# Machine detection of persisting pragmatic linguistic relationships in Monetary Policy and Financial Economics

**Eric A. Daimler**

December, 2013
CMU-ISR-13-109

School of Computer Science
Carnegie Mellon University
Pittsburgh, Pennsylvania

**Thesis Committee**
*Kathleen M. Carley, Carnegie Mellon University (co-Chair)*
*James H. Morris, Carnegie Mellon University (co-Chair)*
*Carolyn P. Rosé, Carnegie Mellon University*
*James Burnham, Duquesne University*

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the field of Computational Economics

**Keywords**: Dynamic Network Analysis, Social network analysis, Computational Linguistics, Learning Algorithms, Enron, Federal Reserve, Computational Economics

# Abstract

Humans speak to each other in a variety of mediums about their expectations for the future. These conversations have various degrees of preparation, formality, and impact. A central banker's speech may be listened to more carefully than an intra-company email, but both are efforts to set expectations of the future. They all contain biases. Combining recent developments in Network Science, Computational Linguistics, and Machine Learning enables new efforts to measure the impact of human-generated text. Measuring the bias may help to reduce it.

This work considers a new multi-step framework for the analysis of text. The efficacy is explored in the domains of public policy (the monetary policy of central banking) and corporate communications (the equity price of a publicly-traded firm) using machine-enhanced semantic network analysis. The implication of this study may view these techniques through different lenses of information use: central banks, corporate treasury, and investors. In supplying a set of reliable quantitative measurements to previously qualitative information, this study may help to improve both communication and the biases in its interpretation. In studying these issues using different communication modes and contexts, I hope to contribute to a broad analysis of communication concerns.

Classical approaches measure sentiment of these texts most often as bimodal (good/bad, increasing/decreasing, etc.). However, it is in decision making that more information is needed and reliable nuanced analysis become useful. In this study, I present approaches in computer science that address these challenges by explicitly testing the circumstances under which quantitative to qualitative relationships occur in the domain of finance and economics. The approach takes as input qualitative data from various sources in addition to quantitative data in the form of financial data. I develop a meta algorithm for measuring and testing the relationship which helps to identify a causal relationship among different data sets in different circumstances. The approach leads to insights on price movement (asset valuation) for the purposes of public policy but also for corporate management in the domain of the treasury function.

The approach I develop may support the assessment and estimation of financial decision processes in many circumstances.

This range suggests an ability to generalize beyond financial decision-making. I start with qualitative data (text data) in various contexts that are then cleaned of extraneous markings such as date, location, and original distribution location (email, speech, etc.). Second, I use a sequence of steps in Dynamic Network Analysis to extract a semantic network that will be used as the quantitative structure for the best comparison with other quantitative data. Third, I collect appropriate quantitative data external to the text against which to compare the semantic network results. Fourth, I use learning algorithms to identify the degree to which a relationship can be found between the extracted semantic network analysis and the external quantitative data. This trimmed structure should allow for further development in future work of a predictive framework in financial decisions.

Text analysis of even the most basic kind has shown to be beneficial, but new approaches are needed. More adaptive systems, where an intelligent system assesses the text as it occurs and provides feedback when necessary, is a promising area of research that can help provide scaffolding for the interpretation of communication. Little is known about how to build these systems and what effects they might have on our collaboration and learning. In this dissertation, I augment existing semantic analysis systems with a more sophisticated analysis and then design, build, and evaluate a more powerful framework.

*To Melissa Daimler, for whom at least a start may be made toward the cliché ‘words do not express my love’ with the 75,746 contained here*

# Acknolwedgements

To turn a phrase from one of the domains under analysis in this Dissertation: *It takes a village to earn Ph.D.* While this research is completely my own, I have found the whole process of completing a dissertation toward a Ph.D. to be far from my early misconception of an activity mostly taking place in isolation from others. Having only this one experience of pursuing the degree to lean on, this path is warped beyond recognition if not for my advisors, colleagues, friends, and family.

First, my committee deserves acknowledgement for participating in a fundamentally volunteer (read: completely optional) activity. They must commit time and energy to engaging with the material that I have produced. Through their effort, they are making a contribution to the next generation of researchers. The images that I have held of my writing always being an engaging read have undoubtedly been fantasy. I am nonetheless committed to continual improvement. For the sake of my committee, and especially my advisors, let the eighteenth draft of this document stand as an apology for the first draft.

I have a large debt to my advisors, Jim Morris and Kathleen Carley. Dr. Morris has stood by me from beginning to end. Jim's open and honest communication is something from which we can all learn. Dr. Carley is the Franz Liszt of academic research in this domain. So good is she at sight-reading long and complex documents that while others are re-reading, she is already ad-libbing off the foundation. If I applied her insights as fast as she offered them, I would have been done a long time ago.

To my colleagues within the University, the School of Computer Science, the department and the program: if not for your example, your feedback, and your friendship this would not have been possible. Edo Airoldi, Michael Ashworth, Virginia Bedford, Michael Benisch, Marcelo Cataldo, George Davis, Jana Diesner, Serge Egelman, Ian Fette, Terrill Frantz, Luke Gerdes, Andrew Gilpin, Brian Hirshman, Laurie Jones, Peter Landwehr, Jamie Olson, Ralph Gross, Patrick Kelly, Brad Malin, Ian McCulloh, Il-Chul Moon, Craig Schreiber, Boris Sofman, Jessie St. Charles, and Patrick Wagstrom. My life is richer for our interactions, big and small.

I have been fortunate to see first-hand the strong character that Ian McCulloh demonstrates in service to our Country; I am forever grateful to the wisdom and friendship that he shared with me. Connie Herold, thank you for facilitating the process. You helped to keep me on track as you do daily for all of us.

To the many other faculty with whom I had the pleasure of interacting outside of my immediate program, I have been helped immensely in bringing down the intial hand-waving to something tractable with rigorous research. Ray Bareiss, Joy Beatty, Lenore Blum, Tina Carr, Lorrie Cranor, James Fadiman, David Farber, Marvin Goodfriend, Martin Griss, Isa Gucciardi, James Herbsleb, Ajay Jain, Thupten Jinpa, Philip Lehman, Allan Meltzer, Norman Sadeh, Roger Schank, Mary Shaw, Herb Simon, Eswaran Subrahmanian, Laurence Summers, and Luis Von Ahn. Thank you.

Outside of my Academic setting, I benefited from many people giving me a perspective that dramatically impacts how I think about a research direction. I owe a special debt to Mark Arnold and Kevin Warsh with whom I had early conversations about generating Monetary Policy within US Federal Reserve and specifically the dynamics of the voting members of the FOMC. These discussions unknowingly seeded my ideas for the research direction represented here. I also benefits from multiple conversations with others: David Allen, Rahul Arora, Michael Barnes, Richard Berman, Jonathan Betz, Sabeer Bhatia, Jeff Bezos, Drew Buechley, Phil Buonadonna, Eric Chiu, Dan Cohen, Joy Covey, Tom Dyal, Bill Ford, Ed Frank, Sameer Gandhi, Jon Goldick, Don Head, Arif Inayatullah, Ray Lane, Gerald Leitner, Chris Maeda, Samir Master, Pranab Nag, Eric O'Brien, Udai Puramsetti, Nader Rahimizad, Scott Sandell, Mark Siegel, Boris Sofman, Jeff Skoll, and Daniel Strachman. I have benefited from your insights, your inspiration, and your encouragement.

Thank you to Bill Bland, Mary Bland, Cindy Corrigan, Debbie Sandona, Pat Wahlton, and Ray Wahlton for their continual encouragement. Ed Bland: sorry you missed this. Melissa Daimler: I am sure that many times you wished that you had.

# Table of Contents

# List of Figures

# List of Equations

# List of Tables

# 1 Introduction

## 1.1 Importance

### 1.1.1 Assumptions

We assume that communication by a person, whether representing themselves or an organization is intended to forward understanding of some sort by transferring information (Wilks, Catizone, Worgan, & Turunen, 2011). This information may be about understanding the past, but may often be about expectations for the future (Alexandersson, Maier, & Reithinger, 1995). This type of exchange for the purpose of prediction has a strong role in commerce and in public policy (Laurance, 1990).

Financial and Non-financial decisions can both be dependent upon the same qualitatively represented expectations. The premise is that decisions are telegraphed in advance as a way of setting expectations (Rubin & Morreale, 1996). Decisions are not made in a vacuum. In consideration of the context in which decisions are made, they are often ‘tested’ with the impacted audience before being implemented, (Barney, 1993; Chermack, 2004; Eisenhardt & Zbaracki, 1992; Foner, 1986; White, Williams, & Greenberg, 1961). The artifacts of these decisions are the discussions in speeches, news, email and other venues.

### 1.1.2 Temporal Importance

It has often been difficult to collect detailed data about the impact of words and correlate these to reliable quantitative measurements. The proliferation of publishing on the web of both qualitative information and time-stamped financial market data makes the data collection task feasible. With data that is both more abundant and more reliable, the analysis can be done on what combinations and proportions, temporal and relational factors in language may govern a process.

The pursuit of this research, if not the understanding of its methodology and conclusions, uses a framework that benefits from both the understanding of the combination of techniques used on the data sets but also some degree of knowledge of the domains in which the data is generated. These combinations of approaches benefit from designating specific problem spaces, recent developments in the research, and the increased power of domain-specific tools. Computers additionally make improvement in the analysis of such text.

### 1.1.3 Sentiment & Bias

Discussions in text can suggest changes in sentiment or even cause changes (Fisher & Statman, 2000; Singer & Radloff, 1963). Quantitative data can often be a representation designed to measure deviation from baseline and therefore doesn't predict the approach of a dramatic shift or 'cliff' event. From housing to art to fixed-income securities, the balance of asset price supply and demand can have a qualitative aspect (Bernanke & Blinder, 1988; Heikkilä, 2002).

Large volumes of text require a person to develop expediencies to process; call those conscious biases. Even careful attention to text can foster unconscious biases, especially if as part of a routine. By placing a reliable quantitative measurement on a body of text, objectification of the corpus might suggest bias over time. With more quantitative data, longitudinal analysis may have more consistency; perhaps at least the inconsistency can be measured. Further, taking a dynamic network analysis approach to analyzing textual data, progress can be made in linking among events, the descriptions of the events, and even the outcomes.

### 1.1.4 Impact

This research may contribute to better judgments being made in areas of public policy, corporate governance, and financial securities valuation. This research is important because text has a large role in economics and finance. This research is important because clarifying the relationship between text and its interpretation can have real consequence in:

- Currency intervention
- In short term interest rates
- In the future of a company

This research is important because the relationship between text and quantitative data in this context is poorly understood. There is news (which generates text in some form) and then there is reporting on news (which generates still more text). Financial security prices move on such reports (Fleming & Remolona, 1999b). News is then again generated on the price moves in a perpetual cycle (Andersson, Overby, & Sebestyén, 2009; Balduzzi, Elton, & Green, 2001; Bomfim, 2003; De Bondt & Thaler, 1984; Green, 2004; Jones, Lamont, & Lumsdaine, 1998; Veronesi, 1999).

Predictable and transparent behavior of organizations around financial decisions is important to the stability of our financial system (Eichengreen, 2004; Meltzer, 2000; Olsen, 1996). Large organizations themselves become complex systems, which include different types of entities and mandates to perform complex tasks that evolve over time (Heiner, 1989). Accurately interpreted feedback received overtime can help to alter the organization mandate and alter execution.

Analyzing increasing amount of text is integral to our lives. However, the volume of the information requires different approaches for analysis. In addition, different speakers may be unaware of each other's communication intention and interpretations. Such asymmetry in motivation and information leads to suboptimal allocation of attention and, in this context, of financial resources. It further contributes to growing problems of information overload as analysis is misplaced and opportunities for more effective resource allocation are missed (Soroka, 2006).

Text has large role in Economics & Finance Psychology (Klibanoff, Lamont, & Wizman, 1998; McKenna & Seidman, 2005) and Behavioral Economics (Barberisa, Shleiferb, & Vishny, 1997; Borch, 2006). Progress in this research might help to increase visibility on a financial organization's future with a successful scientific analysis such as developed in this thesis. Possibly both the performance and the predictability of financial organizations would increase with a follow-on benefit of lower financial market volatility.

Data Mining is the extraction of implicit, previously unknown, and potentially useful information from data. Machine learning provices the technical basis of data mining. Some applications of machine learning focus on prediction: forecasting what will happen in new situations from data that describe what happened in the past, often by guessing the classification of new examples (Witten & Frank, 2005). This Dissertation uses techniques from machine learning to identify patterns. The work of forecasting is placed squarely in the opportunity for future work.

## 1.2 Finance as a research testbed

Financial transactions are timed based as much on the reality of a situation as on the perception of that reality (Baker, Ruback, & Wurgler, 2007; Fama, 1998; Ritter, Constantinides, Harris, & Stulz, 2003). As a test bed for exploring the efficacy of a new framework using machines, the domain of finance serves the purpose well. This is because by any measure, the financial industry produces an enormous amount of data. This data can then be helpful in working to objectively judge any new approach. Additionally, many analysts work on the timing and scale of financial transactions so as to maximize the benefit to an entity. These studies are those that show performance relative to its perception. The result is work that benefits the stability of a company or a country. The benefit of developing better measures for the exchange of information is improvements in the stability of these financial decisions. From the scope of a company with stakeholders that extend beyond shareholders to the scope of a central bank, financial decisions by large organizations can have larger societal impact. This work increases predictability. Increased predictability and lower volatility are hallmarks of a mature and efficient financial system.

Financial decisions are strongly influenced by the external reaction to them (Blinder, Ehrmann , Fratzscher, De Haan, & Jansen, 2008; Fisher & Statman, 2000). Financial decisions of consequence are often signaled in advance through words (Boukus & Rosenberg, 2006). This communication uses a variety of mediums (Burkhart & Fischer, 2008). These conversations also have various degrees of preparation, formality, and impact (Danker & Luecke, 2005; Eisenhardt & Zbaracki, 1992). A central banker's speech may be listened to more carefully than an intra-company email, but both are efforts to set expectations of the future (Eisenhardt & Zbaracki, 1992).

Future behavior is important to analyze because the decision maker desires to identify the timing and magnitude of such future behavior while hedging the risk that a decision could be wrong. At best the decision maker might know historic activities and most of the current background thinking. In the context of decision making for a larger entity, these decisions impact more people.

### 1.2.1 The difficulties in analysis: Interpretations of what is heard

There is substantial benefit in making this decision process more effective. In the framework here, it involves a more efficient exchange of information. The efficiency is increased if the speaker more clearly understands how the listener is interpreting the information. For many organizations, teams of professionals work to both consume and aggregate information for the benefit of an organization's decision-making. Still more teams work to disseminate information. If the decisions are financial in nature, the stakes are raised further.

The listener will be making decisions from the information is received. The speaker may make decisions based on how the information is received. Therefore either the listener, the speaker, or both could be altering their communication based on these decisions. This feedback loop makes the analysis of conversations very difficult. Classical approaches measure sentiment of these texts most often as binary (Mullen & Malouf, 2006; Wilson, Wiebe, & Hoffmann, 2005).

### 1.2.2 The difficulties in analysis: Matching qualitative and quantitative data

The obstacles in analyzing speech effectively require an approach to match the data and the research question. To assign variables for this study to be even more clear, the inherently qualitative $\varphi$ will be compared to a series of quantitative $\delta$. Humans individually or in a group work to interpret text $\varphi$ , but that in itself generally creates more text. The problems created are manifold, the most obvious of which is the difficulty in find finding an appropriate quantitative measure δ.

This research seeks to establish a framework that may help to determine if a relationship exists between qualitative data and quantitative data than be measured. For any given data set φ, and data set δ, a question is if there exists a non-random relationship γ and χ in the following equation:

$$\forall(\varphi,\delta),\varphi = \chi\delta + \gamma$$

If there exists a non-random γ and χ, then can γ and χ be measured? If there is a non-random relationship and it can be measured quantitatively, can

$$\forall(\varphi,\delta),\varphi_{t=0} = \chi\delta_{t=0} + \gamma$$

help give

$$\forall(\varphi,\delta),\varphi_{t+1} = \chi\delta_{t+1} + \gamma\ ?$$

If there is a non-random relationship measured with γ and χ between φ, and δ, the next question is the degree to which a non-random relationship can be established and the circumstances of such a relationship. For purposes of this research, since speech can be converted to text, all spoken and written words will be referred to as text.

## 1.3 Further benefits

Security prices reflect values of the present and expectations of the future. Financial institutions are always working to communicate their intention. For example, they will hold press conferences and give speeches to help set expectations about the future and smooth out the prices tied to the firm's assets. However, they produce a lot of data and interpreting it is hard. In the case of the company, the communications are too numerous. The analysis is even more difficult in the central bank because the communications themselves need to be even more careful. Therefore, interpretation of an organization's direction is one of a series of anecdotal evidence as the most important tactics in real financial decisions.

However, coming up with such a system of data interpretation has several major problems. First getting at the data is hard. Second, cleaning the data is hard. Third, calibrating the language for the different speakers is hard. Hence, figuring out key personnel, information and resource to interpret that data is often beyond the hope of human intuition or anecdotal evidence because of the diverseness and scale of the structures. Second, the organizations are adaptive. They restructure themselves over time and adjust based on their own readings of external events.

The overall question of this work is to determine if there is a relationship between qualitative data and quantitative data that can be measured. Said another way: Can quantitative outcomes be predicted on qualitative data? If relationships are found, the next question is to look in to the degree of the relationship and the circumstances under which predictive qualities exist.

This research takes the more focused approach of being concerned with enabling the early stages of a framework for the machine-enhanced analysis of text in specific contexts. The outcome of this work can then be compared to appropriate quantitative data to test for a relationship. This approach is increasingly useful as Machine-read language activities have advanced. Unfortunately, the current approaches can often be too vague to be useful. A new approach that can suggest the reaction to text and help both sides understand the impact of communication is a promising area of research that can improve effectiveness. This research can help in setting expectations of the future by making opaque situations more clear and in other times clarifying the level of opacity. This will be done by finding the degree and circumstances under which quantitative outcomes might be predicted based on qualitative data. The first step is to find if there is such a relationship that can be measured.

Clarifying the relationship between qualitative data and quantitative data may have real consequence on the direction of short-term interest rates (Baker & Stein, 2004), (Shleifer & Summers, 1990) and currency interventions as well as for the future of a company (Brown & Cliff, 2005), (Frambach & Schillewaert, 1999), (Geroski, 2000). While this clarity may be useful, the relationship is poorly understood. News reports are generated on numbers (Aizawa, 2000), (Godbole, Srinivasaiah, & Skiena, 2007), but also numbers move on news reports (M. W. Berry & Browne, 2005), (Hwang & Salmon, 2008). Questions that may be worth asking are when and where is there a relationship between qualitative and quantitative data or if the relationship has predictive qualities. If a relationship is discovered, it may be only one-way or otherwise signify trends. The relationship may also show consistency or inconsistency in certain circumstances. The data may also show the ways in which the relationship is volatile or the ways in which text influences events.

# 2 Background and problem description

There has been increasing interest in machine-enabled language interpretation to continually sense, collect, and analyze language (Carley & Kaufer, 1993; Carley, Diesner, & De Reno, 2006; Chuang, Tiyyagura, Yang, & Giuffrida, 2000; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1988; Godbole et al., 2007; Jing, Huang, & Shi, 2003). Yet despite all the attention this approach is receiving, the methods remain neither widely adopted nor broadly effective (Lucca & Trebbi, 2009; Luss & d'Aspremont, 2008; Nasukawa & Yi, 2003). One often cited barrier is that many approaches do not adequately add value without substantial manual effort thus mitigating the value of an automated approach (Reeves & Sawicki, 2007; Reinhart & Sack, 2006; Rosa, 2007).

The increasing interest in text comes from the explosion of content available in recent years. The increasing content has been followed by increasing analysis of such content. This analysis includes how individuals relate to the content (Woods, 1975) and language-oriented analysis versus cognitive-oriented analysis (Borge-Holthoefer & Arenas, 2010).

For example, social psychologists, notably Roger Schank and Robert Abelson, have shown how much stories and storytelling, especially human-interest stories, motivate much of human behavior (Abelson & Schank, 1995a, 1995b; Gershon & Page, 2001; Schank, 1990). These stories can count for much more than abstract calculation. In the context of economics and finance, people's moods are largely based on the stories that people tell themselves and tell each other that are related to the subject (Abelson & Schank, 1995b). There is potential value in extracting semantic networks from text to explore this conversation. What is the relationship between public news about a company and its related security prices? What is the relationship between the Monetary Policy of Central Bankers and their public statements?

Approaches to get at solutions have been tried by using classification algorithms on email and public policy documents, bag-of-words approaches and other techniques to get at sentiment on speeches. A link is being explored between these text databases and quantitative data. While rare events might have the most impact, they remain hard to predict. The importance of a potential solution invites further study. Some events in finance & economics have been studied with their associated text: LSA and Investor Sentiment (Barberisa et al., 1997; Boukus & Rosenberg, 2006; Gennheimer, 2002), Copula statistic and associated rare events—asset, credit bubbles, public manias (Gennheimer, 2002; D. Li & Peng, 2009; Mikosch, 2006; Poon, Rockinger, & Tawn, 2004).

Much quantitative data is representation designed to measure deviation from baseline and therefore doesn't predict abrupt changes. Sharp directional turns may remain difficult to detect. Fortunately, in monetary policy and financial decision making, the absolute value is often less important than considering the direction of a trend; even more important may be changes in the trend, especially reversals of a trend. In the case of monetary policy, the direction is determined by expectations of inflation (and in some countries unemployment) and to some extent those expectations can create a feedback loop to effect inflation. The extent of the impact is the point at which the behavior is a social construct.

There is also an issue here of a more general nature: It is possible to approximate quantitative data from qualitative data - for instance, asking people to rate their perception of a sensation on a Likert scale.

This dissertation studies the strengths and weaknesses of using automated semantic analysis for interpretation. Can it help in financial decisions? How can automation be used to detect issues? Part 1 of the dissertation examines if this approach can predict the future. Part 2 studies if this approach is generalizable on email.

## 2.1 Studies of Monetary Policy communications by Central Banks (Part 1)

Central banks have been studied from several angles. Classification of documents especially in public policy and further future study (Y. H. Li & Jain, 1998; Purpura & Hillard, 2006). Communication Policy in Central Banks is closely monitored (Blinder et al., 2008; Burkhart & Fischer, 2008). Only about ¼ of the speeches are about Monetary Policy, but FOMC minutes show some affect on the market (Boukus & Rosenberg, 2006; Purpura & Hillard, 2006). Fed Funds Futures have been suggested as a gauge of future policy actions by the FOMC (Krueger & Kuttner, 1996).

Discussions in text can suggest changes in sentiment or even cause changes. Such approaches benefit from specific problem spaces, developments in research, and the increased power of domain-specific tools. New approaches can be used to explore predictive power.

From housing to art to company stock to government debt, asset prices' balance of supply and demand can have a qualitative aspect that is subject of much study. Taking a dynamic network analysis approach to analyzing textual data, we can make progress in linking among events, the descriptions of the events, and the outcomes

The field of Behavioral Economics directly targets the inquiry in which some agents display human limitations and complications in asset pricing (Mullainathan & Thaler, 2000).There exist many studies of investor perception on stock price behavior (A. W. Berry, 2011; Chan & Lakonishok, 1994) (Iqbal & Shetty, 1995). These suggest the input of emotion into the quantitative world of securities valuation.

## 2.2 Studies of Email, especially within a large corporation (Part 2)

Email has also been studied from many different angles: Anomaly Detection (Priebe, Conroy, Marchette, & Park, 2005); Analyzing Large Scale Networks (Carley & Skillicorn, 2005); Electronic Groups (McKenna & Seidman, 2005); Structure of Enron Data Set (Keila & Skillicorn, 2005); Summarization (Muresan, Tzoukermann, & Klavans, 2001). Asset Prices have also been studied from many different angles: Investor Sentiment (Barberisa et al., 1997); Market liquidity as a sentiment indicator (Baker & Stein, 2004); Clear evidence of Beta herding based on sentiment (Hwang & Salmon, 2008).

## 2.3 Graph Theory, Dynamic Network Analysis, and Semantic Networks

Graph theory developed with studies in the 1950s. Dynamic Network Analysis ("DNA") enriched this work with the explanation of using special relations.

### 2.3.1 Quadratic Assignment Procedure ("QAP") and Multiple Regression QAP ("MRQAP")

The presence of Autocorrelation in Social Networks Analysis of Social Networks motivated the development of QAP (Mantel, 1967) (Hubert, 1987) (Krackhardt, 1987) and exponential random graph models (Dekker, Krackhardt, & Snijders, 2005). The Autocorrelation problem particularly afflicts the analysis of cognitive networks (i.e., networks mapping perceptions). One solution to the autocorrelation problem is to use non-parametric tests to determine whether independent variables are significant predictors of the dependent variable (Kilduff & Tsai, 2005).

### 2.3.2 Semantic Networks

Semantic network analysis is the use of network analytic techniques on paired associations based on shared meaning as opposed to paired associations of behavioral or perceived communication links (Doerfel, 1998); they are graphs on the structure of meaning in language (Lehmann, 1992). Doerfel (Doerfel, 1998) argues that the very definition of semantic networks had become muddled. Some described the essence of the semantic network as the analysis of text to measure the relationship among words (Rice & Danowski, 1993) while others' conceptualization of semantic networks is described as associations based on shared interpretations (Monge & Eisenberg, 1987). The distinction can matter because the different methods can produce different results (Carley, 1993).

Carley's (Carley, 1993) concepts of semantic networks demonstrated differences between semantic analysis using maps compared to simple analysis of the presence and frequency of words; documents with different meanings could have the same concepts but with different meanings. Semantic Networks can be helpful in communicating ideas and in learning (Feghali, 1991).

The last ten years have also seen several studies regarding language complexity (Baldwin & Carpenter, 2012) (Bales & Johnson, 2005) (Belov et al., 2009). Developments in complex networks can be seen as motivating much of this research. These were first focused on abstract and general overviews of language complexity; few of them looked at how language can affect cognition (Daimler, 2009). Borge-Holthoefer (Borge-Holthoefer & Arenas, 2010) even claims a shift in research from language-oriented work to that which has a more cognitive-oriented point of view.

# 3 Methodology

Figure 1 is a visual representation of the framework presented in this Dissertation. It shows the sources of the data, the high-level processing of the data, and the analysis of the outcome. It shows the clarity of thought brought to the evaluation of the effectiveness of the framework. Each of the six segments presented within Figure 1 are expanded in later sections within this Dissertation.

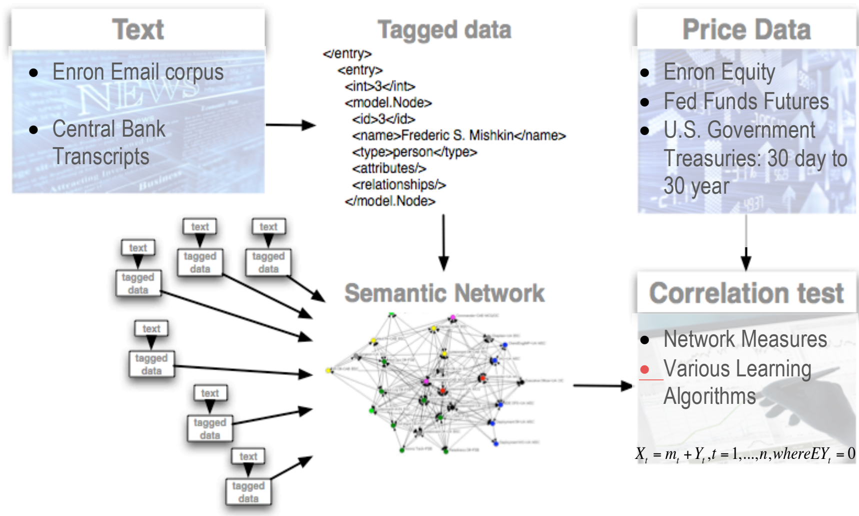


**Figure 1: Framework under consideration**

## 3.1 Objectives and overview of proposed framework

As a practical demonstration of this research, experiments are conducted relating to various combinations of datasets in two domains. The first dataset consists of various texts (written material and transcriptions of orally presented material) from the US Federal Reserve. The second is a corporate email corpus. The models and algorithms are applied to for tasks in these domains: objective and repeatable exploration of correlation, of prediction, a baseline comparision with classical approaches, and some implications of the findings. In the course of giving solutions to these problems, theoretical and empirical results are developed using a framework intended to make them easily applicable to other domains.

To ground this research, three data sources are used: Central Bank public communication, Public market financial data on U.S. Central Bank actions, and corporate emails from Enron. My research will contribute to a unifying framework for using qualitative financial information to match quantitative financial information, a partially automated intelligence analysis capability which can meet the financial decision making in the real world, bridge dynamic network analysis and computational finance, and reduce the time and cost of financial decision making.

This research seeks to investigate the following phenomena:

- Is there a relationship between qualitative data and quantitative data that can be measured?

- Can quantitative outcomes be predicted based on qualitative data? The benefit of this research is the more effective setting of expectations for the future and reducing bias by objectively measuring qualitative information.

## 3.2 Assumptions on Methodology

In analyzing the pragmatics of speech, we assume that the speaker and listener, writer and reader are working together to some level of understanding. Writing and speaking is often done with the intent of influence. Measuring the extent to which writing and speaking influence behavior is difficult. This limits exploration of their interactions both for people and for machines analyzing the conversation. The proliferation of raw computing power applied to these problems has given limited results. Cheap computation is helped by three additional factors: 1) new mathematical approaches, especially in Network Science and 2) domain expertise, and 3) new openness of the Internet and of government to make public data previously hidden.

If the purpose of the communication is to exchange information then we must assume that at least the speaker wishes success. Even with this strong assumption, perceived expertise and attention of the audience has an impact on the communication; these variables change over time. The complexity and adaptively of the listener make the decision more complex to predict the reactions to their behavior and therefore the effectiveness of the information exchange. Combining recent developments in Network Science, Computational Linguistics, and Machine Learning enables new efforts to measure the impact of the human-generated text.

I adopt Network science models as the most flexible path for representing the textual data in quantitative terms and introduce a system for regularizing the speeches appropriate for the analysis. In comparing the data to the output, I argue for both the network science method and against latent semantic analysis, suggesting a focus on more generalizable, useful approaches to studying the relationship between text and the desirable outcome leading to a variety of applications in the real world

A potential solution is to employ automated approaches to support information interpretation. Machines are better at allowing individuals to interpret information efficiently. Applying machines to language, individuals can use the signals to help judge the importance of the communication to which further study is needed. This can reduce interruption costs and information overload.

This research uses security prices as the quantitative data test case against which to compare the qualitative data. This is in many ways an ideal data set because security prices are generally expected to reflect values of the present and expectations of the future (Basu, 1977; Chen & Yeh, 2002; LeRoy, 2010; Malkiel, 2003). It is in decision making that these analyses become most useful.

Decisions in many domains, from public policy to commerce interact with written and spoken communication in increasing quantities. Despite an increasing importance of numerical literacy, communication through writing and speaking continues to increase. The creation of this text may or may not be done with care, but the volume to be consumed suggests an opportunity for assistance by machines.

This computational-centric approach addresses these challenges by explicitly testing the circumstances under which quantitative to qualitative relationships occur in a specific domain: financial economics.  The approach takes as input qualitative data from various sources in addition to quantitative data in the form of asset prices. I develop a meta algorithm for measuring and testing the relationship which helps to identify a true causal relationship among many data sets in many circumstances. The approach leads to insights on asset valuation for the purposes of public policy but also for corporate management in the domain of securities issuance.  The techniques are generally applicable and as an example I will present the algorithm two contexts: corporate malfeasance at Enron and Central bank communcations.

## 3.3 Steps to develop framework

Figure 2 shows the visual representation of the the step-by-step process used in the framework for the analsysis of the texts. These are expanded in text later in this section (3.3), but Figure 2 shows the clear, linear progression of analysis at a higher level.

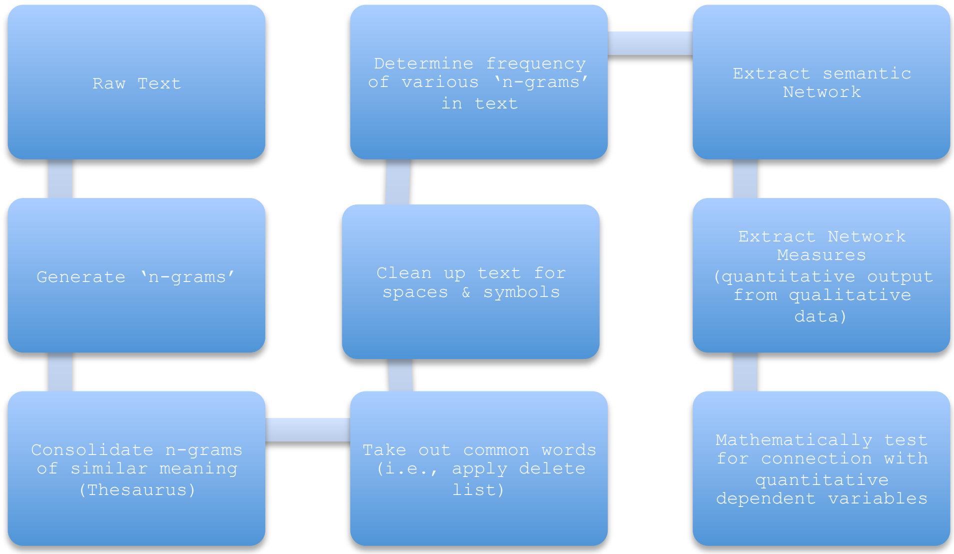


**Figure 2: Detailed look at framework under consideration**

### 3.3.1 Qualitative Data

The qualitative data collected from each data set will, of course, vary in its raw form. However, in the processing, the variables generated will be identical.

***Part I: Acquire and clean qualitative data***

*Step 1: Acquire raw text in various formats (e.g.,.HTML, .pdf, .txt).*
In the case of this study, the data sets include the public statements of the US Federal Reserve found in the speech transcripts, Congressional Testimony, Minutes of the FOMC meetings and Conference Calls. It also includes the publc email corpus of Enron.

*Step 2: Separate useful text from noise and neutralize formatting.*
In the case of this study, the standard disclaimers given by The Fed are identical in most, if not all, documents; these are removed. The information on the person making the speech as well as the venue are likewise deleted. Headers and other technical transmission information are deleted from emails.

### 3.3.2 Quantitative Data

***Part II: Acquire and clean quantitative data***

*Step 3: Acquire appropriate dependent variables (financial data).*
For much of the appropriate financial data, the time interval of the prices (the granularity) can be many times per second. Studying data at these time intervals can be useful only if matched to the time stamp on the qualitative data. This match must be done if the research question involves intraday relationships. For effects lasting more than one day, the choice is between looking at various weightings of average daily prices (e.g., trade volume-weighted data or time-weighted data) and opening or closing day prices.

Closing prices are likely the best choice for measuring effects between days because this is when markets generally experience the highest liquidity (volume). Therefore, the final price is likely to be the most representative of the day. There are several non-systematic gaps in the Fed Funds Futures Data if not in the yield curve.

For those dates where there is no data, analysis is performed using three approaches: using previous data, using an average to the two surrounding data points, and lastly, using the date empty. Further, each Futures contract represents two years. This data needs to be averaged. For example, there is a contract for December 2005. The dates on this contract overlap with the 2006 and 2004 contracts. The dates need to all line up, then the average is taken.

All securities data is treated independently. Specifically, the term structure of interest rates represented by the Yield curve on U.S. government debt is treated as eleven separate instruments. Similarly, the futures contracts are also treated separately from one another even when they overlap.

### 3.3.3 Text processing

***Part III: Transformation of text***

Included in these steps are minor transformations such as the removal of punctuation and capitalization. For this part, we use the Automap Dynamic Network Analysis Tool, "AutoMap", (Carley, Columbus, Bigrigg, & Kunkel, 2011).

*Step 4*: Generate a delete list

Delete lists allow the user to remove non-content words from text, such conjunctions, articles, and other noise words. The format of the list is a plain text file, containing a list of words to be deleted, one on each line (Nimick, 2011). The raw text contains many words that add noise to further analysis. Examples of these words are the, of, I, in, and for. This simple review of words were selected based upon a review of the literature for incorporating delete lists in AutoMap.

*Step 5*: Add or delete space after deleted words

In the figure, 'xxx' is replacing the deleted words. The methodology may also support deleting the words present in the delete list without replacing the word with any substitute. This transforms "In my remarks today, I would like…" to either "xxx xxx remarks xxx , xxx would like…" or "remarks would like". Since "remarks would like" and similar such phrasing may produce erroneous conclusions in this type of study, the choice is to use the substitute characters 'xxx' for those words eliminated using the delete list. The delete used was used elsewhere (Diesner, Frantz, & Carley, 2005) and not generated specifically for this research. Further work could be done using a different approach here.

*Step 6*: Generate list of N-grams

This is a list of short word combinations such as 'financial disruptions' that may be useful in generating meaning from the text. Various lengths of word combinations may be used. These are manually generated. You may reference the n-grams used for this research in the appendix.

*Step 7*: Generate Thesaurus

A speaker or writer may use several different words or phrases to relate to one concept. A frequent example is 'Fed' and 'Federal Reserve' which are synonyms for the US Central Bank (another synonym). To generate a Thesuarus, I start with standard, or widely used, list (Landwehr, 2012), then manually generate additional domain-specific terms to disambiguate similar meaning phrases. See the Appendix for a full accounting of the Thesaurus used.

Given that the thesaurus is drawn from multiple usages elsewhere and not generated independently just for the purposes of this sresearch, it is likely to contain many omissions specific to these datasets. These gaps have been intended to be filled through the manual process. Given the nature of email as a communication medium, there likely remain many spelling errors not comphrensively captured by the thesaurus used.

*Step 8*: Apply N-grams

These are applied from the protocol established in step six.

*Step 9*: Apply Thesaurus

Through an automated process from within Automap, I endeavored to stem words as appropriate (e.g., wording and words are simplified to ‘word’). See the figure below for a visualization of the text transformation.

Figure 3 (below) shows a visualization of the text as it is processed through the stages of the framework. It makes clear what is meant by terms suchs as ‘n-grams’ and ‘delete list’ (the details of which may also be found in the appendix).

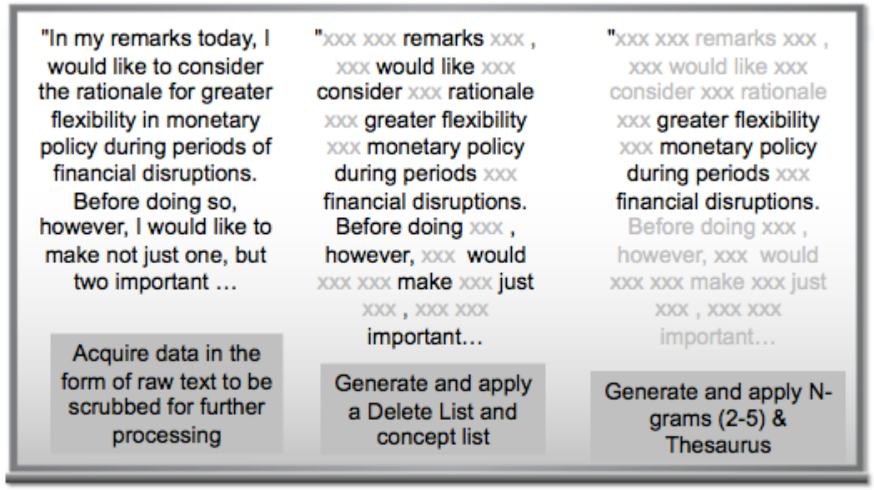


**Figure 3: Visualization of early stages of text processing**

### 3.3.4 Network Measurements

***Part IV: Create network measures***

*Step 10*: Create network measures

For this part, the tool Organization Risk Analyzer, “ORA”, (Carley, Reminga, Storrick, & Columbus, 2011) is used. ORA facilitates the creation of network measures of social networks and that can then be applied to the transformed text created from Steps 1-9. The model is run in ORA to create 86 network measures. The table below shows the names of the network measurements although at this stage, the labeling has no impact.

The list in Table 1 (below) represents a sample of common measurements for exercises as performed in this research. Wasserman & Faust (Wasserman & Faust, 1997) provides a review of the definitions of these network measures so that they will not be repeated here. While the density measures have dominated semantic networks and additional dimensions provide useful measures of connectivity (Carley & Kaufer, 1993), this work is generally concerned with the quantitative output and less concerned with the structures of the network and therefore a deeper inquiry into the definitions.

| NETWORK MEASURES FOR PROCESSING IN FIRST CUT AS CANDIDATE INDEPENDENT VARIABLES |
|---|
| AverageDistanceSemanticNetwork |
| BouarySpannerPotentialSemanticNetworkAverage |
| BouarySpannerSemanticNetworkAverage |
| BreadthColumnSemanticNetwork |
| BreadthRowSemanticNetwork |
| CapabilitySemanticNetworkAverage |
| CentralityAuthoritySemanticNetworkAverage |
| CentralityBetweennessSemanticNetworkAverage |
| CentralityBonacichPowerSemanticNetworkAverage |
| CentralityClosenessSemanticNetworkAverage |
| CentralityColumnDegreeSemanticNetworkAverage |
| CentralityEigenvectorPerComponentSemantic-NetworkAverage |
| CentralityEigenvectorSemanticNetworkAverage |
| CentralityHubSemanticNetworkAverage |
| CentralityInDegreeSemanticNetworkAverage |
| CentralityInClosenessSemanticNetworkAverage |
| CentralityInformationSemanticNetworkAverage |
| CentralityInverseClosenessSemanticNetwork-Average |
| CentralityOutDegreeSemanticNetworkAverage |
| CentralityRowDegreeSemanticNetworkAverage |
| CentralityTotalDegreeSemanticNetworkAverage |
| CliqueCountSemanticNetworkAverage |
| CognitiveDistinctivenessAverage |
| CognitiveExpertiseAverage |
| CognitiveResemblanceAverage |
| CognitiveSimilarityAverage |
| CommunicationHammingDistance |
| CommunicativeNeedSemanticNetwork |
| ComponentCountStrongSemanticNetwork |
| ComponentCountWeakSemanticNetwork |
| ComponentMembersWeakSemanticNetworkAverage |
| ConnectednessSemanticNetwork |
| ConstraintBurtSemanticNetworkAverage |
| CorrelationDistinctivenessSemanticNetworkAverage |
| CorrelationExpertiseSemanticNetworkAverage |
| CorrelationResemblanceSemanticNetworkAverage |
| CorrelationSimilaritySemanticNetworkAverage |

| NETWORK MEASURES FOR PROCESSING IN FIRST CUT AS CANDIDATE INDEPENDENT VARIABLES |
|---|
| CountColumnSemanticNetwork |
| CountNodeSemanticNetwork |
| CountRowSemanticNetwork |
| DensityClusteringCoefficientSemanticNetwork-Average |
| DensitySemanticNetwork |
| DiameterSemanticNetwork |
| DiffusionSemanticNetwork |
| EffectiveNetworkSizeBurtSemanticNetwork-Average |
| EfficiencyGlobalSemanticNetwork |
| EfficiencyLocalSemanticNetwork |
| EfficiencySemanticNetwork |
| ExclusivityCompleteSemanticNetworkAverage |
| ExclusivitySemanticNetworkAverage |
| FragmentationSemanticNetwork |
| HierarchySemanticNetwork |
| InterdepeenceSemanticNetwork |
| InterlockersSemanticNetworkAverage |
| IsolateCountSemanticNetwork |
| KnowledgeHammingDistance |
| LinkCountLateralSemanticNetwork |
| LinkCountPooledSemanticNetwork |
| LinkCountReciprocalSemanticNetwork |
| LinkCountSemanticNetwork |
| LinkCountSequentialSemanticNetwork |
| LinkCountSkipSemanticNetwork |
| MetaMatrixHammingDistance |
| NetworkCentralizationBetweennessSemanticNetwork |
| NetworkCentralizationClosenessSemanticNetwork |
| NetworkCentralizationColumnDegreeSemanticNetwork |
| NetworkCentralizationEigenvectorSemanticNetwork |
| NetworkCentralizationInDegreeSemanticNetwork |
| NetworkCentralizationInClosenessSemanticNetwork |
| NetworkCentralizationOutDegreeSemanticNetwork |
| NetworkCentralizationRowDegreeSemanticNetwork |
| NetworkCentralizationTotalDegreeSemanticNetwork |
| NetworkLevelsSemanticNetwork |
| NumberofConceptnodes |

| NETWORK MEASURES FOR PROCESSING IN FIRST CUT AS CANDIDATE INDEPENDENT VARIABLES |
|---|
| OverallComplexity |
| RadialsSemanticNetworkAverage |
| ReduancyColumnSemanticNetwork |
| ReduancyRowSemanticNetwork |
| SimmelianTiesSemanticNetworkAverage |
| SpanOfControlSemanticNetwork |
| SpeedAverageSemanticNetwork |
| SpeedMinimumSemanticNetwork |
| TaskHammingDistance |
| TransitivitySemanticNetwork |
| TriadCountSemanticNetworkAverage |
| UpperBouednessSemanticNetwork |

**Table 1: All Network Measures generated for initial processing step**

A choice does exist on this part of the framework development on whether to choose analysis by node (in this case a word or phrase) or a graph level measure. (Other network measures such as those pertaining to 'risk' have more meaning in the traditional sense of network analysis). The Figure therefore represents just graph level measures.

The definitions for many of these measurements may be aligned with the intuitive sense of how a non-computational approach to the measurement of a document for meaning. Taking the measurement 'Density: Semantic Network', for example. Density is the ratio of the number of edges versus the maximum possible edges for a network with output between 0 and 1 (Carley et al., 2011). If analyzing manually, trends in density good be a good metric to investigate. Fortunately for the purposes of developing this framework, the concern is with inquiring into the possibility and nature of the relationship between these variables and the quantitative financial measures. These are investigated in Part V.

### 3.3.5 Learning Algorithms

**Part V: Application of learning algorithms: The pipeline performed on a given input file**

The Figure below shows the process flow for the statistical analysis of correlation. After the text is processed, This Figure makes clear the high-level statistical process underwhich the results will be analyzed.

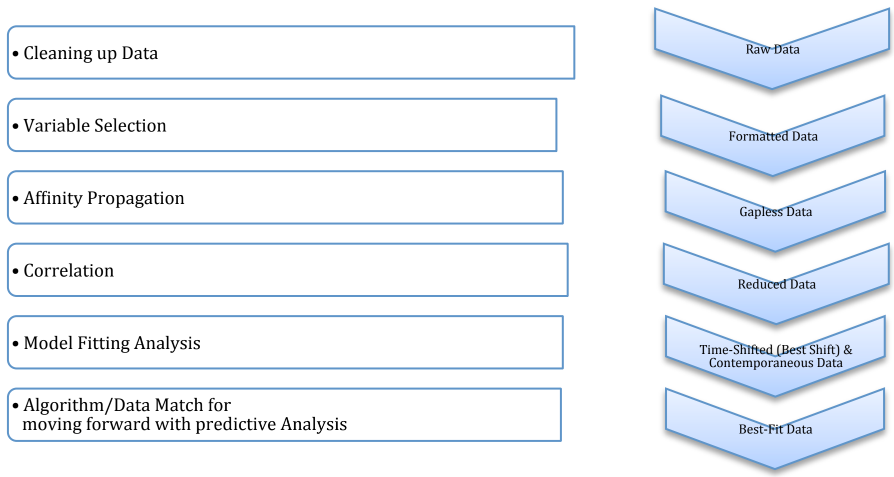


**Figure 4: Process Flow for statistical analysis of correlation**

This section presents the processing steps performed on each file. When a specific file is referred to, it is the aggregate of the Public Policy data, but the generalized considerations hold true for the other datasets within this study. Only the summary file discussed later summarizes the performances of the different models on all input datasets within a study.

*Step 11: Measurement Aggregation*
From the very nature of the data, the quantitative data is available only on days in which the market is open. Weekends and holidays recognized by the US markets make quantitative data unavailable.

Qualitative data occurs with less predictability. The public policy data occurred primarily during days (but not necessary times) in which the US markets were open. The corporate email data was, of course by its nature, dispersed in time of week and time of day. When more than one qualitative data point was available in any one day, these measurements were averaged to create a daily average measure for each Independent Variable.

*Step 12: Temporal Variable Matching*
The Dependent and Independent variables are then matched by time. The result is that each variable (dependent or independent) had only one measurement for each day.

*Step 13: Eliminate data 'breaks'*

- Fill in all empty cells with a standard symbol (e.g., 'ND')
- Remove non-numerical data (e.g., '…', 'ND')
- Rename all columns eliminating the spaces between names and replace them with a standard character (e.g., '_')

This process allows a file then readable by statistical tools. In the example of Public Policy data, the aggregate output file *(#1)* has dimensions 5844x103 (observations x variables), with the first column being the observation date.

*Step 14: Variable selection*
For Public Policy Data, the characteristics of the data is as follows:
Initial number of columns: 103. (Variables plus date)
Initial number of rows: 5844. (Dates)

Initial number of dependent variables: 16
Initial number of independent variables: 86

The first analysis is how many ND's as a percentage of each column (see Figure below). This is important because the majority of rows have 80% or more of ND's. These 'NDs' represent the absence of Public Policy Data data, not any artificial separation of data for the purpose of analysis.

