurable goods	PMM
Current coincident index	PMM

Appendix F: Results of experiment 1

Experiment 1.1

RQ1.1	
Feature	**Coefficient**
Intercept	0.00230
cs_t	0.00018
cs_{t+1}	0.00006
cs_{t+2}	0.00004
cs_{max}	0.00006
cs_{min}	0.00001
cs_{σ^2}	0.00092
$cs_{\bar{x}}$	0.00011
cs_{median}	0.00023
cs_{CR}	-0.00019
cs_{CQD}	-0.00029
bp_t	0.00229

Elastic Net coefficients RQ1.1

RQ1.1	
Feature	**Importance**
cs_t	7.460
cs_{t+1}	2.219
cs_{t+2}	1.492
cs_{max}	2.403
cs_{min}	-
cs_{σ^2}	40.009
$cs_{\bar{x}}$	4.208
cs_{median}	9.682
cs_{CR}	7.670
cs_{CQD}	12.276
bp_t	100.000

Elastic Net feature importance RQ1.1

RQ1.1	
Feature	**Importance**
cs_t	71.211
cs_{t+1}	81.072
cs_{t+2}	79.080
cs_{max}	65.682
cs_{min}	67.426
cs_{σ^2}	98.415
$cs_{\bar{x}}$	73.329
cs_{median}	73.616
cs_{CR}	85.855
cs_{CQD}	89.923
bp_t	70.350

Random Forest feature importance RQ1.1

Experiment 1.2

RQ1.2	
Feature	**Coefficient**
Intercept	-0.00012
cs_t	-
cs_{t+1}	-
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	-
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
bp_t	-

Elastic Net coefficients RQ1.2

RQ1.2	
Feature	**Importance**
cs_t	-
cs_{t+1}	-
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	-
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
bp_t	-

Elastic Net feature importance RQ1.2

RQ1.2	
Feature	**Importance**
cs_t	30.655
cs_{t+1}	20.624
cs_{t+2}	46.361
cs_{max}	21.741
cs_{min}	31.887
cs_{σ^2}	28.235
$cs_{\bar{x}}$	22.810
cs_{median}	23.275
cs_{CR}	25.958
cs_{CQD}	28.758
bp_t	-6.333

Random Forest feature importance RQ1.2

Appendix G: Results of experiment 2

Experiment 2.1

RQ2.1	
Feature	**Importance**
cs_t	0
cs_{t+1}	0
cs_{t+2}	0
cs_{max}	0
cs_{min}	1
cs_{σ^2}	0
$cs_{\bar{x}}$	0
cs_{median}	1
cs_{CR}	1
cs_{CQD}	0
rst	1

582PART feature importance RQ2.1

RQ2.1	
Feature	**Importance**
cs_t	14.733
cs_{t+1}	17.830
cs_{t+2}	17.781
cs_{max}	547.502
cs_{min}	571.022
cs_{σ^2}	17.582
$cs_{\bar{x}}$	532.179
cs_{median}	476.344
cs_{CR}	18.150
cs_{CQD}	22.460
rst	1277.657

Bagging feature importance RQ2.1

RQ2.1	
Feature	**Importance**
cs_t	28.899
cs_{t+1}	29.911
cs_{t+2}	28.133
cs_{max}	107.108
cs_{min}	138.206
cs_{σ^2}	18.001
$cs_{\bar{x}}$	88.650
cs_{median}	59.538
cs_{CR}	20.517
cs_{CQD}	19.236
rst	773.681

Random Forest feature importance RQ2.1

RQ2.1	
Feature	**Importance**
cs_t	56.60
cs_{t+1}	59.05
cs_{t+2}	60.53
cs_{max}	61.21
cs_{min}	61.89
cs_{σ^2}	-
$cs_{\bar{x}}$	62.54
cs_{median}	62.13
cs_{CR}	27.07
cs_{CQD}	27.43
rst	100.00

k-NN feature importance RQ2.1

Experiment 2.2

RQ2.2	
Feature	**Importance**
cs_t	0
cs_{t+1}	0
cs_{t+2}	0
cs_{max}	0
cs_{min}	1
cs_{σ^2}	0
$cs_{\bar{x}}$	0
cs_{median}	1
cs_{CR}	1
cs_{CQD}	0
rst	0