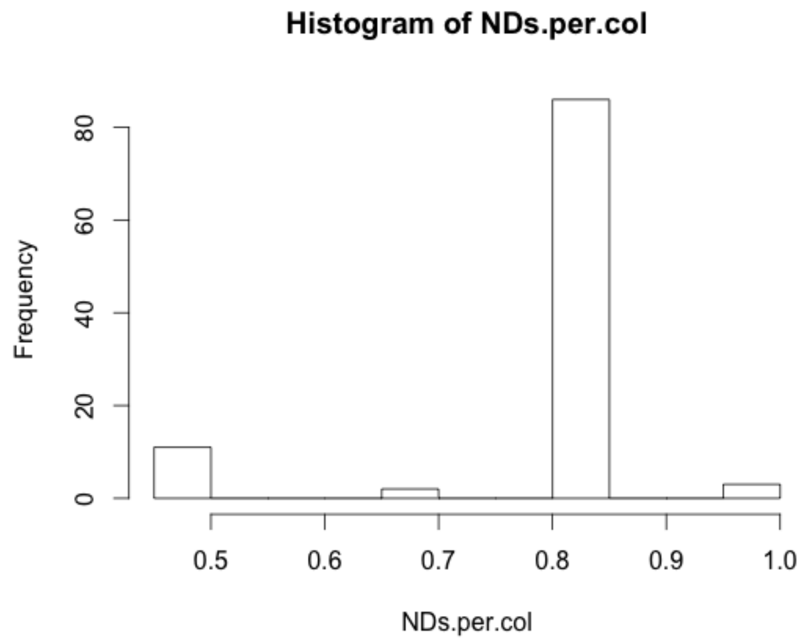


**Figure 5: Distribution of percentage of NDs over the entire original dataset (per row)**

I next need to select a subset of variables for which there exists numerical information. This is particularly important for the independent variables ("IVs") since they are used as predictors. I then investigate specifically the percentage of ND's per row, considering only the IVs. The figure below shows how the majority of rows of independent variables do not present data at all (% of ND's equal to 1), while the remaining rows are fully represented throughout the independent variables (strictly discrete distribution). It is clear that each observation either present ND's for all IVs or presents values for all of them. Note that this is a peculiarity to the Public Policy data that I used, not necessarily the corporate data.

> Total number of observations with values for all independent variables: 1040

Spanning of such observations:
13Jun1996 – 03Dec2009

Dataset dimensions: dependent variables (1040x16), independent variables (1040x86).

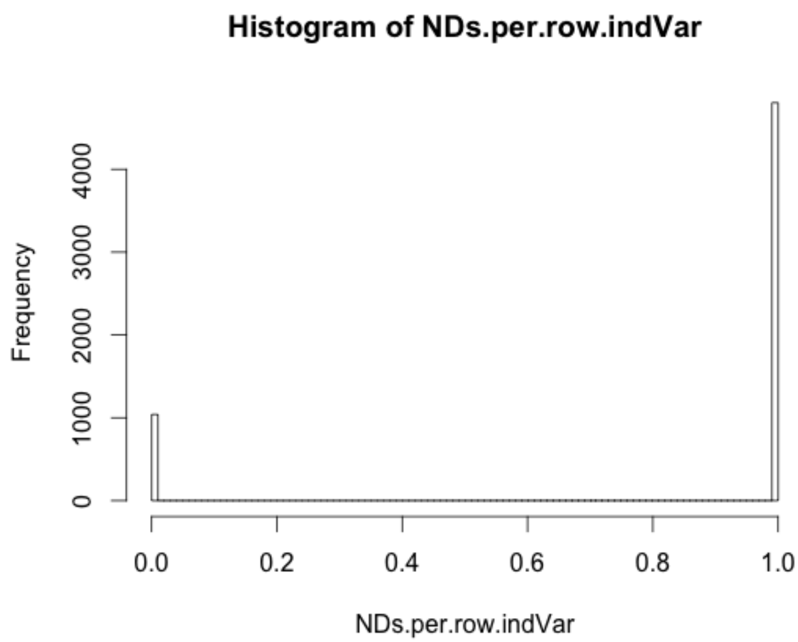


**Figure 6: Distribution of percentage of ND's over the IVs (per row).**

The Figure below shows the results: most of the rows (11 out of 16) of the pre-selected observations (pre-selected based on the independent variables) present only 10% of ND's (equivalent to 112 observations out of 1040), three dependent variables presented no data (i.e., 100% of ND's, dependent variables 3, 4, and 5: increase, decrease and level), and two dependent variables (1 and 2: Fed Funds Futures) presented approximately 50% of observations without values. Table 2 (below) summarizes the percentages of ND's per dependent variable.

| Futures | | Fed Funds | | | US Treasuries | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ctr1 | ctr2 | In-crease | De-crease | Lvl | 1 Mo | 3 Mo | 6 Mo | 1 Yr | 2 Yr | 3 Yr | 5 Yr | 7 Yr | 10 Yr | 20 Yr | 30 Yr |
| 0.50 | 0.54 | 1 | 1 | 1 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |

**Table 2: Percentage of NDs across rows per DV**

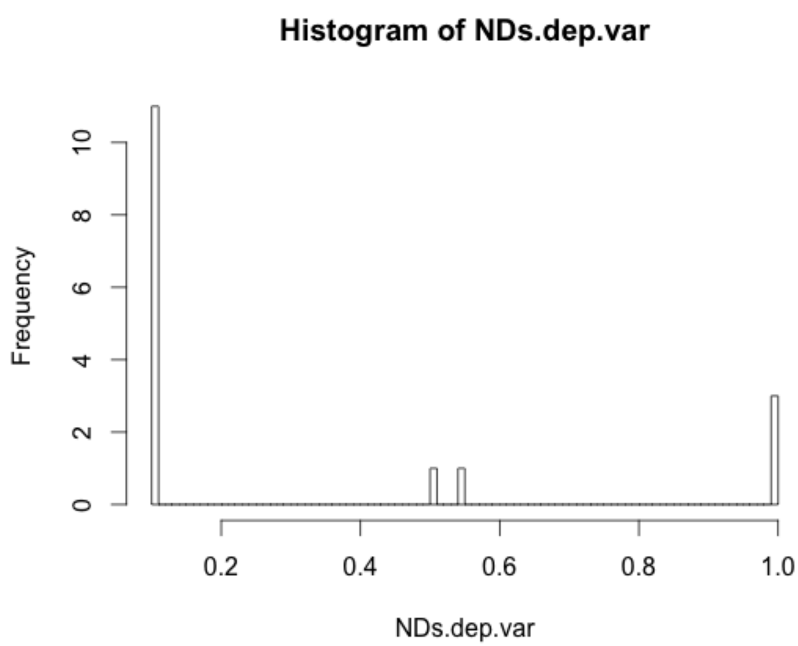


**Figure 7: Distribution of percentage of ND's over the DVs, and limited to the observations previously selected on the IVs.**

*Step 15: Selecting representative metrics through clustering*

There are 86 Independent Variables. I want to investigate the degree to which those can be reduced through finding correlations among them. The first step is finding an appropriate algorithm.

One might consider finding group exemplars by randomly choosing an initial subset of data points and then iteratively refining it. However, this only works well if that initial choice is close to a good solution. These so-called k-centers techniques, begin with an initial set of randomly selected examplars and iteratively refine this set to decrease the sum of sqared errors.

A method that might be considered in multivariate data analysis is Principal Component Analysis (PCA). PCA is a Linear method that greatly reduces the number of variables to be monitored based on eigenvalue and eigenvector decomposition of the covariance matrix (Cheng, Zakharov, Dorado, & Zhang, 2009). However, PCA is not a clustering algorithm. PCA is a way to change the coordinate system in which the data is represented. If this is found, then the data can safely be described in a smaller dimension.

More appropriate clustering algorithms for this problem might be k-means or fuzzy k-means. However, those require the user to pre-specify the number of clusters. Another choice, Affinity Propagation (AP) automatically estimates the best number of clusters.

AP (Frey & Dueck, 2007) is a new algorithm that takes as input measures of similarity between pairs of data points and simultaneously considers all data points as potential exemplars. Real-valued messages are exchanged between data points until a high-quality set of exemplars and corresponding clusters gradually emerges. AP identifies a set of centers (exemplars) from actual data points. Contrary to PCA and other k-centers techniques, AP consideres each data oint as a node in a network, and recurseively transmits real-valued messages along edges of the network until a good set of exemplars and corresponding clusters emerges. At any point in time, the magnitude of each message reflects the current affinity that one data point has for choosing another data point as it exemplar (Sakellariou, Sanoudou, & Spyrou, 2012).

In a variety of clustering problems, AP found clusters with much lower error than those found by other methods including PCA (Sakellariou et al., 2012). It can often do this in less than one-hundredth the amount of time (Frey & Dueck, 2007). Because of its simplicity, general applicability, and performance, AP was chosed over PCA.

Bodenhofer (Bodenhofer, Kothmeier, & Hochreiter, 2011) has implemented Affinity Propagation in R in a package called ‘apcluster’. The result is the identification of clusters and then the selection of an exemplar (representative) for each cluster. The 86 independent variables were automatically clustered into 19 groups. To help illustrate what these results can look like, the figure below is a visualization of a simple affinity propagation finding ‘most similar’ variables to form a cluster (Bodenhofer, Kothmeier, & Palme, 2013).

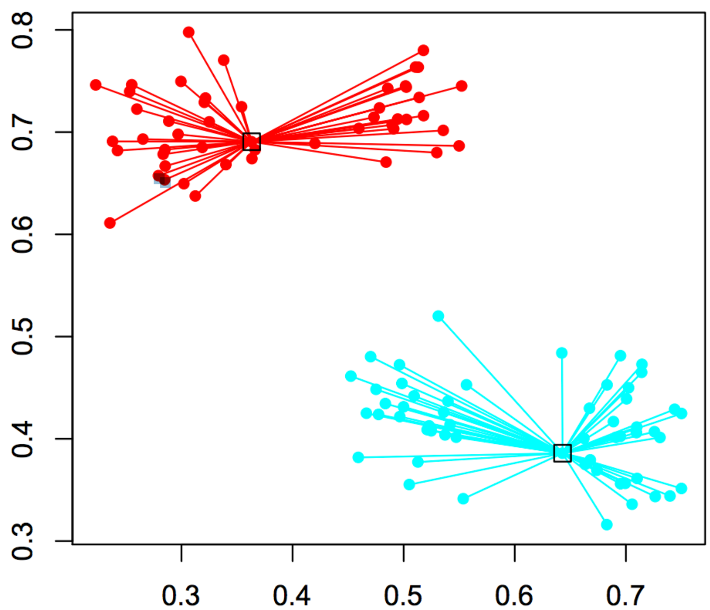


**Figure 8: Visualization of correlations of exemplars among a simple variable cluster**

This affinity propagation reduces the 86 IVs automatically into 19 groups. The new subset of independent variable is thus made up of 1040 observations and 19 independent variables.

*Step16: Temporal Shifts*

This step was performed in order to test whether single IVs would correlate better with single DVs when shifted along time of a few *observations* (i.e., rows). This analysis was only performed for those DVs presenting at least valid numeric value across observations (therefore excluding DVs *Increase*, *Decrease*, and *Level* in the case of the Public Policy Data). If the best shift detected is always the first one being tested (-5), shifting is seen to not improve the single independent variable performances in correlating with single dependent variables. If the best shift equals 0, this also suggests that there is no use in time-shifting the data.

*Step 17: Determine appropriate learning algorithm*

There exist many algorithms available for application to this data. I have chosen to use more than one learning algorithm to capture the different possible ways of finding correlation. For the experiments discussed in this thesis, linear regression is used as a base case against which other models may be compared. Others, in the set of parametric algorithms or non-parametric ensemble methods such as random forests, are also tested. Still other algorithms may be explored in future research.

This thesis covers in the background section how parametric models such as these can suffer from autocorrelation when used on cognitive networks. These can be generated by the self-referential nature of cognitive relationships. Semantic networks do not suffer from the same problems with autocorrelation. However, to strengthen the results of the experiments in this research, both parametric and non-parametric results are used.

This study uses five different models to explore correlations: linear (lm), cart, generalized linear (glm) with Gaussian link function, random forest, and svm (using the radial basis function as kernel). I used the following packages within R to do this: e1071 (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2004), RandomForest (Liaw & Wiener, 2002), and rpart (Ripley, Therneau, & Atkinson, 2013). The model fitting (with nested feature selection) is preceded by an analysis of each DV separately (only for those with at least some numerical information, of course); for each of such DV, I use the following algorithm for building a set of optimal independent variables:

1. Start with an empty selection of IVs.
2. According to the given model, select the IV that produces the best model and add it to the selected set of IVs.
3. Set the current best model fitting to the one obtained with the selected set of IV.

4. Consider again each IV not previously selected. For each of them, assess the new model fitting when the given IV is added to the selected set. The IV that increases the model fitting the most when added to the set is included to the selected set.
5. The procedure stops when either all IVs have been added or no increase in model fitting is possible.

These results are presented in depth for each dataset under review in the form of two different types of tables. The first set of tables show that for each DV, the selected subset of IVs selected by the model fitting (by flags set to 1 for those selected). The second table reports the best performances obtained, for the given DV, with the selected subset of IVs. The performances are R-squared or pseudo R-squared values. The results for the performances are reported in the tables within the corresponding chapters to the dataset being analyzed.

The svm model is generally found to be obtaining the best performances $\forall(DV), pseudoR^2 \geq 0.46$. The tables described above show the percentage of DVs for which a given IV has been selected. This provides a score on the overall importance of any one that independent variable. To help with understanding, an example of a modeled DV is shown in the figure below: here the red line represents the original DV (30-Year Treasury) while the blue one is the fitted model, obtained using corresponding IVs found for that particular iteration. The matching of the two suggests the effectiveness of generated model.

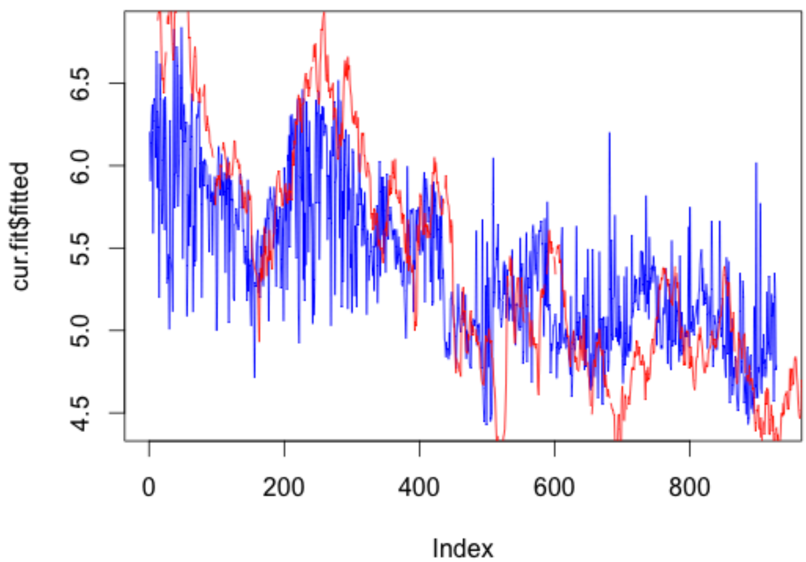


**Figure 9: Original Original (red) and reconstructed (blue) dependent variable time course. SVM model fitting has been used.**

There exist two main types of decision trees used in data mining: Regression trees and classification trees. Useful to address in advance of the details of these algorithms is the exclusion of learning algorithms that might otherwise be thought of as being appropriate for use in such studies. The most notable of these is time-series analysis analysis.

#### *3.3.5.1 Time Series*

The structure of the data may seem to be perfect for a time series analysis. Especially in the domain of mathematical finance, time series analysis is often used for predicting future events based upon a type of sophisticated extrapolation of past data.

**Definition 1** *The classical decomposition model with* **time series** *X and observations 1 to n and no trending*

$$EY_t = 0$$

**Equation 1: Time Series trending**

*Time series X with observations* 1 to $n$ *is given by*

$$X_t = m_t + Y_t$$

**Equation 2: Time Series Obervations**

However, for time series analysis to be an appropriate approach, the data needs to be of temporal uniform density. The data in this research is clustered. There are times with no data, other times sparse data, and then a tighter cluster of data. The profile of that data is not compatible with performing an effective time series analysis. Time series is not a good choice for this data.

### *3.3.5.2 Linear Regression*

**Definition 2** *Given a data set* $\{y_1, x_{i1}, \cdots, x_{ir}\}_{i=1}^{k}$ *of k, where* $y = X\beta + \varepsilon$ *or*

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}, X = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_k \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1r} \\ x_{21} & \cdots & x_{2r} \\ \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{kr} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_2 \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix}$$

**Equation 3: Linear Regression assumption**

*the relationship between y and x is linear.*

For each combination of independent variables and dependent variables, we have a data set of

$$\left\{ y_i, \quad x_{i1}, \quad \cdots, \quad x_{ik} \right\}_{i=1}^{n} \text{ where } y = X\beta + \varepsilon,$$

**Equation 4: Linear Regression**

or expressed in stacked form:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \hline x_{21} & \cdots & x_{2k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

**Equation 5: Linear Regression (stacked form)**

Where y is one or more of the dependent variables (the financial data), $\widehat{x_i}$ is one or more of the independent variables (the semantic network measures), and $\beta_i$ and $\varepsilon_i$ is the slope and intercept, respectively.

#### *3.3.5.3 CART and C5.0*

C5.0 is a simple classification algorithm, a statistical classifier developed by Ross Quinlan that was improved from the landmark C4.5 which itself was built upon his ID3 algorithm (Quinlan, 1986). Multivariate tests were introduced to the framework provided by C4.5 and ID3 with the *classification and regression trees (CART) system for learning decision trees* (Breiman, Friedman, Olshen, & Stone, 1984). For purposes of this research, CART is used in the experiments as the non-parametric algorithm.

CART is a binary recursive partitioning procedure capable of processing continuous and nominal attributes as targets and predictors. It is a learning algorithm in the form of a recursive partitioning method that builds classification and regression trees for predicting dependent variables. The CART mechanism is intended to produce not one tree, but a sequence of nested pruned trees, each of which is a candidate to be the optimal tree. CART can assist in the analysis of the relationship between the semantic network measures and financial data.

**Definition 3** *The Gini coefficient to calculate the homogeneity of each split*

$$g(t) = \sum_{j=1} p(j/t)p(i/t)$$

**Equation 6: Gini coefficient**

*where the sum extends over all categories, p(j/t) is the probability of category j at the node t and p(i/j) is the probability of misclassifying a category j case as category i.*

CART works by looking at the sample space and dividing that in two. Then it looks at each of those two parts and divides each of them, and so on. How it makes the choice of where to cut is based on how similar (or homogeneous) each half would be for each potential cutting place. The Gini coefficient then can calculate how homogenous each resulting group is. The tree generated then uses a pruning mechanism based strictly on the training data that begins with a cost-complexity measure defined below.

**Definition 4** *The cost-complexity measure defined as*

$$Ra(T) = R(T) + a|T|$$

**Equation 7: Cost-complexity measure**

*where R(T) is the training sample cost of the tree, |T| is the number of terminal nodes in the tree and a is a penalty imposed on each node.*

The Gini coefficient, combined with the pruning, comprises the CART algorithm which is also an appropriate analysis technique for this data.

#### *3.3.5.4 Random Forests*

Random Forests is a learning algorithm characterized as an ensemble method as it is an ensemble of decision trees. The diversity introduced in these decision trees is through randomization, training subsets, and feature subsets.

*Step 16*: Apply appropriate learning algorithm

What is clear at this point is that the qualitative data is not coded in any quantitative way. The qualitative text is analyzed using the methodology described above. The resulting quantitative outputs are then fed directly into the appropriate learning algorithms. The output is then analyzed for significance. These are revealed in the results section.

## 3.4 Datasets, collection, and processing

Table 3 (below) enumerates the Descriptive statistics on Qualitative and Quantitative Data. This important to get clear on nature, sources, and quatity of the data used.

| Name | Original Source | Source as-used | Inst-ances ($n$) | Time Period | Qual. or Quant Data | Source of scrubbing (if necessary) |
|---|---|---|---|---|---|---|
| Fed Speeches | Fed | Fed | 867 | 1996-2008 | Qual. | Author |
| Fed Minutes | Fed | Fed | 96 | 1996-2008 | Qual. | Author |
| Fed Conference Calls | Fed | Fed | 14 | 1996-2008 | Qual. | Author |
| Fed Testimony | Fed | Fed | 280 | 1996-2008 | Qual. | Author |
| Enron Emails | Enron | CASOS | 619,446 | 1997-2004 | Qual. | CASOS / Author |
| Derivatives (Fed Funds Futures) | Bloom-berg | Bloom-berg | 304 | 1996-2008 | Quant | Bloom-berg / Author |
| Debt (US Treasury) | Treasury | Bloom-berg | 304 | 1996-2008 | Quant | Bloom-berg / Author |
| S&P 500 | S&P | Bloom-berg | 3,020 | 1997-2008 | Quant | Bloom-berg / Author |
| Equity (Enron) | NYSE | Bloom-berg | 6,135 | 1980-2004 | Quant | Author |

**Table 3: Descriptive statistics on Qualitative and Quantitative Data**

Table 4 (below) shows the raw qualitative data, the texts under evaluation. This is important in noting the varied data available for the public policy study and the large amount of data available through the Enron email corpus.

| STUDY | DATASET | TOTAL DATA AVAILABLE | DATA USED FOR ANALYSIS | DAYS UNDER EACH DATA SET |
|---|---|---|---|---|
| Enron | Enron Emails | 619,446 | 449,442 | 2,191 |
| Federal Reserve (Fed) | Speeches | 867 | 867 | 730 |
| | Conference Calls | 14 | 14 | 14 |
| | Minutes | 96 | 96 | 96 |
| | Congressional Testimony | 280 | 280 | 258 |
| | Aggregate Fed Data | 1,257 | 1,257 | 1,040 |

**Table 4: Raw Data used in correlation analysis**

### 3.4.1 1st dataset: Public communications by members of the Federal Open Market Committee.

#### *3.4.1.1 Source*

The speeches are first acquired through publicly available data from the US Federal Reserve Board of Governors (BoardOfGovernorsOfTheFederalReserveSystem, 1996-2011). Speeches by officers of the Federal Reserve system started to become public in the 1990 through an act of the Fed itself. A list of the Fed speeches can be found in Appendix I. A full sample of a speech transcript from a Fed Official can be found in Appendix II.

#### *3.4.1.2 Preprocessing*

There are three ways in which the data was processed.

1. Speeches were eliminated if they are off-topic or less than two transcribed pages. This study started with 145 speeches of which 19 were eliminated to leave 126.

2. Extraneous information is stripped out that is peculiar to the Fed's transcriptions. These include items such as the standard disclaimers given by The Fed (when present), the information on the person making the speech as well as the venue for the speech.

3. Using the software tools available through Automap (Carley, Columbus, et al., 2011), a list of common words (a 'delete list', see Appendix III) is then mapped onto the collection of speeches. In automap, punctuation is removed, lowercase is forced, then a thesaurus is mapped to avoid duplication of similar words (see thesaurus used in appendix IV). With only those words in the thesaurus remaining, a semantic network is created using a bi-directional window size of seven.

For those days where multiple speeches are given, the network measures are averaged. For those dates that occur when there are no financial data available (e.g., when the bond market is closed), the next day is used.

### 3.4.2 2nd Dataset: Enron Emails

This is a data set that was made public after the collapse of Enron. This data was collected from that release. Substantial work by other researchers has made much of this data clean enough for use on a variety of research questions. The data used for the questions raised in this thesis will be collected from the results of this data cleaning work that has already been preformed. This research will attack one such question and will be processed in a method appropriate to the research question. The full presentation of research using this data is contained within Chapter five. A further description of this data is presented in that chapter.

### 3.4.3 3rd Dataset: Financial Data

This data set is collected in the manner described earlier in the specific experiments. For clarity in this document, if not the final thesis, some of that information on the actual collection will be repeated here.

There are two types of financial data collected: prices of Fed Funds Futures and prices of US government debt. Financial market data is collected directly from the Bloomberg Financial Markets Data service and Monetary Policy data from the Fed. This data is taken as generally reliable, however screens were performed against data provided by other sources (Reuters and Dow Jones) for random and non-random checks of consistency. With this method, no errors were found in the data collection process.

Both the US Debt and Fed Funds Futures are transformed from prices to yield. $y_t = p_t - 100$ from $t = 1$ to $n$, where $n$ is the number of observations. The 30-year Bond was not included as a dependent variable because it was included only halfway through the study (i.e., in 2007). Although the prices of these derivatives are taken throughout each trading day, the concern of this research is inter-day movements so the results are also inter-day.

The term ‘Fed Funds’ (rate) refers to a publicly announced interest rate target established, and regularly reviewed, by the Federal Reserve Open Market Committee. Fed Funds is the interest rate banks charge each other for loans. Derivatives of these Fed Funds are a mechanism through which market participants can express their view on the future direction of interest rate decisions by the FOMC. These ‘Fed Funds Futures’ are derivatives that indicate the public market perception (Robertson & Thornton, 1997) on the future Fed Funds rate.

Fed Funds Futures are traded as contracts with two year maturities. Two sample contacts used for this experiment are those those expiring in December 2007 (ticker FFZ7) and December 2008 (ticker FFZ8). Fed Funds Futures will be in a tight band around the actual Fed Funds rate. These derivatives differ in nature from those with underlying assets in equities or commodities. Price movements reflect this difference in the nature of the security. The prices of the Fed Funds Futures contracts used in this study are listed in the Figures below. These are taken in the manner described earlier in this Dissertation. They are presented here to sho the varied trends against which the framework results will ultimately be compared. No contract has price trends that match any other contact. The details of the contracts and the numerical price histories are listed in Appendix V.

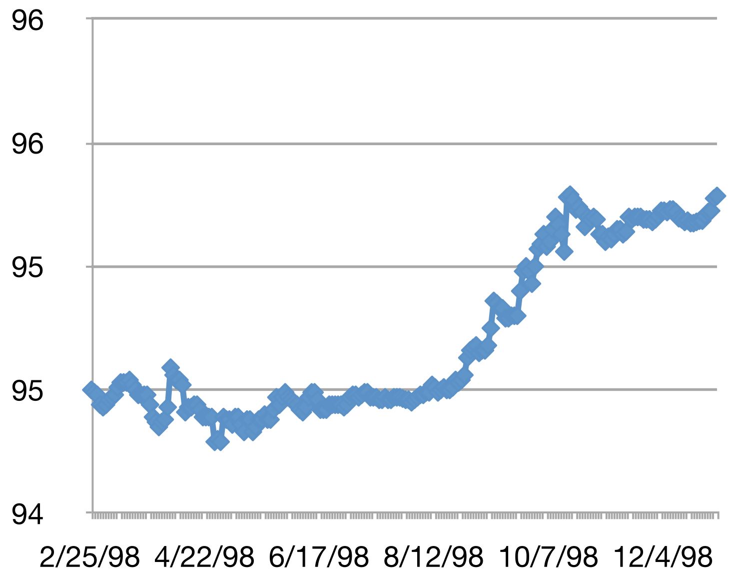


**Figure 10: Fed Funds Futures Contract expiring December 1998 (y-axis represent closing-day contract price in USD)**

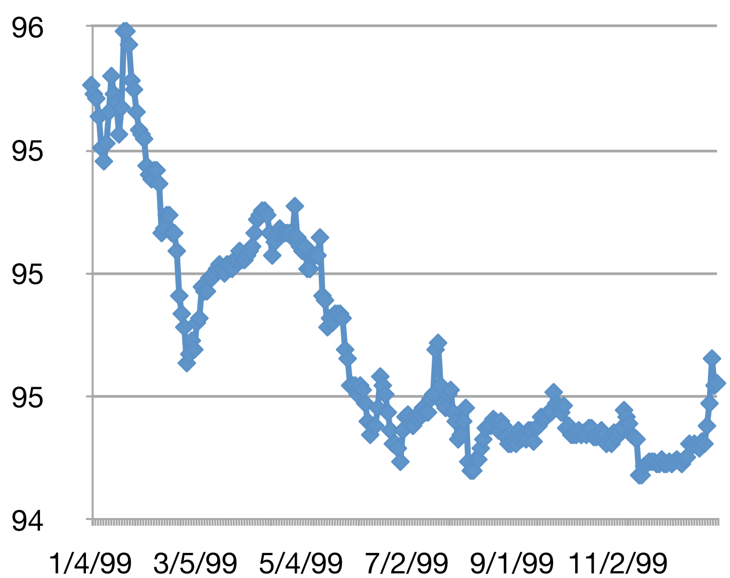


**Figure 11: Fed Funds Futures contract expiring December 1999 (y-axis represent closing-day contract price in USD)**

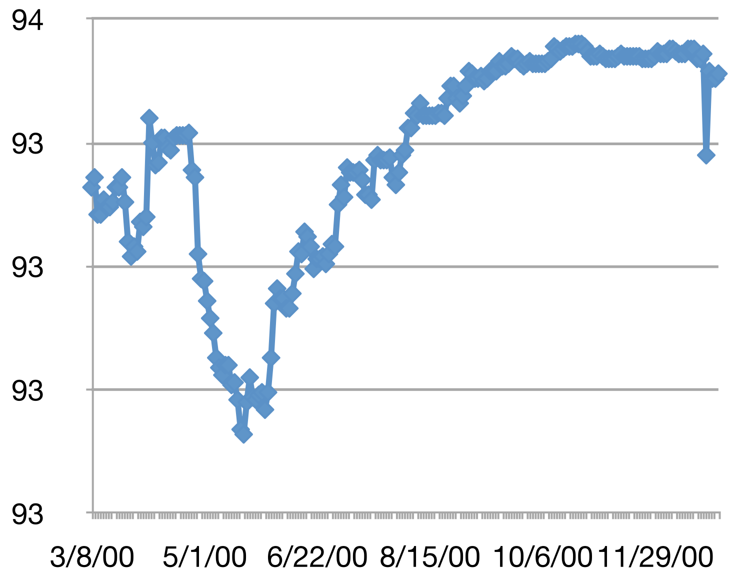


**Figure 12: Fed Funds Futures contract expiring December 2000 (y-axis represent closing-day contract price in USD)**

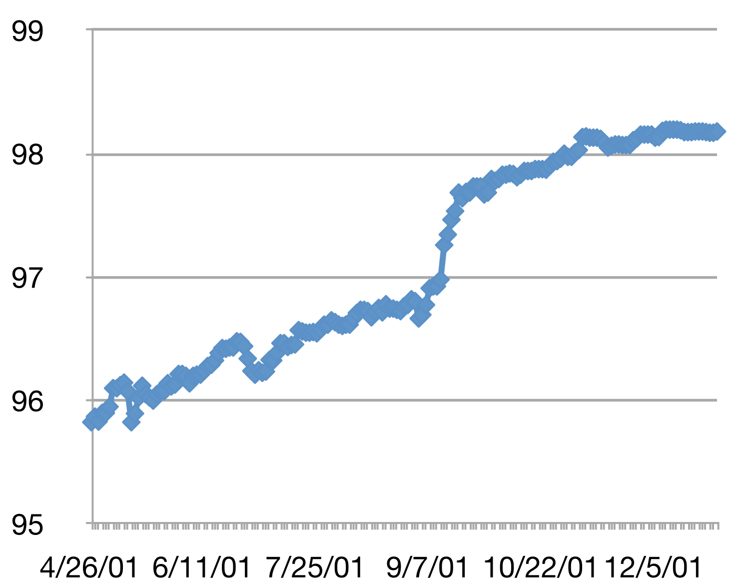


**Figure 13: Fed Funds Futures contract expiring December 2001 (y-axis represent closing-day contract price in USD)**

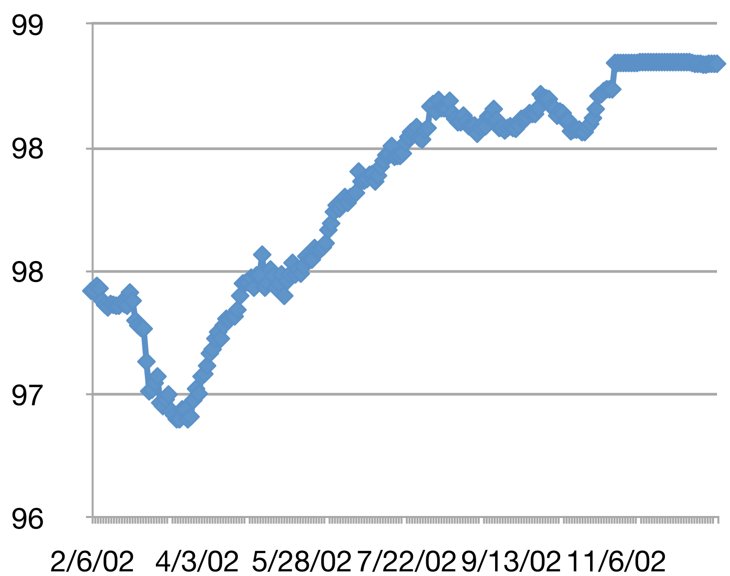


**Figure 14: Fed Funds Futures contract expiring December 2002 (y-axis represent closing-day contract price in USD)**

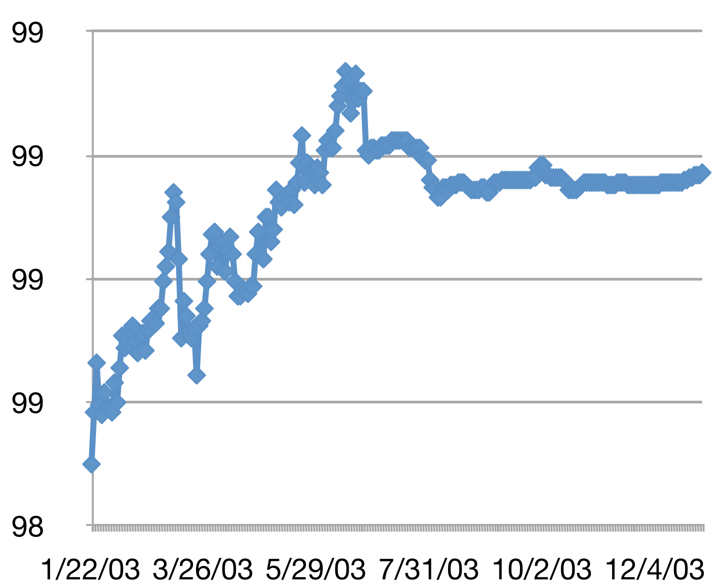


**Figure 15: Fed Funds Futures contract expiring December 2003 (y-axis represent closing-day contract price in USD)**

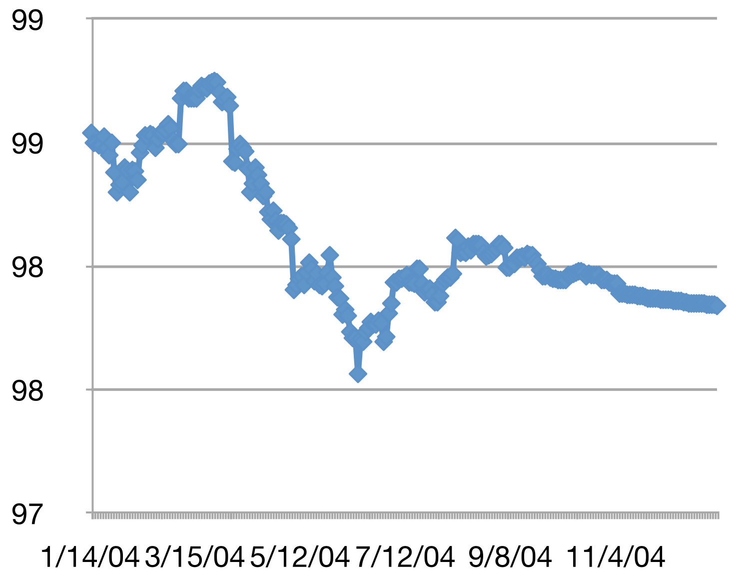


**Figure 16: Fed Funds Futures contract expiring December 2004 (y-axis represent closing-day contract price in USD)**

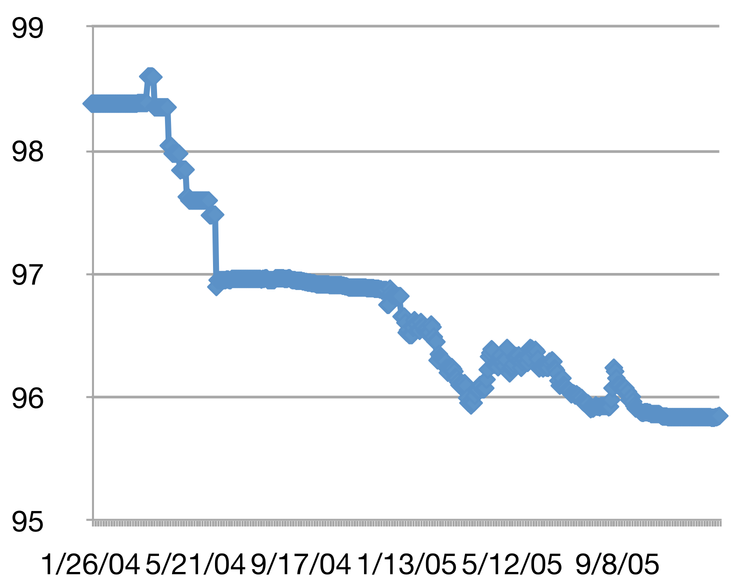


**Figure 17: Fed Funds Futures contract expiring December 2005 (y-axis represent closing-day contract price in USD)**

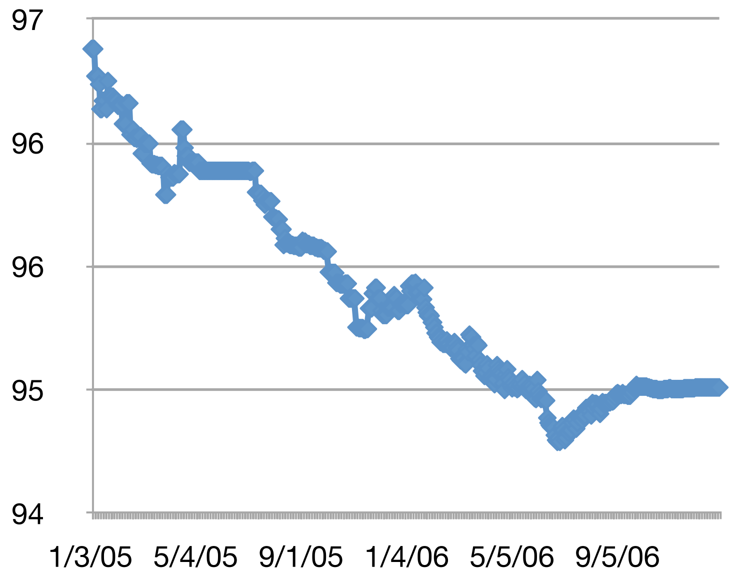


**Figure 18: Fed Funds Futures contract expiring December 2006 (y-axis represent closing-day contract price in USD)**

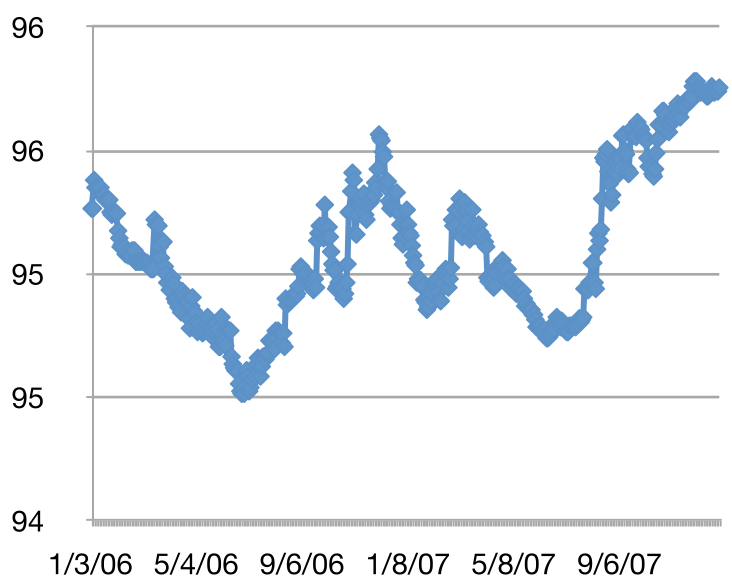


**Figure 19: Fed Funds Futures Contract expiring December 2007 (y-axis represent closing-day contract price in USD)**

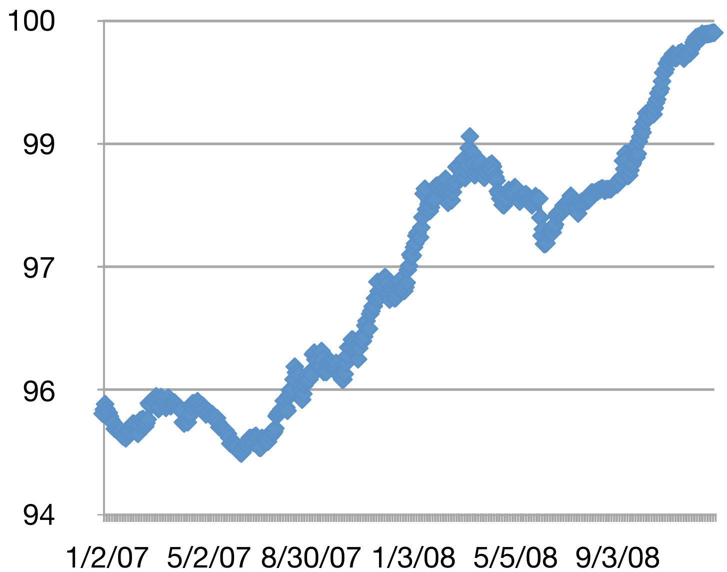


**Figure 20: Fed Funds Futures contract expiring December 2008 (y-axis represent closing-day contract price in USD)**

One set of dependent variables is U.S. treasury debt with maturities from six months to thirty years. These prices are taken from the closing printed price every day that debt markets are open in the United States. Of course the prices are inversely related to their the yield and the short-term debt yield is closely correlated to the prices on the Fed Funds and Fed Funds Futures contracts.

# 4 Analyzing public policy statements with infrequent speech transcripts compared to financial data

Speeches given by decision makers within Central Banks are subject to frequent and careful analysis. However, a systematic process for their evaluation has remained elusive. This chapter introduces a methodology for a systematic process in the form of a semantic network that can be used to augment existing approaches. The approach suggests a correlation between the new systematic method and public market securities data.

## 4.1 Introduction to study of Public Policy Data

There are many studies that attempt to find a correlation between the public pronouncements and monetary policy or the U.S. Treasury Yield curve (Fleming & Remolona, 2001), (Fleming & Remolona, 1997), (M. W. Berry & Browne, 2005). The study of these speeches are interesting for at least a few reasons:

- They are already widely followed by the public
- The subject matter of the speeches lends itself to study because they are tightly constrained around U.S. Monetary Policy (Issing, 2005).
- The speeches are given at regular intervals (See Appendix I).
- The speeches are given by a small and predictable group that make up the membership of the Fed Governing Body (Fox et al., 2005).
- The group itself looks to understand the characteristics of the speeches' impact. (Lucca & Trebbi, 2009)
- There are related or tangential policy bodies against which future research may apply findings (Rosa, 2007), (Reeves & Sawicki, 2007).

Some approaches in computational linguistics such as Latent Semantic Analysis have been applied to central bank speeches , but the results are "nuanced," (Boukus & Rosenberg, 2006). This chapter concerns itself with establishing a system for analyzing the texts that can be routinely applied to speeches given by the central bank officials. Using just the public speeches, the approach seeks to find a correlation with security prices.

## 4.2 Background on study of Public Policy Data

### 4.2.1 Background on US Federal Reserve

The Central Bank of the United States is called the US Federal Reserve ("The Fed"). The Fed is comprised of twelve regional Banks and a central administrative body based in Washington. The decision making body of the Fed is the Federal Open Market Committee ("FOMC") whose annually rotating voting membership is comprised of a combination of seven presidential appointees (known as Fed Governors, the posts for which are not always full as they require Senate approval) who work from the Washington Headquarters and five of the twelve regional Fed Presidents (Fox et al., 2005).

The members of the Fed give public speeches at regular intervals throughout the year (see appendix I) (Fox et al., 2005). As opposed to FOMC meeting minutes, the speeches are intended to provide security market participants some insight into the direction of Monetary Policy (Danker & Luecke, 2005). The Fed has many mechanisms for expressing Monetary Policy and many studies have been done on the degree of efficacy of these actions (Fleming & Remolona, 1997). The primary vehicle for expressing monetary policy remains the setting interest the Fed Funds rate around which many other interest rates are linked (Fox et al., 2005).

The minutes of each FOMC meeting are released to the public on the Thursday following the regularly scheduled meeting. The lag between a meeting and the release of the minutes is about six weeks. Transcripts of meetings for an entire year are release to the public with a five-year lag (Richmond, 2012).

### 4.2.2 Financial Data Background

Some research suggests that the degree to which The Fed is effective in setting expectations on Monetary Policy is the extent to which the Treasury Yield curve retains an upward sloping shape (Gong & Remolona, 1996). Changes in the Fed Funds rate has the biggest effect on the securities with the shortest maturities, or at the 'short end of the curve' (Fleming & Remolona, 1999a), (Gong & Remolona, 1996).

The Fed Funds rate itself is expressed as a target rate (e.g., 3.25%) around which little fluctuation occurs. For purposes of this research, it is treated as a fixed rate that changes only in increments of 25 basis points (i.e., 0.25%). Investors can express views on the probability of a change in the Fed Funds through a mechanism of Fed Funds Futures. These are contracts traded publicly that come into existence each month with a two-year expiration.

Using a variety of techniques, communications from the Fed have been studied in many different from the degree of correlation to Bond prices (Fleming & Remolona, 1997), to the volume of trading in the debt market (Fleming & Remolona, 1999b), to how the voting within the FOMC effects prices (Boukus & Rosenberg, 2006), (Havrilesky & Schweitzer, 1990).

Other linguistic approaches have been brought to bear on the analysis of FOMC speeches from innovations in the application (Nasukawa & Yi, 2003) to innovations in the processing of text itself (Blei, Ng, & Jordan, 2003), (Joachims, 1999). Some of this work centers on the feedback loop of the announcements changing response to the markets which itself has an impact on the markets (Brown & Cliff, 2005), (Barth III, Remolona, & Wooldridge, 2001).

The analysis of the speeches themselves is against a backdrop of a tension within the FOMC to focus on strict rules of market engagement versus more subtle actions (Meade, 2002), (Reinhart & Sack, 2006).

While some research has attempted to develop a sophisticated interpretation of a systemic analysis such ‘a term structure of announcements’ (Fleming & Remolona, 2001), other research takes a large data set of speeches from many different speakers and concludes that the results are ‘nuanced’ (Boukus & Rosenberg, 2006).

### 4.2.3 Semantic Networks & Sentiment Classifiers

While many explore ways to make market bets on sentiment (Hofmann, 1999), (Nasukawa & Yi, 2003) or other forms of analysis (Luss & d'Aspremont, 2008) of qualitative Central Bank communications, the results have not been strong (Reeves & Sawicki, 2007), (Rosa, 2007). Some difficulty in sentiment classification in this domain (Blitzer, Dredze, & Pereira, 2007), (Mani & Bloedorn, 1999), (Pang, Lee, & Vaithyanathan, 2002), (Wang, Joshi, & Rosé, 2007) is from the confusion among domain experts (Frendreis & Tatalovich, 2000). A different approach could be useful.

Semantic network analysis is the use of network analytic techniques on paired associations based on shared meaning as opposed to paired associations of behavioral or perceived communication links (Doerfel, 1998). Semantic Networks have been applied been explored in a variety of circumstances from large-scale news reporting (Godbole et al., 2007) to email (Berry & Browne, 2005), (Woods, 1975), (Diesner & Carley, 2005). To help with understanding, the Figure below presents a classic simple semantic network.

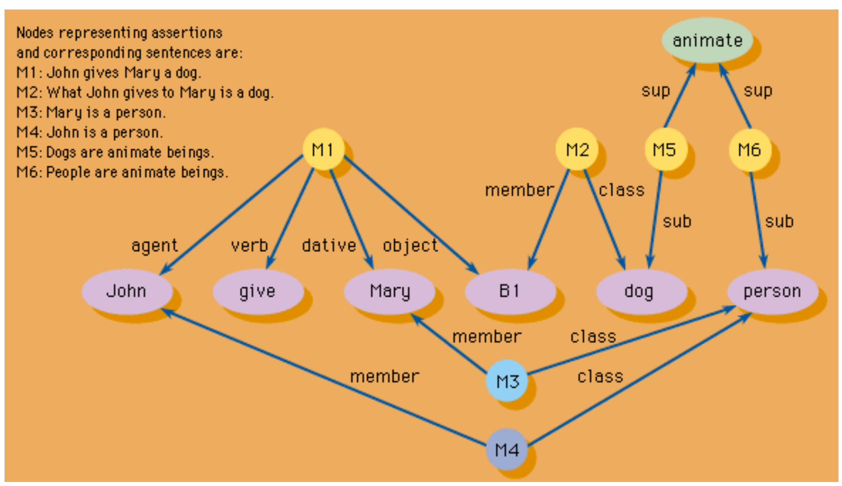


**Figure 21: Classic Simple Semantic Network**

The approaches in Semantic Network Analysis vary depending upon the research question. For some applications, the appropriate methodology is to tag the words as having characteristics such as people or places (Diesner et al., 2005). Other approaches use the method of looking at the relationship of words to each other (Doerfel, 1998). The analysis of Semantic

Networks itself getting richer with analytical tools to measure the network (Carley & Kaufer, 1993). This richness combined with the relational dependence inherent in Semantic Networks suggests a better path toward a systematic analysis of public policy speeches.

## 4.3 Methodology of Public Policy Data Study

The approach toward developing a systematic method of evaluating public policy speeches is to develop a semantic network for a select group of speeches. The network measures generated from a semantic network are then compared to financial data around which the speeches most related.

The speeches collected are from the FOMC in two select years in which the speeches are publicly available. The data from that analysis of these speeches are then compared to various public market interest rate indicators around which the FOMC speeches have the most direct influence.

Since its creation in 1913, The Fed did not release qualitative information about Monetary Policy. This changed in 1996. However, only since the Chairmanship of The Fed transferred to the current Chair, Ben Bernanke, in February of 2006, did The Fed explicitly express a willingness to become more transparent in its communications. Measurements of the semantic network are then compared to quantitative financial data to determine if there is a relationship and if so, the nature of the relationship.

First, qualitative data is collected in the form of speeches. Speeches by FOMC members have the benefit of being clearly labeled in the body of the text for date, location, speaker, and topic. Some texts are excluded from this study: prepared congressional testimony, answers under congressional questioning, FOMC board meeting minutes, and speeches given about bank regulatory matters. Second, financial data is collected in the form of the full U.S. Treasury Yield curve and Fed Funds Futures contracts; all of these contracts expire in the last day of each calendar year (Morgan & Kogan, 2010).

### 4.3.1 Preprocessing financial data

There are two types of financial data collected: prices of Fed Funds Futures and prices of US government debt. Financial market data is collected off of the financial news vendor Bloomberg. Monetary Policy data is taken directly from the US Federal Reserve. See Appendix V for the data used. This data is taken as generally reliable, however screens were performed against data provided by other sources (Bloomberg, Reuters, and Dow Jones) for random and non-random checks of consistency.

Date and time data can be presented in non-standard formats both between suppliers and over time as standards have changed. Despite most financial data being very accurate, even occasional errors in omission, duplication, or formatting can occur as the data is moved from current to archived data. Sometimes the conversion of formatting can warp data on the edges. Anomolies can be checked and then compared with other data sources. There is rarely a ‘source of record’.

Other issues in formatting are treated differently in preparation for further analysis. In this study, both the US Debt and Fed Funds Futures are transformed from prices to yield. $y_t = p_t - 100$ from $t = 1$ to $n$, where $n$ is the number of observations. The 30-year Bond was not included as a dependent variable because it was included only halfway through the study (i.e., in 2007).

#### *4.3.1.1 Fed Funds Futures*

The term ‘Fed Funds’ (rate) refers to a publicly announced interest rate target established, and regularly reviewed, by the Federal Reserve Open Market Committee. Fed Funds is the interest rate banks charge each other for loans. Derivatives of these Fed Funds are a mechanism through which market participants can express their view on the future direction of interest rate decisions by the FOMC. These ‘Fed Funds Futures’ are derivatives that indicate the public market perception (Robertson & Thornton, 1997) on the future Fed Funds rate.

Fed Funds Futures will be in a tight band around the actual Fed Funds rate. These derivatives differ in nature from those with underlying assets in equities or commodities. Price movements reflect this difference in the nature of the security. Graphs of these price movements may be seen in Figures 6 and 7. For this data there is exactly one data point for every day the market is open during the year.

#### *4.3.1.2 US government debt*

One set of dependent variables is U.S. treasury debt with maturities from one month to thirty years. These prices are taken from the closing printed price every day that debt markets are open in the

United States. Of course the prices are inversely related to their the yield and the short-term debt yield is closely correlated to the prices on the Fed Funds and Fed Funds Futures contracts.

### 4.3.2 Creating Semantic Network measures

A semantic network is created using the Network Analysis Tool *ORA*. From this network, 86 measures are taken (see section 3.3.3). These become the candidate independent variables. Further information on those measures themselves are detailed in the 'data' section below and in the 'Steps to develop framework' subsection of the 'Proposed Methods' section of the main document.

### 4.3.3 Relationships with Linear Regression

Two learning algorithms are applied to the data. The first is Linear Regression. With the creation of 86 independent variables from the semantic network and 15 dependent variables created from the financial data, the dates are included in the independent variable count by transforming them into a series (i.e., 1, 2, 3, etc.).

### 4.3.4 Relationships with CART

The analysis covers all combinations of dependent variables against which are taken all 19 independent variables.
The $R^2$ for the CART output is generated as
$R^2 = 1 - SSE / SST, where\ SSE / SST = CART$ Relative Error

### 4.3.5 Data

#### *4.3.5.1 Independent Variables*

Created from the measurements of the semantic network are the independent variables. *See section 3.3.3 for the comphrensive list of candidate Independent Variables from which the reduced set is to be chosen. See section 4.4 for reduced (clustered) data IV data set.* The Attritubes of the qualitative data from which the independent variables are generated are categorized in the Figure below in order to highlight the different data under analysis.

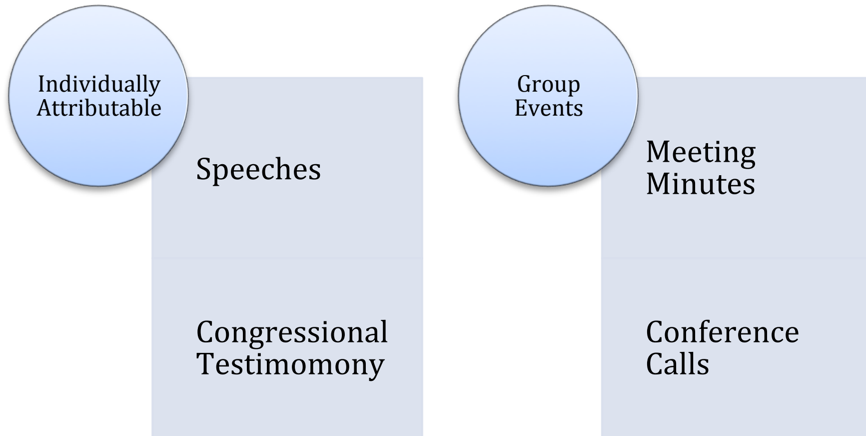


**Figure 22: Attribution of Public Policy qualitative data**

#### *4.3.5.2 Dependent Variables*

The source and nature of the data that makes up the dependent variables is described in detail later in this document and in this Dissertation under the 'Datasets' section. There are fifteen dependent variables considered. Two of those might seem to be particularly inappropriate: Changes in Fed Funds target rate and the actual Fed Funds were excluded as the first one changes too infrequently and the second one has daily moves that are without regard to interest rate expectations. The table below lists all of the dependent variables.

| DEPENDENT VARIABLES: Financial Data x-x |
|---|
| Changes in Fed Funds target rate (increase or decrease) |
| Fed Funds rate target |
| Fed Funds Futures contract expiring in December of current |

| DEPENDENT VARIABLES: Financial Data x-x |
|---|
| year |
| Fed Funds Futures contract expiring in December of subsequent year |
| 1 month Treasury Bill |
| 3 month Treasury Bill |
| 6 month Treasury Bill |
| 1 year Treasury Note |
| 2 year Treasury Note |
| 3 year Treasury Note |
| 5 year Treasury Note |
| 7 year Treasury Note |
| 10 year Treasury Note |
| 20 year Treasury Note |
| 30 year Treasury Note |

**Table 5: Dependent Variables**

### 4.3.6 Text Processing for FOMC data

*See Section 3.3.1*

## 4.4 Public Policy Study Results

Among the various contexts under which the qualitative data for this section is collected, the stability varies. Where there exist the constantly changing (e.g., the date), the slowly changing (e.g., the FOMC Chair) and the ever-present (the institution itself (i.e., the FOMC), the other variables do not fit so neatly into this continuum in comparison to each other. Some venues are constant (e.g., for congressional testimony) while speeches are not. For example, while conference calls exist and are included in analysis, they are not considered by themselves because of the small sample size. Other examples may be examined in future work such as the relative power of FOMC members other than the chairman. Other examples of future work are described in more depth in that section of this document. Appropriately capturing the possible combinations for analysis generated twenty files seen in the tables below.

All of the sources are aggregated in file 1. This file therefore contains the largest amount of data. Even in this file, there is insufficient qualitative data to match all of the available financial data. After 1) all non-numerical data is removed; 2) standard symbols are filled into empty cells; and 3) all the column names as placeholders for the IVs are replaced, the file is readable by the appropriate statistical tools (such as R, which was used in this case). The initial column count is 103 (including the date). The categories of qualitative data available for analysis in this Dissertation is listed in Table 6 (below).

Initial number of columns 103 (including the date)
Initial number of rows: 5844

Initial number of dependent variables: 19
Initial number of independent variables: 86

The Initial row count of 5844 (observations) is reduced to 1040 spanning the dates 13June1996-12March2009. We now have 1040x16 dependent variables and 1040x86 independent variable candidates.

| CATEGORIES OF QUALITATIVE DATA AVAILABLE ON FOMC | |
|---|---|
| i | All Speeches |
| ii | Congressional Testimony (of FED Chairman) |
| iii | Minutes of regular FOMC meetings |
| iv | FOMC Conference Calls |

**Table 6: Categories of qualitative data available for analysis of FOMC**

It is worth considering the degree to which specific sources of qualitative information are more valuable. Toward that end, the qualitative information has been separated out according to the following schedule *Cuts of statistical Analysis for FOMC Qualitative data*.

Further, some days may have more than one measurement. For these days, the measures were averaged according to the schedule below. Other days may have no data. Two versions of each of the ten files were created in the schedule below (Table 7). One has blanks for the missing data. The other ten give inferential data to the spaces where blanks did not exist. Inference generated: Data from most recent observation carried over until next observation The blanks were filled in with the previous data until new days appears. For example, if there was data '4.6' on May 8, '4.5' on May 9, and nothing more until May 20, May 10-19 was filled in with '4.5'. This inferential method was considered, but discarded in the final analysis as possibly confounding the results. With the qualitative data collected coming from various sources, the first and last dates for data collection vary accordingly. These are captured into two of the columns.

| CUTS OF STATISTICAL ANALYSIS FOR FOMC QUALITATIVE DATA | | *Data Start* | *Data End* | *Indep. Var.* | *n* |
|---|---|---|---|---|---|
| File 1 | Combine all measures (i)-(iv) averaged together per date | 13Jun96 | 3Dec09 | 19 | 1040 |

| CUTS OF STATISTICAL ANALYSIS FOR FOMC QUALITATIVE DATA | | *Data Start* | *Data End* | *Indep. Var.* | *n* |
|---|---|---|---|---|---|
| File 2 | Measure (i) by itself, Speeches | 13Jun96 | 11May09 | 17 | 730 |
| File 3 | Measure (ii) by itself, Testimony | 25Jun96 | 3Dec09 | 16 | 258 |
| File 4 | Measure (iii) by itself, Minutes | 5Feb97 | 16Dec08 | 18 | 96 |
| File 5 | Measure (iv) by itself, Conference Calls | 3Jan01 | 7Oct08 | 16 | 14 |
| File 6 | All measures (i)-(iv), but only from Chairman Greenspan or Chairman Bernanke ("G or B") | 13Jun96 | 3Dec09 | 13 | 400 |
| File 7 | Measure (i) just for G or B, Speeches G or B | 13Jun96 | 11May09 | 17 | 293 |
| File 8 | Measure (ii) just for G or B, Testimony, G or B | 26Jul96 | 3Dec09 | 15 | 110 |
| File 9 | Combine measure (i) and (ii) just for G or B | 3Jun96 | 3Dec09 | 13 | 400 |
| File 10 | All measures (i)-(iv), except from G or B in (i) and (ii) | 18Jun96 | 30Nov09 | 18 | 753 |

**Table 7: Cuts of Statistical Analysis for FOMC qualitative data**

Visualized in the Figure below, File 5 is clearly untenable as a subject for analysis by itself. The raw data is presented in this report for purposes of completeness, but the measures of fit for any of the models in this file are too rare to be of value by themselves.

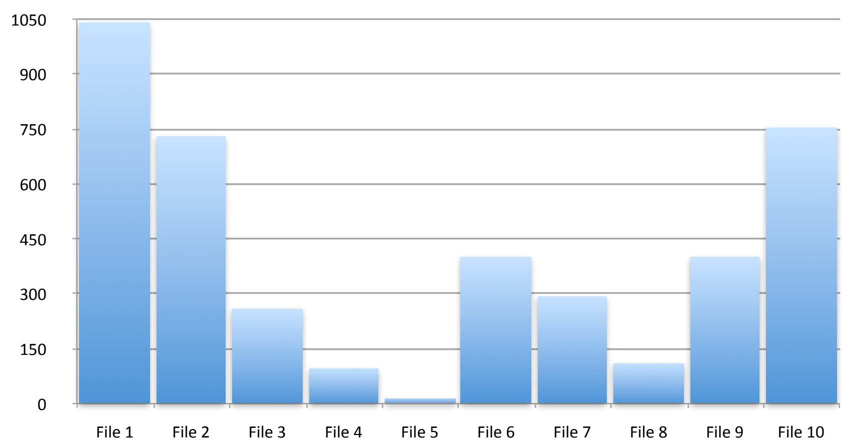


**Figure 23: Number of observations for file separations chosen in analysis of FOMC qualitative data**

This analysis ultimately gives us both the degree to which any independent variables are correlated to any combination of dependent variables, but also the combination of variables themselves. The detail of each analysis is then summarized for both the $R^2$ and the independent variables found.

Another early step in performing analysis between the dependent and independent variables is the identification of correlated, or redundant, variables. In order to reduce the number of independent variables, pair-wise correlation has clustered them with an exemplar for each group. For this, 'apcluster' in 'R' is used (Bodenhofer et al., 2011; Frey & Dueck, 2007) to perform affinity propagation based on negative distance matrix built upon the pair-wise Pearson's correlation coefficients. The 86 independent variables clustered into 19 groups. Of the 86 considered the tables below list those that were found to be constant and those that were correlated with each other. The new subset of independent variable is thus made up of 1040 observations and 19 independent variables

The next table (Table 8, below) lists those Independent variables by the group with which they are correlated. For these, one of the variables is used while the others are eliminated from further analysis. These are then summarized in Table 9 (below).

| CORRELATED INDEPENDENT VARIABLE MEASURES BY GROUP WITH WHICH THEY ARE CORRELATED | |
|---|---|
| INDEPENDENT VARIABLE **(REPRESENTATIVE INDEPENDENT VARIABLE IN BOLD)** | **GROUP** |
| | |
| **NumberofConceptnodes** | **1** |
| Count.Column.SemanticNetwork | 1 |
| LinkCount.SemanticNetwork | 1 |
| Count.Node.SemanticNetwork | 1 |
| Count.Row.SemanticNetwork | 1 |
| | |
| **OverallComplexity** | **2** |
| DensityClusteringCoefficientSemantic-NetworkAverage | 2 |
| CognitiveDistinctivenessAverage | 2 |
| CognitiveExpertiseAverage | 2 |
| CognitiveResemblanceAverage | 2 |
| CognitiveSimilarityAverage | 2 |
| ReduancyColumn.SemanticNetwork | 2 |
| Correlation.Distinctiveness.Semantic-NetworkAverage | 2 |
| Correlation.ExpertiseSemanticNetworkAverage | 2 |
| Correlation.Resemblance.SemanticNetwork.-Average | 2 |
| CorrelationSimilarity.SemanticNetworkAverage | 2 |
| Density.SemanticNetwork. | 2 |
| Efficiency.SemanticNetwork. | 2 |
| Centrality.InformationSemanticNetworkAverage | 2 |
| BouarySpanner.Potential.SemanticNetwork.-Average | 2 |
| Reduancy.Row.SemanticNetwork. | 2 |
| SimmelianTies.SemanticNetwork.Average | 2 |
| | |
| **Meta.MatrixHammingDistance** | **3** |
| CommunicationHammingDistance | 3 |
| KnowledgeHammingDistance | 3 |
| TaskHammingDistance | 3 |
| | |
| **Centrality.Authority.SemanticNetwork.Average** | **4** |
| Centrality.Eigenvector.SemanticNetwork.-Average | 4 |

| CORRELATED INDEPENDENT VARIABLE MEASURES BY GROUP WITH WHICH THEY ARE CORRELATED | |
|---|---|
| INDEPENDENT VARIABLE **(REPRESENTATIVE INDEPENDENT VARIABLE IN BOLD)** | **GROUP** |
| Centrality.EigenvectorPer-Component.SemanticNetwork.Average | 4 |
| Network-Centralization.Eigenvector.SemanticNetwork | 4 |
| Centrality.Hub.SemanticNetwork.Average | 4 |
| Interlockers.SemanticNetwork.Average | 4 |
| Transitivity.SemanticNetwork. | 4 |
| | |
| AverageDistance.SemanticNetwork. | 5 |
| **Speed.Average.SemanticNetwork.** | **5** |
| Efficiency.Global.SemanticNetwork. | 5 |
| Centrality.InverseCloseness.Semantic-NetworkAverage | 5 |
| Speed.Minimum.Semantic_Network. | 5 |
| Network_Levels.Semantic_Network. | 5 |
| | |
| **Breadth.Column.Semantic_Network.** | **6** |
| Diffusion.Semantic_Network. | 6 |
| Link_Count.Pooled.Semantic_Network. | 6 |
| Breadth.Row.Semantic_Network. | 6 |
| Bouary_Spanner.Semantic_Network._Average | 6 |
| | |
| Centrality.BetweennessSemantic_Network_Average | 7 |
| Capability.Semantic_Network._Average | 7 |
| **CentralityColumn_Degree.Semantic_Network._ Average** | **7** |
| Exclusivity.Semantic_Network._Average | 7 |
| Exclusivity.Complete.Semantic_Network._Average | 7 |
| Centrality.In_Degree.Semantic_Network._Average | 7 |
| Interdepeence.Semantic_Network. | 7 |
| Centrality.Out_Degree.Semantic_Network_Average | 7 |
| Radials.Semantic_Network._Average | 7 |
| CentralityRow_Degree.Semantic_Network._Average | 7 |
| Centrality.Total_Degree.Semantic_Network._ Average | 7 |
| | |
| **Communicative_Need.Semantic_Network.** | **8** |
| | |

| CORRELATED INDEPENDENT VARIABLE MEASURES BY GROUP WITH WHICH THEY ARE CORRELATED | |
|---|---|
| INDEPENDENT VARIABLE **(REPRESENTATIVE INDEPENDENT VARIABLE IN BOLD)** | **GROUP** |
| CentralityBonacich_Power.Semantic_Network_Average | 9 |
| Clique_Count.Semantic_Network._Average | 9 |
| Constraint.Burt.Semantic_Network._Average | 9 |
| **Effective_Network_Size.Burt.Semantic_Network_Average** | **9** |
| Efficiency.Local.Semantic_Network. | 9 |
| Span_Of_Control.Semantic_Network. | 9 |
| Triad_Count.Semantic_Network._Average | 9 |
| | |
| **Hierarchy.Semantic_Network.** | **10** |
| | |
| **Centrality.In.Closeness.Semantic_Network._Average** | **11** |
| | |
| **Network_Centralization.In.Closeness.Semantic_Network** | **12** |
| | |
| Network_Centralization.Betweenness.Semantic_Network | 13 |
| Network_Centralization.Column_Degree.Semantic_Network | 13 |
| **Network_Centralization.In_Degree.Semantic_Network** | **13** |
| Network_Centralization.Out_Degree.Semantic_Network | 13 |
| NetworkCentralizationRowDegreeSemanticNetwork | 13 |
| NetworkCentralizationTotalDegreeSemanticNetwork | 13 |
| | |
| Centrality.Closeness.Semantic_Network._Average | 14 |
| Network_Centralization.Closeness.Semantic_Network | 14 |
| Connectedness.Semantic_Network | 14 |
| Diameter.Semantic_Network | 14 |
| Fragmentation.Semantic_Network | 14 |
| **Isolate_Count.Semantic_Network** | **14** |
| Component_Count.Strong.Semantic_Network | 14 |
| Component_Count.Weak.Semantic_Network | 14 |
| | |
| **Link_CountLateralSemantic_Network** | **15** |
| Component_MembersWeakSemantic_Network_Average | 15 |

| CORRELATED INDEPENDENT VARIABLE MEASURES BY GROUP WITH WHICH THEY ARE CORRELATED | |
|---|---|
| INDEPENDENT VARIABLE **(REPRESENTATIVE INDEPENDENT VARIABLE IN BOLD)** | **GROUP** |
| | |
| **Link_Count.Reciprocal.Semantic_Network** | **16** |
| | |
| **Link_Count.Sequential.Semantic_Network** | **17** |
| | |
| **Link_Count.Skip.Semantic_Network** | **18** |
| | |
| **Upper_Bouedness.Semantic_Network** | **19** |

**Table 8: Correlated Independent Variables in study of Public Policy Documents**

| SUMARY OF INDEPENDENT VARIABLE CANDIDATES | |
|---|---|
| Independent Variables Label | Group |
| NumberOfConceptNodes | 1 |
| OverallComplexity | 2 |
| MetaMatrixHammingDistance | 3 |
| CentralityAuthoritySemanticNetworkAverage | 4 |
| SpeedAverageSemanticNetwork | 5 |
| BreadthColumnSemanticNetwork | 6 |
| CentralityColumnDegreeSemanticNetworkAverage | 7 |
| CommunicativeNeedSemanticNetwork | 8 |
| EffectiveNetworkSizeBurtSemanticNetworkAverage | 9 |
| HierarchySemanticNetwork | 10 |
| CentralityInClosenessSemanticNetworkAverage | 11 |
| NetworkCentralizationInClosenessSemanticNetwork | 12 |
| NetworkCentralizationInDegreeSemanticNetwork | 13 |
| IsolateCountSemanticNetwork | 14 |
| LinkCountLateral.SemanticNetwork | 15 |
| LinkCountReciprocalSemanticNetwork | 16 |
| LinkCountSequentialSemanticNetwork | 17 |
| LinkCountSkipSemanticNetwork | 18 |
| UpperBoundednessSemanticNetwork | 19 |

**Table 9: Summary of Representative Independent Variables after clustering**

With the measurements of the qualitative data categorized and simplified for appropriate statistical analysis, the financial data presents an opportunity for equivalent scrutiny. The US Treasury Yield curve exists along a temporal continuum. This continuum is weighted toward earlier maturity securities. For example, there are several bonds under ten years of maturity, but only one with a maturity of greater than ten years. For all of these reasons, the degree, not the existence of the correlation between the securities is worth examining within the time period that we study. The relationship discovered is presented in Figure 18 (below).


**Figure 24: US Treausry Yield Curve relationships**

This table confirms how the relationships might be expected to behave; adjacent securities on the yield curve have a linear relationship while the farther apart on the yield curve the securities are, the more muddled the relationship.

The relationship under investigation for the purposes of this study is contemporaneous. That is, does a relationship exist between the independent variables and dependent variables within the same time period. There is substantial future work available regarding the degree to which the independent variables may express predictive power over the dependent variables. This is investigated with the methods of this report through the shifting of time periods. Single independent variables might correlate better with single dependent variables when temporally shifted. The investigation was performed for all of the files, but remains relatively consistent. The best shift detected was almost always the first one tested (-5, or a five time period lag or the independent variables). The conclusion is that shifting does not improve the single independent variable performances in correlating with single dependent variables. In those limited spaces where the shift does differ, from -5, it almost always equals 0. This also suggests that there is no use in shifting the data. The visualization of these shifts are shown in the Figure below.

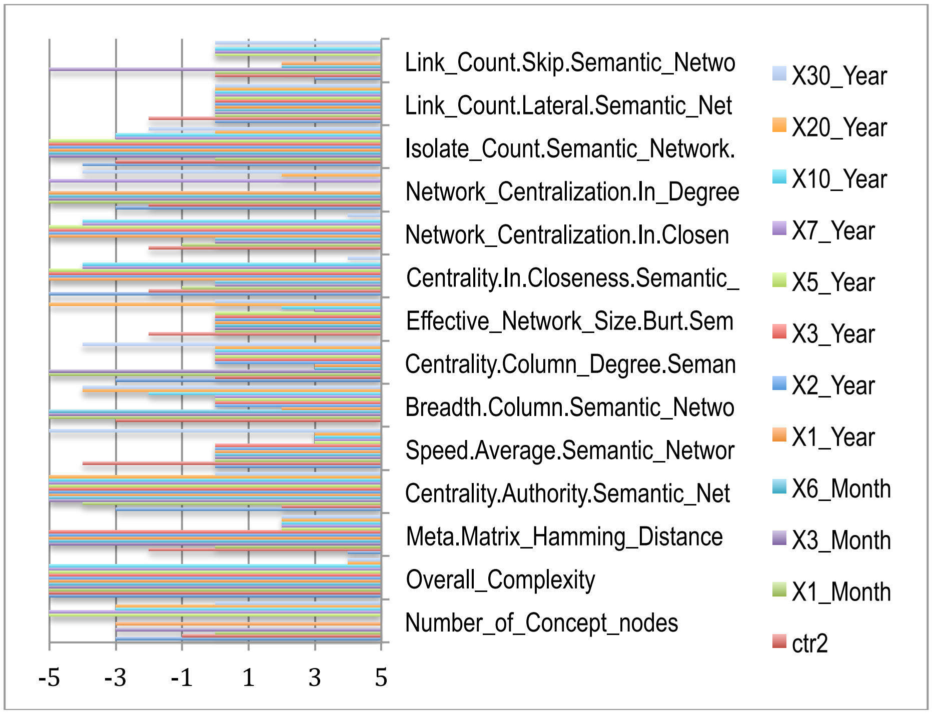


**Figure 25: Results from exploration of time shifting versus contemporaneous comparsions on Federal Reserve Data**

For the analysis of relationships between dependent and independent variables, different models were tested: linear (lm), cart, generalized linear (glm) with Gaussian link function, random forest, and svm (with radial basis function as kernel). The model fitting (with nested feature selection) proceeded by analyzing each dependent variable separately (only for those with at least some numerical information); for each of such dependent variable, a set of optimal independent variables was built as follows:

1. Start with an empty selection of independent variable.
2. According to the given model, select the independent variable that produce the best model and add it to the selected set of independent variables.
3. Set the current best model fitting to the one obtained with the selected set of independent variables.
4. Now start considering again each of the independent variable not previously selected: for each of them, assess the new model fitting when the given independent variable is added to the selected set. The independent variable that, added to the set, increases the model fitting the most is included to the selected set.
5. The procedure stops when either all independent variables have been added or no increase in model fitting is possible.

This output produces two tables: the first one shows, for each dependent variable, the selected subset of independent variables selected by the model fitting (in terms of flags set to 1 for those having been selected). The table demonstrates the degree to which some independent variables are selected as being appropriate matches for analysis to the dependent variables. The analysis has been completed 1) For each of the ten Files for analysis within the Federal Reserve Data; 2) Within each of the five learning models used for analysis; and 3) for both the 'no-shift' and 'shift' scenarios. With each of the ten files for analysis undergoing tests using five models both with and without shifts in time, there are (10x5x2) one hundred Independent Variable Matrices; these are found in their entirely within Appendix VII. The figure below represents the visualization of the very first of these matrices: a no-shift linear model of File 1.

The Figure below shows that for each file, represented by each line, the percentage of dependent variables for which a given independent variable has been chosen. We see some variables such as 'Overall Complexity' being matched with each dependent variable. Other variables such as 'Hierarchy Semantic Network' are chosen by none. What the visualization makes clear is how only *one* of the Independent Variables, 'Communicative Need, Semantic Network' in this case, is the *only one* that has a more complex interaction with the dependent variables as it is being considered by the algorithms for comparison against the dependent variables. This provides a measure of how important the independent variable is overall.

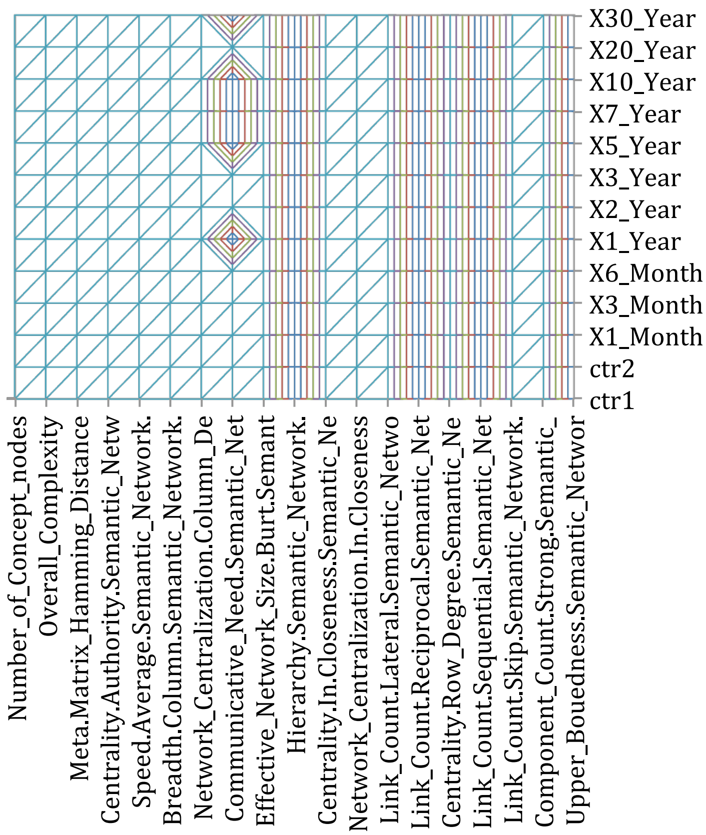


**Figure 26: Example of Independent Variable Selection Results on Federal Reserve Data**

To help with understanding, an example of a modeled DV is shown in the figure below: here the red line represents the original DV (30-Year Treasury) while the blue one is the fitted model, obtained using corresponding IVs found for that particular iteration. The matching of the two suggests the effectiveness of generated model.

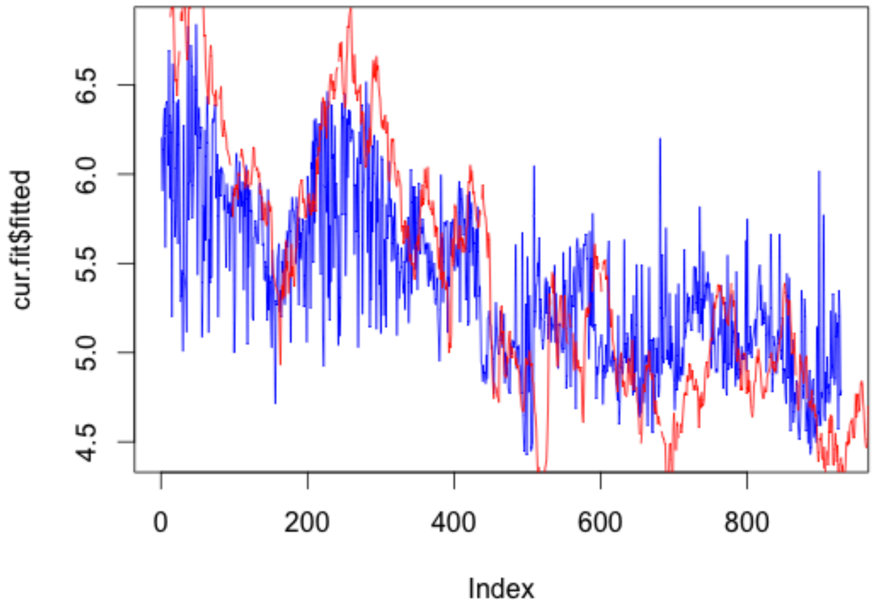


**Figure 27: SVM fitted model (Blue) v. Original Dependent Variable (US 30-Year) over timed observations**

### 4.4.1 Results from Time Shift Analysis

In each of the tables below, the first column gives the name of the dependent variable. The second column gives the performances for the given regression model upon the given dependent variables within the best-shift method. The third column is similar to the second one, but reports the performances for the non-shift model. The performance are pseudo-R-squared values where 'pseudo' refers particularly to the Random Forest model. In that model the actual value can be greater than 1.

### *4.4.1.1 Section (File) 1*

4.4.1.1.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.098301754 | 0.068352615 |
| *ctr2* | 0.089643332 | 0.041579535 |
| *X1_Month* | 0.06177954 | 0.064659425 |
| *X3_Month* | 0.024822444 | 0.022251236 |
| *X6_Month* | 0.024017413 | 0.021331479 |
| *X1_Year* | 0.018754876 | 0.021033308 |
| *X2_Year* | 0.019092568 | 0.022983963 |
| *X3_Year* | 0.019092568 | 0.022983963 |
| *X5_Year* | 0.023746304 | 0.025622521 |
| *X7_Year* | 0.030327741 | 0.03050701 |
| *X10_Year* | 0.037571665 | 0.036386669 |
| *X20_Year* | 0.042129713 | 0.037428552 |
| *X30_Year* | 0.053834254 | 0.049494622 |
| ***mean*** | 0.041778013 | 0.035739607 |
| ***sd*** | 0.026875885 | 0.016297992 |

4.4.1.1.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.36609984 | 0.12311182 |
| *ctr2* | 0.3721548 | 0.36101851 |
| *X1_Month* | 0.27617049 | 0.25040293 |
| *X3_Month* | 0.10579837 | 0.10126494 |
| *X6_Month* | 0.12118571 | 0.11053711 |
| *X1_Year* | 0.1166497 | 0.09056447 |
| *X2_Year* | 0.09031189 | 0.11131129 |
| *X3_Year* | 0.09031189 | 0.11131129 |
| *X5_Year* | 0.0858829 | 0.10826112 |
| *X7_Year* | 0.15926338 | 0.08003965 |
| *X10_Year* | 0.07684369 | 0.11165754 |
| *X20_Year* | 0.0316998 | 0.13352539 |
| *X30_Year* | 0.11715702 | 0.13886956 |
| ***mean*** | 0.15457919 | 0.14091351 |
| ***sd*** | 0.11080846 | 0.07811309 |

4.4.1.1.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.09830175 | 0.06835261 |
| *ctr2* | 0.08964333 | 0.04157953 |
| *X1_Month* | 0.06177954 | 0.06465943 |
| *X3_Month* | 0.02482244 | 0.02225124 |
| *X6_Month* | 0.02401741 | 0.02133148 |
| *X1_Year* | 0.01875488 | 0.02103331 |
| *X2_Year* | 0.01909257 | 0.02298396 |
| *X3_Year* | 0.01909257 | 0.02298396 |
| *X5_Year* | 0.0237463 | 0.02562252 |
| *X7_Year* | 0.03032774 | 0.03050701 |
| *X10_Year* | 0.03757166 | 0.03638667 |
| *X20_Year* | 0.04212971 | 0.03742855 |
| *X30_Year* | 0.05383425 | 0.04949462 |
| ***mean*** | 0.04177801 | 0.03573961 |
| ***sd*** | 0.02687589 | 0.01629799 |

4.4.1.1.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.16902096 | 0.22752913 |
| *ctr2* | 0.19768815 | 0.18575864 |
| *X1_Month* | 0.15441999 | 0.16637053 |
| *X3_Month* | 0.20022655 | 0.20418772 |
| *X6_Month* | 0.20035823 | 0.20887125 |
| *X1_Year* | 0.21218425 | 0.20733948 |
| *X2_Year* | 0.20501338 | 0.21122948 |
| *X3_Year* | 0.2044694 | 0.20760261 |
| *X5_Year* | 0.20355788 | 0.20183796 |
| *X7_Year* | 0.18782649 | 0.20769627 |
| *X10_Year* | 0.20183883 | 0.19112812 |
| *X20_Year* | 0.20840788 | 0.18291285 |
| *X30_Year* | 0.19079475 | 0.19264277 |
| ***mean*** | 0.19506206 | 0.1996236 |
| ***sd*** | 0.01642916 | 0.01557561 |

4.4.1.1.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.66731686 | 0.62797332 |
| *ctr2* | 0.65038926 | 0.63683291 |
| *X1_Month* | 0.45910842 | 0.51367056 |
| *X3_Month* | 0.51482756 | 0.46345258 |
| *X6_Month* | 0.62385188 | 0.4859416 |
| *X1_Year* | 0.50862352 | 0.52509367 |
| *X2_Year* | 0.63555691 | 0.54565414 |
| *X3_Year* | 0.63555691 | 0.54565414 |
| *X5_Year* | 0.4939533 | 0.50867102 |
| *X7_Year* | 0.59688231 | 0.52164656 |
| *X10_Year* | 0.70103324 | 0.66159726 |
| *X20_Year* | 0.60829606 | 0.59709087 |
| *X30_Year* | 0.70169749 | 0.62327611 |
| ***mean*** | 0.59977644 | 0.55819652 |
| ***sd*** | 0.08036193 | 0.0638169 |

#### *4.4.1.2 Section (File) 2*

4.4.1.2.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.06165242 | 0.04507832 |
| *ctr2* | 0.11040552 | 0.09604931 |
| *X1_Month* | 0.10662041 | 0.13405667 |
| *X3_Month* | 0.03504728 | 0.02188696 |
| *X6_Month* | 0.03404813 | 0.02139029 |
| *X1_Year* | 0.0301007 | 0.0195143 |
| *X2_Year* | 0.02742812 | 0.01641673 |
| *X3_Year* | 0.02742812 | 0.01641673 |
| *X5_Year* | 0.02623047 | 0.01749518 |
| *X7_Year* | 0.02136213 | 0.01998868 |
| *X10_Year* | 0.02489078 | 0.02474585 |
| *X20_Year* | 0.02638357 | 0.02604763 |
| *X30_Year* | 0.03905257 | 0.04645584 |
| ***mean*** | 0.04389617 | 0.03888788 |
| ***sd*** | 0.03040329 | 0.03603269 |

4.4.1.2.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.4298069 | 0.29621392 |
| *ctr2* | 0.44715444 | 0.39894924 |
| *X1_Month* | 0.27787796 | 0.26933243 |
| *X3_Month* | 0.26768468 | 0.16085594 |
| *X6_Month* | 0.26515424 | 0.2683542 |
| *X1_Year* | 0.25813562 | 0.25392771 |
| *X2_Year* | 0.25588545 | 0.15319908 |
| *X3_Year* | 0.25588545 | 0.15319908 |
| *X5_Year* | 0.1617325 | 0.14536075 |
| *X7_Year* | 0.24631347 | 0.17182126 |
| *X10_Year* | 0.24139681 | 0.16523483 |
| *X20_Year* | 0.14097631 | 0.20368378 |
| *X30_Year* | 0.21225141 | 0.18335953 |
| ***mean*** | 0.26617348 | 0.21719167 |
| ***sd*** | 0.08681626 | 0.07559469 |

4.4.1.2.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.06165242 | 0.04507832 |
| *ctr2* | 0.11040552 | 0.09604931 |
| *X1_Month* | 0.10662041 | 0.13405667 |
| *X3_Month* | 0.03504728 | 0.02188696 |
| *X6_Month* | 0.03404813 | 0.02139029 |
| *X1_Year* | 0.0301007 | 0.0195143 |
| *X2_Year* | 0.02742812 | 0.01641673 |
| *X3_Year* | 0.02742812 | 0.01641673 |
| *X5_Year* | 0.02623047 | 0.01749518 |
| *X7_Year* | 0.02136213 | 0.01998868 |
| *X10_Year* | 0.02489078 | 0.02474585 |
| *X20_Year* | 0.02638357 | 0.02604763 |
| *X30_Year* | 0.03905257 | 0.04645584 |
| ***mean*** | 0.04389617 | 0.03888788 |
| ***sd*** | 0.03040329 | 0.03603269 |

4.4.1.2.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.20042305 | 0.26418925 |
| *ctr2* | 0.19482825 | 0.22402364 |
| *X1_Month* | 0.15433405 | 0.15738967 |
| *X3_Month* | 0.17542498 | 0.17299837 |
| *X6_Month* | 0.14664057 | 0.17923916 |
| *X1_Year* | 0.15408969 | 0.17509289 |
| *X2_Year* | 0.16620338 | 0.17781426 |
| *X3_Year* | 0.1645277 | 0.21919757 |
| *X5_Year* | 0.21495588 | 0.17623129 |
| *X7_Year* | 0.21340014 | 0.19812973 |
| *X10_Year* | 0.20015906 | 0.14127319 |
| *X20_Year* | 0.20591972 | 0.15198129 |
| *X30_Year* | 0.19133633 | 0.27309917 |
| ***mean*** | 0.18324944 | 0.19312765 |
| ***sd*** | 0.02404429 | 0.0410354 |

4.4.1.2.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.70637418 | 0.7456701 |
| *ctr2* | 0.73884141 | 0.70925701 |
| *X1_Month* | 0.59065524 | 0.56331578 |
| *X3_Month* | 0.66342933 | 0.61978976 |
| *X6_Month* | 0.55401355 | 0.61768359 |
| *X1_Year* | 0.67578068 | 0.62765372 |
| *X2_Year* | 0.68607187 | 0.6374434 |
| *X3_Year* | 0.68607187 | 0.6374434 |
| *X5_Year* | 0.70175175 | 0.65752492 |
| *X7_Year* | 0.75168401 | 0.69327012 |
| *X10_Year* | 0.76727913 | 0.70836294 |
| *X20_Year* | 0.76531415 | 0.73494204 |
| *X30_Year* | 0.71896783 | 0.6871447 |
| ***mean*** | 0.69278731 | 0.66457704 |
| ***sd*** | 0.06342961 | 0.05332111 |

### *4.4.1.3 Section (File) 3*

4.4.1.3.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.17701172 | 0.17489781 |
| *ctr2* | 0.18827708 | 0.14145413 |
| *X1_Month* | 0.18868303 | 0.19923181 |
| *X3_Month* | 0.05079582 | 0.07433595 |
| *X6_Month* | 0.04870737 | 0.07520522 |
| *X1_Year* | 0.04459333 | 0.07863118 |
| *X2_Year* | 0.06546717 | 0.08984258 |
| *X3_Year* | 0.06546717 | 0.08984258 |
| *X5_Year* | 0.06781351 | 0.09420892 |
| *X7_Year* | 0.07660057 | 0.09938459 |
| *X10_Year* | 0.11076111 | 0.10571546 |
| *X20_Year* | 0.09205068 | 0.10369225 |
| *X30_Year* | 0.0792341 | 0.11384172 |
| ***mean*** | 0.09657405 | 0.11079109 |
| ***sd*** | 0.05333572 | 0.03858914 |

4.4.1.3.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.52078122 | 0.5029823 |
| *ctr2* | 0.42164327 | 0.44698346 |
| *X1_Month* | 0.4414789 | 0.46944379 |
| *X3_Month* | 0.4091519 | 0.44422479 |
| *X6_Month* | 0.42173835 | 0.4168775 |
| *X1_Year* | 0.43765126 | 0.42478268 |
| *X2_Year* | 0.44536443 | 0.40183121 |
| *X3_Year* | 0.44536443 | 0.40183121 |
| *X5_Year* | 0.45204529 | 0.44105224 |
| *X7_Year* | 0.44381758 | 0.44963065 |
| *X10_Year* | 0.39866479 | 0.45390997 |
| *X20_Year* | 0.44065668 | 0.44204108 |
| *X30_Year* | 0.43475754 | 0.45418453 |
| ***mean*** | 0.43947043 | 0.44229042 |
| ***sd*** | 0.02901933 | 0.02739338 |

4.4.1.3.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.17701172 | 0.17489781 |
| *ctr2* | 0.18827708 | 0.14145413 |
| *X1_Month* | 0.18868303 | 0.19923181 |
| *X3_Month* | 0.05079582 | 0.07433595 |
| *X6_Month* | 0.04870737 | 0.07520522 |
| *X1_Year* | 0.04459333 | 0.07863118 |
| *X2_Year* | 0.06546717 | 0.08984258 |
| *X3_Year* | 0.06546717 | 0.08984258 |
| *X5_Year* | 0.06781351 | 0.09420892 |
| *X7_Year* | 0.07660057 | 0.09938459 |
| *X10_Year* | 0.11076111 | 0.10571546 |
| *X20_Year* | 0.09205068 | 0.10369225 |
| *X30_Year* | 0.0792341 | 0.11384172 |
| ***mean*** | 0.09657405 | 0.11079109 |
| ***sd*** | 0.05333572 | 0.03858914 |

4.4.1.3.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.40578724 | 0.23745402 |
| *ctr2* | 0.24759166 | 0.40102574 |
| *X1_Month* | 0.15448294 | 0.26834771 |
| *X3_Month* | 0.13832258 | 0.22483522 |
| *X6_Month* | 0.12529171 | 0.22649463 |
| *X1_Year* | 0.12954196 | 0.21903954 |
| *X2_Year* | 0.23146085 | 0.22741002 |
| *X3_Year* | 0.22562594 | 0.24028409 |
| *X5_Year* | 0.29709677 | 0.19146259 |
| *X7_Year* | 0.286037 | 0.20628099 |
| *X10_Year* | 0.24234234 | 0.21423627 |
| *X20_Year* | 0.23035327 | 0.23539266 |
| *X30_Year* | 0.16257042 | 0.22422625 |
| ***mean*** | 0.22126959 | 0.23972998 |
| ***sd*** | 0.08049098 | 0.05178108 |

4.4.1.3.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.85348137 | 0.82033739 |
| *ctr2* | 0.86036116 | 0.83204613 |
| *X1_Month* | 0.78389951 | 0.66657653 |
| *X3_Month* | 0.82462897 | 0.29560757 |
| *X6_Month* | 0.78195498 | 0.74296627 |
| *X1_Year* | 0.79380222 | 0.7735928 |
| *X2_Year* | 0.79998542 | 0.78072701 |
| *X3_Year* | 0.79998542 | 0.78072701 |
| *X5_Year* | 0.79531853 | 0.28003745 |
| *X7_Year* | 0.84725175 | 0.2556789 |
| *X10_Year* | 0.39975265 | 0.25375134 |
| *X20_Year* | 0.85954542 | 0.21671524 |
| *X30_Year* | 0.87680852 | 0.23014796 |
| ***mean*** | 0.79052123 | 0.5329932 |
| ***sd*** | 0.12195059 | 0.27103652 |

### *4.4.1.4 Section (File) 4*

4.4.1.4.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.50095587 | 0.40565156 |
| *ctr2* | 0.31921235 | 0.16207199 |
| *X1_Month* | 0.44219777 | 0.38054594 |
| *X3_Month* | 0.52093434 | 0.39907923 |
| *X6_Month* | 0.46103533 | 0.39042831 |
| *X1_Year* | 0.52886149 | 0.40793656 |
| *X2_Year* | 0.53055349 | 0.40987007 |
| *X3_Year* | 0.53055349 | 0.40987007 |
| *X5_Year* | 0.52762772 | 0.4117429 |
| *X7_Year* | 0.50897783 | 0.40268594 |
| *X10_Year* | 0.54004315 | 0.40395926 |
| *X20_Year* | 0.45712191 | 0.3903068 |
| *X30_Year* | 0.52287103 | 0.38373168 |
| ***mean*** | 0.49161121 | 0.38137541 |
| ***sd*** | 0.06094863 | 0.06669611 |

4.4.1.4.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.55847506 | 0.48027143 |
| *ctr2* | 0.40063695 | 0.35544669 |
| *X1_Month* | 0.56794669 | 0.60512989 |
| *X3_Month* | 0.56850867 | 0.501874 |
| *X6_Month* | 0.61213789 | 0.49978926 |
| *X1_Year* | 0.61125833 | 0.51304601 |
| *X2_Year* | 0.6113377 | 0.52352631 |
| *X3_Year* | 0.6113377 | 0.52352631 |
| *X5_Year* | 0.5370721 | 0.52888809 |
| *X7_Year* | 0.56742757 | 0.5307071 |
| *X10_Year* | 0.44771398 | 0.53141093 |
| *X20_Year* | 0.54178946 | 0.52797025 |
| *X30_Year* | 0.47613412 | 0.57685189 |
| ***mean*** | 0.54705971 | 0.51526447 |
| ***sd*** | 0.06740957 | 0.05776496 |

4.4.1.4.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.50095587 | 0.40565156 |
| *ctr2* | 0.31921235 | 0.16207199 |
| *X1_Month* | 0.44219777 | 0.38054594 |
| *X3_Month* | 0.52093434 | 0.39907923 |
| *X6_Month* | 0.46103533 | 0.39042831 |
| *X1_Year* | 0.52886149 | 0.40793656 |
| *X2_Year* | 0.53055349 | 0.40987007 |
| *X3_Year* | 0.53055349 | 0.40987007 |
| *X5_Year* | 0.52762772 | 0.4117429 |
| *X7_Year* | 0.50897783 | 0.40268594 |
| *X10_Year* | 0.54004315 | 0.40395926 |
| *X20_Year* | 0.45712191 | 0.3903068 |
| *X30_Year* | 0.52287103 | 0.38373168 |
| ***mean*** | 0.49161121 | 0.38137541 |
| ***sd*** | 0.06094863 | 0.06669611 |

#### 4.4.1.4.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.41695677 | 0.37106701 |
| *ctr2* | 0.42349284 | 0.43637646 |
| *X1_Month* | 0.16236145 | 0.10113822 |
| *X3_Month* | 0.28768872 | 0.21889532 |
| *X6_Month* | 0.23842677 | 0.2429871 |
| *X1_Year* | 0.28219017 | 0.24676783 |
| *X2_Year* | 0.31944201 | 0.21399383 |
| *X3_Year* | 0.20364756 | 0.24621727 |
| *X5_Year* | 0.21500518 | 0.19490454 |
| *X7_Year* | 0.26063782 | 0.16660131 |
| *X10_Year* | 0.22876193 | 0.24236022 |
| *X20_Year* | 0.2097402 | 0.23295637 |
| *X30_Year* | 0.258466 | 0.1533706 |
| ***mean*** | 0.26975519 | 0.23597201 |
| ***sd*** | 0.07832841 | 0.08708706 |

#### 4.4.1.4.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.80826979 | 0.76787057 |
| *ctr2* | 0.56392586 | 0.42823755 |
| *X1_Month* | 0.74603433 | 0.68508713 |
| *X3_Month* | 0.76897774 | 0.73629029 |
| *X6_Month* | 0.71713012 | 0.73266122 |
| *X1_Year* | 0.76233578 | 0.75179259 |
| *X2_Year* | 0.77256916 | 0.71808753 |
| *X3_Year* | 0.77256916 | 0.71808753 |
| *X5_Year* | 0.722026 | 0.74466999 |
| *X7_Year* | 0.82235158 | 0.8555162 |
| *X10_Year* | 0.88663415 | 0.89094871 |
| *X20_Year* | 0.88607062 | 0.92758353 |
| *X30_Year* | 0.94644841 | 0.93230791 |
| ***mean*** | 0.78271867 | 0.76070313 |
| ***sd*** | 0.09509771 | 0.13067268 |

### *4.4.1.5 Section (File) 5*

#### 4.4.1.5.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 1 | 1 |
| *ctr2* | 1 | 1 |
| *X1_Month* | 1 | 1 |
| *X3_Month* | 1 | 1 |
| *X6_Month* | 1 | 1 |
| *X1_Year* | 1 | 1 |
| *X2_Year* | 1 | 1 |
| *X3_Year* | 1 | 1 |
| *X5_Year* | 1 | 1 |
| *X7_Year* | 1 | 1 |
| *X10_Year* | 1 | 1 |
| *X20_Year* | 1 | 1 |
| *X30_Year* | 1 | 1 |
| ***mean*** | 1 | 1 |
| ***sd*** | 0 | 0 |

4.4.1.5.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0 | 0 |
| *ctr2* | 0 | 0 |
| *X1_Month* | 0 | 1.11E-16 |
| *X3_Month* | 1.11E-16 | 1.11E-16 |
| *X6_Month* | 1.11E-16 | 0 |
| *X1_Year* | 0 | 0 |
| *X2_Year* | 0 | 0 |
| *X3_Year* | 0 | 0 |
| *X5_Year* | 1.11E-16 | 0 |
| *X7_Year* | 0 | 0 |
| *X10_Year* | 0 | 0 |
| *X20_Year* | 0 | 0 |
| *X30_Year* | 0 | 0 |
| ***mean*** | 2.5621E-17 | 1.708E-17 |
| ***sd*** | 4.8687E-17 | 4.1693E-17 |