PART feature importance RQ2.2

RQ2.2	
Feature	**Importance**
cs_t	11.020
cs_{t+1}	12.405
cs_{t+2}	9.641
cs_{max}	16.998
cs_{min}	16.407
cs_{σ^2}	15.325
$cs_{\bar{x}}$	7.106
cs_{median}	5.312
cs_{CR}	18.763
cs_{CQD}	17.881
rst	-

Bagging feature importance RQ2.2

RQ2.2	
Feature	**Importance**
cs_t	2.389
cs_{t+1}	2.677
cs_{t+2}	2.320
cs_{max}	2.548
cs_{min}	2.948
cs_{σ^2}	5.193
$cs_{\bar{x}}$	2.647
cs_{median}	2.839
cs_{CR}	5.472
cs_{CQD}	5.569
rst	-

Random Forest feature importance RQ2.2

RQ2.2	
Feature	**Importance**
cs_t	82.26
cs_{t+1}	78.36
cs_{t+2}	80.73
cs_{max}	85.33
cs_{min}	93.34
cs_{σ^2}	79.34
$cs_{\bar{x}}$	87.90
cs_{median}	85.19
cs_{CR}	99.08
cs_{CQD}	100.00
rst	-

k-NN feature importance RQ2.2

Appendix H: Results of experiment 3

Experiment 3.1

RQ3.1	
Feature	**Coefficient**
Intercept	0.00248
cs_t	-
cs_{t+1}	-
cs_{t+2}	0.00042
cs_{max}	-
cs_{min}	-
cs_{σ^2}	0.00020
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
bp_t	0.00256

Elastic Net coefficients RQ3.1 - Cluster 1

RQ3.1	
Feature	**Importance**
cs_t	-
cs_{t+1}	-
cs_{t+2}	16.475
cs_{max}	-
cs_{min}	-
cs_{σ^2}	7.922
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
bp_t	100.000

Elastic Net feature importance RQ3.1 - Cluster 1

RQ3.1	
Feature	**Importance**
cs_t	39.455
cs_{t+1}	51.794
cs_{t+2}	53.928
cs_{max}	41.509
cs_{min}	39.952
cs_{σ^2}	44.923
$cs_{\bar{x}}$	44.720
cs_{median}	43.813
cs_{CR}	42.749
cs_{CQD}	45.261
bp_t	52.540

Random Forest feature importance RQ3.1 - Cluster 1

RQ3.1	
Feature	**Coefficient**
Intercept	0.00388
cs_t	0.00033
cs_{t+1}	0.00028
cs_{t+2}	-
cs_{max}	0.00023
cs_{min}	-
cs_{σ^2}	0.00232
$cs_{\bar{x}}$	0.00015
cs_{median}	0.00036
cs_{CR}	-0.00022
cs_{CQD}	-0.00037
bp_t	0.00443

Elastic Net coefficients RQ3.1 - Cluster 2

RQ3.1	
Feature	**Importance**
cs_t	7.463
cs_{t+1}	6.301
cs_{t+2}	-
cs_{max}	5.215
cs_{min}	-
cs_{σ^2}	52.475
$cs_{\bar{x}}$	3.374
cs_{median}	8.107
cs_{CR}	5.114
cs_{CQD}	8.248
bp_t	100.000

Elastic Net feature importance RQ3.1 - Cluster 2

RQ3.1	
Feature	**Importance**
cs_t	36.917
cs_{t+1}	32.903
cs_{t+2}	36.041
cs_{max}	36.177
cs_{min}	34.714
cs_{σ^2}	34.773
$cs_{\bar{x}}$	41.640
cs_{median}	34.401
cs_{CR}	35.380
cs_{CQD}	36.010
bp_t	3.209

Random Forest feature importance RQ3.1 - Cluster 2

RQ3.1	
Feature	**Coefficient**
Intercept	0.00201
cs_t	0.00014
cs_{t+1}	0.00009
cs_{t+2}	-
cs_{max}	0.00008
cs_{min}	0.00005
cs_{σ^2}	0.00004
$cs_{\bar{x}}$	0.00008
cs_{median}	0.00010
cs_{CR}	-
cs_{CQD}	-
bp_t	0.00056