4.4.1.5.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 1 | 1 |
| *ctr2* | 1 | 1 |
| *X1_Month* | 1 | 1 |
| *X3_Month* | 1 | 1 |
| *X6_Month* | 1 | 1 |
| *X1_Year* | 1 | 1 |
| *X2_Year* | 1 | 1 |
| *X3_Year* | 1 | 1 |
| *X5_Year* | 1 | 1 |
| *X7_Year* | 1 | 1 |
| *X10_Year* | 1 | 1 |
| *X20_Year* | 1 | 1 |
| *X30_Year* | 1 | 1 |
| ***mean*** | 1 | 1 |
| ***sd*** | 0 | 0 |

4.4.1.5.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 2.08514397 | 1.15538852 |
| *ctr2* | 1.53000259 | 2.01007497 |
| *X1_Month* | 0.88485133 | 1.0865401 |
| *X3_Month* | 0.85797404 | 1.30205427 |
| *X6_Month* | 0.67106465 | 1.19339734 |
| *X1_Year* | 0.53897377 | 1.1524668 |
| *X2_Year* | 1.13634112 | 1.0891449 |
| *X3_Year* | 1.16519535 | 1.12113565 |
| *X5_Year* | 0.92250254 | 1.05816959 |
| *X7_Year* | 0.67145311 | 1.05520406 |
| *X10_Year* | 1.03011293 | 1.05295723 |
| *X20_Year* | 1.06252872 | 1.03887412 |
| *X30_Year* | 0.78987914 | 1.03950537 |
| ***mean*** | 1.02661717 | 1.18114715 |
| ***sd*** | 0.40949145 | 0.26014645 |

4.4.1.5.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.99572009 | 0.9901832 |
| *ctr2* | 0.99970878 | 0.99981102 |
| *X1_Month* | 0.99749339 | 0.98341559 |
| *X3_Month* | 0.98226985 | 0.82693375 |
| *X6_Month* | 0.9761763 | 0.86633269 |
| *X1_Year* | 0.95452034 | 0.91729934 |
| *X2_Year* | 0.99481039 | 0.85607619 |
| *X3_Year* | 0.99481039 | 0.85607619 |
| *X5_Year* | 0.99549901 | 0.92411518 |
| *X7_Year* | 0.95639665 | 0.95170059 |
| *X10_Year* | 0.99854787 | 0.95687617 |
| *X20_Year* | 0.99863517 | 0.9633716 |
| *X30_Year* | 0.99903426 | 0.98749514 |
| ***mean*** | 0.98797096 | 0.92920667 |
| ***sd*** | 0.01600578 | 0.05969004 |

### *4.4.1.6 Section (File) 6*

4.4.1.6.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.08055451 | 0.03145721 |
| *ctr2* | 0.09296763 | 0.04588962 |
| *X1_Month* | 0.208585 | 0.22120208 |
| *X3_Month* | 0.0221647 | 0.02709307 |
| *X6_Month* | 0.03069201 | 0.02604472 |
| *X1_Year* | 0.02008632 | 0.02421597 |
| *X2_Year* | 0.01853294 | 0.02360475 |
| *X3_Year* | 0.01853294 | 0.02360475 |
| *X5_Year* | 0.01911866 | 0.02550493 |
| *X7_Year* | 0.02494488 | 0.03381535 |
| *X10_Year* | 0.03204275 | 0.04771175 |
| *X20_Year* | 0.04741394 | 0.05563613 |
| *X30_Year* | 0.09911976 | 0.10139633 |
| ***mean*** | 0.05498123 | 0.05285974 |
| ***sd*** | 0.05473613 | 0.05497114 |

4.4.1.6.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.41482333 | 0.40324477 |
| *ctr2* | 0.43224904 | 0.39481312 |
| *X1_Month* | 0.43715361 | 0.45279034 |
| *X3_Month* | 0.36731486 | 0.3772933 |
| *X6_Month* | 0.37770878 | 0.31296399 |
| *X1_Year* | 0.45303069 | 0.29697874 |
| *X2_Year* | 0.4025369 | 0.32164591 |
| *X3_Year* | 0.4025369 | 0.32164591 |
| *X5_Year* | 0.42083633 | 0.34270442 |
| *X7_Year* | 0.33983765 | 0.34007847 |
| *X10_Year* | 0.35257148 | 0.40534824 |
| *X20_Year* | 0.35545368 | 0.4025108 |
| *X30_Year* | 0.37095182 | 0.379902 |
| ***mean*** | 0.39438501 | 0.36553231 |
| ***sd*** | 0.03623062 | 0.04625576 |

#### 4.4.1.6.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.08055451 | 0.03145721 |
| *ctr2* | 0.09296763 | 0.04588962 |
| *X1_Month* | 0.208585 | 0.22120208 |
| *X3_Month* | 0.0221647 | 0.02709307 |
| *X6_Month* | 0.03069201 | 0.02604472 |
| *X1_Year* | 0.02008632 | 0.02421597 |
| *X2_Year* | 0.01853294 | 0.02360475 |
| *X3_Year* | 0.01853294 | 0.02360475 |
| *X5_Year* | 0.01911866 | 0.02550493 |
| *X7_Year* | 0.02494488 | 0.03381535 |
| *X10_Year* | 0.03204275 | 0.04771175 |
| *X20_Year* | 0.04741394 | 0.05563613 |
| *X30_Year* | 0.09911976 | 0.10139633 |
| ***mean*** | 0.05498123 | 0.05285974 |
| ***sd*** | 0.05473613 | 0.05497114 |

#### 4.4.1.6.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.19579763 | 0.14910238 |
| *ctr2* | 0.26007762 | 0.34417178 |
| *X1_Month* | 0.07698715 | 0.07946371 |
| *X3_Month* | 0.17296119 | 0.15646343 |
| *X6_Month* | 0.19792512 | 0.16934768 |
| *X1_Year* | 0.17748618 | 0.17500294 |
| *X2_Year* | 0.16694623 | 0.19036949 |
| *X3_Year* | 0.1705372 | 0.18614889 |
| *X5_Year* | 0.17929538 | 0.17591335 |
| *X7_Year* | 0.14658124 | 0.18974355 |
| *X10_Year* | 0.1737765 | 0.17348562 |
| *X20_Year* | 0.16563493 | 0.18442198 |
| *X30_Year* | 0.14318761 | 0.15174433 |
| ***mean*** | 0.17132261 | 0.17887532 |
| ***sd*** | 0.04046546 | 0.05759622 |

#### 4.4.1.6.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.27217522 | 0.19759657 |
| *ctr2* | 0.33384083 | 0.01255648 |
| *X1_Month* | 0.37887472 | 0.31650333 |
| *X3_Month* | 0.20238047 | 0.18040453 |
| *X6_Month* | 0.17707033 | 0.18163552 |
| *X1_Year* | 0.20004965 | 0.17514337 |
| *X2_Year* | 0.2518885 | 0.16862101 |
| *X3_Year* | 0.2518885 | 0.16862101 |
| *X5_Year* | 0.22592987 | 0.17696669 |
| *X7_Year* | 0.24214654 | 0.16539762 |
| *X10_Year* | 0.2125536 | 0.16262408 |
| *X20_Year* | 0.05188283 | 0.15469289 |
| *X30_Year* | 0.08921424 | 0.16707238 |
| ***mean*** | 0.22229964 | 0.17137196 |
| ***sd*** | 0.08726825 | 0.06300111 |

### *4.4.1.7 Section (File) 7*

4.4.1.7.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.2075899 | 0.08162272 |
| *ctr2* | 0.26278511 | 0.22583355 |
| *X1_Month* | 0.31496584 | 0.38295496 |
| *X3_Month* | 0.09449864 | 0.07136339 |
| *X6_Month* | 0.09603946 | 0.07432002 |
| *X1_Year* | 0.09583984 | 0.07127265 |
| *X2_Year* | 0.09283043 | 0.06924528 |
| *X3_Year* | 0.09283043 | 0.06924528 |
| *X5_Year* | 0.09295646 | 0.07070055 |
| *X7_Year* | 0.09444997 | 0.07724981 |
| *X10_Year* | 0.11582781 | 0.09655799 |
| *X20_Year* | 0.1155942 | 0.10337491 |
| *X30_Year* | 0.15631274 | 0.17850908 |
| ***mean*** | 0.14096314 | 0.12094232 |
| ***sd*** | 0.07437656 | 0.09232785 |

4.4.1.7.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.50494593 | 0.54169603 |
| *ctr2* | 0.51494859 | 0.52234795 |
| *X1_Month* | 0.59771036 | 0.62009827 |
| *X3_Month* | 0.47560942 | 0.4678589 |
| *X6_Month* | 0.446698 | 0.47582789 |
| *X1_Year* | 0.5117121 | 0.42727641 |
| *X2_Year* | 0.45456663 | 0.45934534 |
| *X3_Year* | 0.45456663 | 0.45934534 |
| *X5_Year* | 0.53897345 | 0.44315035 |
| *X7_Year* | 0.51628301 | 0.43146497 |
| *X10_Year* | 0.56414283 | 0.46963899 |
| *X20_Year* | 0.4077923 | 0.46672715 |
| *X30_Year* | 0.54602883 | 0.49052553 |
| ***mean*** | 0.5026137 | 0.48271563 |
| ***sd*** | 0.0531721 | 0.05246989 |

4.4.1.7.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.2075899 | 0.08162272 |
| *ctr2* | 0.26278511 | 0.22583355 |
| *X1_Month* | 0.31496584 | 0.38295496 |
| *X3_Month* | 0.09449864 | 0.07136339 |
| *X6_Month* | 0.09603946 | 0.07432002 |
| *X1_Year* | 0.09583984 | 0.07127265 |
| *X2_Year* | 0.09283043 | 0.06924528 |
| *X3_Year* | 0.09283043 | 0.06924528 |
| *X5_Year* | 0.09295646 | 0.07070055 |
| *X7_Year* | 0.09444997 | 0.07724981 |
| *X10_Year* | 0.11582781 | 0.09655799 |
| *X20_Year* | 0.1155942 | 0.10337491 |
| *X30_Year* | 0.15631274 | 0.17850908 |
| ***mean*** | 0.14096314 | 0.12094232 |
| ***sd*** | 0.07437656 | 0.09232785 |

4.4.1.7.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.22369082 | 0.27312011 |
| *ctr2* | 0.17484482 | 0.37833829 |
| *X1_Month* | 0.33329909 | 0.27737476 |
| *X3_Month* | 0.19849321 | 0.19729494 |
| *X6_Month* | 0.20985381 | 0.19502416 |
| *X1_Year* | 0.22253607 | 0.17967474 |
| *X2_Year* | 0.20568262 | 0.19019861 |
| *X3_Year* | 0.21595554 | 0.19171985 |
| *X5_Year* | 0.20901683 | 0.17053698 |
| *X7_Year* | 0.19324611 | 0.30425279 |
| *X10_Year* | 0.21223764 | 0.27673188 |
| *X20_Year* | 0.25399539 | 0.30083009 |
| *X30_Year* | 0.22242216 | 0.26313356 |
| ***mean*** | 0.22117493 | 0.24601775 |
| ***sd*** | 0.03843048 | 0.06324264 |

4.4.1.7.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.47577674 | 0.8039219 |
| *ctr2* | 0.79760147 | 0.76182705 |
| *X1_Month* | 0.80245141 | 0.59072352 |
| *X3_Month* | 0.74620142 | 0.74607655 |
| *X6_Month* | 0.73786497 | 0.34862581 |
| *X1_Year* | 0.75775424 | 0.74616302 |
| *X2_Year* | 0.82197276 | 0.34059855 |
| *X3_Year* | 0.82197276 | 0.34059855 |
| *X5_Year* | 0.87842292 | 0.75689513 |
| *X7_Year* | 0.90131953 | 0.7858465 |
| *X10_Year* | 0.90824243 | 0.80363678 |
| *X20_Year* | 0.93116317 | 0.82451347 |
| *X30_Year* | 0.93683746 | 0.86208298 |
| ***mean*** | 0.80904471 | 0.67011614 |
| ***sd*** | 0.12166719 | 0.1966665 |

#### *4.4.1.8 Section (File) 8*

4.4.1.8.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.4244938 | 0.2461603 |
| *ctr2* | 0.52374057 | 0.24509971 |
| *X1_Month* | 0.19271575 | 0.17884444 |
| *X3_Month* | 0.26578362 | 0.18974794 |
| *X6_Month* | 0.24789937 | 0.17757758 |
| *X1_Year* | 0.23280694 | 0.16977023 |
| *X2_Year* | 0.2072913 | 0.15979781 |
| *X3_Year* | 0.2072913 | 0.15979781 |
| *X5_Year* | 0.18958657 | 0.15342115 |
| *X7_Year* | 0.229115 | 0.1379074 |
| *X10_Year* | 0.20204322 | 0.13159583 |
| *X20_Year* | 0.200265 | 0.12124998 |
| *X30_Year* | 0.21428899 | 0.12785101 |
| ***mean*** | 0.25671703 | 0.16914009 |
| ***sd*** | 0.10099868 | 0.03989295 |

#### 4.4.1.8.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.52897217 | 0.39610301 |
| *ctr2* | 0.59668999 | 0.48623729 |
| *X1_Month* | 0.39755169 | 0.43589668 |
| *X3_Month* | 0.52288569 | 0.47910584 |
| *X6_Month* | 0.5247801 | 0.50139086 |
| *X1_Year* | 0.52320355 | 0.47938426 |
| *X2_Year* | 0.50962409 | 0.46348171 |
| *X3_Year* | 0.50962409 | 0.46348171 |
| *X5_Year* | 0.53243602 | 0.45699671 |
| *X7_Year* | 0.46748126 | 0.49281825 |
| *X10_Year* | 0.39864535 | 0.49610121 |
| *X20_Year* | 0.47500229 | 0.43262162 |
| *X30_Year* | 0.51516575 | 0.45806926 |
| ***mean*** | 0.50015862 | 0.46474526 |
| ***sd*** | 0.05460643 | 0.02983336 |

#### 4.4.1.8.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.4244938 | 0.2461603 |
| *ctr2* | 0.52374057 | 0.24509971 |
| *X1_Month* | 0.19271575 | 0.17884444 |
| *X3_Month* | 0.26578362 | 0.18974794 |
| *X6_Month* | 0.24789937 | 0.17757758 |
| *X1_Year* | 0.23280694 | 0.16977023 |
| *X2_Year* | 0.2072913 | 0.15979781 |
| *X3_Year* | 0.2072913 | 0.15979781 |
| *X5_Year* | 0.18958657 | 0.15342115 |
| *X7_Year* | 0.229115 | 0.1379074 |
| *X10_Year* | 0.20204322 | 0.13159583 |
| *X20_Year* | 0.200265 | 0.12124998 |
| *X30_Year* | 0.21428899 | 0.12785101 |
| ***mean*** | 0.25671703 | 0.16914009 |
| ***sd*** | 0.10099868 | 0.03989295 |

#### 4.4.1.8.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.38022954 | 0.40765839 |
| *ctr2* | 0.24135456 | 0.49938529 |
| *X1_Month* | 0.31058527 | 0.29149701 |
| *X3_Month* | 0.25998591 | 0.19840555 |
| *X6_Month* | 0.26431163 | 0.20760695 |
| *X1_Year* | 0.19681758 | 0.22750691 |
| *X2_Year* | 0.22863996 | 0.22404665 |
| *X3_Year* | 0.22511548 | 0.23953502 |
| *X5_Year* | 0.25408744 | 0.24864265 |
| *X7_Year* | 0.28792402 | 0.21655938 |
| *X10_Year* | 0.32192953 | 0.25736358 |
| *X20_Year* | 0.2971752 | 0.24223101 |
| *X30_Year* | 0.3666564 | 0.34020324 |
| ***mean*** | 0.27960096 | 0.27697243 |
| ***sd*** | 0.05476133 | 0.08870975 |

4.4.1.8.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.61080691 | 0.60641447 |
| *ctr2* | 0.90108155 | 0.55321045 |
| *X1_Month* | 0.49142846 | 0.33359485 |
| *X3_Month* | 0.60116363 | 0.47299052 |
| *X6_Month* | 0.56665966 | 0.46349195 |
| *X1_Year* | 0.57543696 | 0.46892071 |
| *X2_Year* | 0.59823184 | 0.49132929 |
| *X3_Year* | 0.59823184 | 0.49132929 |
| *X5_Year* | 0.59500416 | 0.49527296 |
| *X7_Year* | 0.587829 | 0.46930787 |
| *X10_Year* | 0.56458837 | 0.46091926 |
| *X20_Year* | 0.60367805 | 0.42755573 |
| *X30_Year* | 0.48124388 | 0.45170253 |
| ***mean*** | 0.59810649 | 0.47584922 |
| ***sd*** | 0.09990588 | 0.06292508 |

### *4.4.1.9 Section (File) 9*

4.4.1.9.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.08055451 | 0.03288493 |
| *ctr2* | 0.09296763 | 0.03702093 |
| *X1_Month* | 0.208585 | 0.21345924 |
| *X3_Month* | 0.0221647 | 0.02524272 |
| *X6_Month* | 0.03069201 | 0.02421184 |
| *X1_Year* | 0.02008632 | 0.02314709 |
| *X2_Year* | 0.01853294 | 0.02375999 |
| *X3_Year* | 0.01853294 | 0.02375999 |
| *X5_Year* | 0.01911866 | 0.02597571 |
| *X7_Year* | 0.02494488 | 0.03430479 |
| *X10_Year* | 0.03204275 | 0.04746917 |
| *X20_Year* | 0.04741394 | 0.05431119 |
| *X30_Year* | 0.09911976 | 0.09558932 |
| ***mean*** | 0.05498123 | 0.05085669 |
| ***sd*** | 0.05473613 | 0.05281552 |

4.4.1.9.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.41482333 | 0.40324477 |
| *ctr2* | 0.43224904 | 0.36165081 |
| *X1_Month* | 0.43715361 | 0.41984305 |
| *X3_Month* | 0.36731486 | 0.3772933 |
| *X6_Month* | 0.37770878 | 0.31296399 |
| *X1_Year* | 0.45303069 | 0.31274836 |
| *X2_Year* | 0.4025369 | 0.32178298 |
| *X3_Year* | 0.4025369 | 0.32178298 |
| *X5_Year* | 0.42083633 | 0.34270442 |
| *X7_Year* | 0.33983765 | 0.34007847 |
| *X10_Year* | 0.35257148 | 0.38611323 |
| *X20_Year* | 0.35545368 | 0.3916109 |
| *X30_Year* | 0.37095182 | 0.38155796 |
| ***mean*** | 0.39438501 | 0.3594904 |
| ***sd*** | 0.03623062 | 0.03646664 |

4.4.1.9.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.08055451 | 0.03288493 |
| *ctr2* | 0.09296763 | 0.03702093 |
| *X1_Month* | 0.208585 | 0.21345924 |
| *X3_Month* | 0.0221647 | 0.02524272 |
| *X6_Month* | 0.03069201 | 0.02421184 |
| *X1_Year* | 0.02008632 | 0.02314709 |
| *X2_Year* | 0.01853294 | 0.02375999 |
| *X3_Year* | 0.01853294 | 0.02375999 |
| *X5_Year* | 0.01911866 | 0.02597571 |
| *X7_Year* | 0.02494488 | 0.03430479 |
| *X10_Year* | 0.03204275 | 0.04746917 |
| *X20_Year* | 0.04741394 | 0.05431119 |
| *X30_Year* | 0.09911976 | 0.09558932 |
| ***mean*** | 0.05498123 | 0.05085669 |
| ***sd*** | 0.05473613 | 0.05281552 |

4.4.1.9.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.18893207 | 0.16021317 |
| *ctr2* | 0.23832456 | 0.3505768 |
| *X1_Month* | 0.07962232 | 0.11534489 |
| *X3_Month* | 0.18244761 | 0.16242885 |
| *X6_Month* | 0.22677538 | 0.16338384 |
| *X1_Year* | 0.17968728 | 0.15901599 |
| *X2_Year* | 0.15448701 | 0.14295837 |
| *X3_Year* | 0.1691722 | 0.1470111 |
| *X5_Year* | 0.1692121 | 0.13372412 |
| *X7_Year* | 0.14139175 | 0.13495223 |
| *X10_Year* | 0.17869809 | 0.16380561 |
| *X20_Year* | 0.16715463 | 0.16794689 |
| *X30_Year* | 0.13921528 | 0.18501918 |
| ***mean*** | 0.17039387 | 0.16818316 |
| ***sd*** | 0.03953569 | 0.05769313 |

4.4.1.9.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.27217522 | 0.19027636 |
| *ctr2* | 0.33384083 | 0.28071316 |
| *X1_Month* | 0.37887472 | 0.30203748 |
| *X3_Month* | 0.20238047 | 0.16637164 |
| *X6_Month* | 0.17707033 | 0.16879802 |
| *X1_Year* | 0.20004965 | 0.16928058 |
| *X2_Year* | 0.2518885 | 0.17011583 |
| *X3_Year* | 0.2518885 | 0.17011583 |
| *X5_Year* | 0.22592987 | 0.16862237 |
| *X7_Year* | 0.24214654 | 0.16539762 |
| *X10_Year* | 0.2125536 | 0.16262408 |
| *X20_Year* | 0.05188283 | 0.14976672 |
| *X30_Year* | 0.08921424 | 0.16092457 |
| ***mean*** | 0.22229964 | 0.18654187 |
| ***sd*** | 0.08726825 | 0.04754786 |

### *4.4.1.10 Section (File) 10*

4.4.1.10.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.108583 | 0.09674392 |
| *ctr2* | 0.08384846 | 0.04214162 |
| *X1_Month* | 0.05125622 | 0.04103781 |
| *X3_Month* | 0.06460015 | 0.04338642 |
| *X6_Month* | 0.06031789 | 0.04112119 |
| *X1_Year* | 0.05487838 | 0.04018956 |
| *X2_Year* | 0.03962869 | 0.0404017 |
| *X3_Year* | 0.03962869 | 0.0404017 |
| *X5_Year* | 0.04143372 | 0.04203687 |
| *X7_Year* | 0.03971443 | 0.04252465 |
| *X10_Year* | 0.0460185 | 0.04469444 |
| *X20_Year* | 0.04536396 | 0.04179747 |
| *X30_Year* | 0.0400985 | 0.04459861 |
| ***mean*** | 0.05502851 | 0.04623661 |
| ***sd*** | 0.02066329 | 0.01524773 |

4.4.1.10.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.42656656 | 0.41758819 |
| *ctr2* | 0.39412046 | 0.40497059 |
| *X1_Month* | 0.31811925 | 0.22719581 |
| *X3_Month* | 0.26520217 | 0.25680084 |
| *X6_Month* | 0.23518265 | 0.23184622 |
| *X1_Year* | 0.21221947 | 0.22420151 |
| *X2_Year* | 0.28799718 | 0.12031159 |
| *X3_Year* | 0.28799718 | 0.12031159 |
| *X5_Year* | 0.24409118 | 0.19146692 |
| *X7_Year* | 0.26125954 | 0.24388222 |
| *X10_Year* | 0.21085003 | 0.24928796 |
| *X20_Year* | 0.18074434 | 0.2583174 |
| *X30_Year* | 0.25403162 | 0.29590599 |
| ***mean*** | 0.27526013 | 0.2493913 |
| ***sd*** | 0.07045349 | 0.08800428 |

4.4.1.10.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.108583 | 0.09674392 |
| *ctr2* | 0.08384846 | 0.04214162 |
| *X1_Month* | 0.05125622 | 0.04103781 |
| *X3_Month* | 0.06460015 | 0.04338642 |
| *X6_Month* | 0.06031789 | 0.04112119 |
| *X1_Year* | 0.05487838 | 0.04018956 |
| *X2_Year* | 0.03962869 | 0.0404017 |
| *X3_Year* | 0.03962869 | 0.0404017 |
| *X5_Year* | 0.04143372 | 0.04203687 |
| *X7_Year* | 0.03971443 | 0.04252465 |
| *X10_Year* | 0.0460185 | 0.04469444 |
| *X20_Year* | 0.04536396 | 0.04179747 |
| *X30_Year* | 0.0400985 | 0.04459861 |
| ***mean*** | 0.05502851 | 0.04623661 |
| ***sd*** | 0.02066329 | 0.01524773 |

#### 4.4.1.10.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.19438419 | 0.2751031 |
| *ctr2* | 0.23526743 | 0.19961883 |
| *X1_Month* | 0.21226368 | 0.21009178 |
| *X3_Month* | 0.15560392 | 0.17472536 |
| *X6_Month* | 0.16404976 | 0.18907601 |
| *X1_Year* | 0.15553625 | 0.18156603 |
| *X2_Year* | 0.21979047 | 0.18110531 |
| *X3_Year* | 0.21517634 | 0.17556062 |
| *X5_Year* | 0.22203938 | 0.16763389 |
| *X7_Year* | 0.16241598 | 0.16005097 |
| *X10_Year* | 0.16852208 | 0.13414832 |
| *X20_Year* | 0.18035552 | 0.19215176 |
| *X30_Year* | 0.20269371 | 0.19618332 |
| ***mean*** | 0.19139221 | 0.18746272 |
| ***sd*** | 0.02829496 | 0.03266933 |

#### 4.4.1.10.5 SVM rbf

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.7080891 | 0.75117224 |
| *ctr2* | 0.70653878 | 0.70042356 |
| *X1_Month* | 0.63595348 | 0.58439505 |
| *X3_Month* | 0.69421688 | 0.58082697 |
| *X6_Month* | 0.69512274 | 0.60685139 |
| *X1_Year* | 0.60862913 | 0.62734045 |
| *X2_Year* | 0.62532551 | 0.62413981 |
| *X3_Year* | 0.62532551 | 0.62413981 |
| *X5_Year* | 0.63663525 | 0.61647298 |
| *X7_Year* | 0.67161778 | 0.68900977 |
| *X10_Year* | 0.7313251 | 0.7018668 |
| *X20_Year* | 0.73312036 | 0.70537559 |
| *X30_Year* | 0.74102964 | 0.75938041 |
| ***mean*** | 0.67791763 | 0.65933806 |
| ***sd*** | 0.04655791 | 0.06110984 |

### 4.4.2 Variable Choices in Stationary and Time-shifted analysis

The tables below show the dynamics of the clustering apparent in each file. For example, in File 1, the correlations between Independent Variables ("IV") produces 19 clusters of IVs. The '1' and '0' represent the flag on that IV having been chosen by the given Dependent variable. These are presented first by File aggregation, then by Algorithm, and lastly by temporality (i.e., best shift or no-shift).

#### *4.4.2.1 Section (File) 1*

##### 4.4.2.1.1 Linear Model

*4.4.2.1.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.S* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| emantic_Network. | | | | | | | | | | | | | |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Skip.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.1.1.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number_of_Concept_nodes | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Overall_Complexity | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Meta.Matrix_Hamming_Distance | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Centrality.Authority.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Speed.Average.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Breadth.Column.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Network_Centralization.Column_Degree.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Communicative_Need.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| Effective_Network_Size.Burt.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Hierarchy.Semantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Centrality.In.Closeness.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Network_Centralization.In.Closeness.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Lateral.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Centrality.Row_Degree.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Skip.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Component_Count.Strong.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.1.2 CART

##### 4.4.2.1.2.1 *Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number_of_Concept_nodes | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Overall_Complexity | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Meta.Matrix_Hamming_Distance | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| Centrality.Authority.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Speed.Average.Semantic_Network. | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Breadth.Column.Semantic_Network. | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Centrality.Column_Degree.Semantic_Network._Average | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Communicative_Need.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Effective_Network_Size.Burt.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Hierarchy.Semantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Centrality.In.Closeness.Semantic_Network._Average | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Network_Centralization.In.Closeness.Semantic_Network. | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| Network_Centralization.In_Degree.Semantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Isolate_Count.Semantic_Network. | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Lateral.Semantic_Network. | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Skip.Semantic_Network. | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### 4.4.2.1.2.2 *No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number_of_Concept_nodes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Overall_Complexity | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Meta.Matrix_Hamming_Distance | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Row_Degree.Semantic_Network._Average* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Component_Count.Strong.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.1.3 GLM

##### *4.4.2.1.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| *Effective_Network_Size.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Burt.Semantic_Network._ Average* | | | | | | | | | | | | | |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In_Degree.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_CountSequentialSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_CountSkipSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_BouednessSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.1.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *MetaMatrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityAuthoritySemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *SpeedAverageSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *BreadthColumnSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_CentralizationColumn_DegreeSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Row_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Component_Count.Strong.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

4.4.2.1.4 Random Forests

*4.4.2.1.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Overall_Complexity* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### *4.4.2.1.4.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Overall_Complexity* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Row_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Count.Strong.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.1.5 SVM rbf

##### *4.4.2.1.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *Meta.Matrix_Hamming_Distance* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.1.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Row_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Component_Count.Strong.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.2 Section (File) 2*

#### 4.4.2.2.1 Linear Model

##### *4.4.2.2.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### *4.4.2.2.1.2 No Time Shift*

| **Dependent Variable** | **ctrl** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_ Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Se mantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree. Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Sema ntic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| *Effective_Network_Size. Burt.Semantic_Network._ Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_ Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.S emantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization. In.Closeness.Semantic_N etwork.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization. Out_Degree.Semantic_Net work.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Seman tic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.S emantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.S emantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semanti c_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Members.Weak.Se mantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

##### 4.4.2.2.2 CART

###### *4.4.2.2.2.1 Time Shift*

| **Dependent Variable** | **ctrl** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| *Overall_Complexity* | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_ Network.* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Fragmentation.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.2.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| *Overall_Complexity* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Out_Degree.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Members.Weak.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

#### 4.4.2.2.3 GLM

##### *4.4.2.2.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.2.3.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Out_Degree.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Members.Weak.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

#### 4.4.2.2.4 Random Forests

*4.4.2.2.4.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Overall_Complexity* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Fragmentation.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.2.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *Overall_Complexity* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Fragmentation.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Network_Centralization.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Out_Degree.Semantic_Network.* | | | | | | | | | | | | | |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Members.Weak.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.2.5 SVM rbf

##### *4.4.2.2.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.2.5.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Fragmentation.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Out_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_MembersWeakSemantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

### *4.4.2.3 Section (File) 3*

#### 4.4.2.3.1 Linear Model

##### *4.4.2.3.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.3.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *MetaMatrixHamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.3.2 CART

##### *4.4.2.3.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityClosenessSemantic_Network._Average* | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *SpanOfControlSemanticNetwork* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Total_Degree.Semantic_Network* | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.3.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| *MetaMatrix_Hamming_Distance* | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| *EfficiencySemantic_Network* | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |

4.4.2.3.3 GLM

*4.4.2.3.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *MetaMatrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *CentralityBetweennessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| *Connectedness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.3.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *MetaMatrixHammingDistance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *SpeedAverageSemanticNetwork* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *NetworkCentralization.Column_Degree.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| *Connectedness.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Reciprocal. Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential. Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semant ic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semant ic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.3.4 Random Forests

##### *4.4.2.3.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Dis tance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_ Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Betweenness.Se mantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Se mantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree. Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Sema ntic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_ Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *EfficiencySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Se mantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.S emantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.S emantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semanti c_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization. Total_Degree.Semantic_N etwork.* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semanti c_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.3.4.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centralit.Betweenness.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.3.5 SVM rbf

*4.4.2.3.5.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Total_Degree.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.3.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.4 Section (File) 4*

4.4.2.4.1 Linear Model

*4.4.2.4.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Average_Distance.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| *Connectedness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityIn.ClosenessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Minimum.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *RadialsSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_ControlSemantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_BouednessSemantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### 4.4.2.4.1.2 *No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Centrality.Authority.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Average_Distance.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Network_Centralization.Column_Degree.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Communicative_Need.Semantic_Network | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| Connectedness.Semantic_Network | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Efficiency.Semantic_Network | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Exclusivity.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Exclusivity.Complete.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HierarchySemantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CentralityInClosenessSemantic_Network_Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Lateral.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Speed.Minimum.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| RadialsSemantic_Network_Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Count.Row.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Span_Of_Control.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### 4.4.2.4.2 CART

###### 4.4.2.4.2.1 *Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number_of_Concept_nodes | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| Centrality.Authority.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Average_Distance.Semantic_Network. | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network_Average* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *Speed.Minimum.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *RadialsSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_ControlSemantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### *4.4.2.4.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Average_Distance.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityIn.ClosenessSemantic_Network_Average* | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speed.Minimum.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Radials.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CountRowSemantic_Network | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Span_Of_Control.Semantic_Network. | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.4.3 GLM

##### *4.4.2.4.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Number_of_Concept_nodes | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Centrality.Authority.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Average_Distance.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Network_Centralization.Column_Degree.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Communicative_Need.Semantic_Network. | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Connectedness.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Efficiency.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Exclusivity.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Exclusivity.Complete.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HierarchySemantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CentralityInClosenessSemantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Lateral.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Speed.Minimum.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Radials.Semantic_Network._Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Span_Of_Control.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### *4.4.2.4.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Average_Distance.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| *Connectedness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Minimum.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Radials.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CountRowSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### 4.4.2.4.4 Random Forests

###### *4.4.2.4.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| *Average_Distance.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network.* | | | | | | | | | | | | | |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network_Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| *Speed.Minimum.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Radials.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### *4.4.2.4.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Average_Distance.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Connectedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network_Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Speed.Minimum.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Radials.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CountRowSemantic_Network* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_ControlSemantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_BouednessSemantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

4.4.2.4.5 SVM rbf

*4.4.2.4.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| *Average_Distance.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Connectedness.Semantic_Network.* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInCloseness.Semantic_Network_Average* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Minimum.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| *Radials.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.4.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Average_Distance.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Connectedness.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Exclusivity.Complete.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Minimum.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Radials.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CountRowSemantic_Network* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.5 Section (File) 5*

4.4.2.5.1 Linear Model

*4.4.2.5.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *Cognitive_Expertise_Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.5.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Correlation.Expertise.Semantic_Network_Average* | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

4.4.2.5.2 CART

*4.4.2.5.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Cognitive_Expertise_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Total_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.5.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CorrelationExpertise.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.5.3 GLM

*4.4.2.5.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *Cognitive_Expertise_Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityTotal_DegreeSemantic_Network_Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.5.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CorrelationExpertise.Semantic_Network._Average* | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.5.4 Random Forests

*4.4.2.5.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *Cognitive_Expertise_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityTotal_DegreesSemantic_Network_Average* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

###### *4.4.2.5.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CorrelationExpertise.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_CountReciprocalSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.5.5 SVM rbf

##### *4.4.2.5.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Cognitive_Expertise_Average* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityTotal_DegreeSemantic_Network_Average* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.5.5.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *MetaMatrix_Hamming_Distance* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Betweenness.Semantic_Network.* | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network* | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *CorrelationExpertise.Semantic_Network_Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_CountSemantic_Network.* | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_CountReciprocalSemantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_CountSequentialSemantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_ControlSemantic_Network* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_BouednessSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.6 Section (File) 6*

#### 4.4.2.6.1 Linear Model

*4.4.2.6.1.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Communicative_Need.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.6.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.6.2 CART

##### *4.4.2.6.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| *CentralityBetweennessSemantic_Network_Average* | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.6.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *SpeedAverageSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Lateral.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.6.3 GLM

##### *4.4.2.6.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.6.3.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.6.4 Random Forests

##### *4.4.2.6.4.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *SpeedAverageSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityBetweenness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average* | | | | | | | | | | | | | |
| *Efficiency.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.6.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Betweenness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.6.5 SVM rbf

##### *4.4.2.6.5.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speed.Average.Semantic_Network. | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| CentralityBetweenness.Semantic_Network_Average | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Centrality.Closeness.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Communicative_Need.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Effective_Network_Size.Burt.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EfficiencySemantic_Network | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| HierarchySemantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Centrality.Hub.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Lateral.Semantic_Network. | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Link_Count.Pooled.Semantic_Network. | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Upper_Bouedness.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.6.5.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Centrality.Authority.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Speed.Average.Semantic_Network. | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Centrality.Betweenness.Semantic_Network._Average | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Centrality.Closeness.Semantic_Network._Average | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Communicative_Need.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Effective_Network_Size.Burt.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Efficiency.Semantic_Network. | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| HierarchySemantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.7 Section (File) 7*

#### 4.4.2.7.1 Linear Model

##### *4.4.2.7.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.7.1.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

4.4.2.7.2 CART

*4.4.2.7.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| *Overall_Complexity* | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.7.2.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| *Meta.Matrix_Hamming_Distance* | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_Degree Semantic_Network._Average* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_BouednessSemantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

4.4.2.7.3 GLM

*4.4.2.7.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_NeedSemantic_Network.* | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_BouednessSemantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.7.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBonacich_PowerSemantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_DegreeSemantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### 4.4.2.7.4 Random Forests

###### *4.4.2.7.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Overall_Complexity* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Meta.Matrix_Hamming_Distance* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Eigenvector.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.7.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Overall_Complexity* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *MetaMatrixHammingDistance* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *NetworkCentralizationEigenvectorSemanticNetwork* | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.7.5 SVM rbf

##### *4.4.2.7.5.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.In_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.7.5.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Overall_Complexity* | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Meta.Matrix_Hamming_Distance* | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Eigenvector.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_BouednessSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.8 Section (File) 8*

#### 4.4.2.8.1 Linear Model

##### *4.4.2.8.1.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *ExclusivityComplete.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_CentralizationIn_DegreeSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityOut_DegreeSe* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *mantic_Network_Average* | | | | | | | | | | | | | |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Count.Weak.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*4.4.2.8.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *ExclusivityCompleteSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Component_Count.Strong.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.8.2 CART

##### *4.4.2.8.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *ExclusivityComplete.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityOut_DegreeSemantic_Network_Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *Span_Of_Control.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Count.Weak.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

##### *4.4.2.8.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| *Centrality.Betweenness.Semantic_Network._Average* | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Centrality.Column_Degree.Semantic_Network._Average | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| Communicative_Need.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Efficiency.Semantic_Network. | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| ExclusivityCompleteSemantic_Network_Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hierarchy.Semantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Network_CentralizationIn_DegreeSemantic_Network | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Skip.Semantic_Network. | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Span_Of_ControlSemantic_Network | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Component_Count.Strong.Semantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Upper_BouednessSemantic_Network | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.8.3 GLM

##### *4.4.2.8.3.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average_Distance.Semantic_Network | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Centrality.Betweenness.Semantic_Network._Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Breadth.Column.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Communicative_Need.Semantic_Network. | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| Efficiency.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ExclusivityComplete.Semantic_Network_Average | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hierarchy.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Network_Centralization.In_Degree.Semantic_Network. | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| CentralityOut_DegreeSemantic_Network_Average | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Link_Count.Reciprocal.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Link_Count.Sequential.Semantic_Network. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Count.Weak.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*4.4.2.8.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweennessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *ExclusivityComplete.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Component_Count.Strong.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.8.4 Random Forests

##### *4.4.2.8.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *Centrality.Betweenness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *ExclusivityComplete.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityOut_DegreeSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Count.Weak.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.8.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *CentralityBetweennessSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Column_Degree.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *ExclusivityCompleteSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Span_Of_Control.Semantic_Network.* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Component_Count.Strong.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.8.5 SVM rbf

##### *4.4.2.8.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *CentralityBetweennessSemantic_Network_Average* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *ExclusivityCompleteSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *CentralityOut_DegreeSemantic_Network_Average* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Skip.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_Count.Weak.Semantic_Network.* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*4.4.2.8.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Column_Degree.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *ExclusivityCompleteSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Skip.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Span_Of_Control.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Component_Count.Strong.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *4.4.2.9 Section (File) 9*

#### 4.4.2.9.1 Linear Model

##### *4.4.2.9.1.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweennessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityEigenvectorSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.9.1.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_MembersWeakSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

#### 4.4.2.9.2 CART

##### *4.4.2.9.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 |
| *Centrality.BetweennessSemantic_Network_Average* | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *CentralityEigenvectorSemantic_Network_Average* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.9.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_MembersWeakSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |

#### 4.4.2.9.3 GLM

*4.4.2.9.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityEigenvector.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.9.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_MembersWeak.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

#### 4.4.2.9.4 Random Forests

##### *4.4.2.9.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityEigenvectorSemantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.9.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Link_Count.Pooled.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_MembersWeakSemantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.9.5 SVM rbf

##### *4.4.2.9.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Speed.Average.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityEigenvector.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.9.5.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityBetweenness.Semantic_Network_Average* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Pooled.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Component_MembersWeakSemantic_Network._Average* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### *4.4.2.10 Section (File) 10*

4.4.2.10.1 Linear Model

*4.4.2.10.1.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInCloseness.Semantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Row.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.10.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *EfficiencySemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityInClosenessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Network_CentralizationInClosenessSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_CountSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_CountLateralSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.10.2 CART

##### *4.4.2.10.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network._Average* | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| *Isolate_Count.Semantic_Network.* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Row.Semantic_Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.10.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *CentralityInClosenessSemantic_Network._Average* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |
| *Isolate_Count.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.10.3 GLM

*4.4.2.10.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Row.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.10.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *CentralityHubSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *CentralityInClosenessSemantic_Network_Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Network_CentralizationInClosenessSemantic_Network* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### 4.4.2.10.4 Random Forests

###### *4.4.2.10.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *BreadthRowSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_CountSequentialSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_BouednessSemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*4.4.2.10.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityIn.ClosenessSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Lateral.Semantic_Network.* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### 4.4.2.10.5 SVM rbf

##### *4.4.2.10.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Centrality.Authority.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Speed.Average.Semantic_Network.* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.Column_Degree.Semantic_Network.* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size.Burt.Semantic_Network._Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.InClosenessSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Breadth.Row.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *4.4.2.10.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Number_of_Concept_nodes* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Speed.Average.Semantic_Network.* | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *Centrality.Closeness.Semantic_Network._Average* | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Breadth.Column.Semantic _Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *CentralityColumn_DegreeSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization. Column_Degree.Semantic_ Network.* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effective_Network_Size. Burt.Semantic_Network._ Average* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Efficiency.Semantic_Network.* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *HierarchySemantic_Network* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Centrality.Hub.Semantic _Network._Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *CentralityInClosenessSemantic_Network_Average* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Network_Centralization. In.Closeness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Isolate_Count.Semantic_ Network.* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Lateral.Semantic_Network.* | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Upper_Bouedness.Semantic_Network.* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### 4.4.3 Summary Results from correlation study of public policy data

The Table below summarizes the best performances obtained for each Dependent Variable with this appropriate subset of independent variables as selected in the previous step. It represents a sampling of the R-squared results using the SVM model.

| DEPENDENT VARIABLE | SHIFT | NOSHIFT |
|---|---|---|
| ctr1 | 0.667316855135585 | 0.627973315207854 |
| ctr2 | 0.650389256871564 | 0.636832912384701 |
| X1_Month | 0.459108422906905 | 0.513670559348339 |
| X3_Month | 0.514827564741221 | 0.463452579475976 |
| X6_Month | 0.623851875896093 | 0.48594160027166 |
| X1_Year | 0.50862352073609 | 0.525093668326345 |
| X2_Year | 0.63555691159017 | 0.545654140181347 |
| X3_Year | 0.63555691159017 | 0.545654140181347 |
| X5_Year | 0.493953298579009 | 0.508671018221193 |
| X7_Year | 0.59688231317501 | 0.521646563126267 |
| X10_Year | 0.701033243584885 | 0.661597255848017 |
| X20_Year | 0.60829605892706 | 0.597090870929685 |
| X30_Year | 0.70169749152022 | 0.623276112040356 |
| | | |
| Mean | 0.599776440404152 | 0.558196518118699 |
| S.D. | 0.0803619266711078 | 0.0638169023286193 |

**Table 10: Results from using SVM on FOMC data**

The SVM model obtains the best performance for this particular data set. For each dependent variable the R-squared value is at least 0.46. This is presented in the summary tables and visuals for all 100 of the learning algorithm summary results presented in the previous section. In the graphical Figures presented below, the learning algorithms are represented in order below as LM, CART, GLM, RF, and SVM. These show clearly the effectiveness of which learning algorithms on which files. The first Figure is for no-shift data, the second is for time-shifted data.

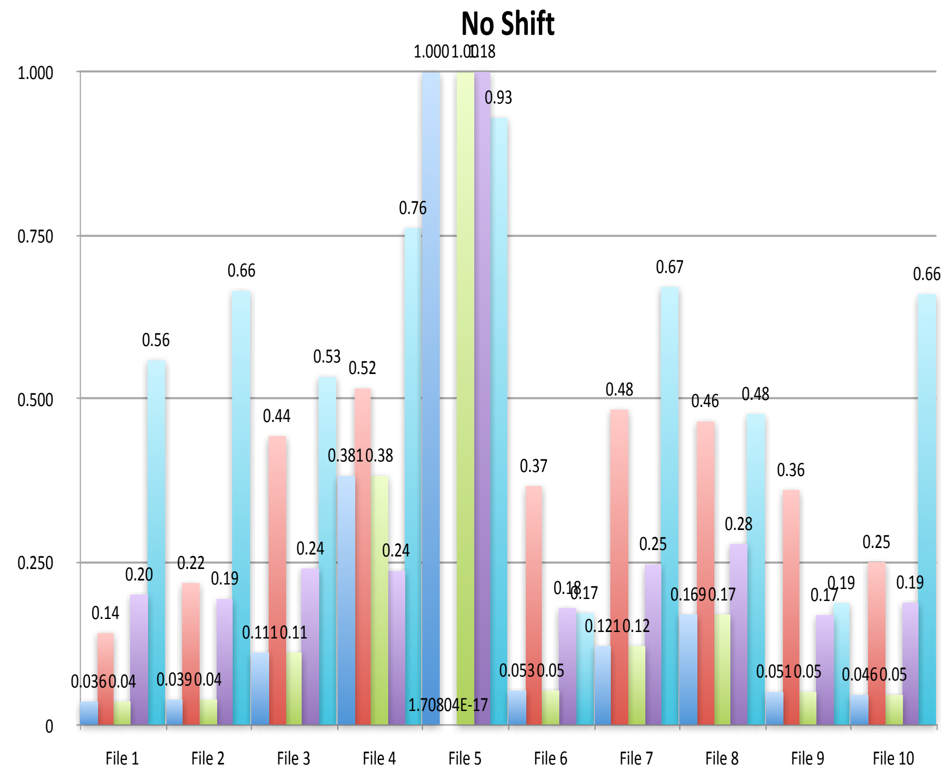


**Figure 28: No Time Shift Summary results from five learning algorithms on ten files; learning algorithms LM, CART, GLM, RF, and SVM (Numerical scale represents pseudo R-squard)**

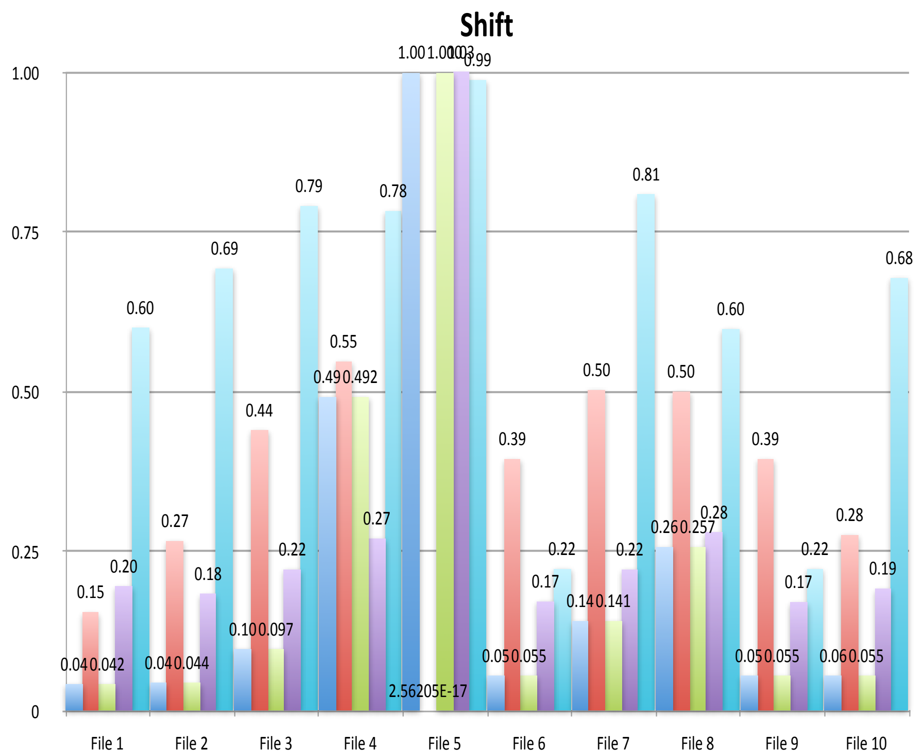


**Figure 29: Time Shift (best Shift) Summary results from five learning algorithms on ten files, learning algorithms LM, CART, GLM, RF, and SVM (Numerical scale represents pseudo R-squard)**

In the Table below, the summary statistics are presented for each of the network measurements used. These numbers represent the total range seen for all networks considered in the final analysis of this Dissertation. Especially as interpretations are posited on the value of attention paid to each measure or combination of measures, the absolute numbers can be useful to reference as well as the magnitude.

| Exemplar Independent Variable | arithmetic mean | median | mode | Min. | Max. |
|---|---|---|---|---|---|
| Number of Concept nodes | 861.10 | 866.50 | 779.00 | 93.00 | 1939.00 |
| Overall Complexity | 0.02 | 0.02 | 0.01 | 0.01 | 0.08 |
| Meta-Matrix Hamming Distance | 0.01 | 0.01 | 0.01 | 0.00 | 0.04 |
| Centrality-Authority Average | 0.03 | 0.02 | 0.02 | 0.00 | 0.10 |
| Speed-Average | 0.33 | 0.33 | 0.34 | 0.18 | 0.38 |
| Breadth-Column | 1.00 | 1.00 | 1.00 | 0.98 | 1.00 |
| Centrality-Column Degree Average | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 |
| Communicative Need | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Effective Network Size-Burt Average | 11.25 | 11.02 | 12.02 | 3.74 | 23.39 |
| Hierarchy | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Centrality-In-Closeness Average | -797.50 | 0.32 | 0.32 | -814613.78 | 0.83 |
| Network Centralization-In-Closeness | 1599.43 | 0.26 | 0.00 | 0.00 | 1631396.38 |
| Network Centralization-In Degree | 0.02 | 0.02 | 0.02 | 0.00 | 0.12 |
| Isolate Count | 0.45 | 0.00 | 0.00 | 0.00 | 15.00 |
| Link Count-Lateral | 0.55 | 0.54 | 0.54 | 0.43 | 1.00 |
| Link Count-Reciprocal | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Link Count-Sequential | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Link Count-Skip | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 |
| Upper Boundedness | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

**Table 11: Summary Statisics on Independent Variables in Public Policy Data**

Excluding those measurements in the table above that report either all zeros or all 1s, the Figure below is produced for median, mode,

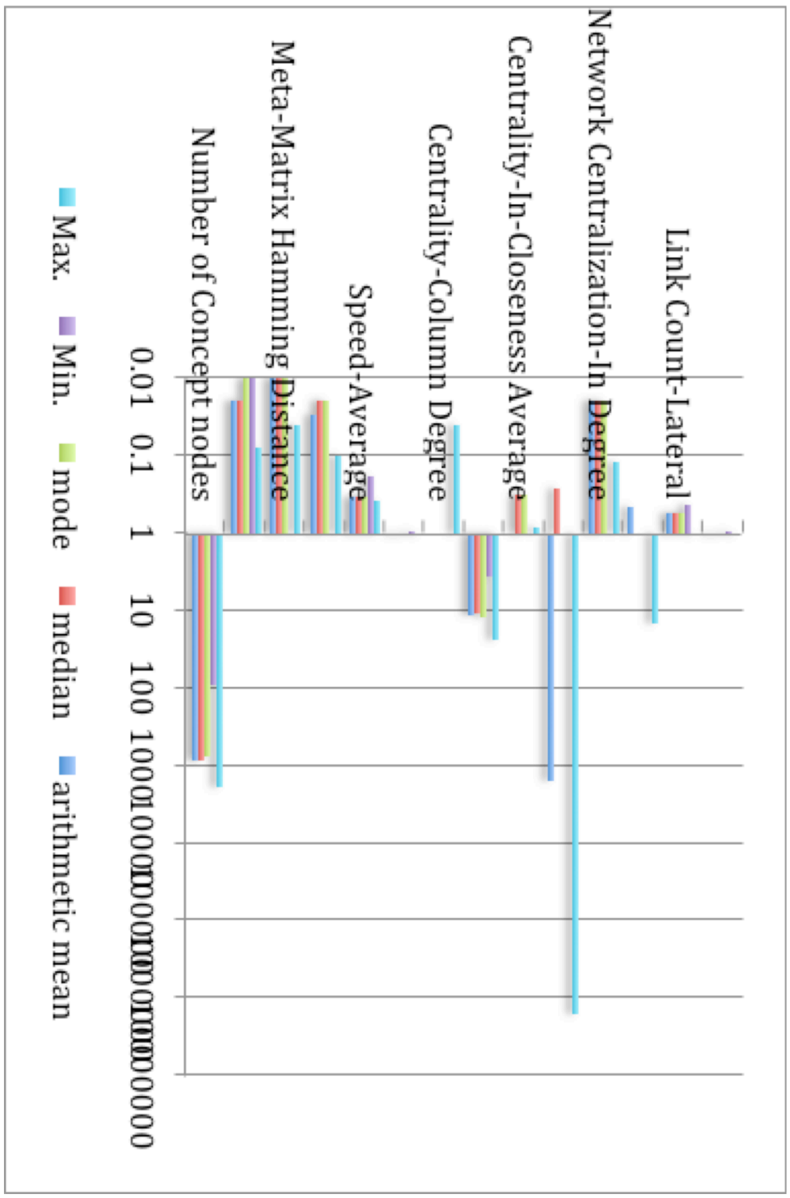


**Figure 30: Summary Statistics for Public Policy Data**

& average. Combined with the additional Figure below, these visualizations may help to better frame the results for the implication discussions that follow later in this Dissertation.

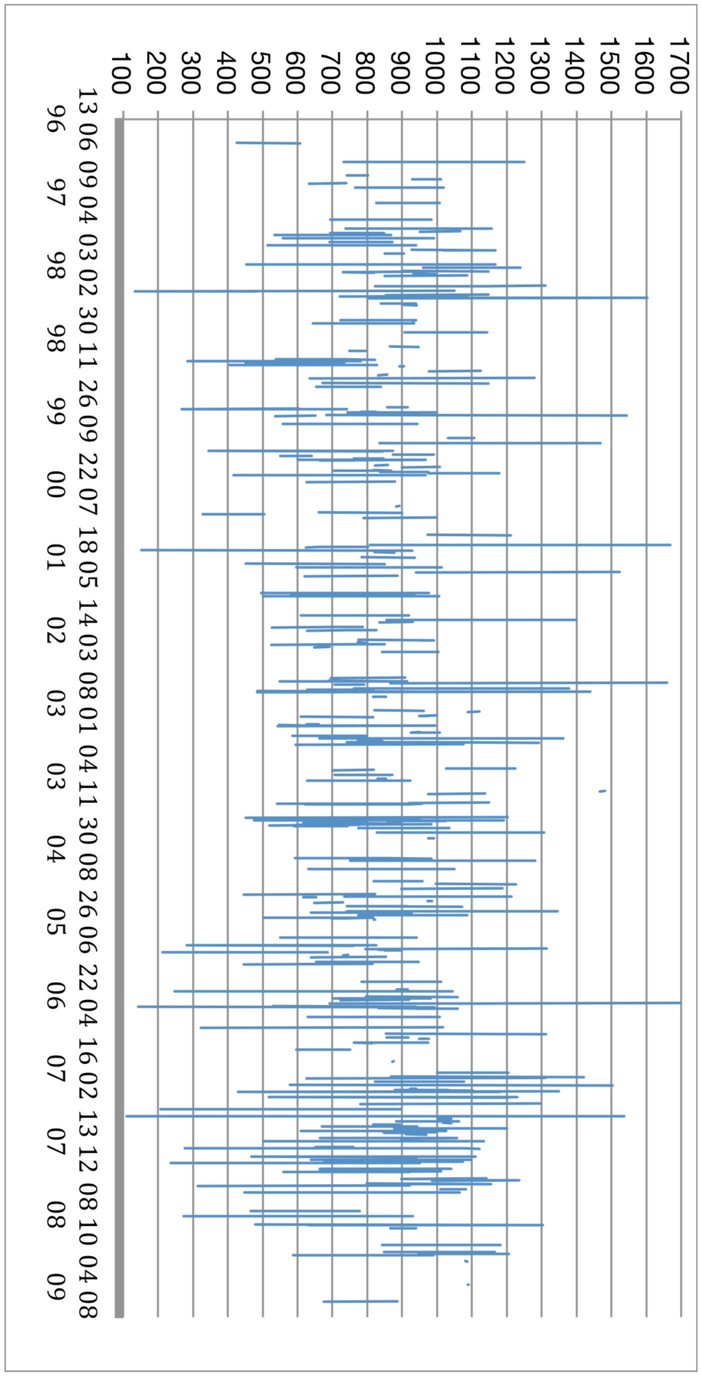


**Figure 31: Number of concept nodes over time for public policy dataset**

| SHIFT | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **File 1** | **File 2** | **File 3** | **File 4** | **File 5** | **File 6** | **File 7** | **File 8** | **File 9** | **File 10** |
| **lm** | 0.04 | 0.044 | 0.097 | 0.49 | 1.0 | 0.055 | 0.14 | 0.26 | 0.055 | 0.055 |
| **cart** | 0.15 | 0.27 | 0.44 | 0.55 | 0.00 | 0.39 | 0.50 | 0.50 | 0.39 | 0.28 |
| **glm** | 0.042 | 0.044 | 0.097 | 0.49 | 1.0 | 0.055 | 0.14 | 0.26 | 0.055 | 0.055 |
| **rf** | 0.20 | 0.18 | 0.22 | 0.27 | 1.0 | 0.17 | 0.22 | 0.28 | 0.17 | 0.19 |
| **svm** | 0.60 | 0.69 | 0.79 | 0.78 | 0.99 | 0.22 | 0.81 | 0.60 | 0.22 | 0.68 |
| NO SHIFT | | | | | | | | | | |
| | **File 1** | **File 2** | **File 3** | **File 4** | **File 5** | **File 6** | **File 7** | **File 8** | **File 9** | **File 10** |
| **lm** | 0.036 | 0.039 | 0.11 | 0.38 | 1 | 0.053 | 0.12 | 0.17 | 0.051 | 0.05 |
| **cart** | 0.14 | 0.22 | 0.44 | 0.52 | 0.00 | 0.37 | 0.48 | 0.46 | 0.36 | 0.25 |
| **glm** | 0.04 | 0.039 | 0.11 | 0.38 | 1 | 0.053 | 0.12 | 0.17 | 0.051 | 0.046 |
| **rf** | 0.20 | 0.19 | 0.24 | 0.24 | 1.2 | 0.18 | 0.25 | 0.28 | 0.17 | 0.19 |
| **svm** | 0.56 | 0.66 | 0.53 | 0.76 | 0.93 | 0.17 | 0.67 | 0.48 | 0.19 | 0.66 |

**Table 12: Summary Results from five learning algorithms on ten files both best-shifted and non-shifted for time**

Table 12 (above) shows the average performance across dependent variables for each regression model. It shows both the best-shift and non-shift models. These are the summary numbers from the tables earlier. They are important in showing which combination of data (the files), analysis (the algorithms) and treatment (shift or no-shift) produces the best results.

Table 13 (below) shows the independent variables chosen most often by the dependent variables for modeling a relationship. This is important in both the interpretation and in the consideration of future work.

| |
|---|
| Number_of_Concept_nodes |
| Meta.Matrix_Hamming_Distance |
| Centrality.Authority.Semantic_Network._Average |

| |
|---|
| Speed.Average.Semantic_Network. |
| Communicative_Need.Semantic_Network. |
| Effective_Network_Size.Burt.Semantic_Network._Average |
| Isolate_Count.Semantic_Network. |
| Link_Count.Lateral.Semantic_Network. |
| Link_Count.Skip.Semantic_Network. |

**Table 13: Independent Variables most frequently modeled by Dependent variables**

The regression models used include Linear Regression, CART, GLM (Gaussian link function), random forests, and support vector machines (with radial basis functions kernel). No tuning on regression model parameters was performed, leaving the default values in R (e.g., the default values for standard deviation and penalizing factor in the SVM model).

The shift models considered are two: no-shift and shift. In 'no shift', each independent variable was kept contemporaneous with the date. In 'shift', each pair of independent-dependent variable pair lead to a best shift of the independent variable within a range of +/- five observations. It is important to note that the shifting is unstable given the sparseness of the data for this particular study; shifting of one position in time does not necessarily mean shifting of one time-unit. Additionally, shifting either forward or backward in the time series (i.e., the independent variable) means introducing elements at the beginning or end of the time series. This further reduces the observations for which it is possible to calculate correlation in the regression model. For these reasons, characterized by sparse information, the no-shift model is the most appropriate one.

## 4.5 Conclusion from applying framework to public policy data

The approach presented in this paper is a systematic analysis of public policy speeches given by central bankers in the U.S. The analysis suggests some correlation between relevant financial data and the semantic networks approach presented. Between the two different approaches for analyzing the correlation and the combinations of twelve dependent variables, there appears to be some consistency in the independent variables. For example, Network Centralization (Column Degree) is an independent variable in nine cases in the regression analysis. Among the four models of the CART analysis, only five independent variables are unique to one model.

There is also some consistency of results among the multiple analysis methodologies. Average Distance played a part in seven of the Regression models and half of the CART models. It is important to note that the numbers themselves do not necessary mean anything. The numbers are relative to each other.

Together, the results suggest that in some circumstances, there exists a correlation between financial data and a systematic approach using semantic networks to analyzing public policy speeches. However, the conclusions are limited in several ways.

First, the speeches are for those by U.S. Central Bankers. These have benefits as outlined earlier, but the conclusions may prove difficult to generalize to other Central Banks and the public policy pronouncements of other officials.

Second, while there are many documents produced by the U.S. Central Bank, this research looks at only the speeches as stated earlier. The minutes of the FOMC board meeting minutes could be another study.

Third, this study only covers the years 2006-2007. While there are good reasons for this limitation as described earlier, the effectiveness of the conclusions may vary over other years.

Fourth, this correlation does not predict the outcomes of the results in any way. Prediction of any sort, for example, either binary (i.e., the numbers will go up or down) or in direction (i.e., the numbers will stop going up) would be very interesting research by itself. Others have begun to explore this (Robertson &

Thornton, 1997), (Luss & d'Aspremont, 2008), (Fleming & Remolona, 1999a)

Fifth, the results of a Semantic Network approach are inherently impacted by qualitative decisions made early in the process such as the development of the delete list and the Thesaurus.

Sixth, there are other dependent variables that could be included in further study such as U.S. GDP growth or the 34 other Fed Funds Futures expiring between 2007 and 2008. Treatment of dependent variables could also vary such as normalization to equity prices or equity derivatives.

# 5 Exploring corporate email as a basis for predicting financial events

## 5.1 Introduction to study of large email corpus

Despite the massive volume of email communication, privacy concerns may have limited studies of emails to analysis in the aggregate. The public availability of years of emails from the demise of Enron has generated many new studies previously unavailable (Diesner et al., 2005; Klimt & Yang, 2004; Klimt & Yiming, 2004) (Keila & Skillicorn, 2005). New studies have looked at relationships, speaking style (Sabater, Turney, & Fleta, 2008), and even patterns of behavior (Qian, Zhang, & Yang, 2006).

This chapter introduces a new systematic methodology for the analysis of this substantial email corpus. This analysis is interesting for at least a few reasons:

- The vast scale of email usage. It is used both widely and frequently.
- The range of usage. Email is used both a communication tool of individuals and organizations.
- Email captures both informal and formal communication
- This new process allows for an inquiry into correlations with other data.
- Email content can be studied alongside sender/receiver relationships
- Email volume is generally continuous

With a correlation between changes in behavior captured at an organizational level, enormous new opportunities for study open:

- Personnel engagement may be judged based on email communication
- Organizational health may be measured in a new way

- Communications effectiveness may not only be evaluated, but coached

- Treasury departments may be have a new metric against which to judge the timing of corporate financial actions

This chapter concerns itself with establishing a system for analyzing the email corpus in such a way that can be routinely applied.

## 5.2 Background on studies of email

Implications from the work on the Enron corpus have been far reaching. The email communication can be seen as a test bed for text classification (Wang et al., 2007), a study in the network of relationships (Diesner et al., 2005) or an analysis of discourse. The Figure below shows the stock price of Enron as a context for the data under consideration. It suggests the relevance of studying the more volatile time periods included in this Dissertation.

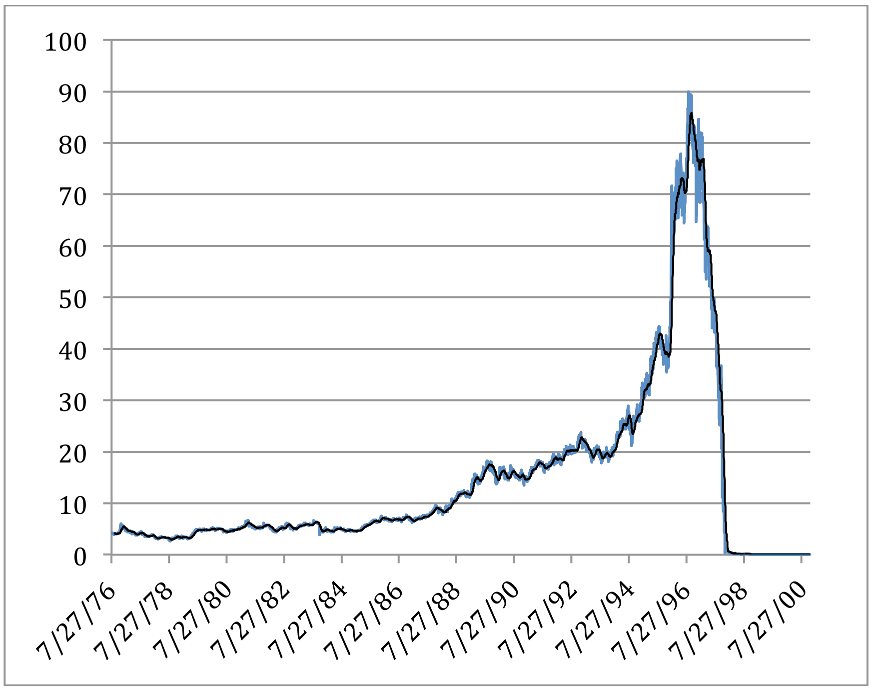


**Figure 32: Enron closing day and 30-day moving average equity price 1980-2004 (linear-scale y-axis as closing price of Enron Equity in USD)**

This framework is influenced by, and influences, all three. While email is, by volume, mostly a one-way communication mechanism, its origins were in 1:1 communication. While some argue that language and discourse has been changed by email (Judge, 2012) (Walther, 2012) (Algeo & Pyles, 2009), others argue that discourse and language have changed the nature of email (Herring, 2012). The study of sentiment, perspective, and opinion in email are social aspects to text open to social interpretation (Rosé, 2012).

## 5.3 Methodology of email study

There are some important differences in the methods for the acquisition, processing, and analysis of email data. The nature of the scale of the analysis requires some changes to the approach.

### 5.3.1. Qualitative Data

***Part I: Acquire and clean qualitative data***

*Step 1*: Acquire raw text.

A large set of email messages, the Enron corpus, was made public during the legal investigation concerning the Enron corporation (Gervasio, 2004).

*Step 2*: Separate useful text from noise and neutralize formatting.

Further details on the Enron corpus are provided by Diesner (Diesner & Carley, 2005), Klimt (Klimt & Yang, 2004), Keila (Keila & Skillicorn, 2005), and Priebe (Priebe et al., 2005): The raw corpus contains 619,446 messages belonging to 158 users. Deleting duplicate messages gives gives 498,849 and considering threads of length greater than one and those messages within the date under consideration for this study (1997-2002) gives 449,442. Other emails outside of this date range have been identified as either intentionally misleading (marketing messages that might be characterized as email spam) or with characteristics that make another date characterization impossible (i.e., there is not identifying information to place the email at any *other* date).

### 5.3.2 Quantitative Data

***Part II: Acquire and clean quantitative data***

*See Section 3.3.2*

### 5.3.3 Text Transformation

***Part III: Transformation of text***

*See Section 3.3.3*

## 5.4 Results from study of email corpus

The next step is to map the equity prices (shown in the Figure below) to the text analysis methods. This log-axis plot of the equity price shows most clearly the linear increase in prices before the collapse.

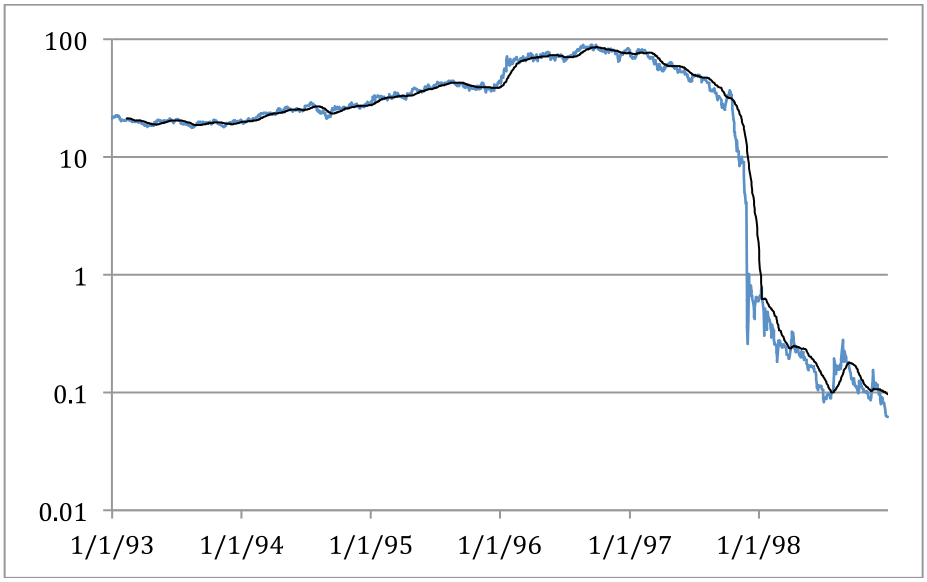


**Figure 33: Enron closing day and 30-day moving average equity price during period of this study: 1999-2002 (log-scale y-axis as closing price of Enron Equity in USD)**

In the table below, the summary statistics are presented for each of the network measurements used. These numbers represent the total range seen for all networks considered in the final analysis of this Dissertation. Especially as interpretations are posited on the value of attention paid to each measure or combination of measures, the absolute numbers can be useful to reference as well as the magnitude.

| Exemplar Independent Variable | arithmetic mean | median | mode | minimum | maximum |
|---|---|---|---|---|---|
| Number of Concept nodes | 119.32 | 109.85 | 27.50 | 17.50 | 1858.50 |
| Overall Complexity | 0.19 | 0.19 | 0.35 | 0.01 | 0.48 |
| Meta-Matrix Hamming Distance | 0.08 | 0.08 | 0.13 | 0.01 | 0.26 |
| Centrality-Authority Average | 0.10 | 0.10 | 0.24 | 0.01 | 0.29 |
| Speed-Average | 0.40 | 0.41 | 0.55 | 0.16 | 0.62 |
| Breadth-Column | 0.93 | 0.99 | 1.00 | 0.33 | 1.00 |
| Centrality-Column Degree Average | 0.04 | 0.03 | 0.13 | 0.00 | 0.19 |
| Communicative Need | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Effective Network Size-Burt Average | 13.58 | 14.79 | 6.03 | 2.13 | 43.15 |
| Hierarchy | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Centrality-In-Closeness Average | -36954.22 | 0.35 | 0.56 | -11311939.83 | 0.64 |
| Network Centralization-In-Closeness | 3627.59 | 0.25 | 0.31 | 0.00 | 877187.64 |
| Network Centralization-In Degree | 0.04 | 0.03 | 0.13 | 0.00 | 0.19 |
| Isolate Count | 0.49 | 0.19 | 0.00 | 0.00 | 5.45 |
| Link Count-Lateral | 0.65 | 0.63 | 0.59 | 0.33 | 1.00 |
| Link Count-Reciprocal | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Link Count-Sequential | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Link Count-Skip | 0.93 | 0.98 | 0.99 | 0.28 | 1.00 |
| Upper Boundedness | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

**Table 14: Summary Statistics on Independent Variables in email data**

The following two Figures present a visualization of the total count studied and of the above Table in order to help make the numbers more clear for understanding in the context of the implication discussion elsewhere in this Disseration.

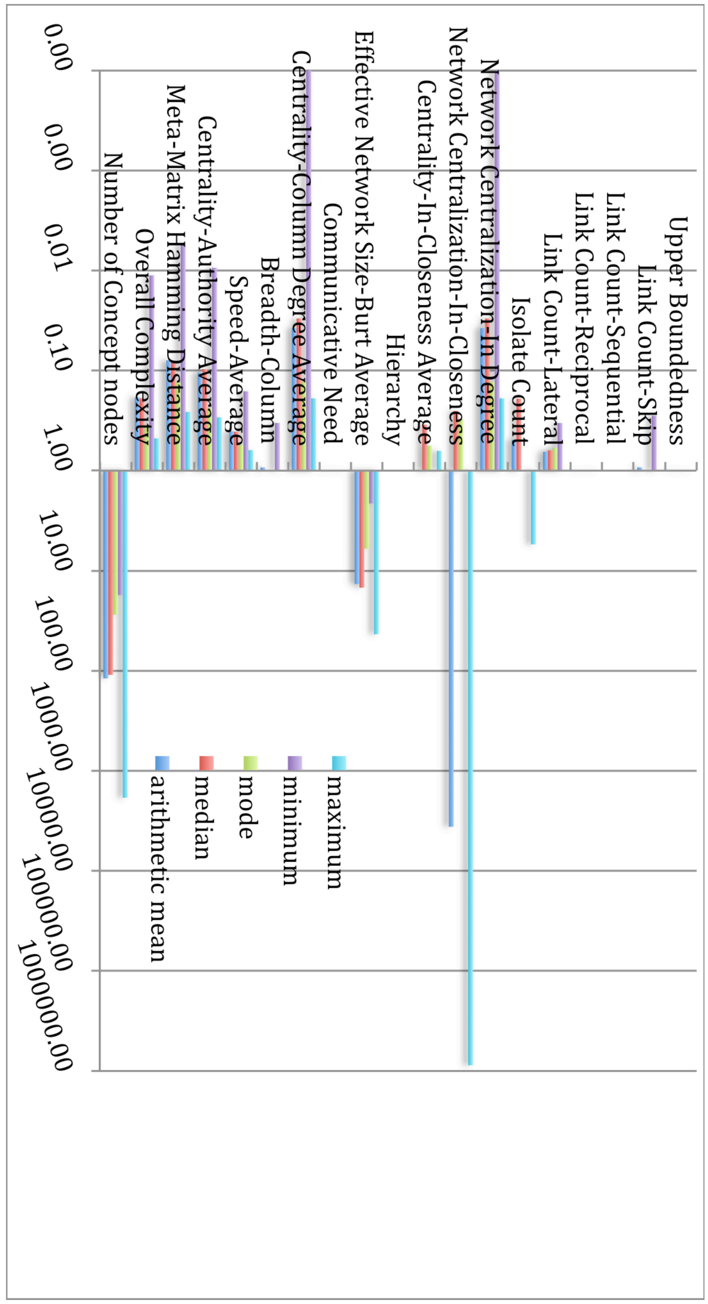


**Figure 34: Summary Statistics for Corporate Finance Data**

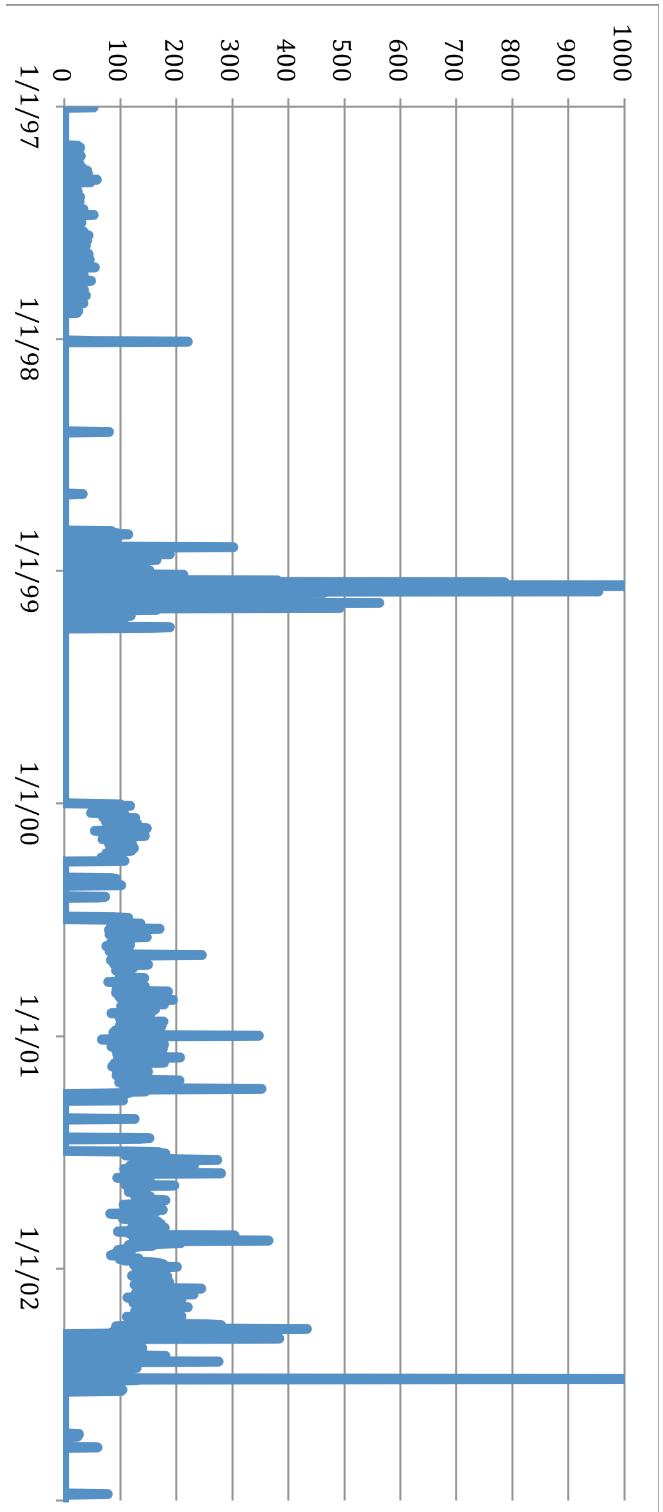


**Figure 35: Number of Nodes in processed email dataset**

### 5.4.1 Statistical Results for each learning algorithm

Below are the summary results for each learning algorithm. These are presented in summary form taken from other tables within this Dissertation just to make the results more clear. The Figure below presents a visual representation of the same data as another form aiding in interpretation of the results.

#### 5.4.1.1 Best Shift

| | ***R2*** |
|---|---|
| ***lm*** | 0.3057198 |
| ***cart*** | 0.80834252 |
| ***glm*** | 0.3057198 |
| ***rf*** | 0.88128138 |
| ***svm*** | 0.79029317 |

#### 5.4.1.2 No Shift

| | ***R2*** |
|---|---|
| ***lm*** | 0.30625858 |
| ***cart*** | 0.80196355 |
| ***glm*** | 0.30625858 |
| ***rf*** | 0.87727112 |
| ***svm*** | 0.79097243 |

#### 5.4.1.3 No. of Observations being considered

| | ***ds4*** |
|---|---|
| ***N.obs*** | 2190 |
| ***Starting Day*** | 2-Jan-97 |
| ***Ending Day*** | 31-Dec-02 |

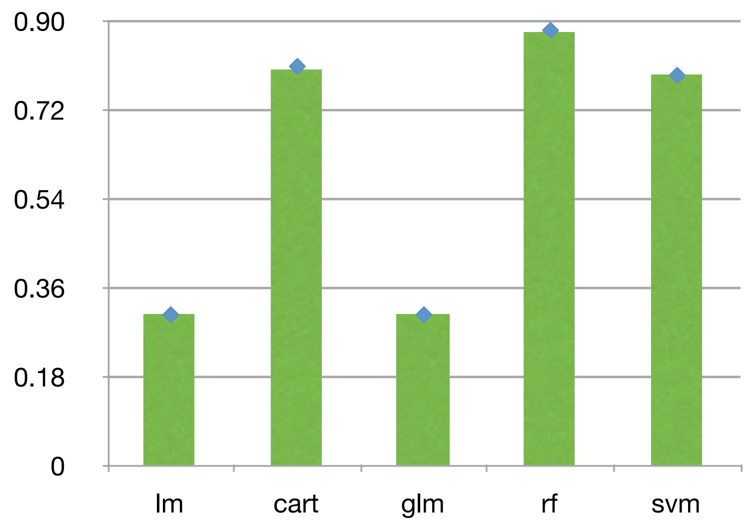


**Figure 36: Statistical Results from time-shifted analysis (green bars) and contemporaneous (blue dots) of corporate email data using framework. (Numerical scale represents pseudo R-squared)**

### 5.4.2 Results from approach applied to Corporate email data

#### *5.4.2.1 Variable Selection on Corporate email*

5.4.2.1.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *Enron* | 0.19359093 | 0.19391205 |
| *SP* | 0.41784867 | 0.41860511 |
| ***mean*** | 0.3057198 | 0.30625858 |
| ***sd*** | 0.15857417 | 0.15888198 |

*5.4.2.1.1.1 Time Shift*

| Dependent Variable | Enron | SP |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |

| | | |
|---|---|---|
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 1 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

*5.4.2.1.1.2 No Time Shift*

| **Dependent Variable** | **Enron** | **SP** |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 1 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

5.4.2.1.2 Cart

5.4.2.1.2 Cart

| DV | Shift | NoShift |
|---|---|---|
| *Enron* | 0.79752118 | 0.79282952 |
| *SP* | 0.81916387 | 0.81109757 |
| ***mean*** | 0.80834252 | 0.80196355 |
| ***sd*** | 0.01530369 | 0.01291747 |

*5.4.2.1.2.1 Time Shift*

| **Dependent Variable** | **Enron** | **SP** |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 0 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 0 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 0 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 1 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

*5.4.2.1.2.2 No Time Shift*

| **Dependent Variable** | **Enron** | **SP** |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 0 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 0 | 0 |

| | | |
|---|---|---|
| *Breadth.Column.Semantic_Network.* | 1 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 0 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 0 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 1 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

5.4.2.1.3 GLM

5.4.2.1.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *Enron* | 0.19359093 | 0.19391205 |
| *SP* | 0.41784867 | 0.41860511 |
| ***mean*** | 0.3057198 | 0.30625858 |
| ***sd*** | 0.15857417 | 0.15888198 |

*5.4.2.1.3.1 Time Shift*

| Dependent Variable | Enron | SP |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 |
| *Count.Node.Semantic_Network.* | 1 | 1 |

| | | |
|---|---|---|
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 1 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

*5.4.2.1.3.2 No Time Shift*

| **Dependent Variable** | **Enron** | **SP** |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 1 | 1 |
| *Communicative_Need.Semantic_Network.* | 1 | 1 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 1 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

5.4.2.1.4 Random Forest

5.4.2.1.4 Random Forest

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *Enron* | 0.85361775 | 0.85137037 |
| *SP* | 0.90894501 | 0.90317188 |
| ***mean*** | 0.88128138 | 0.87727112 |
| ***sd*** | 0.03912229 | 0.0366292 |

*5.4.2.1.4.1 Time Shift*

| Dependent Variable | Enron | SP |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 1 |
| *Hierarchy.Semantic_Network.* | 1 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 0 |
| *Upper_Boundedness.Semantic_Network.* | 1 | 0 |

*5.4.2.1.4.2 No Time Shift*

| Dependent Variable | Enron | SP |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 1 | 1 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 0 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 1 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 1 | 1 |
| *Count.Node.Semantic_Network.* | 1 | 1 |

| | | |
|---|---|---|
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 0 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

5.4.2.1.5 SVM (radial basis)

5.4.2.1.5 SVM (radial basis)

| DV | Shift | NoShift |
|---|---|---|
| *Enron* | 0.7612759 | 0.76131746 |
| *SP* | 0.81931044 | 0.8206274 |
| ***mean*** | 0.79029317 | 0.79097243 |
| ***sd*** | 0.04103661 | 0.04193846 |

*5.4.2.1.5.1 Time Shift*

| Dependent Variable | Enron | SP |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 0 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 0 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

*5.4.2.1.5.2 No Time Shift*

| **Dependent Variable** | **Enron** | **SP** |
|---|---|---|
| *Average_Distance.Semantic_Network.* | 0 | 0 |
| *Centrality.Bonacich_Power.Semantic_Network._Average* | 1 | 1 |
| *Network_Centralization.Closeness.Semantic_Network.* | 1 | 1 |
| *Cognitive_Expertise_Average* | 1 | 1 |
| *Breadth.Column.Semantic_Network.* | 0 | 0 |
| *Communicative_Need.Semantic_Network.* | 0 | 0 |
| *Hierarchy.Semantic_Network.* | 0 | 0 |
| *Centrality.Hub.Semantic_Network._Average* | 1 | 1 |
| *Centrality.In.Closeness.Semantic_Network._Average* | 0 | 0 |
| *Network_Centralization.In.Closeness.Semantic_Network.* | 0 | 0 |
| *Count.Node.Semantic_Network.* | 1 | 1 |
| *Link_Count.Reciprocal.Semantic_Network.* | 0 | 0 |
| *Link_Count.Sequential.Semantic_Network.* | 0 | 0 |
| *Triad_Count.Semantic_Network._Average* | 1 | 0 |
| *Upper_Boundedness.Semantic_Network.* | 0 | 0 |

## 5.5 Conclusions from email study

Under some circumstances there does appear to be a correlation between the email corpus and the financial data. The shifting of time does not appear to make a material difference.

# 6 Comparitive analysis using a baseline approach of sentiment analysis

## 6.1 Introduction to baseline comparisons

The developments represented in earlier chapters of this document are described in their effectiveness on an absolute basis. The framework is presented and then analyzed using two different datasets. The judgement of value from these absolute measurements may be given perspective from comparison with another standard of analysis. The comparison considered in this chapter is that of basic sentiment analysis and its level of effectiveness in detecting a similar correlation under consideration from Chapter four and five.

## 6.2 Background on baseline comparisons

Sentiment analysis involves classifying opinions in text into categories like ‘positive’ or ‘negative’ often with an implicit category of ‘neutral’. A classic sentiment application would be tracking down what bloggers are saying about a brand such as Apple. Sentiment analysis is also called opinion mining (Baldwin & Carpenter, 2012).

Sentiment analysis at this level can look to be quite basic as having positive or negative sentiment. A sample of short positive sentences such as

- I love this home;
- This weather is amazing;
- I feel great right now;
- I am so excited about the dinner;
- She is my best friend;

can be easily contracted with short negative sentences such as

- I do not like this home;
- This weather is terrible;
- I feel tired right now;
- I do not look forward to this dinner;
- She is my enemy.

These types of sentences can be used to train the classifiers used in sentiment analysis. With longer sentences, the algorithms differ in their approaches and their effectiveness. For example, the an application by Laurent Luce uses Python and the Natural Language Toolkit (TLTK) to optimize for just such short sentences in the analysis of Tweets (Luce, 2012).

## 6.3 Methods for baseline comparisions

In this study, an appropriate baseline measurement of simple observations is sought. While sentiment analysis can represent nuances (Uijlings, Smeulders, & Scha, 2009) (Tirilly, Claveau, & Gros, 2008), this study uses two popular algorithms as baselines: LingPipe (Alias-i, 2008), which might be the most appropriate representation of the approach (Carpenter, 2004) (Carpenter, 2006) (Carpenter, 2007) (Konchady, 2008) and Sentiwordnet (Esuli & Sebastiani, 2006). The two data sets explored in Chapters four and five are both included in this comparison. Based on the results in time shifts from the earlier studies of the data, some time-shifted data was excluded to simplify comparisons. To ensure equal comparison, the full text of each email is included and multiple measurements on a single day are averaged into one measurement per day.

## 6.4 Results of baseline comparisons using LingPipe

### 6.4.1 Summary of Statistical relationships of sentiment analysis on Public Policy Data

#### 6.4.1.1 Shift

| | ***speeches_only*** | ***minutes_only*** | ***combined*** |
|---|---|---|---|
| ***lm*** | 0.02507089 | 0.14287396 | 0.02408234 |
| ***cart*** | 0.11068188 | 0.4062863 | 0.07694316 |
| ***glm*** | 0.02514788 | 0.14322739 | 0.02408931 |
| ***rf*** | 0.37082467 | 0.46336863 | 0.38912912 |
| ***svm*** | 0.14233192 | 0.38614606 | 0.13503587 |

#### 6.4.1.2 No Shift

| | ***speeches_only*** | ***minutes_only*** | ***combined*** |
|---|---|---|---|
| ***lm*** | 0.0295992 | 0.0457892 | 0.02320204 |
| ***cart*** | 0.13428867 | 0.31691777 | 0.09039193 |
| ***glm*** | 0.03012585 | 0.04600033 | 0.02363621 |
| ***rf*** | 0.37472643 | 0.52018652 | 0.3865248 |
| ***svm*** | 0.07608528 | 0.25590367 | 0.06649528 |

#### 6.4.1.3 No. of Observations being considered

| | ***speeches_only*** | ***minutes_only*** | ***combined*** |
|---|---|---|---|
| ***N.obs*** | 672 | 96 | 767 |
| ***Starting Day*** | 05_01_97 | 05_02_97 | 05_01_97 |
| ***Ending Day*** | 08_12_08 | 16_12_08 | 16_12_08 |

## (ANALYSIS 1)

### *6.4.1.1 Variable selection on Speeches Only*

#### 6.4.1.1.1 Linear Model

*6.4.1.1.1.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |

*6.4.1.1.1.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *X._Neutral_Sentences* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

#### 6.4.1.1.2 CART

*6.4.1.1.2.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

*6.4.1.1.2.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Negative_Sentences* | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| *Neutral_Sentences* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |

#### 6.4.1.1.3 GLM

*6.4.1.1.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*6.4.1.1.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

6.4.1.1.4 Random Forests

*6.4.1.1.4.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Positive_Sentences* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *X._Negative_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Neutral_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

*6.4.1.1.4.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *X._Positive_Sentences* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Neutral_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### 6.4.1.1.5 SVM (radial basis)

###### *6.4.1.1.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |

###### *6.4.1.1.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### *6.4.1.2. Variable Selection on Minutes Only*

#### 6.4.1.2.1 Linear Model

*6.4.1.2.1.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |

*6.4.1.2.1.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

#### 6.4.1.2.2 CART

*6.4.1.2.2.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Neutral_Sentences* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *X._Neutral_Sentences* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

###### *6.4.1.2.1.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *X._Negative_Sentences* | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| *X._Neutral_Sentences* | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

##### 6.4.1.2.3 GLM

###### *6.4.1.2.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*6.4.1.2.3.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

6.4.1.2.4 Random Forests

*6.4.1.2.4.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| *X._Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Negative_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *X._Neutral_Sentences* | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |

*6.4.1.2.4.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *X._Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| *X._Neutral_Sentences* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |

#### 6.4.1.2.5 SVM (radial basis)

##### *6.4.1.2.5.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

##### *6.4.1.2.5.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

### *6.4.1.3. Variable selction on Combined Data*

#### 6.4.1.3.1 Linear Model

##### *6.4.1.3.1.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

##### *6.4.1.3.1.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

#### 6.4.1.3.2 CART

##### *6.4.1.3.2.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *Negative_Sentences* | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| *X._Positive_Sentences* | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| *X._Negative_Sentences* | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

##### *6.4.1.3.2.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *X._Positive_Sentences* | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| *X._Negative_Sentences* | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |

6.4.1.3.3 GLM

*6.4.1.3.3.1 Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*6.4.1.3.3.2 No Time Shift*

| **Dependent Variable** | **ctr1** | **ctr2** | **X1_Month** | **X3_Month** | **X6_Month** | **X1_Year** | **X2_Year** | **X3_Year** | **X5_Year** | **X7_Year** | **X10_Year** | **X20_Year** | **X30_Year** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

6.4.1.3.4 Random Forests

*6.4.1.3.4.1 Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Positive_Sentences* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *X._Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Neutral_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

*6.4.1.3.4.2 No Time Shift*

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Positive_Sentences* | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| *X._Negative_Sentences* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| *X._Neutral_Sentences* | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

#### 6.4.1.3.5 SVM (radial basis)

##### 6.4.1.3.5.1 Time Shift

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| *Negative_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| *Neutral_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| *X._Negative_Sentences* | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

##### 6.4.1.3.5.2 No Time Shift

| Dependent Variable | ctr1 | ctr2 | X1_Month | X3_Month | X6_Month | X1_Year | X2_Year | X3_Year | X5_Year | X7_Year | X10_Year | X20_Year | X30_Year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Positive_Sentences* | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *Negative_Sentences* | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| *Neutral_Sentences* | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *X._Neutral_Sentences* | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

#### *6.4.1.1. Statistical Relationships of Speeches Only data set*

6.4.1.1.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.01759973 | 0.0328862 |
| *ctr2* | 0.01691962 | 0.01118617 |
| *X1_Month* | 0.04648381 | 0.04906992 |
| *X3_Month* | 0.01319452 | 0.02502083 |
| *X6_Month* | 0.01256812 | 0.02447051 |
| *X1_Year* | 0.01367335 | 0.02293652 |
| *X2_Year* | 0.0255131 | 0.02494518 |
| *X3_Year* | 0.0255131 | 0.02494518 |
| *X5_Year* | 0.02630808 | 0.02646164 |
| *X7_Year* | 0.02956915 | 0.03128239 |
| *X10_Year* | 0.03157518 | 0.03505746 |
| *X20_Year* | 0.03178379 | 0.03548636 |
| *X30_Year* | 0.03522005 | 0.04104123 |
| ***mean*** | 0.02507089 | 0.0295992 |
| ***sd*** | 0.01008537 | 0.00945153 |

6.4.1.1.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.26961595 | 0.26155651 |
| *ctr2* | 0.2654184 | 0.21395308 |
| *X1_Month* | 0.15894542 | 0.11140857 |
| *X3_Month* | 0.10236519 | 0.08959809 |
| *X6_Month* | 0.05266194 | 0.09182476 |
| *X1_Year* | 0.02206403 | 0.10321851 |
| *X2_Year* | 0.02516151 | 0.12092153 |
| *X3_Year* | 0.02516151 | 0.12092153 |
| *X5_Year* | 0.04676166 | 0.13363116 |
| *X7_Year* | 0.11501381 | 0.13951853 |
| *X10_Year* | 0.11236161 | 0.12194128 |
| *X20_Year* | 0.12694449 | 0.11616432 |
| *X30_Year* | 0.11638889 | 0.12109487 |
| ***mean*** | 0.11068188 | 0.13428867 |
| ***sd*** | 0.08259 | 0.04905182 |

6.4.1.1.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *ctr1* | 0.01759973 | 0.03401742 |
| *ctr2* | 0.01691962 | 0.01555027 |
| *X1_Month* | 0.04748296 | 0.05026788 |
| *X3_Month* | 0.01319452 | 0.02505149 |
| *X6_Month* | 0.01256812 | 0.02449994 |
| *X1_Year* | 0.01367335 | 0.02295395 |
| *X2_Year* | 0.0255131 | 0.02497061 |
| *X3_Year* | 0.0255131 | 0.02497061 |
| *X5_Year* | 0.02630808 | 0.02647335 |
| *X7_Year* | 0.02956915 | 0.03128332 |

| | | |
|---|---|---|
| *X10_Year* | 0.03157518 | 0.03505799 |
| *X20_Year* | 0.03178379 | 0.03548637 |
| *X30_Year* | 0.03522173 | 0.04105278 |
| ***mean*** | 0.02514788 | 0.03012585 |
| ***sd*** | 0.01026451 | 0.00904812 |

6.4.1.1.4 Random Forests

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.39665329 | 0.45682945 |
| *ctr2* | 0.38694565 | 0.37521278 |
| *X1_Month* | 0.34174558 | 0.35037694 |
| *X3_Month* | 0.38400029 | 0.39058438 |
| *X6_Month* | 0.37935134 | 0.37861839 |
| *X1_Year* | 0.38748706 | 0.37948219 |
| *X2_Year* | 0.37985552 | 0.36954236 |
| *X3_Year* | 0.37396167 | 0.37801647 |
| *X5_Year* | 0.37193658 | 0.37044738 |
| *X7_Year* | 0.36070718 | 0.37123518 |
| *X10_Year* | 0.36429805 | 0.36443139 |
| *X20_Year* | 0.35484247 | 0.34960411 |
| *X30_Year* | 0.338936 | 0.3370625 |
| ***mean*** | 0.37082467 | 0.37472643 |
| ***sd*** | 0.39665329 | 0.45682945 |

6.4.1.1.5 SVM (radial basis)

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.19457095 | 0.10720703 |
| *ctr2* | 0.17214914 | 0.06171241 |
| *X1_Month* | 0.15394519 | 0.11716006 |
| *X3_Month* | 0.1633082 | 0.06341605 |
| *X6_Month* | 0.16409261 | 0.06321869 |
| *X1_Year* | 0.15999542 | 0.06269215 |
| *X2_Year* | 0.10543781 | 0.06271709 |
| *X3_Year* | 0.10543781 | 0.06271709 |
| *X5_Year* | 0.10738359 | 0.06334159 |
| *X7_Year* | 0.12318908 | 0.0727383 |
| *X10_Year* | 0.13940722 | 0.07720202 |
| *X20_Year* | 0.13976292 | 0.08299045 |
| *X30_Year* | 0.12163501 | 0.09199574 |
| ***mean*** | 0.14233192 | 0.07608528 |
| ***sd*** | 0.02847818 | 0.01871008 |

### *6.4.1.2. Minutes Only*

6.4.1.2.1 Linear Model

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.22456176 | 0.10216318 |
| *ctr2* | 0.10787205 | 0.12085409 |
| *X1_Month* | 0.3395627 | 0.10046191 |
| *X3_Month* | 0.06903972 | 0.02930206 |
| *X6_Month* | 0.07032451 | 0.02332737 |
| *X1_Year* | 0.07184233 | 0.02142176 |
| *X2_Year* | 0.08722291 | 0.01662829 |
| *X3_Year* | 0.08722291 | 0.01662829 |
| *X5_Year* | 0.08495536 | 0.01798558 |

| | | |
|---|---|---|
| *X7_Year* | 0.12019593 | 0.02248008 |
| *X10_Year* | 0.15630871 | 0.03045191 |
| *X20_Year* | 0.17816159 | 0.03113472 |
| *X30_Year* | 0.26009098 | 0.0624204 |
| ***mean*** | 0.14287396 | 0.0457892 |
| ***sd*** | 0.08542705 | 0.03754515 |

##### 6.4.1.2.2 CART

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.38398057 | 0.3278003 |
| *ctr2* | 0.38933108 | 0.38714272 |
| *X1_Month* | 0.48184962 | 0.27890364 |
| *X3_Month* | 0.40807906 | 0.28133792 |
| *X6_Month* | 0.39571688 | 0.29283898 |
| *X1_Year* | 0.38346058 | 0.3176192 |
| *X2_Year* | 0.42393094 | 0.2750967 |
| *X3_Year* | 0.42393094 | 0.2750967 |
| *X5_Year* | 0.35545613 | 0.27067268 |
| *X7_Year* | 0.37272972 | 0.32550254 |
| *X10_Year* | 0.37643617 | 0.33429381 |
| *X20_Year* | 0.41908714 | 0.38088808 |
| *X30_Year* | 0.46773303 | 0.37273772 |
| ***mean*** | 0.4062863 | 0.31691777 |
| ***sd*** | 0.03684855 | 0.04238864 |

##### 6.4.1.2.3 GLM

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.22456176 | 0.10300067 |
| *ctr2* | 0.10787205 | 0.1209569 |
| *X1_Month* | 0.3395627 | 0.10211477 |
| *X3_Month* | 0.06903972 | 0.02932414 |
| *X6_Month* | 0.07032451 | 0.0233399 |
| *X1_Year* | 0.07184233 | 0.02143138 |
| *X2_Year* | 0.0874167 | 0.01662853 |
| *X3_Year* | 0.0874167 | 0.01662853 |
| *X5_Year* | 0.08576371 | 0.01798684 |
| *X7_Year* | 0.12096831 | 0.02248207 |
| *X10_Year* | 0.15714641 | 0.03045251 |
| *X20_Year* | 0.17898336 | 0.03115151 |
| *X30_Year* | 0.26105781 | 0.06250658 |
| ***mean*** | 0.14322739 | 0.04600033 |
| ***sd*** | 0.08549403 | 0.03786987 |