Elastic Net coefficients RQ3.1 - Cluster 3

RQ3.1	
Feature	**Importance**
cs_t	24.026
cs_{t+1}	16.636
cs_{t+2}	-
cs_{max}	14.143
cs_{min}	9.507
cs_{σ^2}	6.811
$cs_{\bar{x}}$	14.405
cs_{median}	18.528
cs_{CR}	-
cs_{CQD}	-
bp_t	100.000

Elastic Net feature importance RQ3.1 - Cluster 3

RQ3.1	
Feature	**Importance**
cs_t	89.817
cs_{t+1}	86.980
cs_{t+2}	88.329
cs_{max}	72.722
cs_{min}	80.380
cs_{σ^2}	120.125
$cs_{\bar{x}}$	83.692
cs_{median}	80.356
cs_{CR}	113.053
cs_{CQD}	117.660
bp_t	117.701

Random Forest feature importance RQ3.1 - Cluster 3

RQ3.1	
Feature	**Coefficient**
Intercept	0.00187
cs_t	0.00016
cs_{t+1}	0.00019
cs_{t+2}	0.00003
cs_{max}	0.00015
cs_{min}	0.00002
cs_{σ^2}	0.00059
$cs_{\bar{x}}$	0.00014
cs_{median}	0.00027
cs_{CR}	-
cs_{CQD}	-
bp_t	0.00147

Elastic Net coefficients RQ3.1 - Cluster 4

RQ3.1	
Feature	**Importance**
cs_t	10.913
cs_{t+1}	13.026
cs_{t+2}	2.117
cs_{max}	9.991
cs_{min}	1.079
cs_{σ^2}	40.204
$cs_{\bar{x}}$	9.674
cs_{median}	18.041
cs_{CR}	-
cs_{CQD}	-
bp_t	100.000

Elastic Net feature importance RQ3.1 - Cluster 4

RQ3.1	
Feature	**Importance**
cs_t	25.177
cs_{t+1}	32.803
cs_{t+2}	30.940
cs_{max}	22.138
cs_{min}	28.151
cs_{σ^2}	36.063
$cs_{\bar{x}}$	22.662
cs_{median}	27.743
cs_{CR}	30.797
cs_{CQD}	33.887
bp_t	9.663

Random Forest feature importance RQ3.1 - Cluster 4

Experiment 3.2

RQ3.2

Feature	Coefficient
Intercept	0.00126
cs_t	-
cs_{t+1}	-
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	-
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
ci_t	-0.01594

Elastic Net coefficients RQ3.2 - Cluster 1

RQ3.2

Feature	Importance
cs_t	-
cs_{t+1}	-
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	-
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
ci_t	100.000

Elastic Net feature importance RQ3.2 - Cluster 1

RQ3.2

Feature	Importance
cs_t	39.896
cs_{t+1}	36.362
cs_{t+2}	46.558
cs_{max}	36.385
cs_{min}	37.466
cs_{σ^2}	37.892
$cs_{\bar{x}}$	40.229
cs_{median}	39.030
cs_{CR}	36.579
cs_{CQD}	34.459
ci_t	58.357

Random Forest feature importance RQ3.2 - Cluster 1

RQ3.2

Feature	Coefficient
Intercept	-0.00041
cs_t	-
cs_{t+1}	-0.04367
cs_{t+2}	0.03794
cs_{max}	0.00192
cs_{min}	0.06428
cs_{σ^2}	0.05880
$cs_{\bar{x}}$	-
cs_{median}	-0.07384
cs_{CR}	-0.04943
cs_{CQD}	-
ci_t	-0.03451

Elastic Net coefficients RQ3.2 - Cluster 2

RQ3.2

Feature	Importance
cs_t	-
cs_{t+1}	59.145
cs_{t+2}	51.377
cs_{max}	2.594
cs_{min}	87.059
cs_{σ^2}	79.630
$cs_{\bar{x}}$	-
cs_{median}	100.000
cs_{CR}	66.944
cs_{CQD}	-
ci_t	46.740

Elastic Net feature importance RQ3.2 - Cluster 2

RQ3.2

Feature	Importance
cs_t	24.080
cs_{t+1}	23.418
cs_{t+2}	21.541
cs_{max}	19.196
cs_{min}	18.977
cs_{σ^2}	22.545
$cs_{\bar{x}}$	22.346
cs_{median}	22.986
cs_{CR}	22.543
cs_{CQD}	23.462
ci_t	19.543

Random Forest feature importance RQ3.2 - Cluster 2

RQ3.2

Feature	Coefficient
Intercept	0.00018
cs_t	-0.00203
cs_{t+1}	-0.00104
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	0.05051
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-0.05214
ci_t	-0.04133

Elastic Net coefficients RQ3.2 - Cluster 3

RQ3.2

Feature	Importance
cs_t	3.895
cs_{t+1}	1.994
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	96.871
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	100.000
ci_t	79.270

Elastic Net feature importance RQ3.2 - Cluster 3

RQ3.2

Feature	Importance
cs_t	62.331
cs_{t+1}	53.597
cs_{t+2}	61.540
cs_{max}	62.584
cs_{min}	58.320
cs_{σ^2}	42.867
$cs_{\bar{x}}$	63.918
cs_{median}	60.370
cs_{CR}	52.421
cs_{CQD}	52.225
ci_t	131.339

Random Forest feature importance RQ3.2 - Cluster 3

RQ3.2

Feature	Coefficient
Intercept	0.00060
cs_t	-
cs_{t+1}	0.00265
cs_{t+2}	-
cs_{max}	-
cs_{min}	-
cs_{σ^2}	-
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
ci_t	-0.01978

Elastic Net coefficients RQ3.2 - Cluster 4

RQ3.2

Feature	Importance
cs_t	-
cs_{t+1}	-
cs_{t+2}	13.37
cs_{max}	-
cs_{min}	-
cs_{σ^2}	-
$cs_{\bar{x}}$	-
cs_{median}	-
cs_{CR}	-
cs_{CQD}	-
ci_t	100.000

Elastic Net feature importance RQ3.2 - Cluster 4

RQ3.2

Feature	Importance
cs_t	55.299
cs_{t+1}	57.078
cs_{t+2}	46.606
cs_{max}	49.063
cs_{min}	52.805
cs_{σ^2}	50.079
$cs_{\bar{x}}$	52.304
cs_{median}	56.769
cs_{CR}	48.915
cs_{CQD}	48.120
ci_t	53.226

Random Forest feature importance RQ3.2 - Cluster 4

Appendix I: Experiment 3.1 with 70/30 partitioning

Experiment 3.1 (10^{-4})

Cl.	Model	MSE train		MSE test		Parameters
1.	Baseline	6.382	-	8.053	-	-
	Elastic Net	2.820	(55.81%)	4.382	(45.56%)	*alpha* = 0.2954246; *lambda* = 0.001006681
	SVM	3.226	(49.45%)	4.572	(43.23%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 468*
	Random Forest	0.890	(86.05%)	0.433	(94.62%)	*ntree* = 5000; *importance* = TRUE
2.	Baseline	15.766	-	4.126	-	-
	Elastic Net	5.470	(65.31%)	1.928	(53.27%)	*alpha* = 0.1503563; *lambda* = 0.001694307
	SVM	5.524	(64.96%)	2.065	(49.95%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 233*
	Random Forest	2.472	(84.32%)	2.915	(29.35%)	*ntree* = 5000; *importance* = TRUE
3.	Baseline	0.963	-	0.846	-	-
	Elastic Net	0.433	(55.04%)	0.405	(52.13%)	*alpha* = 0.02349696; *lambda* = 0.001108925
	SVM	0.456	(52.65%)	0.431	(49.05%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1001*
	Random Forest	0.112	(88.37%)	0.201	(76.24%)	*ntree* = 5000; *importance* = TRUE
4.	Baseline	2.843	-	0.696	-	-
	Elastic Net	0.920	(67.64%)	0.320	(54.02%)	*alpha* = 0.1480629; *lambda* = 0.001153624
	SVM	0.928	(67.36%)	0.329	(52.73%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 731*
	Random Forest	0.479	(83.15%)	0.331	(52.44%)	*ntree* = 5000; *importance* = TRUE