##### 6.4.1.2.4 Random Forests

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.46092086 | 0.66641841 |
| *ctr2* | 0.5714039 | 0.36420694 |
| *X1_Month* | 0.46871807 | 0.39187568 |
| *X3_Month* | 0.4591757 | 0.54978094 |
| *X6_Month* | 0.45349029 | 0.59359527 |
| *X1_Year* | 0.47745879 | 0.57628256 |
| *X2_Year* | 0.49546198 | 0.57910286 |
| *X3_Year* | 0.49468345 | 0.5597329 |
| *X5_Year* | 0.45784543 | 0.55279895 |
| *X7_Year* | 0.46456383 | 0.51687202 |

| | | |
|---|---|---|
| *X10_Year* | 0.42317669 | 0.50377505 |
| *X20_Year* | 0.45098266 | 0.47471424 |
| *X30_Year* | 0.34591058 | 0.43326894 |
| ***mean*** | 0.46336863 | 0.52018652 |
| ***sd*** | 0.04968874 | 0.08543508 |

6.4.1.2.5 SVM (radial basis)

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.49388001 | 0.33871532 |
| *ctr2* | 0.34949178 | 0.35067595 |
| *X1_Month* | 0.66444912 | 0.25614967 |
| *X3_Month* | 0.33188587 | 0.24896098 |
| *X6_Month* | 0.32406822 | 0.23821186 |
| *X1_Year* | 0.32699301 | 0.23477069 |
| *X2_Year* | 0.3672909 | 0.22952498 |
| *X3_Year* | 0.3672909 | 0.22952498 |
| *X5_Year* | 0.33300247 | 0.23103123 |
| *X7_Year* | 0.3452428 | 0.23460631 |
| *X10_Year* | 0.36290596 | 0.25017848 |
| *X20_Year* | 0.36420206 | 0.23253668 |
| *X30_Year* | 0.38919567 | 0.25186064 |
| ***mean*** | 0.38614606 | 0.25590367 |
| ***sd*** | 0.09440116 | 0.04052411 |

#### *6.4.1.3. Statistical relationship of Combined public policy data from sentiment analysis study*

6.4.1.3.1 Linear Model

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.01476103 | 0.02964442 |
| *ctr2* | 0.02513971 | 0.01168737 |
| *X1_Month* | 0.03930665 | 0.03229607 |
| *X3_Month* | 0.02404474 | 0.02100246 |
| *X6_Month* | 0.02340513 | 0.02068466 |
| *X1_Year* | 0.02110564 | 0.01922209 |
| *X2_Year* | 0.02138179 | 0.02023213 |
| *X3_Year* | 0.02138179 | 0.02023213 |
| *X5_Year* | 0.02157901 | 0.02111953 |
| *X7_Year* | 0.02429459 | 0.02439126 |
| *X10_Year* | 0.02548306 | 0.02650891 |
| *X20_Year* | 0.02523566 | 0.02632317 |
| *X30_Year* | 0.02595156 | 0.02828239 |
| ***mean*** | 0.02408234 | 0.02320204 |
| ***sd*** | 0.00545274 | 0.00542899 |

6.4.1.3.2 CART

| **DV** | **Shift** | **NoShift** |
|---|---|---|
| *ctr1* | 0.29939957 | 0.19991818 |
| *ctr2* | 0.12588381 | 0.06206556 |
| *X1_Month* | 0.09490258 | 0.05463357 |
| *X3_Month* | 0.01540962 | 0.03250724 |
| *X6_Month* | 0.01579948 | 0.01579948 |
| *X1_Year* | 0.01663302 | 0.06565013 |
| *X2_Year* | 6.91E-02 | 1.28E-01 |

| | | |
|---|---|---|
| X3_Year | 6.91E-02 | 1.28E-01 |
| X5_Year | 0.09141169 | 0.08568261 |
| X7_Year | 0.02056755 | 0.11528642 |
| X10_Year | 0.02120558 | 0.08628222 |
| X20_Year | 0.0594021 | 0.09982323 |
| X30_Year | 0.10144086 | 0.10144086 |
| **mean** | 0.07694316 | 0.09039193 |
| **sd** | 0.07686125 | 0.04769945 |

6.4.1.3.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| ctr1 | 0.01476103 | 0.03036883 |
| ctr2 | 0.02513971 | 0.01538175 |
| X1_Month | 0.03930665 | 0.03326331 |
| X3_Month | 0.02404474 | 0.0210575 |
| X6_Month | 0.02340513 | 0.02073156 |
| X1_Year | 0.02110564 | 0.01925332 |
| X2_Year | 0.02138179 | 0.02027121 |
| X3_Year | 0.02138179 | 0.02027121 |
| X5_Year | 0.02157901 | 0.02114393 |
| X7_Year | 0.02429459 | 0.0244015 |
| X10_Year | 0.02548306 | 0.0265134 |
| X20_Year | 0.02532458 | 0.0263308 |
| X30_Year | 0.02595337 | 0.02828239 |
| **mean** | 0.02408931 | 0.02363621 |
| **sd** | 0.00545442 | 0.00505809 |

6.4.1.3.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| ctr1 | 0.40626451 | 0.41734127 |
| ctr2 | 0.42558075 | 0.37847238 |
| X1_Month | 0.37468678 | 0.37660976 |
| X3_Month | 0.4098617 | 0.41327322 |
| X6_Month | 0.40295789 | 0.40549749 |
| X1_Year | 0.4011862 | 0.39859858 |
| X2_Year | 0.39045758 | 0.38244874 |
| X3_Year | 0.38272627 | 0.39640453 |
| X5_Year | 0.37994209 | 0.38913479 |
| X7_Year | 0.3738146 | 0.37206007 |
| X10_Year | 0.36443698 | 0.36097722 |
| X20_Year | 0.37572792 | 0.36442474 |
| X30_Year | 0.37103534 | 0.3695796 |
| **mean** | 0.38912912 | 0.3865248 |
| **sd** | 0.01842948 | 0.01846735 |

6.4.1.3.5 SVM (radial basis)

| DV | Shift | NoShift |
|---|---|---|
| ctr1 | 0.1697257 | 0.09131814 |
| ctr2 | 0.19857053 | 0.06332888 |
| X1_Month | 0.14775474 | 0.1042096 |
| X3_Month | 0.11746474 | 0.05566748 |
| X6_Month | 0.11828647 | 0.05272335 |
| X1_Year | 0.12482073 | 0.05377732 |
| X2_Year | 0.09762549 | 0.05467129 |
| X3_Year | 0.09762549 | 0.05467129 |

| | | |
|---|---|---|
| *X5_Year* | 0.15299225 | 0.05483599 |
| *X7_Year* | 0.16180642 | 0.05993311 |
| *X10_Year* | 0.15211568 | 0.06897901 |
| *X20_Year* | 0.11112643 | 0.06846523 |
| *X30_Year* | 0.10555166 | 0.08185795 |
| ***mean*** | 0.13503587 | 0.06649528 |

### 6.4.2 Results from Sentiment analysis of corporate email data

#### *6.4.2.1 Variable selection from corporate email*

##### 6.4.2.1.1 Linear Model

###### *6.4.2.1.1.1 Time Shift*

| **Dependent Variable** | **S.P500** | **ENRQ** |
|---|---|---|
| *Positive_Sentences* | 1 | 1 |
| *Negative_Sentences* | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 |

###### *6.4.2.1.1.2 No Time Shift*

| **Dependent Variable** | **S.P500** | **ENRQ** |
|---|---|---|
| *Positive_Sentences* | 1 | 1 |
| *Negative_Sentences* | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 |

#### 6.4.2.1.2 CART

##### *6.4.2.1.2.1 Time Shift*

| Dependent Variable | S.P500 | ENRQ |
|---|---|---|
| *Positive_Sentences* | 0 | 0 |
| *Negative_Sentences* | 0 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 |
| *X._Negative_Sentences* | 1 | 0 |
| *X._Neutral_Sentences* | 1 | 0 |

##### *6.4.2.1.2.2 No Time Shift*

| Dependent Variable | S.P500 | ENRQ |
|---|---|---|
| *Positive_Sentences* | 1 | 0 |
| *Negative_Sentences* | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 0 | 0 |
| *X._Negative_Sentences* | 0 | 0 |
| *X._Neutral_Sentences* | 0 | 0 |

#### 6.4.2.1.3 GLM

##### *6.4.2.1.3.1 Time Shift*

| Dependent Variable | S.P500 | ENRQ |
|---|---|---|
| *Positive_Sentences* | 1 | 1 |
| *Negative_Sentences* | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 |

*6.4.2.1.3.2 No Time Shift*

| **Dependent Variable** | **S.P500** | **ENRQ** |
|---|---|---|
| *Positive_Sentences* | 1 | 1 |
| *Negative_Sentences* | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 |

6.4.2.1.4 Random Forests

*6.4.2.1.4.1 Time Shift*

| **Dependent Variable** | **S.P500** | **ENRQ** |
|---|---|---|
| *Positive_Sentences* | 0 | 0 |
| *Negative_Sentences* | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 |
| *X._Positive_Sentences* | 0 | 0 |
| *X._Negative_Sentences* | 1 | 0 |
| *X._Neutral_Sentences* | 0 | 1 |

*6.4.2.1.4.2 No Time Shift*

| **Dependent Variable** | **S.P500** | **ENRQ** |
|---|---|---|
| *Positive_Sentences* | 0 | 0 |
| *Negative_Sentences* | 0 | 0 |
| *Neutral_Sentences* | 0 | 0 |
| *X._Positive_Sentences* | 0 | 0 |
| *X._Negative_Sentences* | 1 | 0 |
| *X._Neutral_Sentences* | 0 | 1 |

6.4.2.1.5 SVM (radial basis)

*6.4.2.1.5.1 Time Shift*

| Dependent Variable | S.P500 | ENRQ |
|---|---|---|
| *Positive_Sentences* | 1 | 0 |
| *Negative_Sentences* | 0 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 1 |
| *X._Negative_Sentences* | 1 | 1 |
| *X._Neutral_Sentences* | 1 | 1 |

*6.4.2.1.5.2 No Time Shift*

| Dependent Variable | S.P500 | ENRQ |
|---|---|---|
| *Positive_Sentences* | 1 | 0 |
| *Negative_Sentences* | 1 | 1 |
| *Neutral_Sentences* | 1 | 1 |
| *X._Positive_Sentences* | 1 | 0 |
| *X._Negative_Sentences* | 1 | 0 |
| *X._Neutral_Sentences* | 1 | 1 |

### *6.4.2.2. Statistical relationships of sentiment analysis on Corporate email Data*

6.4.2.2.1 Shift Means

| | ***file3*** |
|---|---|
| ***lm*** | 0.0837588 |
| ***cart*** | 0.27387306 |
| ***glm*** | 0.0837588 |
| ***rf*** | 0.30235446 |
| ***svm*** | 0.23969063 |

6.4.2.2.2 No Shift Means

| | ***file3*** |
|---|---|
| ***lm*** | 0.08828001 |
| ***cart*** | 0.26113073 |
| ***glm*** | 0.08828001 |
| ***rf*** | 0.3032603 |
| ***svm*** | 0.20751688 |

6.4.2.2.3 No. of Observations being considered

| | ***file3*** |
|---|---|
| ***N.obs*** | 972 |
| | |
| ***Starting Day*** | 30-Oct-98 |
| ***Ending Day*** | 12-Jul-02 |

The results are from the tables above are summarized in visual form in the following four Figures below. They make clear which learning algorithms under which circumstances were deomonstratably more effective in the analysis with the datasets under investigation. The implications of the findings are discussed elsewhere in this Dissertation.

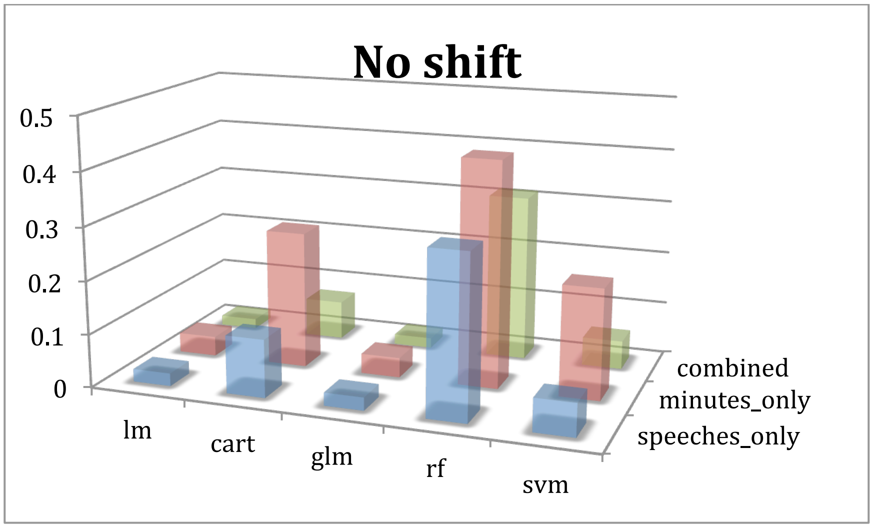


**Figure 37: Statistical Results from contemporanious sentiment analysis (LingPipe) of corporate email data**
**(Numerical scale represents pseudo R-squared)**

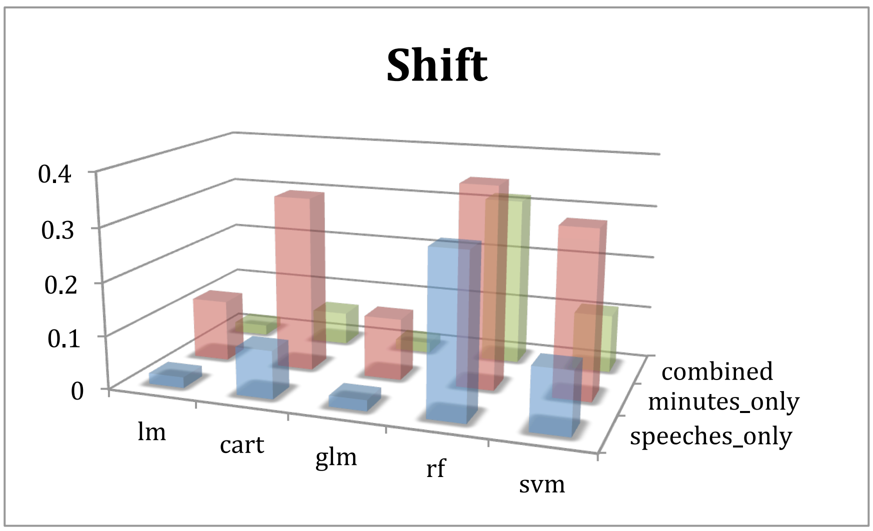


**Figure 38: Statistical Results from time-shifted (best shift) sentiment analsysis (LingPipe) of Public Policy data**
**(Numerical scale represents pseudo R-squared)**

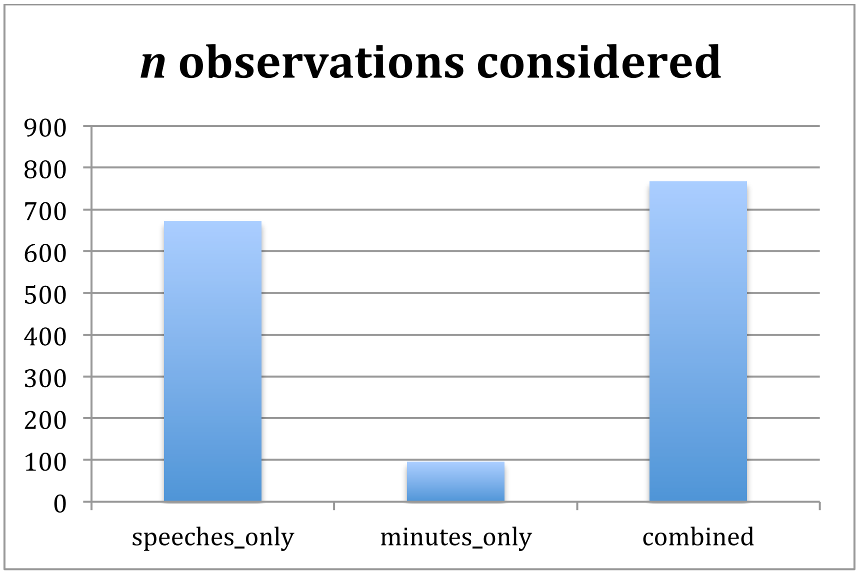


**Figure 39: Number of Observations for sentiment analysis of Public Policy Data**

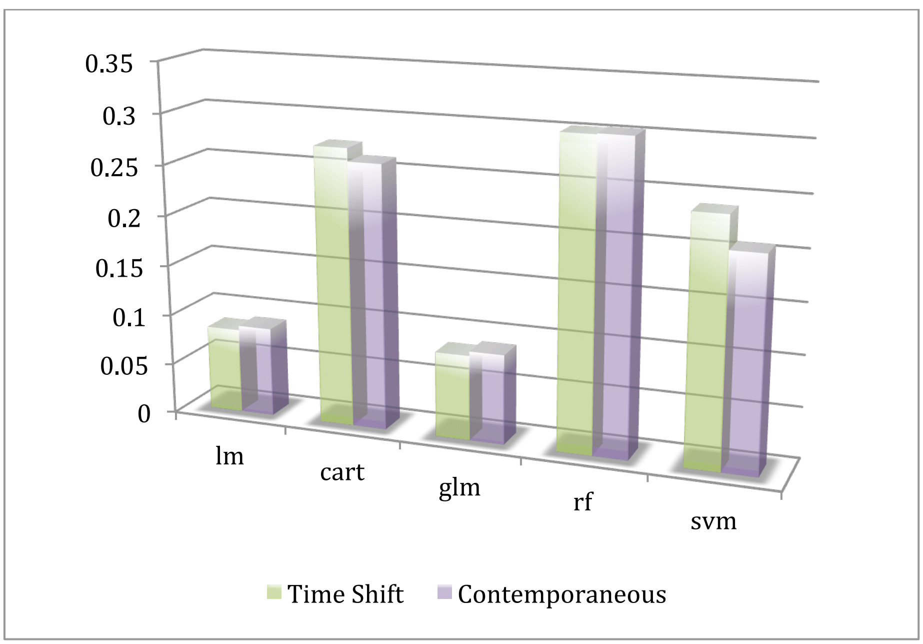


**Figure 40: Statistical results from both contemporaneous and time-shifted analysis of Corporate email data using baseline approach (Numerical scale represents pseudo R-squared)**

### 6.4.2 Summary of Statistical Relationship Summary of Corporate email data using Sentiment Analysis.

6.4.2.1.1 Linear Model

| DV | Shift | NoShift |
|---|---|---|
| *S.P500* | 0.03101154 | 0.0356285 |
| *ENRQ* | 0.13650607 | 0.14093153 |
| ***mean*** | 0.0837588 | 0.08828001 |
| ***sd*** | 0.0745959 | 0.07446049 |

6.4.2.1.2 CART

| DV | Shift | NoShift |
|---|---|---|
| *S.P500* | 0.18573685 | 0.16857795 |
| *ENRQ* | 0.36200928 | 0.35368352 |
| ***mean*** | 0.27387306 | 0.26113073 |
| ***sd*** | 0.12464343 | 0.13088941 |

6.4.2.1.3 GLM

| DV | Shift | NoShift |
|---|---|---|
| *S.P500* | 0.03101154 | 0.0356285 |
| *ENRQ* | 0.13650607 | 0.14093153 |
| ***mean*** | 0.0837588 | 0.08828001 |
| ***sd*** | 0.0745959 | 0.07446049 |

6.4.2.1.4 Random Forests

| DV | Shift | NoShift |
|---|---|---|
| *S.P500* | 0.31828135 | 0.32556496 |
| *ENRQ* | 0.28642757 | 0.28095565 |
| ***mean*** | 0.30235446 | 0.3032603 |
| ***sd*** | 0.02252403 | 0.03154355 |

6.4.2.1.5 SVM (radial basis)

| DV | Shift | NoShift |
|---|---|---|
| *S.P500* | 0.17449055 | 0.1202917 |
| *ENRQ* | 0.30489072 | 0.29474206 |
| ***mean*** | 0.23969063 | 0.20751688 |
| ***sd*** | 0.09220684 | 0.12335504 |

## 6.5 Results from sentiment analysis using Sentiwordnet

The inquiry into the effectiveness of Sentiwordnet allows for visualizations of the sentiment in addition to the numerical representations. The visualization gives clear relief to the data points outside of the time span under consideration: there are email dates in the future and distant past.

### 6.5.1 Summary of statistical relationships from Public Policy dataset using Sentiwordnet

The following two Figures below are visualizations of the sentiment results from the positive & negative assessments Sentiwordnet placed on the public policy data set. In the first Figure, the speeches by themselves, then the meeting minutes in the next Figure. Lastly, the two combined in the last Figure.

The regular pattern to the communications from the Fed may suggest a degree of planning. There are fewer communications in January and July; much fewer in August and December. The meeting minutes, of course, are more regularly distributed by their very nature.

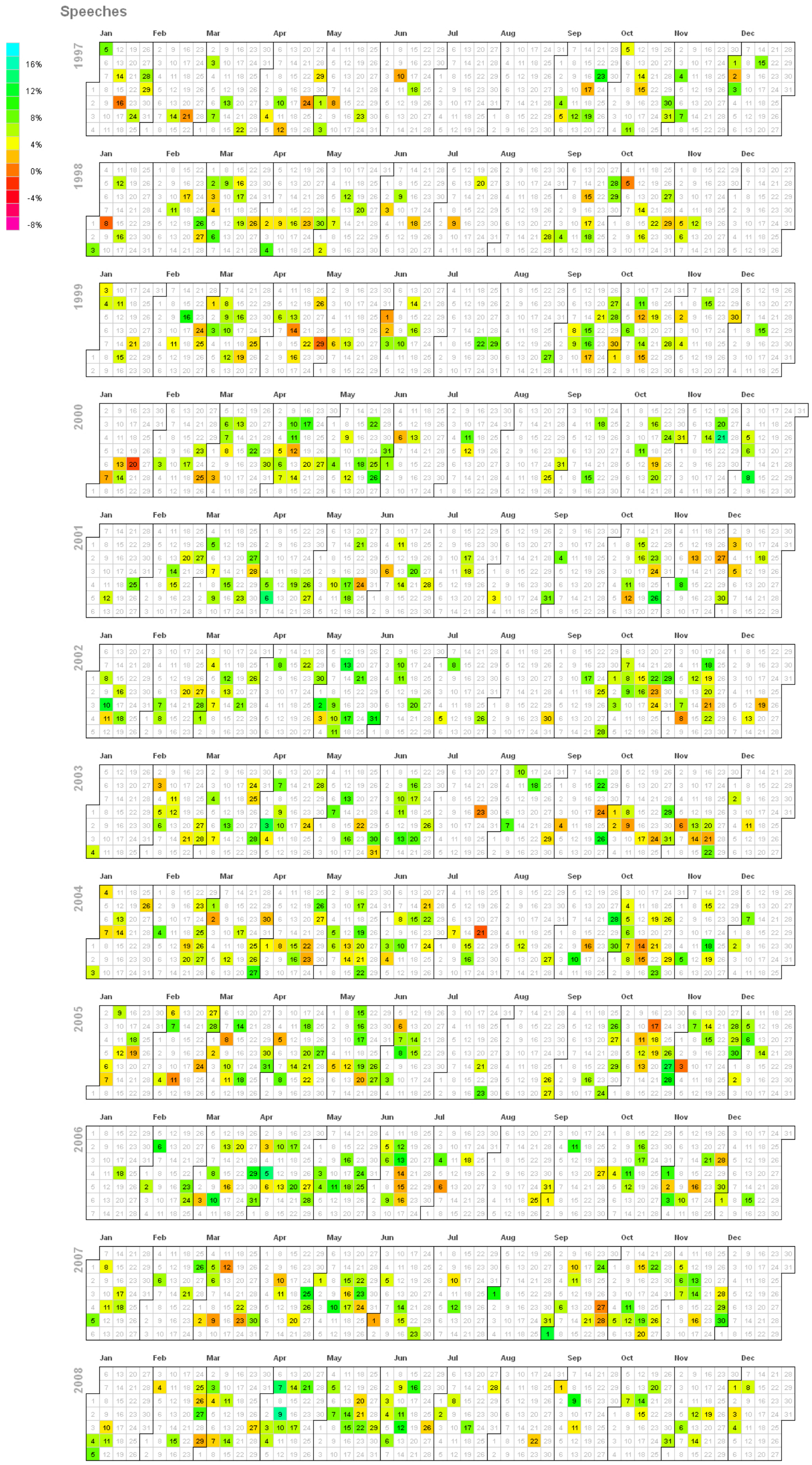


**Figure 41: Visualization of Sentiwordnet analysis of Fed Speeches**

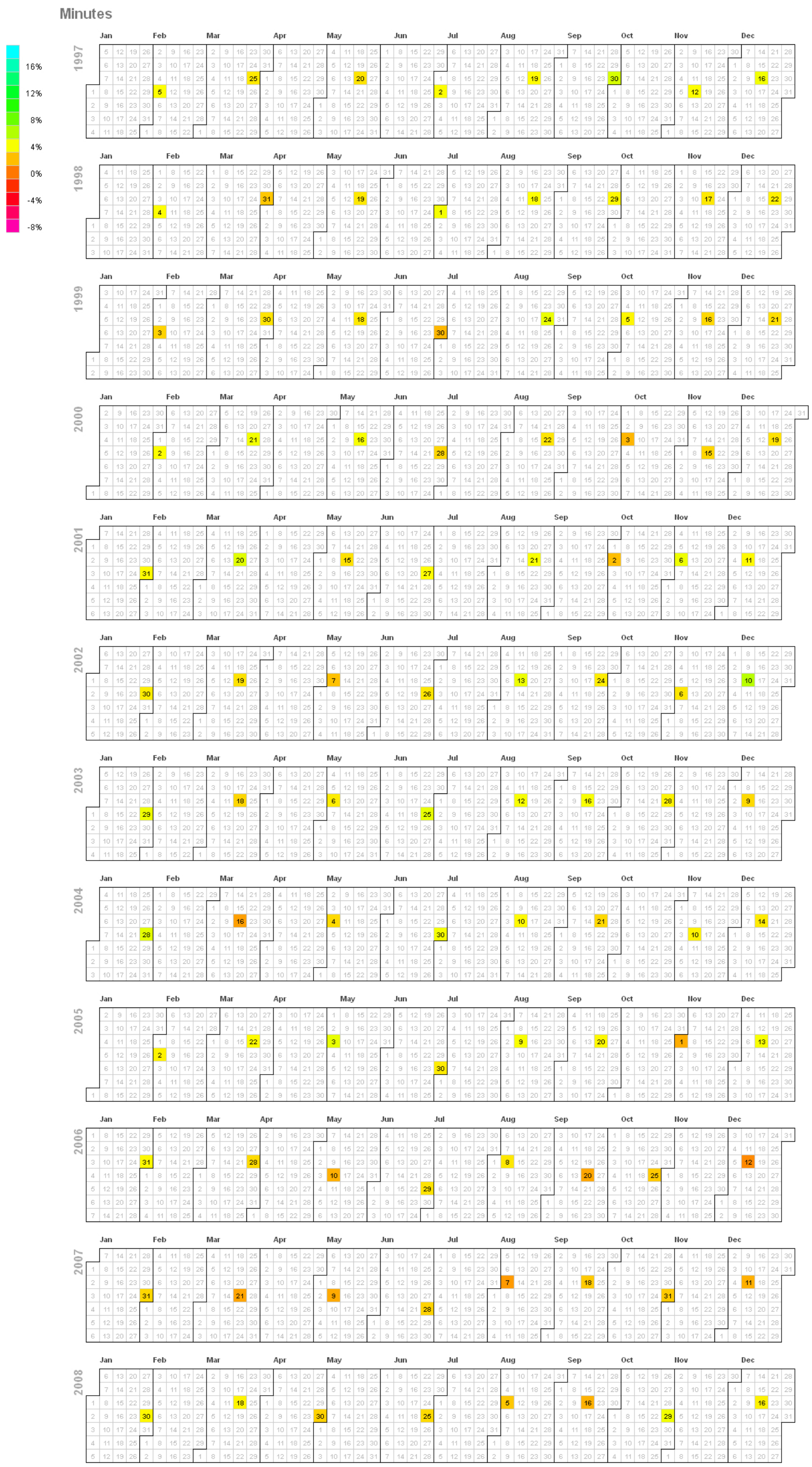


**Figure 42: Sentiwordnet results from Fed Meeting minutes**

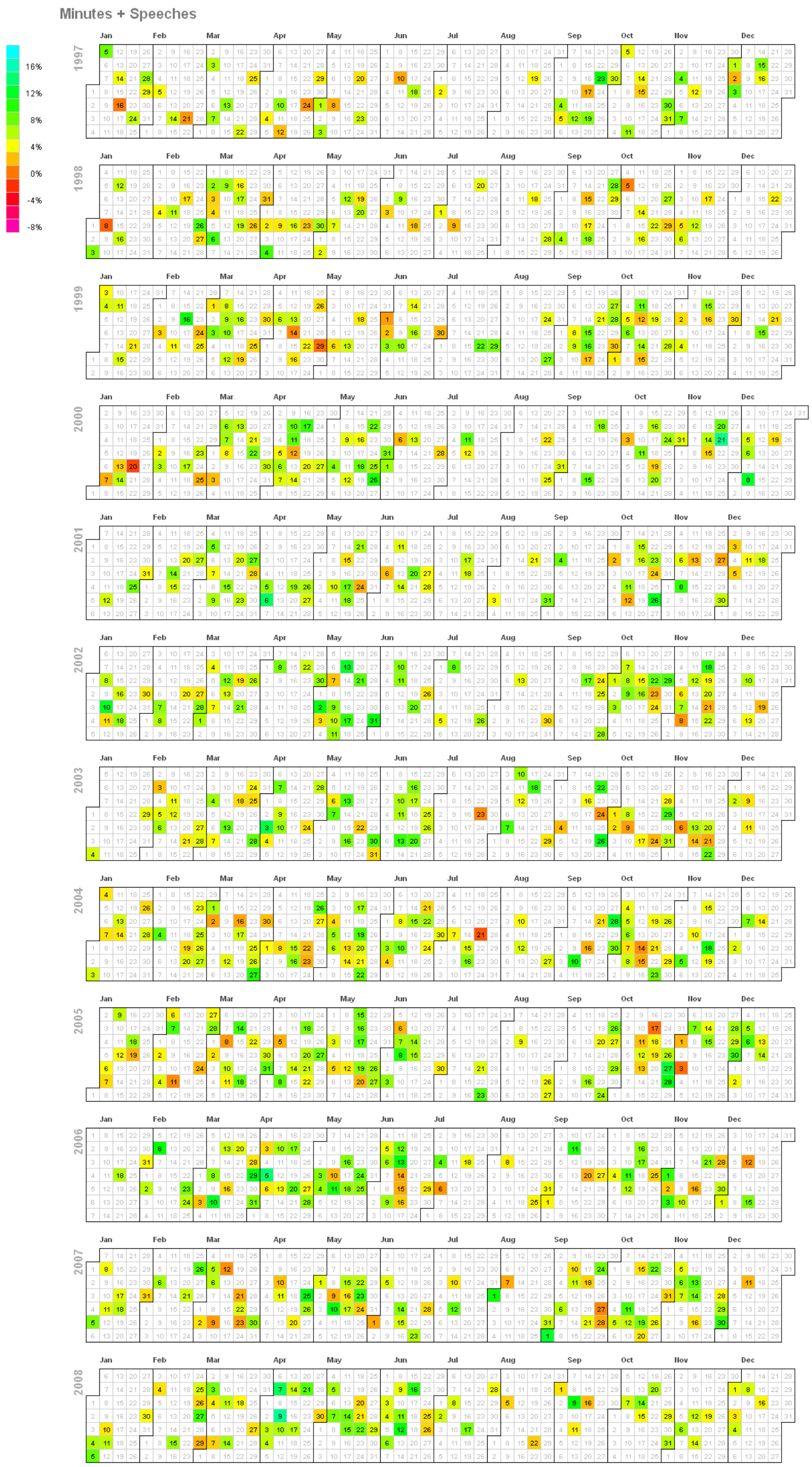

Figure 43: Sentiwordnet results from aggregate Public Policy dataset

### 6.5.2 Results from Corporate email dataset using sentiwordnet

Cropping out the visualization of dates that occur in the distant future still leaves a view into the emails from the distant past. The remaining emails that were under consideration in the date range specified elsewhere in this study show materially fewer negative measurements. This is visualized in the Figure below.

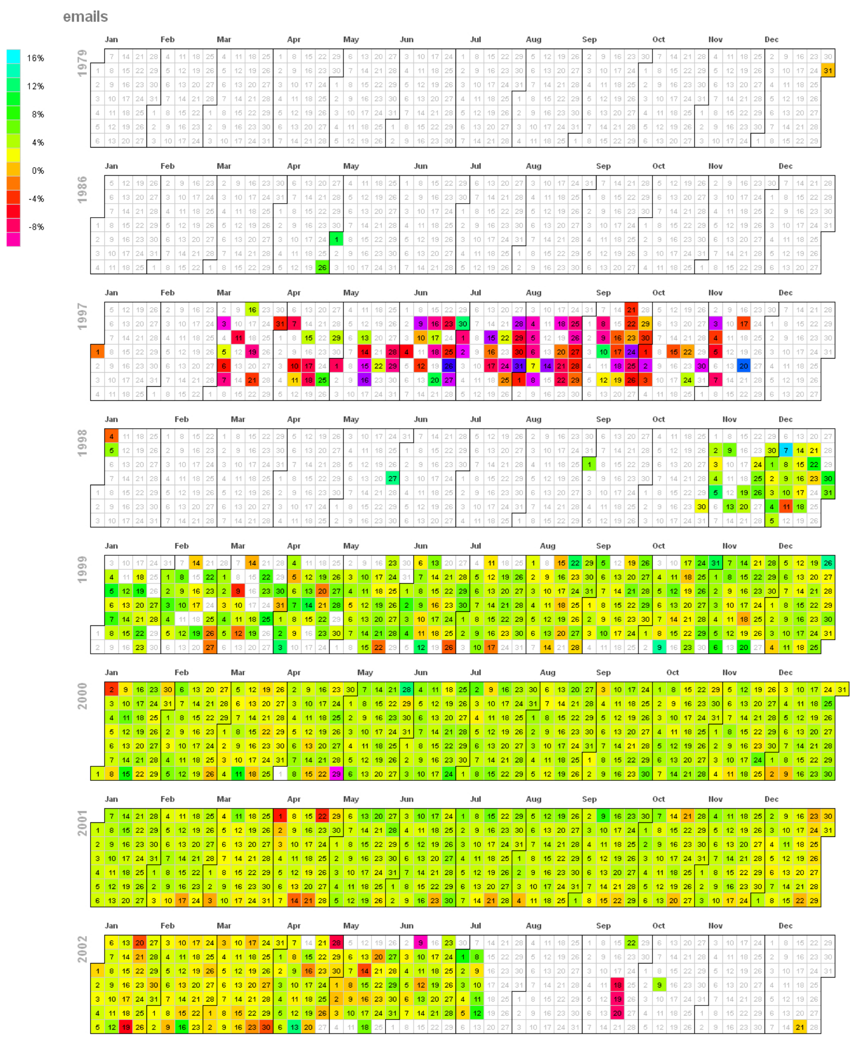


**Figure 44: Sentiwordnet results on email corpus**

## 6.6 Conclusion of baseline comparision

In this chapter, I present a sentiment analysis of data from both Public Policy Documents and a large email corpus. I used the same raw data, in the same time frame as used in Chapters Four and Five. I presented visualizations of this data and some of the email data that was excluded from consideration. The purpose of this inquiry is to use an established methodology for measuring sentiment against which any new approaches might be compared for efficacy. The results from this study suggest a correlation between the results of a sentiment analysis and the quantitative data under consideration. However, the analysis reveals a connection that might likely be considered insufficient to justify further study using this approach. Both of the classic sentiment algorithms used showed a weak correlation under all cases and extremely weak correlations under all but one case among both data sets.

# 7 Predictive Value

Although this framework is offered as a possilble method for inquiring into the relationship between text and numbers, this research document may find within its scope some limited explortation into the predictive value for the approach. Toward that end, this chapter takes the larger data set (in time, if not in absolute numbers) to explore the degree to which results may be predicted from an intelligent applicaton of the proposed framework.

## 7.1 Predicting Dependent Variables (i.e., the numbers): Contraints on predictions

The datasets under investigation do not present numbers unrelated to those immediately predeeding them. Said another way, the best indicator of tomorrow's number is today's number. Therefore, any additional data used in the prediction of tomorrow's number can easily be confused with the predictive capacity already present. If today's number is '1', then tomorrow's number is much more likely to be '1' than '2'.

A straightforward constraint to measure the additional predictive value of additional information is to use the absolute difference in the number from that of the previous day. Therefore the first predcion is the actual change each day, positive or negative.

In these examples, the data moves in one one-hundreth of a point (i.e., 0.01) increments. Some movement around a baseline may indicate nothing at all. Therefore, the second degree of predction asks if the number moves only above the threshold of 0.02. Movement $n$ in either direction is ignored where $n \leq 0.02$.

To relax preduction ability further, two additional experiments are run: a) Can *any* directional movement at all be detected; and even easier b) detecting movement $n$ in either direction only where $n > 0.02$.

In previous chapters, the total set of Independent Variables was reduced through a rigorous selection criteria. However, in this prediction model, we will always want to take out some observations for cross validation and test the conclusions on the data that has been excluded. For this reason, the data is split randomly one hundred times into training and test data and only at each split is the clustering algorithm applied (and applied only to the training set) to identify the exemplar variables for that reduced set.

I present five tests of predictive power. These five tests , summarized in the table below, are important for interpreting the pattern recognition results in this chapter and in the consideration of expansions of this research in future work.

| Case | Description |
|---|---|
| 0 | Predict numerical difference from previous day |
| 1 | Case 0, but only IFF $n > 0.02$ |
| 2 | Predict movement (+/-), but not magnitude of movement. |
| 3 | Case 2, but only IFF $n > 0.02$ |
| 4 | Predictive power of individual Independent Variables |

**Table 15: Prediction scenarios**

## 7.2 Methods for finding predictive power

### 7.2.1 For cases including all Independent Variables

For each Dependent Variable, I iterate 100 times to split on a test and training set. I retain 20% for a test set in each iteration. For each training set, I again cluster the Independent Variables (IVs) that are Exemplars for that chosen set. In the results, I present those IVs with the number of times that they have been selected as Exemplars among those 100 random subsets. I then present the average error for each prediction (along with the standard deviation of each error) over those 100 iterations.

Cases 0 & 1 in the table above are also presented with the average error for each prediction. However, since these cases are binary (i.e., the numbers either did or did not change), a direct comparison of percentage error can be misleading. I perform two exercises to make more direct comparisions. First, I present all errors as a percentage of the total. Also, I present cases 2 & 3 with outcome measurement that are discrete (+/-). This is done by rounding the numbers. An example is presented in the table (Table 16) below. The method for determining predictive qualities of Grouped Independent Variables is presented as a flow chart following that Table. The results from that analysis are presented in the subsequent sections.

| Case | Short Description | Actual number from data | Guessed number | Error |
|---|---|---|---|---|
| 0 | Numerical difference | 1.1 | -1.0 | 2.1 |
| 1 | Case 0 IFF $n > 0.02$ | 1.1 | 0.9 | 0 |
| 2 | Predict movement (+/-), but not magnitude of movement. | 1.1 | 0.9 | -1 |
| 2 disc. | Case 2, but rounded to an integer. | 1.1 | 0.9 | 0 |
| 3 | Case 2 IFF $n > 0.02$ | 1.1 | 0.8 | 1 |
| 3 disc. | Case 3, but rounded to an integer. | 1.1 | 0.7 | 0 |
| 4 | Case 3, discrete, but for individual IVs | 1.1 | 1.5 | 0 |

**Table 16: Examples of cases 0-4 handling predictive data**

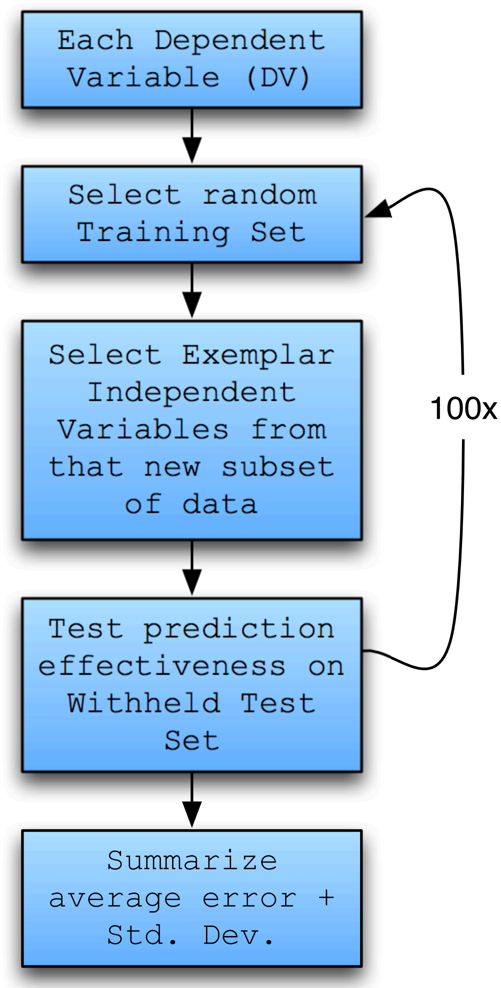


**Figure 45: Method for determining predictive qualities of Grouped Independent Variables**

### 7.2.2 Restricting cases to predictive power of Individual Independnet Variables

In all previous experiments, I have clustered the Independent Variables. In this last experiment, I look to see the degree to which each Independent Variable, on its own, can supply predictive capacity, regardless of the presence of any other Independent Variables.

## 7.3 Presentation of results from predictions

The following Figures serve as visualizations of the results from each case. The measurements are taken for each Dependent Variable with complete data (i.e., excluding the change in actual Fed Funds targets) with the Y-Axis reporting on the measurement 0-100% of the error

### 7.3.1 Predicting numerical differences from previous day

The first case presented in the Figure below is run 100 times with exemplar IVs predicting tomorrow's data from today's. Three of the DVs (actual change in FF and target rate) because of an absence of data for this experiment. This chart may be compared to those that follow.

e

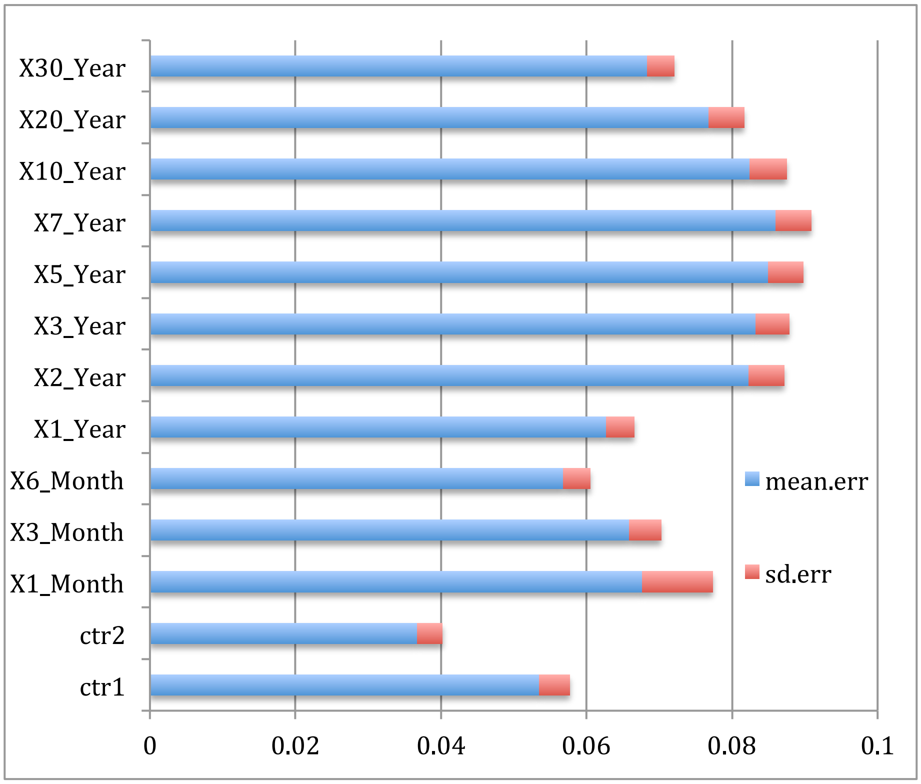


**Figure 46: Mean & Standard Deviation of effectiveness in Predicting absolute numbers (Case 0). (Numerical scale represents absolute Average Error)**

### 7.3.2 Previous day, IFF $n > 0.02$

To limit the inpact of minor changes (i.e., noise) in the daily movements, another analysis is performed that only detects movement iff the movement is greater than two basis points (0.02). The importance and implications are discussed later in this chapter.

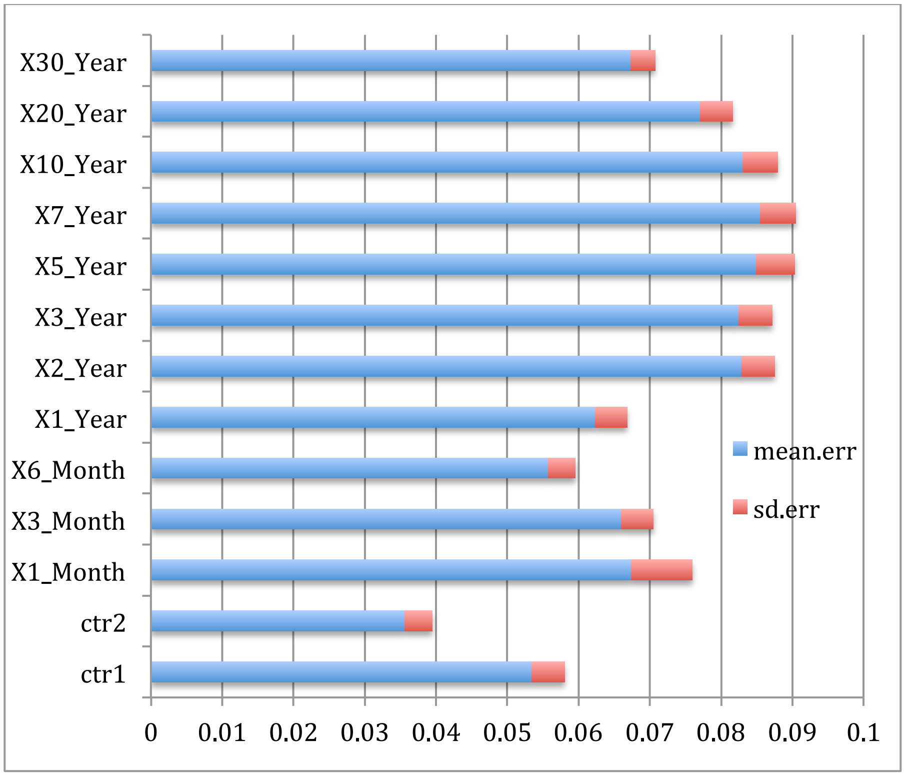


**Figure 47: Mean & Standard Deviation of effectiveness in Predicting absolute numbers outside a range (Case 1) .**
**(Numerical scale represents absolute Average Error)**

### 7.3.3 Predicting presence of movement

Predicting the numbers in cases 0 and 1 above are replaced in this case with just the prediction of any movement in any direction. The visualization of this data in the Figure below suggests a material increase in effectiveness with this methodology.

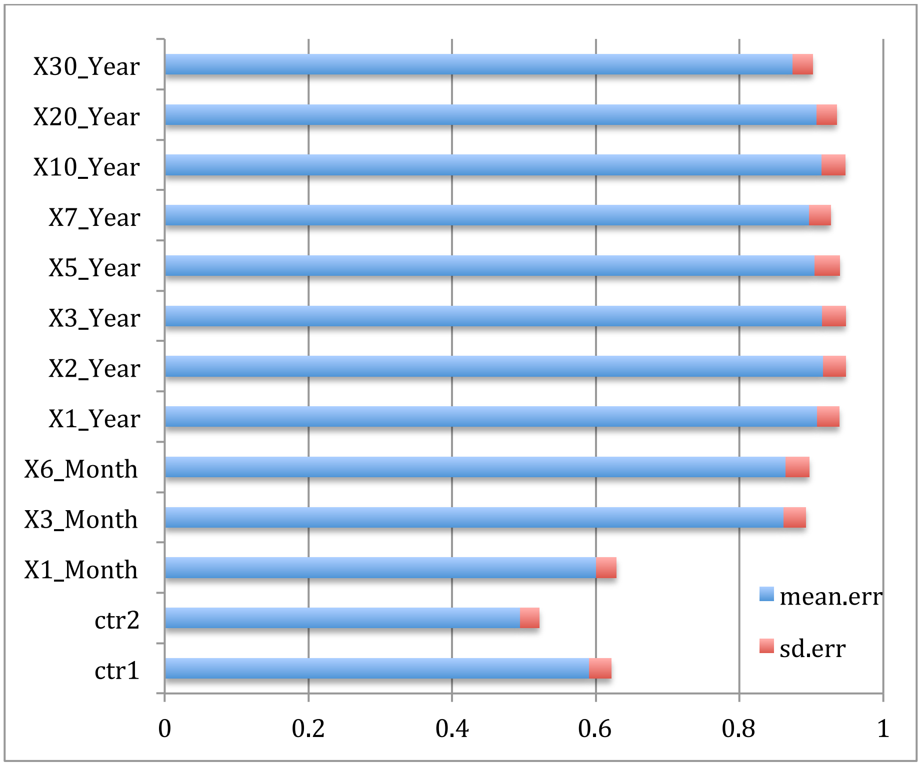


**Figure 48: Mean & Standard Deviation of effectiveness in Predicting any movement in any direction (Case 2) .**
**(Numerical scale represents absolute Average Error)**

### 7.3.4 Predicting movement IFF $n > 0.02$

Like the exercise in Case 1 over Case 0, this case seeks to reduce the effect of noise on the result. This case detects any movement in any direction iff the movement is greater than two basis points (0.02). The correlations with variables ctr1 and ctr2 are reduced while many of the others remain high.