Appendix J: Experiment 3.2 with 70/30 partitioning

Experiment 3.2

Cl.	**Model**	**MSE train**		**MSE test**		**Parameters**
1.	Baseline	0.113	-	0.095	-	-
	Elastic Net	0.041	(63.72%)	0.040	(57.89%)	*alpha* = 0.03792504; *lambda* = 0.009532534
	SVM	0.040	(64.60%)	0.040	(57.89%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1063*
	Random Forest	0.021	(81.41%)	0.043	(54.74%)	*ntree* = 5000; *importance* = TRUE
2.	Baseline	0.119	-	0.159	-	-
	Elastic Net	0.046	(61.34%)	0.069	(56.60%)	*alpha* = 0.8567382; *lambda* = 0.008085317
	SVM	0.044	(63.03%)	0.070	(55.97%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 694*
	Random Forest	0.020	(83.19%)	0.074	(53.46%)	*ntree* = 5000; *importance* = TRUE
3.	Baseline	0.124	-	0.095	-	-
	Elastic Net	0.050	(59.68%)	0.044	(53.68%)	*alpha* = 0.07239164; *lambda* = 0.003337746
	SVM	0.047	(62.10%)	0.043	(54.74%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1684*
	Random Forest	0.026	(79.03%)	0.044	(53.68%)	*ntree* = 5000; *importance* = TRUE
4.	Baseline	0.090	-	0.076	-	-
	Elastic Net	0.034	(62.22%)	0.035	(53.95%)	*alpha* = 0.01827431; *lambda* = 0.006731816
	SVM	0.033	(63.33%)	0.035	(53.95%)	*method = eps-regression; kernel = radial; C = 1; gamma = 0.09090909; epsilon = 0.1; support vectors: 1306*
	Random Forest	0.019	(78.89%)	0.037	(51.32%)	*ntree* = 5000; *importance* = TRUE

Appendix K: Results of experiment 4

Experiment 4.1

RQ4.1

Feature	Coefficient
Intercept	0.03949
bp_t	-
bp_{t+1}	0.00289
bp_{t+2}	0.00226
bp_{max}	-
bp_{min}	-
bp_{σ^2}	-
$bp_{\bar{x}}$	-
bp_{median}	-
bp_{CR}	-0.05817
bp_{CQD}	0.00692
cs_t	-0.30534

Elastic Net coefficients RQ4.1

RQ4.1

Feature	Importance
bp_t	-
bp_{t+1}	0.948
bp_{t+2}	0.741
bp_{max}	-
bp_{min}	-
bp_{σ^2}	-
$bp_{\bar{x}}$	-
bp_{median}	-
bp_{CR}	19.050
bp_{CQD}	2.268
cs_t	100.000

Elastic Net feature importance RQ4.1

RQ4.1

Feature	Importance
bp_t	93.033
bp_{t+1}	93.779
bp_{t+2}	90.071
bp_{max}	87.340
bp_{min}	87.703
bp_{σ^2}	142.421
$bp_{\bar{x}}$	93.767
bp_{median}	94.672
bp_{CR}	102.711
bp_{CQD}	109.283
cs_t	258.137

Random Forest feature importance RQ4.1

Experiment 4.2

RQ4.2

Feature	Coefficient
Intercept	0.03949
rs_t	0.35703
rs_{t+1}	0.00248
rs_{t+2}	-0.54901
rs_{max}	-
rs_{min}	-
rs_{σ^2}	-0.10267
$rs_{\bar{x}}$	-
rs_{median}	0.01632
rs_{CR}	-0.06562
rs_{CQD}	-
cs_t	-0.45370

Elastic Net coefficients RQ4.2

RQ4.2

Feature	Importance
rs_t	65.032
rs_{t+1}	0.452
rs_{t+2}	100.000
rs_{max}	-
rs_{min}	-
rs_{σ^2}	18.699
$rs_{\bar{x}}$	-
rs_{median}	2.973
rs_{CR}	11.952
rs_{CQD}	-
cs_t	82.639

Elastic Net feature importance RQ4.2

RQ4.2

Feature	Importance
rs_t	35.366
rs_{t+1}	14.968
rs_{t+2}	39.025
rs_{max}	25.026
rs_{min}	18.867
rs_{σ^2}	34.862
$rs_{\bar{x}}$	28.787
rs_{median}	14.901
rs_{CR}	34.796
rs_{CQD}	24.016
cs_t	77.292

Random Forest feature importance RQ4.2

Experiment 4.3

RQ4.3	
Feature	**Coefficient**
Intercept	0.03949
ci_t	-0.33595
ci_{t+1}	-0.16706
ci_{t+2}	0.58840
ci_{max}	-0.00005
ci_{min}	0.01761
ci_{σ^2}	0.13447
$ci_{\bar{x}}$	-
ci_{median}	-0.01800
ci_{CR}	-0.02495
ci_{CQD}	0.01279
cs_t	-0.38059

Elastic Net coefficients RQ4.3

RQ4.3	
Feature	**Importance**
ci_t	57.096
ci_{t+1}	28.393
ci_{t+2}	100.000
ci_{max}	0.0084
ci_{min}	2.992
ci_{σ^2}	22.854
$ci_{\bar{x}}$	-
ci_{median}	3.059
ci_{CR}	4.241
ci_{CQD}	2.174
cs_t	64.682

Elastic Net feature importance RQ4.3

RQ4.3	
Feature	**Importance**
ci_t	116.425
ci_{t+1}	107.359
ci_{t+2}	115.332
ci_{max}	99.967
ci_{min}	111.285
ci_{σ^2}	108.662
$ci_{\bar{x}}$	100.335
ci_{median}	107.319
ci_{CR}	105.841
ci_{CQD}	103.137
cs_t	220.702

Random Forest feature importance RQ4.3

Appendix L: Sliding window approach

	Window 1	Window 2	Window 3	Window 4	Window 5
Year	2005 to 2018	2005 to 2018	2005 to 2018	2005 to 2017	2005 to 2017
Month 1	January	February	March	April	May
Month 2	February	March	April	May	June
Month 3	March	April	May	June	July
	Window 6	Window 7	Window 8	Window 9	Window 10
Year	2005 to 2017	2005 to 2017	2005 to 2017	2005 to 2017	2005 to 2017
Month 1	June	July	August	September	October
Month 2	July	August	September	October	November
Month 3	August	September	October	November	December

Appendix M – Software and packages

This appendix provides a global overview of the packages that were used in the experimental procedure. All analyses and experiments were implemented using R Studio.

mice. The "mice" package is used to perform multiple imputation. The *mice::quickpred()* function is used to quick select predictors from the data. The *mincor* parameter that specifies the minimum threshold is set to 0.25. The *mice::mice* is used to replace the missing values. The parameter of the number of imputations *m* is set to 1 with the number of iterations *maxit* set to 1 too. The seed is set to '314159'. *Mice::complete* extracts the subset of complete cases. **TSPred.** "TSPred" is used for the sliding window method. *TSPred::slidingWindows* extracts all possible subsequences of a time series. The parameter *swSize* is set to 3. **zoo.** The "zoo" package is used for the extraction of features from the sliding window data. With *zoo::rollapply* the functions for the construction of the features is applied to rolling margins of the data. The parameter *width* is set to 3. **stats.** The "stats" package is used to perform PCA with the use of *stats::prcomp*. This function performs a principal components analysis on the data and returns the weights of the components. **glmnet.** "glmnet" is used to create an OLS model for the Monte Carlo simulation. The parameter *intercept* is set to TRUE, the parameters *alpha* and *lambda* to 0 and *standardize* to FALSE. **e1071.** The package "e1071" is used for training SVM. The parameters of *e1071::svm* are presented in the results section. **RWeka.** *RWeka::PART* is used for the PART algorithm. **ipred.** *ipred::bagging* is used to implement the bagging classification model. **randomForest.** The "randomForest" package is used for the implementation of the Random Forest algorithm. The parameters of the function *randomForest::randomForest* are listed in the results section. **caret.** "caret" is used to fit Elastic Net and k-NN to the data. *caret::confusionMatrix* is used for calculating cross-tabulations of the observed and predicted classes. *caret::createDataPartition()* is used to partition the data for experiment 3. **ggplot2, tidyr.** Packages used for creating visualizations of the data.