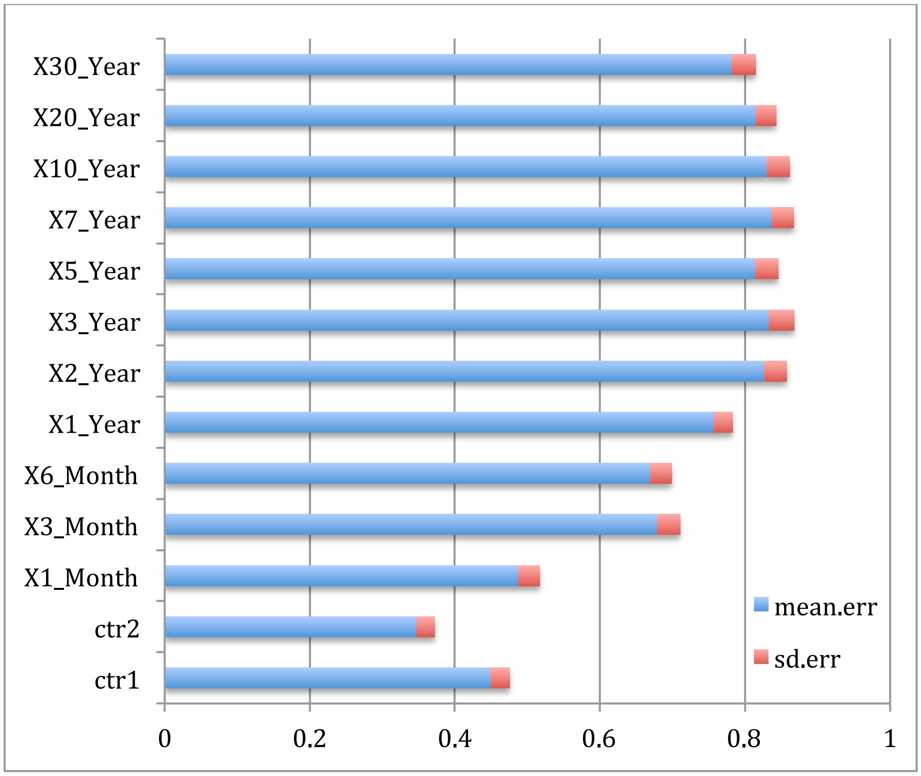


**Figure 49: Mean & Standard Deviation of effectiveness in Predicting any movement in any direction outside of a range (Case 3) . (Numerical scale represents absolute Average Error)**

### 7.3.5 Predictive power of Individual Independent Variables

Where the previous cases find clusters of exemplar Independent Variables, this case explores the value of individual Independent Variables. The results in the table below constrain the predictions to only one Independent Variable. This scenario appears to reduce the correction of all variables.

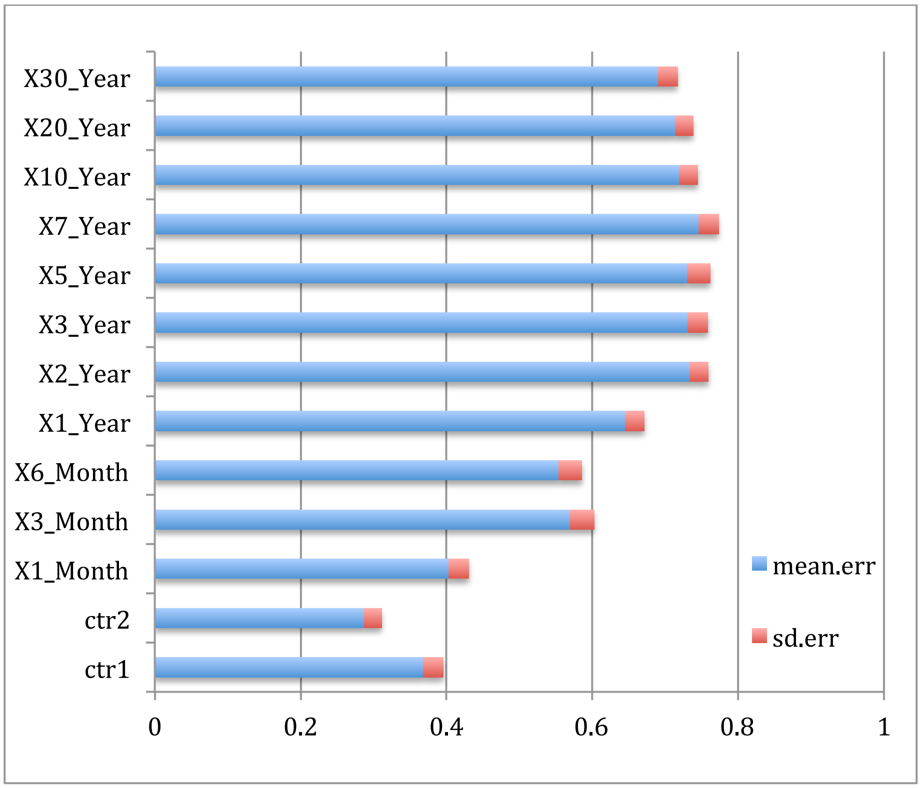


**Figure 50: Mean & Standard Deviation of effectiveness in Predicting any movement in any direction outside of a range, contrainted by individual Independent Variable (Case 4) .**
**(Numerical scale represents absolute Average Error)**

# 8 Concluding Remarks

## 8.1 Model Comparisons

This study presents a framework for the analysis of public policy documents specifically in a context that has implications in the financial markets. Chapter Four investigated the appropriateness of this framework in the U.S. Central Bank. Chapter Five investigated this in the context of corporate email. Chapter Six investigated both data sets using an established method of sentiment analysis. The goal of this analysis is to reveal patterns not observable under previous approaches. The framework suggests an effectiveness that justifies further inquiry both in the particular context explored in this study and in other domains.

The results generated in Chapter Four and Five compare favorably to the baseline results explored in Chapter Six to be used as a comparison.

## 8.2 Interpretation of summary findings of correlation

Each study presents the chosen predictors (subset of IVs) and performances for each combination input file, DV, shift-model, and regression model. Regression models that have been analyzed include: linear, CART, glm (with Gaussian link function), random forests, and support vector machines (with radial basis functions kernel). There has been no tuning on regression model parameters; this leaves the default values (e.g., default values for standard deviation and penalizing factor in the svm model).

Two shift models are considered: no-shift (contemporaneous), in which each IV is kept as it is, and best-shift, in which each pair of IV-DV lead to a best shift of the IV (within a -5 +5 range) in order to increase the correlation between the two. It is important to notice that shifting the IV in time is a rather risky procedure for the Public Policy data. First, the observations have already been pre-selected (see Chapter Three) in order to retain only those rows for which data is available. Therefore, shifting of one position in time does not necessarily mean shifting of one time-unit. Secondly, shifting the time series (i.e., IV) either forward or backward means introducing not-a-number elements at the beginning or end of the time series, hence further reducing the observations for which it is possible to calculate correlation in the regression model. It is because of these reasons that the no-shift model is the most appropriate one to interpret, at least with the current methodology that is characterized by sparse information.

In the appropriate chapter, the regression results are summarized for each input file separately. For each model, there are a number of rows equal to the number of DVs. Concerning the columns:

- the first column gives the name of the DV;
- the second column gives the performances for the given regression model upon the given DV, within the best-shift model;
- the third column is similar to the second one, only reporting the performances for the non-shift model. Performances, both for the second and third columns, are to be interpreted as (pseudo-) R-square value. The "pseudo" refers particularly to the random forest model, in which the actual value is defined as 1 – (mmse)/var(dep.var.) (and hence can be higher than 1);

- there are a number of columns which varies by model. This is due to the clustering of IVs according to their respective correlation change with the given input file. For example, File 1 has 19 IVs. There are therefore 19 columns representing these variables. Each has a zero or a one depending on whether that IV was chosen (1) or not (0) for the subset of best IVs with any given DV. Tables are labeled as referring to either the best-shift model or the non-shift model; and
- for each regression model, the average and standard deviation of the performances across all DVs.

The average performances for all input files, all regression models, and both best-shift and non-shift models are finally summarized in the appropriate chapters.

Each of two tables (one for each shift model) in each appropriate chapter, shows the average performances across DVs for each regression model and for each input file. If we were to base our final decision only upon this information, the set of variables in file five provides the average best results for the svm model. However, it is important to check these results with information regarding the final dataset actually being considered. For each input file, a different subset of observations was found to be usable, depending on how many rows presented real values. I therefore report a summary table showing, for each input file, how many observations (i.e., rows) where actually used, the span in terms of dates of these information, and the number of IV clusters identified. From this radically reduced dataset (now down to 14), the file is clearly not worth considering further. In the case of the Public Policy Dataset, the appropriate input files for pursuing the next step of building a predictive model would be most appropriate to include 1, 2 or 10. Each of these three input files retains a large number of temporal information. For these cases, the best performances are still obtained using the svm model, and are approximately 0.6. There appears to be an advantage in using the best-shift model, but such increase in performances is not striking.

The differences in correlations between sources of text have similar results. The more dense the data available, the better the results. For example, adding the sparsely populated ‘conference call’ information does not materially increase the correlations found.

The Figure below is the summarization of the comparsion of the framework's effectiveness on the meeting minutes part of the Public Policy data relative to baseline. The visualizastion suggestions materially better performance for the framework over baseline methods across four of five learning algorithms; the best results are from the SVM algorithm and time-shifted data. The importance and implications are discussed further elsewhere in this Dissertation.

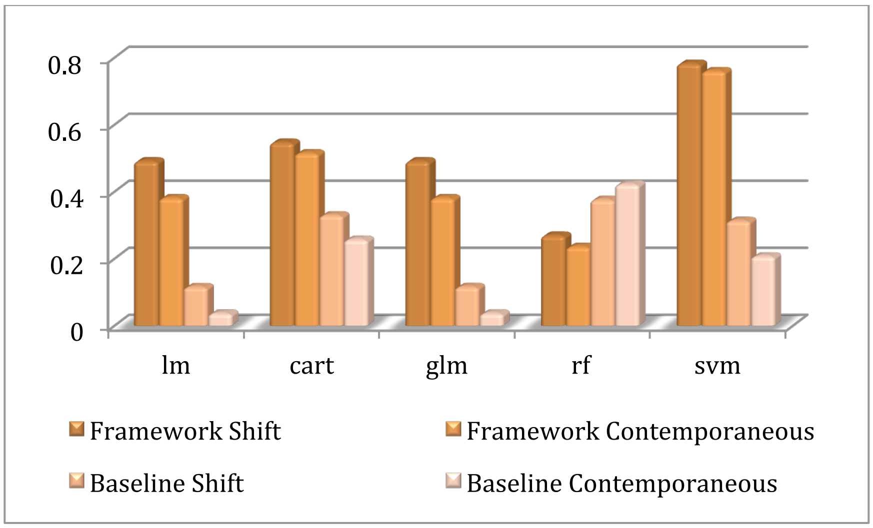


**Figure 51: Comparison of Framework against baseline of summary results from study of public policy dataset on all meeting minutes data (Numerical scale represents pseudo R-squared)**

The Figure below is the summarization of the comparsion of the framework's effectiveness on the Speech part of the Public Policy data relative to baseline. The visualizastion suggestions materially better performance for the framework over baseline methods across four of five learning algorithms; the best results are again from the SVM algorithm and time-shifted data. The importance and implications are discussed further elsewhere in this Dissertation.

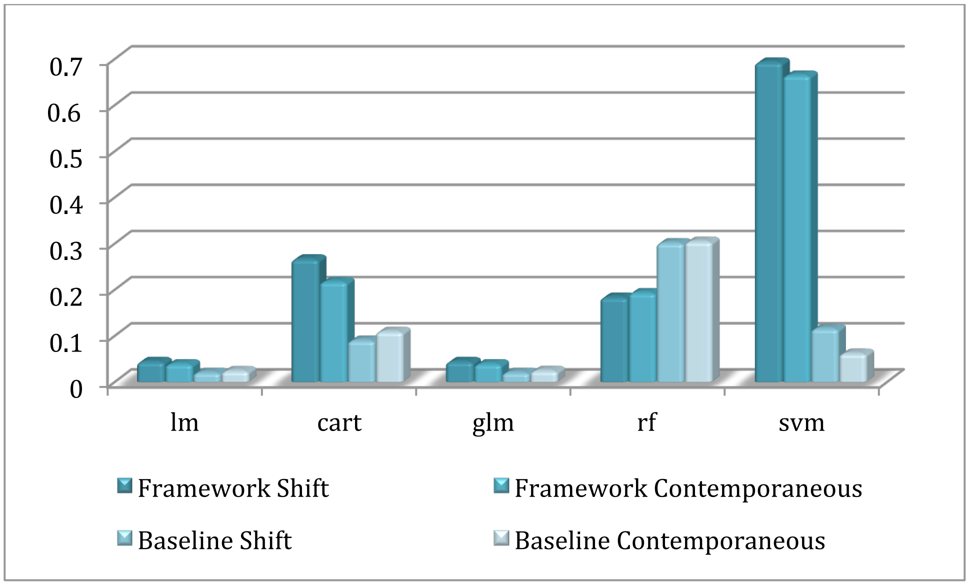


**Figure 52: Comparison of Framework against baseline of summary results from study of public policy dataset on all Speech data (Numerical scale represents pseudo R-squared)**

The Figure below is the summarization of the comparsion of the framework's effectiveness on all parts of the available Public Policy data relative to baseline. The visualizastion suggestions materially better performance for the framework over baseline methods across four of five learning algorithms; the best results are from the SVM algorithm and time-shifted data. The importance and implications are discussed further elsewhere in this Dissertation.

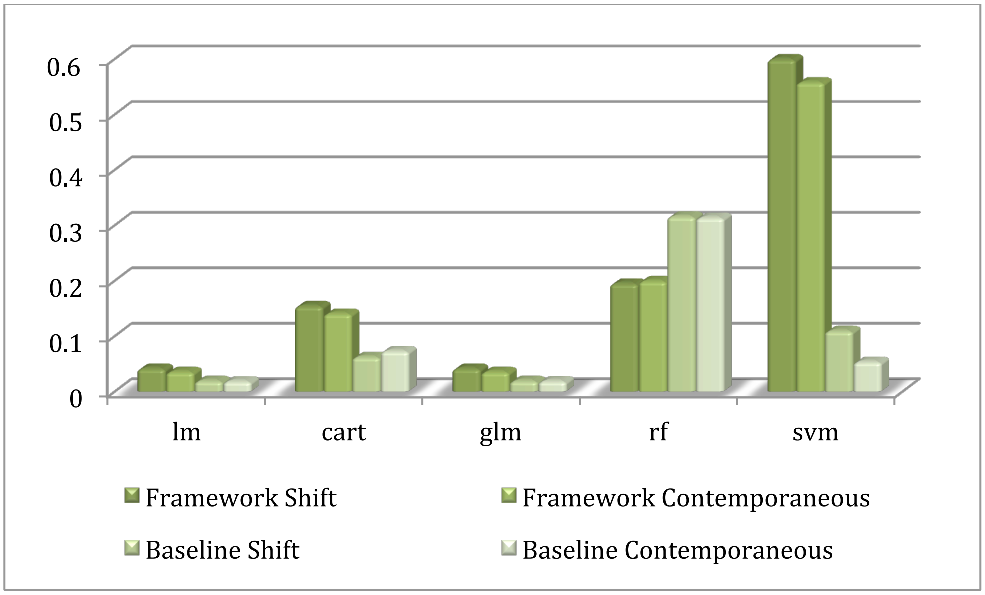


**Figure 53: Comparison of Framework against baseline of summary results from study of public policy dataset on all data combined (Numerical scale represents pseudo R-squared)**

The Figure below is the summarization of the comparsion of the framework's effectiveness on the Corporate email corpus relative to baseline. The visualization suggestions materially better performance for the framework over baseline methods across all learning algorithms; the best results are from the Random Forests algorithm with little differences suggested by either time-shifted or contemporaneous data. The importance and implications are discussed further elsewhere in this Dissertation.

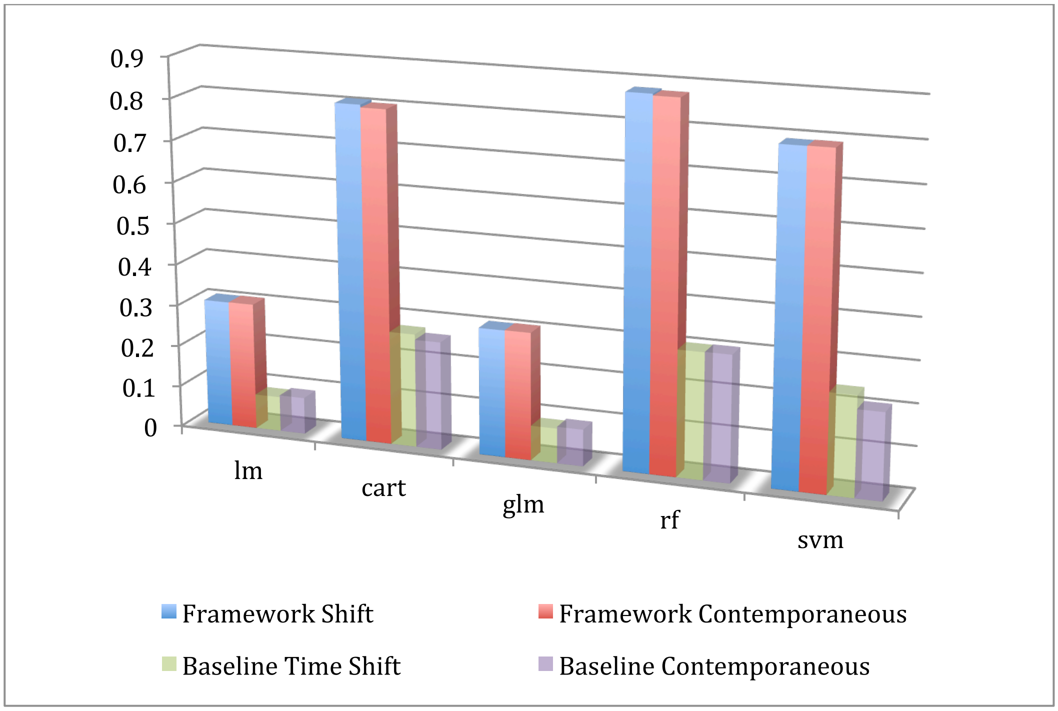


**Figure 54: Comparison of Framework against baseline of summary results from study of corporate email dataset (Numerical scale represents pseudo R-squared)**

## 8.3 Interpretation of Summary Findings of predictive value

The previous chapter explores the predictive value of Independent Variables in five cases. The results may imply the increased effectiveness of predicting the actual number over movement. However, an easier comparison between the efficacy of the different approaches is to look at the percentage of the error.

The case of predicting number may look interesting from the various Figures below. The first two Figures below suggest a relatively low difference in the effectiveness in predicting numbers, but also low absolute level of effectiveness. The implications for this in Future work is discussed further later in this Dissertation.

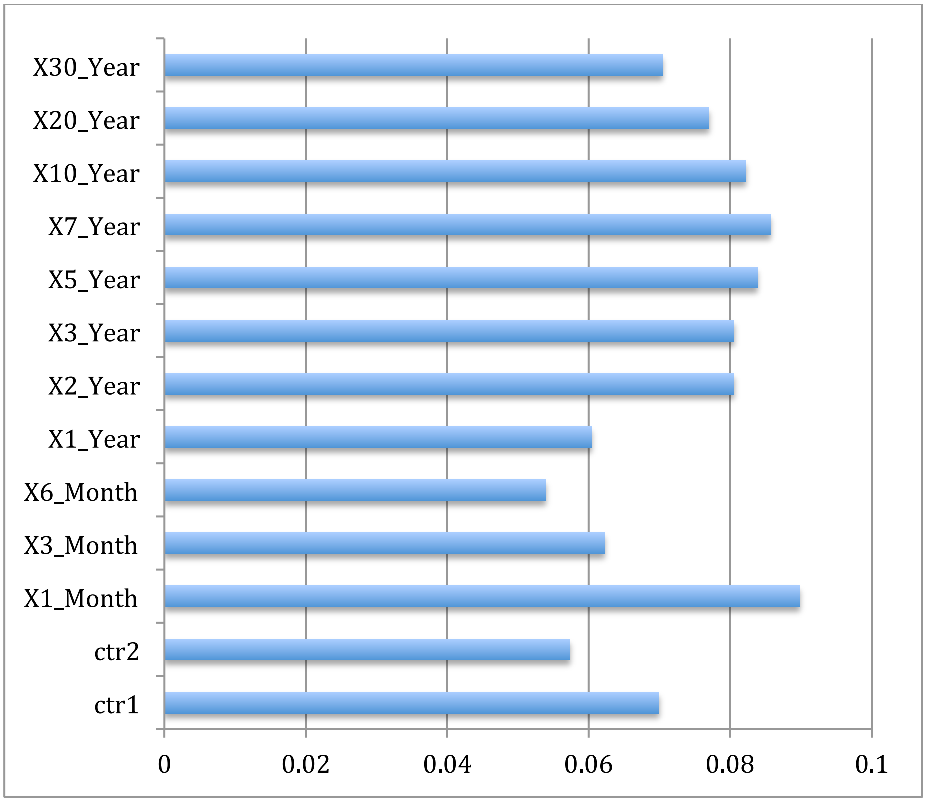


**Figure 55: Mean Delta of effectiveness in predicting number (Case 0) .**
**(Numerical scale represents absolute Average Error)**

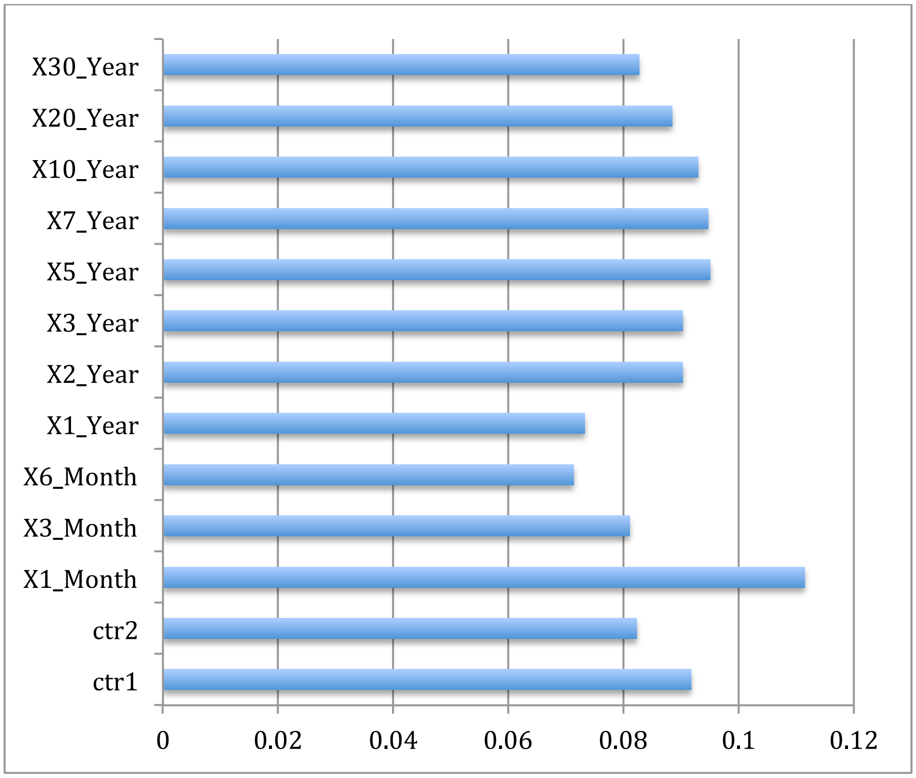


**Figure 56: Mean Delta of effectiveness in Predicting number outside of a range (Case 1). (Numerical scale represents absolute Average Error)**

However, as the above Figures suggest some degree of uniformity in effectivenss, as the Figure below makes clear, the percentage by which the numerical prediction was in error sometimes exceeded 100%. This suggests that the absolute number is unlikely to be effectiveness predicted.

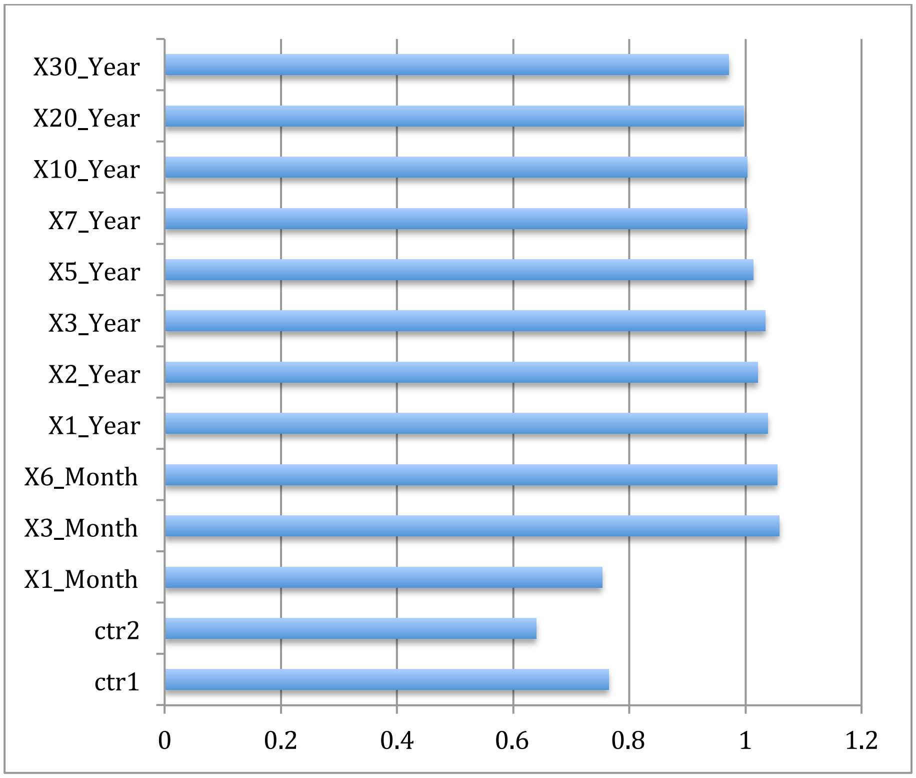


**Figure 57: Average percentage error of effectiveness in predicting number (Case 0). (Numerical scale represents Percent Error)**

Where the percentage error make the comparisons more clear with the later cases, the Figure below also demonstrates an increased effectiveness in predicting movement greater than 0.02. This result suggests that in future work, the framework is more effective with movements above some level that may be characterized as 'noise' or otherwise random daily movements.

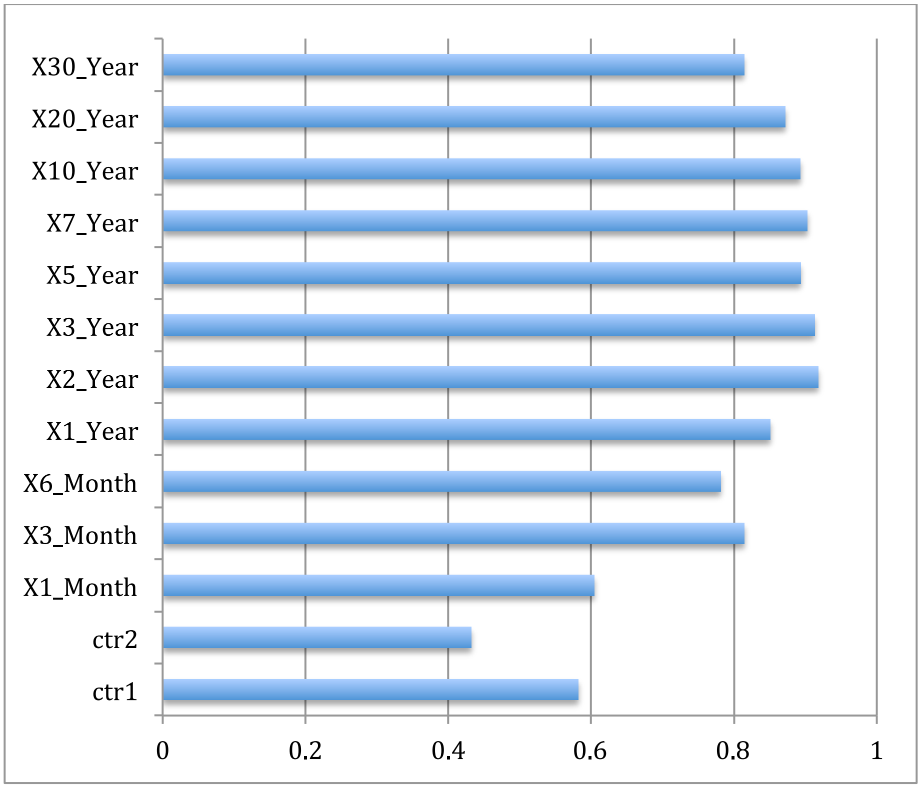


**Figure 58: Average percentage error of effectiveness in predicting number outside of a range (Case 1).**
**(Numerical scale represents Percent Error)**

With both cases 2 & 3, when the estimation is wrong, the result is very wrong as visualized in the Figure below. In every case, the predicted number is outside the range of the error. The result is essentially random.

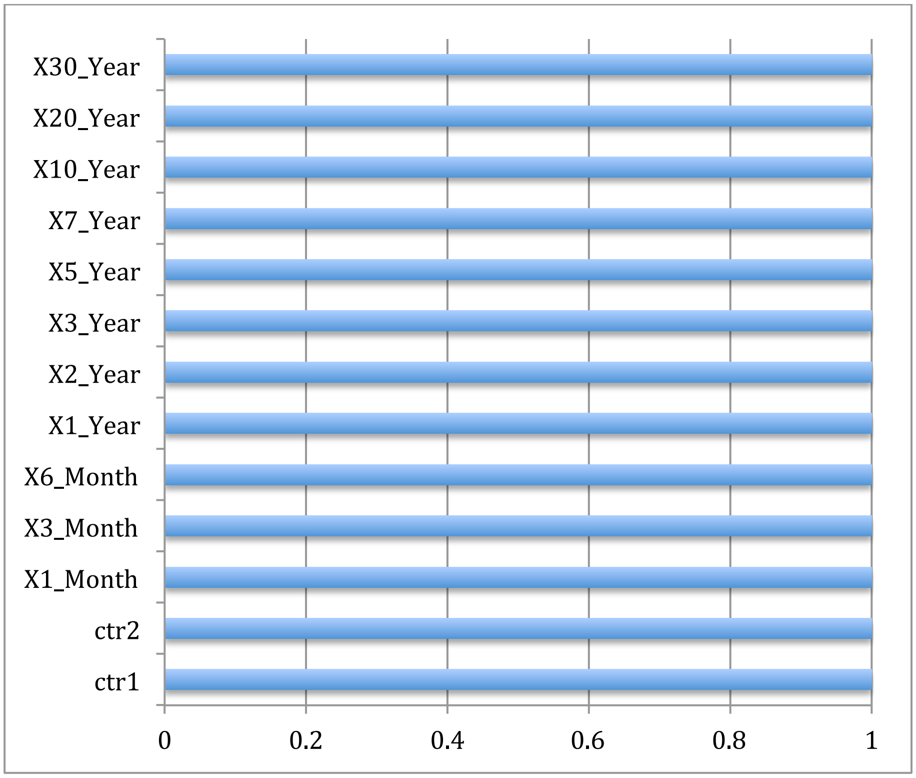


**Figure 59: Mean Delta of effectiveness in Predicting number outside of a range (Case 1). (Numerical scale represents absolute Average Error)**

With the mean delta revealed as a less useful measure for comparison, we look again to the percentage error for cases 2 & 3 presented in the Figures below. The results, importance, and implicsations for future work are discussed later in this chapter.

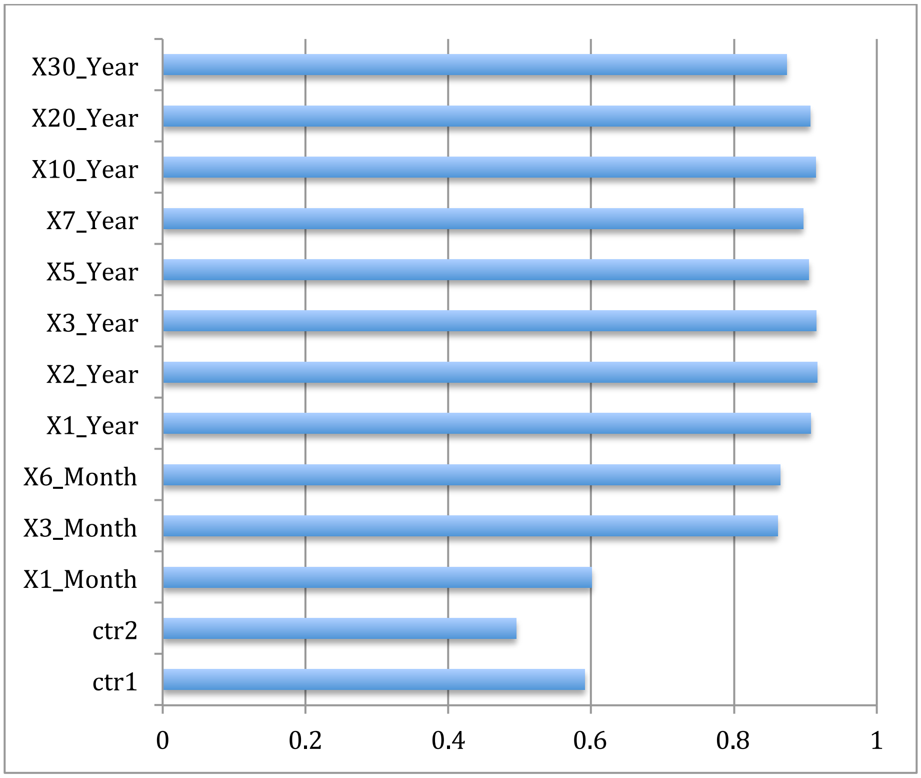


**Figure 60: Average percentage error of effectiveness in predicting number (Case 2). (Numerical scale represents Percent Error)**

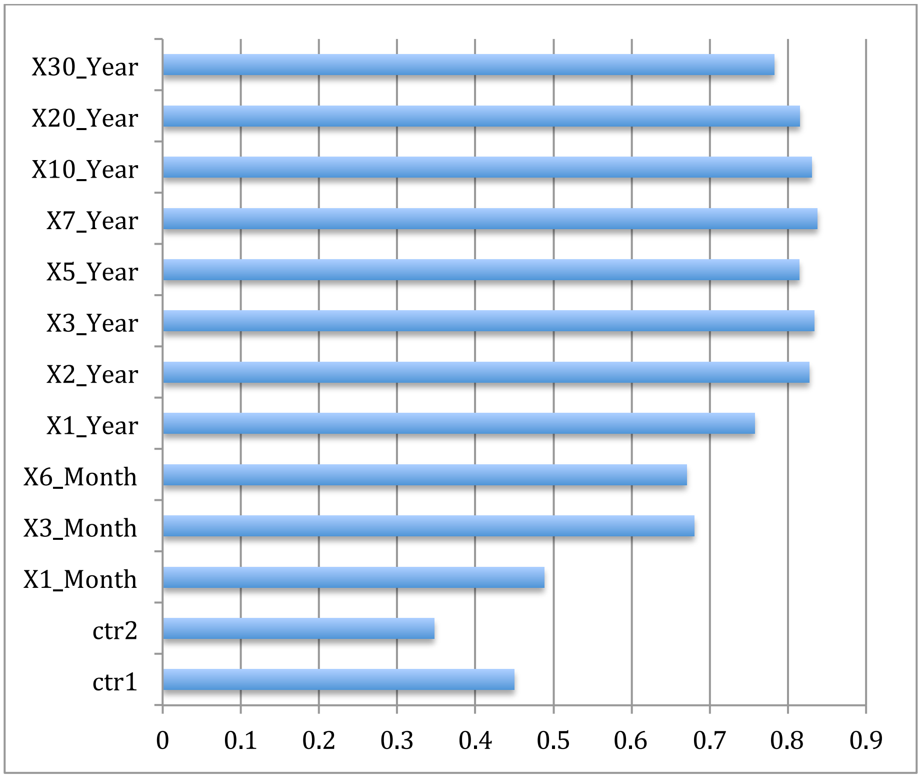


**Figure 61: Average percentage error of effectiveness in predicting number outside of a range (Case 3).**
**(Numerical scale represents Percent Error)**

Where the percent error in cases 2 & 3 look to present substantial improvement over cases 0 & 1, the percent error for cases 2 & 3 may also be presented as a rounded integer as discussed earlier. The Figures below visualize the improvement in rounding the numbers for cases 2 & 3. Presented as cases 2 & 3 *discrete.* The results, importance, and implicsations for future work are discussed later in this chapter.

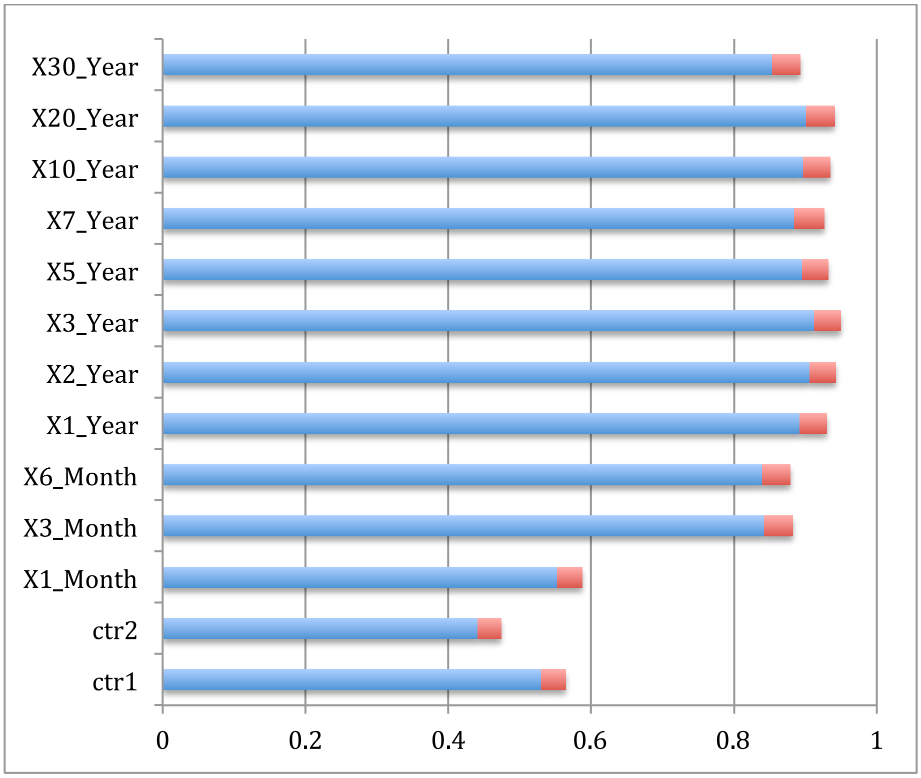


**Figure 62: Average percentage error and standard deviation of effectiveness in predicting movement (Case 2 *discrete*). (Numerical scale represents Percent Error)**

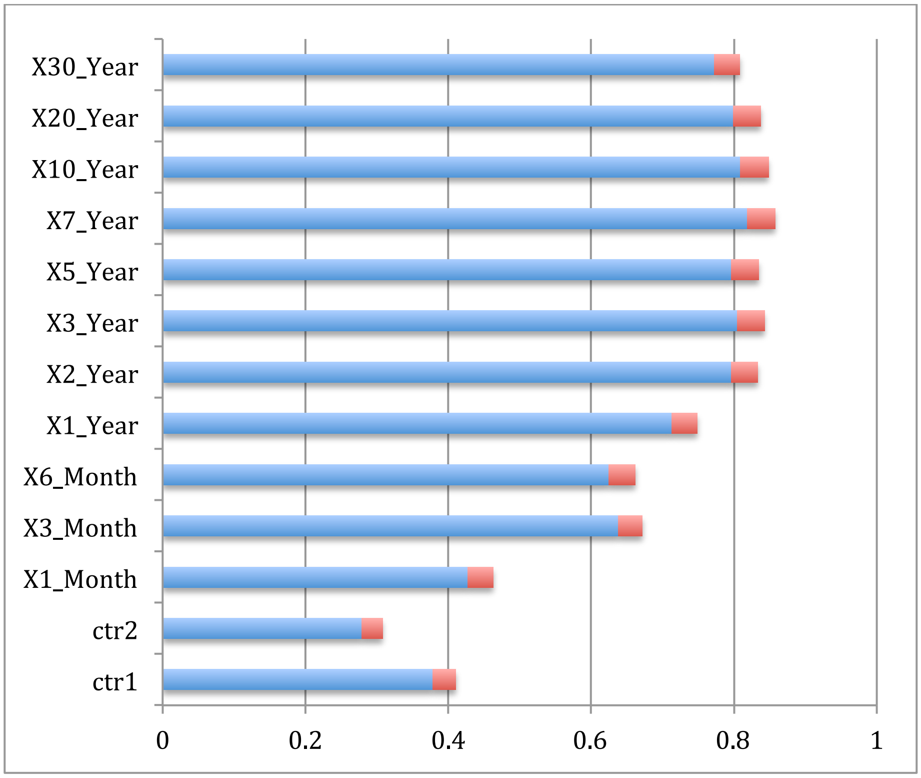


**Figure 63: Average percentage error and standard deviation of effectiveness in predicting movement outside of a range (Case 3 *discrete*). (Numerical scale represents Percent Error)**

While the Figures above tell the narrative, the Figure below visualizes the summary findings:

- Predicting movement of any magnitude in any direction is more effective than predicting an actual number.
- Predicting movement outside of the range around zero is more effective.
- Rounding the results produces fewer errors in the case of predicting movement.
- Two Dependent Variables in the Public Policy Dataset (the financial derivatives) suggest materially superior predictive capacity for the approach than the debt securities longer than thirty days in term.

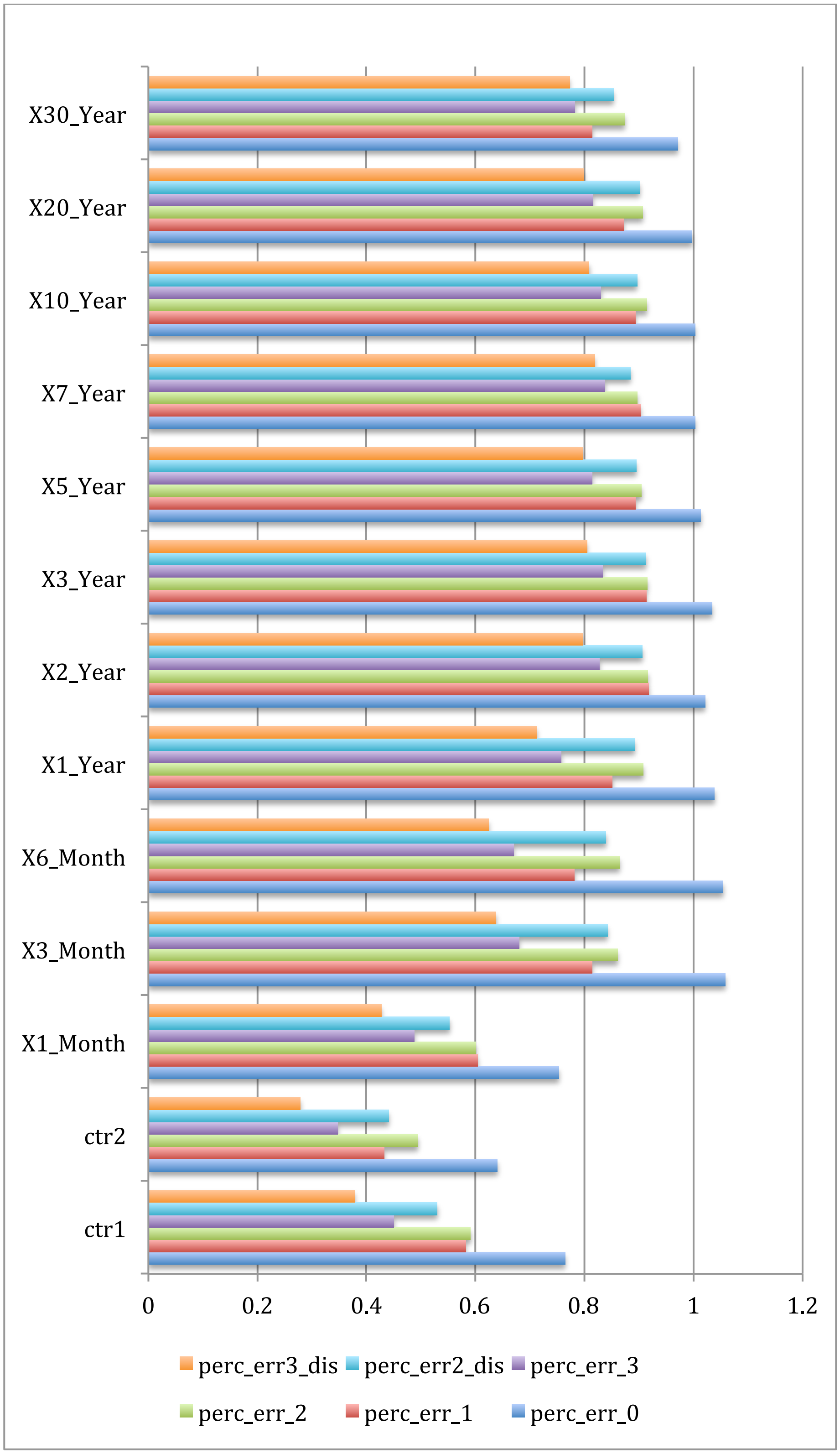


**Figure 64: Average percentage error for Cases 0-3.**

**(Numerical scale represents Percent Error)**

# 9 Implications

Previous chapters have introduced a new framework for the analysis of text and described the potential significance (see Chapters 1-3). This Dissertation measures the effectiveness of this framework is on two datasets with different characteristics and in different domains (Chapters 4 and 5). The results have been compared to classical solutions as a baseline (Chapter 6). The degree to which future work may find predictive value is explored in Chapter 7. Comparisons of these applications are discussed in Chapter 8. This chapter works to explain the implications of these findings.

Below, the suggested impact of significant independent variables on the dependent variable is addressed. I also examine how robust are the findings to alternative behaviors than those observed. Lastly, I provide some experimental data on the impact of variations in the Independent Variables.

## 9.1 The meaning of the Independent Variables as Network Measures (Redux)

### 9.1.1 Independent Variables w/ theoretical DV relationship

*Average Distance*, a graph-level measure, is inversely related to *Closeness* (Carley, 2002) (Carley, 2002) (Carley, Reminga, Storrick, & Columbus, 2011) (Freeman, 1978), which in the semantic networks analyzed, is associated with price changes. *Average Distance* is the average shortest path length between nodes, (excluding infinite distances) (Carley, 2002); it measures how easily a node can be reached from the other vertices. Text may alter this measure by either changing length or variety of phrasing (Borge-Holthoefer & Arenas, 2010).

*Breadth* is graph-level measure that gives the fraction of entities with nodes that have degree greater than one (Carley, Reminga, et al., 2011). Pichl (Pichl, 2010) argues that this measure can be increased by phrases being repeated in a corpus.

The measure of *Density*, a graph-level measure, has also been found to be associated with price changes. It is defined as simply the number of ties in the network divided by the maximum number of ties that are possible (Wasserman & Faust, 1997). Unfortunately, its nature seems to make it difficult to provide reliable guidance toward the measurement's impact from the dynamics of a transcript. While large-scale semantic networks are characterized by sparse connectivity, we cannot necessarily conclude that coherent, simple messages have a higher density (Steyvers & Tenenbaum, 2011). *Density* measures in large networks may be associated with more structural cohesion than higher densities in smaller networks (Bales & Johnson, 2005). Since Density becomes a misleading indicator of structural cohesion when a group has subgroups (Friedkin, 1981), the semantic analysis to reveal subgroups would have to be performed in advance of each iteration of the measurement's use.

While Cantador (Cantador & Castells, 2006) argues that the frequency (or absence) of identical matched phrases in a corpus can move the metric of *Efficiency*, a graph-level measure, (both local and global), Borge-Holthoefer (Borge-Holthoefer & Arenas,

2010) argues that the measure *Efficiency* has similar characteristics to *Density* in the analysis of Semantic Networks. In the context of this research, these variables (*Density* and *Efficiency*) appear to be robust to alternative behaviors. This does not imply that the measurement is not correlated to the dependent variables.

The measures *Link Count* and *Row Count*, both graph-level measures, similarly have a correlation with the dependent variable, but also appear to be robust to alternative behaviors from the viewpoint of this research. Rogers (Rogers, 2006) found that for this particular measure, completely different contexts could be given the same weight. This might be as varied as an essay on a Zebra or a Barber pole in his example; for the research under consideration an example is between monetary policy suggested for different temporalities. The measure remains valid in its application, but problematic analyzing in reverse.

*OutDegree*, a graph-level measure, which measures the influence score of each node (i.e., how many nodes are affected by node i); and *inDegree*, a graph-level measure, which measures how many nodes influence node *i*. are measures that can manifest themselves in phrases that are often (or always) connected (Kenett, Kenett, Ben-Jacob, & Faust). Generic examples of this are words such as *four, dough, baked, bakery.* Both of these measures are associated with changes in the dependent variable. Manipulating the effects of these variables should involve the experimentation with the repetition of phrases of various length and commonality. Phrases can then individually be found to have impact in the measures *inDegree* or *outDegree* and thus important to the emerging ontology of a group seeking to impact the measurements (Hoser, Hotho, Jäschke, Schmitz, & Stumme, 2005).

*Redundancy*, a graph-level measure, captures the robustness of a network: in a highly redundant network, if a random connection is deleted, the deleted link will not alter the likelihood of a connected path between two works (Beckage, Smith, & Hills, 2010). The network measurement of *Redundancy* takes a network *N*, and produces the mean number of non-zero row entries in excess of one in the network's matrix representation. That is, with input *N* of dimensions $m$ x $n$, *Redundancy, Row* produces output $\Re \in [0, (n-1)\times m]$ (Reminga & Carley, 2003). *Redundancy* therefore suggests the fluidity of speech such as being less hampered by the forgetting of a few words. A lower *Redundancy* measure can imply word-finding and word-retrieval difficulties (Beckage et al., 2010). The research findings may not be robust to a subject's experimentation with a broader variety of words.

*Span of Control*, a graph-level measure, is the number of 'subordinates' per 'supervisor'. This is calculated taking the sum of all subordinates then dividing by the number of supervisors (Carley, 2002). The nomenclature definition suggests a motivation for its invention that may not translate to the analysis of semantic networks. While changes in this measure are correlated to changes in the dependent variable, this measure also appears to be robust to alternative behaviors (such as changes in one's speech) than those observed.

*Transitivity*, a graph-level measure, is the percentage of triads *i,j,k* in a square network *N* such that if *(i,j)* and *(j,k)* are in the network, then *(j,k)* is in the network (Reminga & Carley, 2003). Fallucchi (Fallucchi & Zanzotto, 2011) distinguishes a generalization of the distributional hypothesis as an appropriate model for *Transitivity*: "Words that tend to occur in the same contexts tend to have similar meanings." This measure may not be robust to a research subject experimenting with the use of one-word changes in otherwise identical phrasing.

### 9.1.2 Independent Variables w/o theoretical DV relationship

The following measurements are omitted from consideration in this discussion because there is no evidence for their relationship with the dependent variables in the research under consideration: *Clustering Coefficient, Connectedness, Diffusion, Fragmentation, Interdependence,* and *Speed*.

The dependent variables are not individually identified again because they tend to move as a class and in the same direction. Exceptions to this generalization where there are sharp reversals of policy or those times when the yield curve is inverted are outside the scope of this research. Therefore, in the tables below, the dependent variables are characterized as one class.

With regard to the interactions of the independent and dependent variables, direction cannot be immediately determined. Sometimes interest rates will be trending upward while other times will be trending downward. Still other times there is indecision. With the nature of this research, movements of the variable up or down cannot be characterized outside of the context under which they are investigated. Future work would involve dissecting portions of the data based on trend. The most likely outcome is not an absolute direction but rather reversals of trends occurring with stronger, less ambiguous phrasing. The network measures (independent variables) are presented along with their output ranges in the following Table (Table 17).

<table>
<tr><th colspan="3">ACTUAL INTERACTION BETWEEN<br>DEPENDENT AND INDEPENDENT VARIABLES</th></tr>
<tr><th colspan="2">With dependent variables moving together as a class, changes (up or down) in the dependent variables…</th><th rowspan="2">No evidence of a relationship</th></tr>
<tr><th>…suggest changes in the independent variable (up or down).</th><th>…are robust to changes in the independent variable.</th></tr>
<tr><td>Average Distance<br>$\Re \in [0,1]$</td><td>Density<br>$\Re \in [0,1]$</td><td>Clustering Coefficient<br>$\Re \in [0,1]$</td></tr>
<tr><td>Breadth<br>$\Re \in [0,1]$</td><td>Efficiency<br>$\Re \in [0,1]$</td><td>Connectedness<br>$\Re \in [0,1]$</td></tr>
<tr><td>inDegree and OutDegree<br>$\Re \in [0,1]$</td><td>Link Count and Row Count<br>$\Re \in$ [0, number of links/rows in the network]</td><td>Speed<br>$\Re \in [0,1]$</td></tr>
<tr><td>Redundancy<br>$\Re \in [0,(n-1)*m]$ for $N$ dimension $m \times n$</td><td>Span of Control<br>$\Re \in [0,|V|-1]$</td><td>Diffusion<br>$\Re \in [0,1]$</td></tr>
</table>

| ACTUAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES | | |
|---|---|---|
| With dependent variables moving together as a class, changes (up or down) in the dependent variables… | | No evidence of a relationship |
| …suggest changes in the independent variable (up or down). | …are robust to changes in the independent variable. | |
| *Transitivity* $\Re \in [0,1]$ | | *Interdependence* $\Re \in [0,1]$ |
| | | *Fragmentation* $\Re \in [0,1]$ |

**Table 17: Actual Interaction between Dependent and Indepdendent Variables in study of Public Policy Documents**

## 9.2 Theoretical Interaction between Independent and Dependent Variables

### 9.2.1 Generalized

With Table 17 (above) describing the actual findings of the research, Table 18 (below) hypothesizes the effects on a theoretical dependent variable. Based on descriptions of each independent variable, future work may determine the actual interaction between these independent variables and this new dependent variable (or class of dependent variables).

| THEORETICAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES (Generalized) | | |
|---|---|---|
| With theoretical dependent variables indicating economic confidence (simply phrased as 'prices'), moving together as a class, changes in the dependent variables may have the following interactions with the independent variables. | | |
| Changes in the dependent variables may impact the independent variable. | Changes in the dependent variables are robust to changes in the independent variable. | No evidence of a relationship |
| Increases in *Average Distance* may suggest an increase in prices to the extent that longer or more varied phrasing demonstrates optimism for the future. $\Re \in [0,1]$ | *Density* $\Re \in [0,1]$ | *Clustering Coefficient* $\Re \in [0,1]$ |
| Increases in *Breadth* may suggest decreases in prices if increase in phrase repetition demonstrates concern for the future. $\Re \in [0,1]$ | *Efficiency* $\Re \in [0,1]$ | *Connectedness* $\Re \in [0,1]$ |
| Increases in *inDegree* and *OutDegree* may suggest lower prices if longer, repeated phrases suggest concern for the current trajectory. $\Re \in [0,1]$ | *Link Count* and *Row Count* $\Re \in$ [0, number of links/rows in the network] | *Diffusion* $\Re \in [0,1]$ |
| Increases in *Redundancy* may suggest lower prices as word repetition may be expressing concern for the future. $\Re \in [0,(n-1)*m]$ for $N$ dimension $m \times n$ | *Span of Control* $\Re \in [0,\|V\|-1]$ | *Fragmentation* $\Re \in [0,1]$ |

| THEORETICAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES (Generalized) | | |
|---|---|---|
| With theoretical dependent variables indicating economic confidence (simply phrased as 'prices'), moving together as a class, changes in the dependent variables may have the following interactions with the independent variables. | | |
| Changes in the dependent variables may impact the independent variable. | Changes in the dependent variables are robust to changes in the independent variable. | No evidence of a relationship |
| Increases in *Transitivity* may suggest lower prices as repeating even similar phrases may express concern for the future.<br>$\mathfrak{R} \in [0,1]$ | | *Speed*<br>$\mathfrak{R} \in [0,1]$ |
| | | *Interdependence*<br>$\mathfrak{R} \in [0,1]$ |

**Table 18: Theoretical Interaction between dependent and independent variables in public policy study (Part I)**

### 9.2.2 Specific interactions in this study

The tables below go further as a Second and Third Part to the generalized interaction. They describe the predicted effect on the dependent variables under particular conditions. That is, they describe the interaction for this particular study. Most importantly, the conditions are constrained by considering the dependent variables as a class. This can be a reasonable assumption under most circumstances for the research into Central Bank behavior because the objective of much of the communication from the Fed is to effect short-term rates.

The other assumptions used for the tables below are more nuanced. Communications from the Fed are done with full knowledge of market events and therefore trends. The first table below is concerned with those times where communications is intended to continue the current trends. Under this grouping, the trend could be up or down, fast or slow. The issue is the encouragement of the current trends. The second table below outlines the response to the opposite. These communications are meant to discourage current trends: slow, stop, or even reverse them.

| THEORETICAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES (Part II) | | |
|---|---|---|
| With theoretical dependent variables indicating economic confidence (simply phrased as 'prices'), moving together as a class, changes in the dependent variables may have the following interactions with the independent variables.<br><br>Additional assumption is that intent of phrasing is to continue recent trends. | | |
| Changes in the dependent variables may impact the independent variable. | Changes in the dependent variables are robust to changes in the independent variable. | No evidence of a relationship |
| Increases in *Average Distance* may suggest a continuation of current trends to the extent that longer or more varied phrasing demonstrates optimism for the future or the current direction of prices. $\Re \in [0,1]$ | *Density* $\Re \in [0,1]$ | *Clustering Coefficient* $\Re \in [0,1]$ |
| Decreases in *Breadth* may suggest decreases in prices if decreases in phrase repetition demonstrates acceptance of current trends. $\Re \in [0,1]$ | *Efficiency* $\Re \in [0,1]$ | *Connectedness* $\Re \in [0,1]$ |
| Decreases in *inDegree* and *OutDegree* may suggest encouragement of current trends if longer, repeated phrases suggest sanguinity for the current market trajectory. $\Re \in [0,1]$ | *Link Count* and *Row Count* $\Re \in [0,$ number of links/rows in the network$]$ | *Diffusion* $\Re \in [0,1]$ |

| THEORETICAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES (Part II) | | |
|---|---|---|
| With theoretical dependent variables indicating economic confidence (simply phrased as ‘prices’), moving together as a class, changes in the dependent variables may have the following interactions with the independent variables.<br><br>Additional assumption is that intent of phrasing is to continue recent trends. | | |
| Changes in the dependent variables may impact the independent variable. | Changes in the dependent variables are robust to changes in the independent variable. | No evidence of a relationship |
| Decreases in *Redundancy* may suggest continued market trending as less word repetition may be expressing less concern for the current price trajectory.<br>$\Re \in [0,(n-1)*m]$ for *N* dimension<br>$m \times n$ | *Span of Control*<br>$\Re$ □ $[0,\|V\|-1]$ | *Fragmentation*<br>$\Re \in [0,1]$ |
| Decreases in *Transitivity* may suggest encouragement of the current price trajectory as repeating fewer phrases may express less concern for the future.<br>$\Re \in [0,1]$ | | *Speed*<br>$\Re \in [0,1]$ |
| | | *Interdependence*<br>$\Re \in [0,1]$ |

**Table 19: Theoretical Interaction between Dependent and Independent Variables in Public Policy Study (Part II)**

| THEORETICAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES (Part III) | | |
|---|---|---|
| With theoretical dependent variables indicating economic confidence (simply phrased as 'prices'), moving together as a class, changes in the dependent variables may have the following interactions with the independent variables.<br><br>Additional assumption is that intent of phrasing is to slow, stop or reverse recent trends. | | |
| Changes in the dependent variables may impact the independent variable. | Changes in the dependent variables are robust to changes in the independent variable. | No evidence of a relationship |
| Decreases in *Average Distance* may suggest an effort to reverse trends to the extent that shorter phrasing demonstrates concern for the future.<br>$\Re \in [0,1]$ | *Density*<br>$\Re \in [0,1]$ | *Clustering Coefficient*<br>$\Re \in [0,1]$ |
| Increases in *Breadth* may suggest efforts to reverse trends if increase in phrase repetition demonstrates concern for the future.<br>$\Re \in [0,1]$ | *Efficiency*<br>$\Re \in [0,1]$ | *Connectedness*<br>$\Re \in [0,1]$ |
| Increases in *inDegree* and *OutDegree* may suggest efforts to reverse the current trends if longer, repeated phrases suggest concern for the current trajectory.<br>$\Re \in [0,1]$ | *Link Count* and *Row Count*<br>$\Re \in [0,$ number of links/rows in the network$]$ | *Diffusion*<br>$\Re \in [0,1]$ |
| Increases in *Redundancy* may suggest discomfort for the current market trajectory as word repetition may be expressing concern for the future.<br>$\Re \in [0,(n-1)*m]$ for $N$ dimension $m \times n$ | *Span of Control*<br>$\Re \in [0,\|V\|-1]$ | *Fragmentation*<br>$\Re \in [0,1]$ |
| Increases in *Transitivity* may suggest an attempt to temper current trends as repeating even similar phrases may express concern for the current market trajectory.<br>$\Re \in [0,1]$ | | *Speed*<br>$\Re \in [0,1]$ |

| THEORETICAL INTERACTION BETWEEN DEPENDENT AND INDEPENDENT VARIABLES (Part III) | | |
|---|---|---|
| With theoretical dependent variables indicating economic confidence (simply phrased as 'prices'), moving together as a class, changes in the dependent variables may have the following interactions with the independent variables.<br><br>Additional assumption is that intent of phrasing is to slow, stop or reverse recent trends. | | |
| Changes in the dependent variables may impact the independent variable. | Changes in the dependent variables are robust to changes in the independent variable. | No evidence of a relationship |
| | | *Interdependence*<br>$\Re \in [0,1]$ |

**Table 20: Theoretical Interaction between Dependent and Independent Variables in Public Policy Study (Part III)**

## 9.3 Hypothesis for behavior of Exemplar Independent Variables

From *Section 4.4*, the Independent Variable Candidates are summarized for consideration before input into the learning algorithms. They are repeated in Table 21 (below), but this time mapped to the degree to which the variables their relationships to the dependent variables are impacted by their own movements.

| INDEPENDENT VARIABLE CANDIDATES | |
|---|---|
| *MI+*:CHANGES IN DV MAY IMPACT IV<br>*RC-*:CHANGES IN DV ARE ROBUST TO IV CHANGES<br>*NR 0*:NO EVIDENCE OF RELATIONSHIP | |
| BreadthColumnSemanticNetwork | *MI+* |
| CentralityAuthoritySemanticNetworkAverage | *NR 0* |
| CentralityColumnDegreeSemanticNetworkAverage | *NR 0* |
| CentralityInClosenessSemanticNetworkAverage | *NR 0* |
| CommunicativeNeedSemanticNetwork | *NR 0* |
| EffectiveNetworkSizeBurtSemanticNetworkAverage | *RC-* |
| HierarchySemanticNetwork | *RC-* |
| IsolateCountSemanticNetwork | *RC-* |
| LinkCountLateral.SemanticNetwork | *RC-* |
| LinkCountReciprocalSemanticNetwork | *RC-* |
| LinkCountSequentialSemanticNetwork | *RC-* |
| LinkCountSkipSemanticNetwork | *RC-* |
| MetaMatrixHammingDistance | *MI+* |
| NetworkCentralizationInClosenessSemanticNetwork | *MI+* |
| NetworkCentralizationInDegreeSemanticNetwork | *MI+* |
| NumberOfConceptNodes | *RC-* |
| OverallComplexity | *MI+* |
| SpeedAverageSemanticNetwork | *NR 0* |
| UpperBoundednessSemanticNetwork | *RC-* |

**Table 21: Summary of Representative Independent Variables after culustering**

## 9.4 Predictive value of Exemplar Independent variables

Earlier chapters investigating the predictive value to be attained from this framework have found the greatest power in predicting any movement in the dependent variables. This is in contrast to being able to predict the precise value of the dependent variables. Therefore, the experiment on the predictive value of each each Exemplar Independnt variable is used that outcome to frame this inquiry.

Earlier chapters looked at the average predictive value of individual variables over all Dependent Variables. The results from that experiment suggested that the inquiry of predictive capacity is materially more effective under what was termed *case 3 discrete*. That is, the prediction of any movement outside of a range, with the result rounded. Additionally, the results from those experiments suggested that three of the Dependent variables for Data set 1 were materially more effective than the rest. It with those conclusions that the inquiry for this chapter starts and the summarized results are presented in the three Figures below. The importance and implications for future work are discussed below and in more detail in those respective sections.

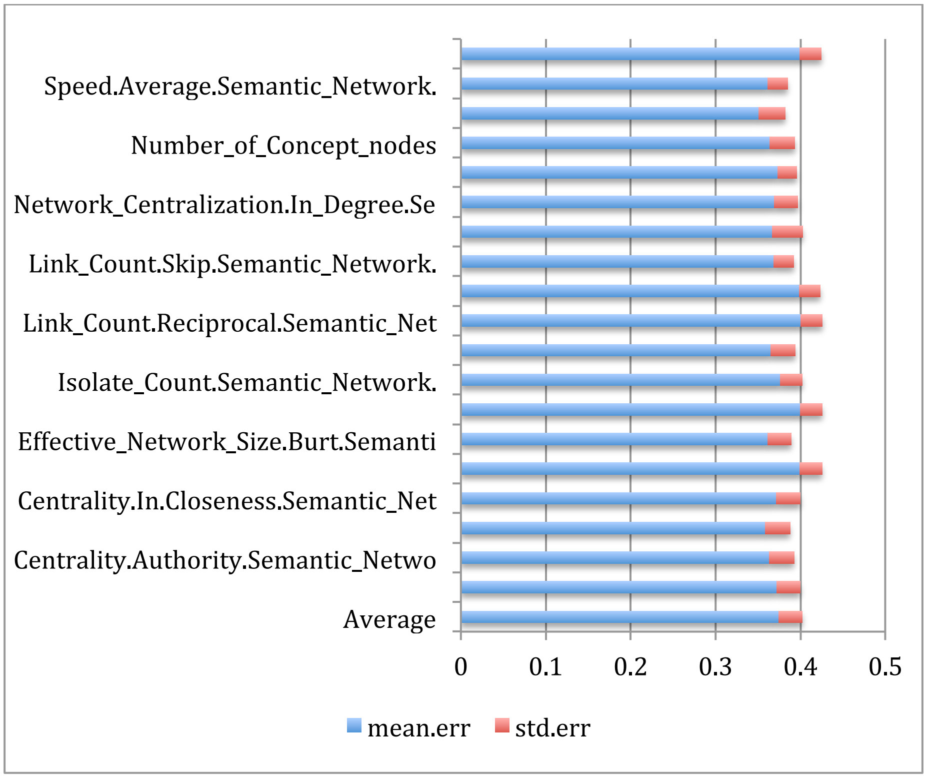


**Figure 65: For DV1 (ctr1): Effectiveness in Predicting movement outside of a range (Case 3 *discrete*). (Numerical scale represents Percent Error)**

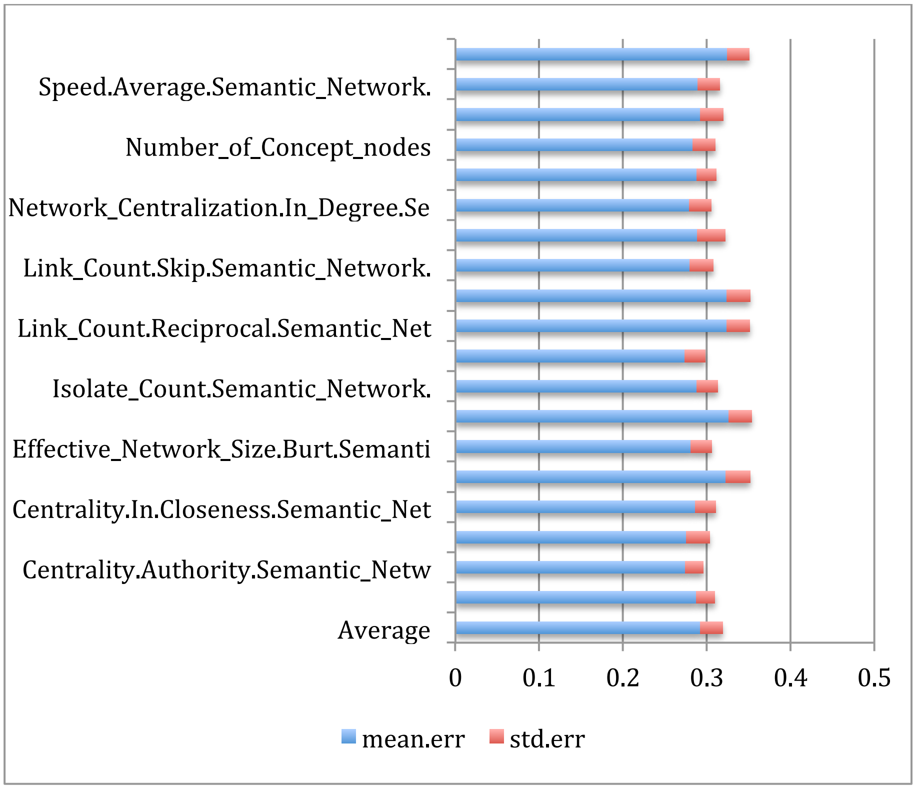


**Figure 66: For DV2 (ctr2): Effectiveness in Predicting movement outside of a range (Case 3 *discrete*). (Numerical scale represents Percent Error)**

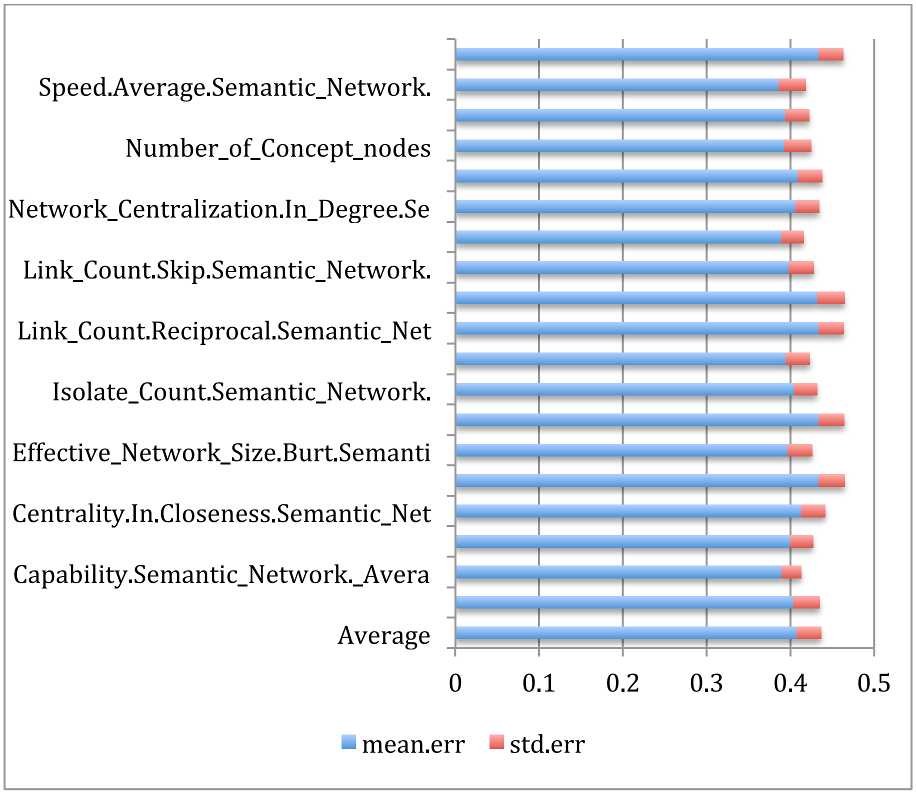


**Figure 67: For DV6 (X1_Month): Effectiveness in Predicting movement outside of a range (Case 3 *discrete*).**
**(Numerical scale represents Percent Error)**

The individual variables all appear to perform in line with each other relative to the averages of all the independent variables presented earlier. The visualization in the Figure below compares all the Exemplar Independent Variables.

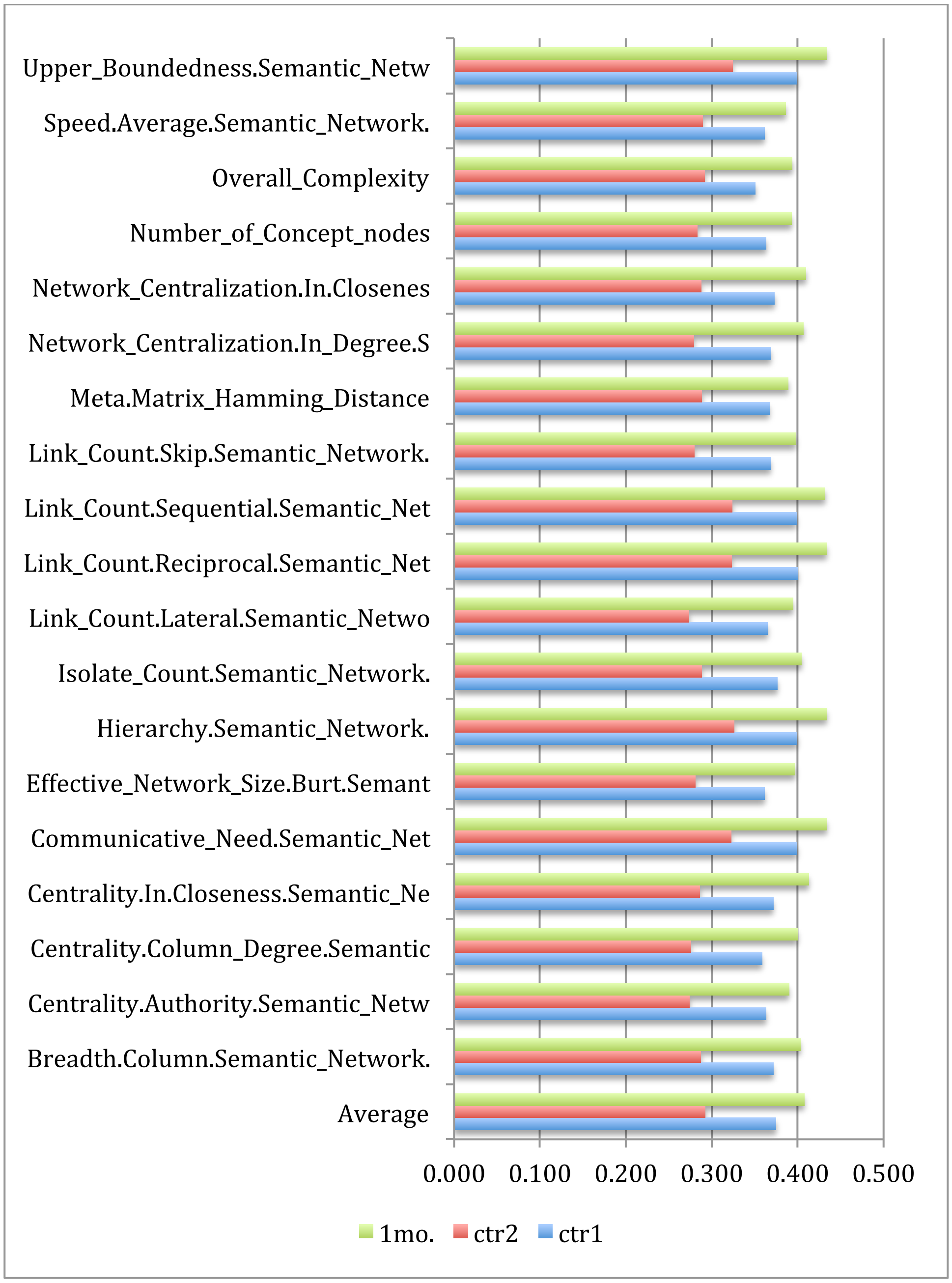


**Figure 68: For DV1, DV2, & DV6: Effectiveness in Predicting movement outside of a range (Case 3 *discrete*). (Numerical scale represents Percent Error)**

While the summarization of the data presented in the Figure above makes clear the effectiveness of each of Exemplar IV by themselves relative to the grouping used earlier, the Table 22 (below) makes clear the effectiveness of each IV.

| Dependent Variable | ctr1 | ctr2 | 1mo |
|---|---|---|---|
| Average | 0.375 | 0.292 | 0.408 |
| BreadthColumn.SemanticNetwork | 0.372 | 0.287 | 0.403 |
| CentralityAuthoritySemanticNetworkAverage | 0.363 | 0.274 | 0.390 |
| CentralityColumnDegreeSemanticNetworkAverage | 0.359 | 0.276 | 0.400 |
| CentralityInClosenessSemanticNetworkAverage | 0.372 | 0.286 | 0.413 |
| CommunicativeNeed.Semantic_Network | 0.399 | 0.322 | 0.434 |
| EffectiveNetworkSizeBurtSemanticNetworkAverage | 0.361 | 0.281 | 0.397 |
| HierarchySemanticNetwork | 0.399 | 0.326 | 0.434 |
| IsolateCountSemanticNetwork | 0.376 | 0.288 | 0.404 |
| LinkCountLateralSemanticNetwork | 0.365 | 0.274 | 0.394 |
| LinkCountReciprocalSemanticNetwork | 0.400 | 0.323 | 0.434 |
| LinkCountSequentialSemanticNetwork | 0.398 | 0.324 | 0.432 |
| LinkCountSkipSemanticNetwork | 0.368 | 0.280 | 0.398 |
| MetaMatrixHammingDistance | 0.367 | 0.289 | 0.389 |
| NetworkCentralizationInDegreeSemanticNetwork | 0.369 | 0.279 | 0.407 |
| NetworkCentralizationInClosenessSemanticNetwork | 0.373 | 0.288 | 0.410 |
| NumberOfConceptNodes | 0.363 | 0.283 | 0.393 |
| OverallComplexity | 0.351 | 0.292 | 0.394 |
| SpeedAverageSemanticNetwork | 0.361 | 0.289 | 0.386 |
| UpperBoundednessSemanticNetwork | 0.399 | 0.324 | 0.434 |

**Table 22: Exemplar Independent Variables percent error effectiveness in Predicting movement outside of a range (Case 3 *discrete*). The color of the cells are on a gradient from Green to Yellow to Red based on their mean error rate.**

Table 22 (above) makes clear that some of the Independent Variables are less effective than others in the errors.

While the experiments suggest that all of the Exemplar Independent Variables have some degree of effectiveness, we may consider the relative effectiveness of each Exemplar IV. With the following Hueristic, I consider the Exemplar IV effectiveness relative to its theoretical impact. These are summarized in Table 23 (below).

| HEURISTIC IN MODELING EFFECTIVESS BETWEEN THEORETICAL AND ACTUAL OUTCOMES | | |
|---|---|---|
| PREDICTION | MATCH TO THEORY | NO MATCH TO THEORY |
| MI+ (impact) | Better than Average | Average or Worse than Average |
| NR 0 (no evidence of a relationship) | Average | Better or Worse than Average |
| RC- (changes in DV are robust to changes in DV | Worse or Average | Better than Average |

**Table 23: Modeling Effectiveness Hueristic**

In Table 24 (below) a comparison is attempted between the average effectiveness of each Exemplar IV relative to its prediction. The second column is the average effectiveness of each IV among the DVs in the table above. In Judging the effectiveness The third column is the effectiveness relative to the average of the group. For this case 'same' is presented as those averages where the difference is less than 0.02.

| | Avg. | Relative to Avg. | Predic -tion (+,0,-) | Match to Theory? |
|---|---|---|---|---|
| Average | 0.358 | | | |
| BreadthColumn.SemanticNetwork | 0.354 | Better | *MI+* | + |
| CentralityAuthoritySemanticNetworkAverage | 0.342 | Better | *0* | - |
| CentralityColumnDegreeSemanticNetworkAverage | 0.345 | Better | *0* | - |
| CentralityInClosenessSemanticNetworkAverage | 0.357 | Same | *0* | + |
| CommunicativeNeed.Semantic_Network | 0.385 | Worse | *0* | - |
| EffectiveNetworkSizeBurtSemanticNetworkAverage | 0.346 | Same | - | + |
| HierarchySemanticNetwork | 0.386 | Worse | - | + |
| IsolateCountSemanticNetwork | 0.356 | Same | - | + |
| LinkCountLateralSemanticNetwork | 0.344 | Better | - | - |
| LinkCountReciprocalSemanticNetwork | 0.386 | Worse | - | + |
| LinkCountSequentialSemanticNetwork | 0.385 | Worse | - | + |
| LinkCountSkipSemanticNetwork | 0.349 | Better | - | - |
| MetaMatrixHammingDistance | 0.348 | Better | *MI+* | + |
| NetworkCentralizationInDegreeSemanticNetwork | 0.351 | Better | *MI+* | + |
| NetworkCentralizationInClosenessSemanticNetwork | 0.357 | Same | *MI+* | - |
| NumberOfConceptNodes | 0.346 | Better | - | - |
| OverallComplexity | 0.345 | Better | *MI* | + |
| SpeedAverageSemanticNetwork | 0.346 | Better | *0* | - |
| UpperBoundednessSemanticNetwork | 0.386 | Worse | - | + |

**Table 24: The peformance of network measurements in this study relative to Theory**

Four of the Independent Variables maybe worth further investigation based upon the results presented.

- BreadthColumSemanticNetwork;
- MetaMatrixHammingDistance;
- NetworkCentralizationInClosenessSemanticNetwork; and
- OverallComplexity.

These produced results in prediction that were better than average and this impact on the prediction was predicted in Theory (*see Section 9.3*). Mapping the above results into the table generated in *Section 9.2.2*, the experiments may support the claims in Table 25 (below, taken from *Section 9.2.2*).

| |
|---|
| **Increases in *Average Distance* may suggest a continuation of current trends to the extent that longer phrasing demonstrates optimism for the future or the current direction of prices.** |
| **Decreases in *Breadth* may suggest decreases in prices if decreases in phrase repetition demonstrates acceptance of current trends.** |

**Table 25: Evidence-supported claims on impact of Independent Variables in experiments on Public Policy Data Set (1).**

From these conclusions, Table 26 (below) is presented as a possible guide to speakers.

| Language Behavior | Suggesting |
|---|---|
| Lengthen Phrasing | Continued up-trend |
| Decrease Repetition | Continued down-trend |

**Table 26: Summary sugguestions for speakers communicating financial information**

# 10 Limitations, Contributions, & Future Work

## 10.1 Limitations and Challenges

The preceding chapters are concerned with a novel approach to analysis of public policy documents. The study has several major limitations. 1) the framework under consideration in this study has been applied to tightly constrained applications in financial markets; 2) the complexity that arises in both the number of steps required in the framework and the multiple choices available at each step make for a vastly expanded set of possible options unavailable for thorough exploration in any one inquiry; 3) there are necessarily rounding errors inherent in choosing a daily time scale; 4) Federal Reserve data and styles of communication are changing; 5) email requires substantial pre-processing that can introduce errors; 6) the framework consumes substantial time and computational power for larger data sets.

### 10.1.1 Domain limitation

This set of experiments in financial decision-making suggests some benefit if the researcher has a hypothesis regarding what relationship exists between variables. In this case, dependent variables were chosen that appropriately match the qualitative data. Other qualitative data may not have such an appropriate match.

### 10.1.2 Combinatorial complexity limitation

This framework is inherently a solution to a combinatorial problem. There is a precise sequence of steps necessary for the proper execution of this framework. Each step must be executed with precision and in the specified order. At scale, this may be easy to get wrong. Additionally, there are choices available at every stage. This study outlined many of these choices and the reasoned approach found to be appropriate for this particular study. An analysis of a scale that is infeasible to execute today for any one study may demonstrate among the vastly expanded set of alternatives that are available, a better combination of choices.

### 10.1.3 Temporality limitation

The data sets used are tightly constrained by time. Email time stamps occur at any time; reliability may be an issue. In this framework, choices must be made on the timescale of investigation and the reliability of this data. The communication from the Central Bank is known to effect financial markets so the data is released with care. Some of this communication occurs during the hours that financial markets are open. The question to round by the nearest second, minute, hour, or day maps against the possibility of also using all available securities in all affected markets. Argued earlier in this study is the reasoning behind the particular temporal selection. This rounding at one day may have produced less robust data than might otherwise be available.

### 10.1.4 Evolving FOMC limitation

As mentioned within this study, this the particular focus of Chapter 4 on the US Central Bank presents an issue with the evolving communication goals of that institution. Famously opaque since its inception earlier in the 20th century, the most recent Chairman demonstrably communicates with more frequency and is more direct about objectives. This may confound both the analysis and the interpretation of the results.

The financial crisis of 2008 also radically altered the policies of the Federal Reserve. Given the purpose of this study is the introduction and analysis of a framework, minimizations of externalities, or at least consistencies in them, are important. A the Federal Reserve, the changes in communication transparency and changes in policy expressed through quantitative easing complicate comparision in this study enough to exclude recent history (Krishnamurthy & Vissing-Jorgensen, 2011).

### 10.1.5 Large Dataset limitation

The email dataset under consideration in this study has been subject to a thorough analysis and therefore a thorough cleaning. This cleaning required such actions as the disambiguation of addresses. However, errors remained (such as inaccurate timestamps, email headers and email spam) that might still confound machine analysis in this framework.

### 10.1.6 Computational limitation

Innovations in computing, in math, and in the development of certain tools taking advantage of both has made this study feasible. However, large datasets overwhelm the current toolset. With 499,442 emails in the Enron Data Set under consideration in Chapter 5 and 6 required the manual processing of hundreds of separate file groupings as even highly resourced computers failed under the load of these tools matched to this dataset under this framework. Specifically, in this particular study, I used many tools at many stages of the execution of the framework. Some of these tools maximized the available CPU resources. Some maximized available memory. I experimented in stages with more powerful local computers, maximizing specifications for processing speed and then memory. I even tried multiple such physical machines running in parallel. I finally opened multiple remote instances of server-quality computers only to find a limited speed improvement in executing my framework. The current tools are not able to utilize the application of raw computing power above their historical limits fixed from their legacy applications. The tools currently do not take advantage of all the memory or processing power available to them. The tool versions in this study were not allowing for threaded processing. This is a limitation that raw computing power (of any magnitude) will not solve, but an evolution of the tools likely will. Commercial versions of these tools, unavailable to this study, may already address this issue.

## 10.2 Contributions

### 10.2.1 Theoretical Contribution

This work makes substantive contributions to theory in Computational Sociology, Public Policy, and applications of Public and Corporate Finance using methods from Machine Learning and Computational Linguistics. Most of the theoretical contributions are made in the process of expanding and interoperating dynamic network analysis, corporate finance, machine learning, and public policy. Where developments in network analysis have spread to domains as diverse as military intelligence (Carley, 2010), (Frantz & Carley, 2009) and healthcare (Effken et al., 2011), this work directly adds to the theoretical base in the domains of finance and public policy.

There may be no other domain where the purpose of text is to effect the movement of numbers. Practitioners speculate about the degree to which phrasing impacts markets. This has been extended to anecdotes on whole speeches. Commentary may even ascribe the movements in prices to some convenient co-occurrence. Beyond whole speeches, even the presence of entire conferences, such as the gathering of Central Bankers in Wyoming's Jackson Hole, can get this treatment. This study has enabled the reasoning using a repeatable and measurable framework for semantic networks. By measuring the impacts between words and numbers, we can theorize that the observed impacts represent different quantifiable characteristics in the semantic network. From these observations of changes in the semantic networks, analysis may now observe, in sufficiently large data sets, changes in discourse that quantifiably change facts outside of those conversations.

The study expands on the dynamic network analysis theory and changes in semantic networks. Analysis of public policy documents have previously used network maps of relationships or their own judgment to predict changes in behavior based on their own experience with the relationships under study or the language being used. Analysis may now look to answer questions related to organizational or even individual behavior with direct, quantifiable measures that are not reliant on anecdotes, human experience, or other indirect measures. By using this framework to assess an individual, team, or large organizations discourse, the study allowed for more measurable and in-depth analysis of future behavior.

Dynamic Network Analysis has developed various metrics that evaluate complex organization structures (Moon, 2008). However, the metrics application in these applications are limited to the assessment of relationships. By using this framework combined with relationship networks, theorists may expand on the notion of influence networks to answer questions about status and changes in network structure.

There is so far no developed framework that so clearly allows comparisons within Public policy and finance. There currently is no example of a complete application of social network analysis to commerce, economics, or finance. This interoperation enables new frontiers for theorists and these new frontiers enable more in-depth and nuanced assessment of behavior of policy-making bodies.

With this research, theorists have a tool to assist in further work to monitor the effects of speech on financial markets and other financial indicators. This represents a quantifiable framework to assess policy decisions. This work additionally points to how communication itself might improve in delivering the intended effects to the listener. With one stable framework, this study suggests how multiple speeches might be better measured with consistency.

The measures of semantic networks brought to bear in this framework can assist others in identifying and modeling the conversational constructs within these, and other, domains. Whereby changes in the qualitative data in this study can be theorized to move the framework's measures in particular directions (as discussed in detail within Chapter 4), the interrelatedness of the measures will help future theorists develop models for conversational and group outcomes.

Lastly, this work also tries to expand the horizon of existing methods. For example, in extracting a semantic network from Fed speeches, a new approach is available for improving the communication of central bank policy if not the policy itself. To the extent that these benefits improve central bank policy, they may additionally expand into other areas of public policy.

### 10.2.2 Technical Methods Contribution

The technical contributions involve an expansion in the body of work demonstrating how text analysis can be used for public policy and the role that semantic analysis can play in these analysis frameworks. With an exploration into corporate email, the work can effect the decision making of organizations that may expand beyond the financial. This work makes other contributions by increasing understanding of the effects of text and speech on behavior. This work adds to the understanding of cognitive and financial decision makers. In two studies, this work demonstrates that the framework improves the quality of the understanding of these decisions. As part of this work, I additionally add to the understanding of semantic analysis by which different types of analyses are done. Overall, this dissertation demonstrates that adaptive semantic analysis can be an appropriate research direction for improving decision making in the context of financial decisions and possibly beyond.

### 10.2.3 Empirical Contribution

By taking a dynamic network analysis approach and focusing on semantic networks relation to quantitative data, we can begin to distinguish the degree to which the language suggests future behavior. We can begin to make headway in reasoning about complexity and adaptation of future behavior. Using learning algorithms applied to a map of organization theory and the pragmatics branch of language studies, this work creates framework for working with complexity and adaptation in terms of properties of nodes and relations for identifying complex behavior and the reaction to speaking and writing. These templates enable a new set of capabilities linking previously separate techniques such as network analysis and semantic analysis with news analysis. The result is a technology that has more analytical power than any one method alone. I use the results from this combined technology to identify potential decision points for the purchase or sale of financial securities. Based on a combination of network analytics, organizational theory, mathematical finance, and pragmatics, this approach represented in this body of work makes solid contributions to the advancement of knowledge in these domains.

In Chapter 4, I applied the developed framework to the datasets of the U.S. Federal Reserve. Through this analysis and that comparison baseline Chapter 6, I generated a way of looking at the organization's behavior and the relationship of its participants that did not exist prior to my inquiry. The results end up being easy-to-review semantic network measurements.

In Chapter 5, I applied the developed framework to the datasets of the Enron email corpus. By applying this framework to that dataset and comparing to an often-used approach presented in Chapter 6, I  present simple tool that may be used in applications as dramatic as the detection of corporate malfeasance to the everyday measurement of employee engagement. These tools can be helpful in a wide range of concerns from corporate security to corporate treasury actions.

## 10.3 Future Work

Substantial future work is available using the framework presented here. These explorations may be categorized as:

1) data sets in the existing domains;

2) data sets in different, but similar domains;

3) changes (major and minor) to existing methodology;

4) forecasting and predictions; and

5) applications of new methods.

Questions to be asked: Have these relationships always held? What is the trend in these relationships? Another way of looking at the variable of time is to ask questions on the time decay of the relationships. Boukus (Boukus & Rosenberg, 2006) used the approach of LSA to look at the communications of the Federal Reserve using a time window decay. Future work could look at the nature of the relationships under different time windows. If this study looked at daily time movements, there are studies available of similar scope that might investigate shorter and shorter time windows. The experience evident in this study is that such an inquiry requires not only a substantial computational investment but also a high degree of human judgment in the cleaning of the available financial data. This study has found that financial data older than the initial sample captured in the earlier versions of this research to be often materially incomplete, suspicious in its consistency, or both. For example, the Fed Funds Futures contracts have are gaps in the available data prior to 2004.

This work is are also constrained from just a few qualitative sources. An analysis of news about the topic under consideration can have a dramatic effect on the data available for analysis. Obviously there is a additional source of data that is virtually infinite in the form of the quantitative financial data against which this can all be compared. This work has necessarily been tightly focused on the most direct link between the qualitative and quantitative data. Future work could look at the circumstances under which relationships gained or lost strength. For example us on the quantitative side, does an expansion or contraction (of the economy, of the stock market, of the bond market, of the currency market) affect the relationship? This path of exploration could help answering questions such as the degree to which big events in financial prices can be predicted (e.g., >20% moves). On the qualitative side, do the individual speakers matter or do their titles/positions matter more? Increasing openness as a communication policy of the Federal Reserve under Chairman Bernanke presents opportunities for further future work.

At the risk of confounding the research with foreign language translation, there are many questions on the nature of these relationships in the same domain across other languages or cultures. For example, this framework can help future work in exploring the degree to which information discrimination may exist in simultaneous translations of actions by the European Central Bank. News analysis would have similar questions in every financial domain. This work has been focused on a single language in one policy making body.

This framework uses a sequence of tools and methods that have been demonstrated as effective. However, implicit in the results are choices that have been outlined that could be changed and explored for their individual effectiveness: delete list, thesaurus, window size and nature, and manual versus automated tagging are a few of the constraints used that be changed in future work. The field of pragmatics may suggest a push toward more analysis of the phrasing in addition to, or a substitute for, the more automated approach suggested as a direction of the framework developed here.

Interesting integration with models of faster-changing environments is also possible (Belov et al., 2009). Synthesizing all future work might make the framework robust enough to express a generalized hypothesis by domain and algorithmically use all available data to determine the nature and circumstances of a relationship between text and the impact of the text.

This study represents a focused exploration of past data to discern patterns. I used the techniques of machine learning to identify patterns, but did not predict the future on different data. This forcasting represents a material expansion of the work not explored in this Dissertation, but available in future work. Ultimately, we might enjoy an expansion of the questions that can address the extent and circumstances under which we might predict the future.

LIST OF REFERENCES

Abelson, Robert, & Schank, Roger. (1995a). Knowledge and Memory: The Real Story. In Robert S. Wyer (Ed.), *Advances in social cognition: Knowledge and memory: the real story* (Vol. 8, pp. 1-86). Hillsdale, New Jersey: Lawrence Eribaum Associates, Inc.

Abelson, Robert, & Schank, Roger. (1995b). So All Knowledge Isn't Stories? In Robert S. Wyer (Ed.), *Advances in social cognition: Knowledge and memory: the real story* (Vol. 8, pp. 227-249). Hillsdale, New Jersey: Lawrence Eribaum Associates, Inc.

Aizawa, Akiko. (2000). *The feature quantity: an information theoretic perspective of Tfidf-like measures*. Paper presented at the Annual ACM conference on research and development in information retrieval, Athens, Greece.

Alexandersson, Jan, Maier, Elisabeth, & Reithinger, Norbert. (1995). *A robust and efficient three-layered dialogue component for a speech-to-speech translation system*. Paper presented at the Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics, Dublin, Ireland.

Algeo, J., & Pyles, T. (2009). *The origins and development of the English language* (Vol. 1): Wadsworth Publishing Company.

Andersson, Magnus, Overby, Lars Jul, & Sebestyén, Szabolcs. (2009). Which News Moves the Euro Area Bond Market? *German Economic Review, 10*(1), 1-31. doi: 10.1111/j.1468-0475.2008.00439.x

Baker, Malcolm, Ruback, Richard S., & Wurgler, Jeffrey. (2007). Behavioral Corporate Finance. In B. Espen Eckbo (Ed.), *Handbook of Corporate Finance: Empiral Corporate Finance* (Vol. 1, pp. 146-168): Elsevier.

Baker, Malcolm, & Stein, Jeremy. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets, 7*(3), 271-299.

Balduzzi, Pierluigi, Elton, Edwin J., & Green, T. Clifton. (2001). Economic News and Bond Prices: Evidence from the U.S. Treasury Market. *Journal of Financial and Quantitative Analysis, 36*(04), 523-543. doi: doi:10.2307/2676223

Baldwin, Breck, & Carpenter, Bob. (2012). What is Sentiment Analysis? *LingPipe*. Retrieved Jan 10, 2013, from http://alias-i.com/lingpipe/demos/tutorial/sentiment/read-me.html

Bales, Michael E., & Johnson, Stephen B. (2005). Graph theoretic modeling of large-scale semantic networks. *Journal of Biomedical Informatics*, 26.

Barberisa, Nicholas, Shleiferb, Andrei, & Vishny, Robert. (1997). A model of investor sentiment. *Journal of Financial Economics, 49*(3), 307-343. doi: 10.1016/S0304-405X(98)00027-0

Barney, Jay B. (1993). Looking inside for Competitive Advantage. *The Academy of Management Executive, 9*(4), 49-61.

Barth III, M, Remolona, E, & Wooldridge, P. (2001, October 2001). *Changes in market functioning and central bank policy: an overview of the issues*. Paper presented at the BIS Working Papers, Basel, Switzerland.

Basu, S. (1977). Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *The Journal of Finance, 32*(3), 663-682.

Beckage, Nicole M., Smith, Linda B., & Hills, Thomas. (2010). *Semantic Network Connectivity is Related to Vocabulary Growth Rate in Children*. Paper presented at the 32nd Annual Conference of the Cognitive Science Society, Portland, Oregon.

Belov, Nadya, Martin, Michael K., Patti, Jeff, Reminga, Jeff, Pawlowski, Angela, & Carley, Kathleen M. (2009). Dynamic Networks: Rapid Assessment of Changing Scenarios

Social Computing and Behavioral Modeling (pp. 1-9): Springer US.

Bernanke, Ben S., & Blinder, Alan S. (1988). Credit, Money, and Aggregate Demand: Is it money or credit, or both, or neither? *The American Economic Review, 78*(2, Papers and Proceedings of the One-Hundredth Annual Meeting of the American Economic Association ), 5.

Berry, A.W. (2011). The role of behavioral science in stock market investing. *Helium*, 1. Retrieved from Helium website: http://www.helium.com/items/2151778-how-behavioral-finance-influences-financial-decisions

Berry, Michael W., & Browne, Murray. (2005). Email Surveillance using Non-negative Matrix Factorization. *Computational & Mathematical Organizational Theory, 11*(249-264), 16.

Blei, David M, Ng, Andrew Y, & Jordan, Michael I. (2003). Latent Direchlet Allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Blinder, Alan S., Ehrmann , Michael, Fratzscher, Marcel, De Haan, Jakob, & Jansen, David-Jan. (2008). Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence.

Blitzer, J, Dredze, M, & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *Annual Meeting-Association For Computational Linguistics, 45*(1), 440.

BoardOfGovernorsOfTheFederalReserveSystem. (1996-2011). Speeches of Federal Reserve Officials. Retrieved 2011, 2011, from http://www.federalreserve.gov/newsevents/speech/2011speech.htm

Bodenhofer, Ulrich, Kothmeier, Andreas, & Hochreiter, Sepp. (2011). APCluster: an R package for affinity propagation clustering. *Bioinformatics, 27*(17), 2463-2464. doi: 10.1093/bioinformatics/btr406

Bodenhofer, Ulrich, Kothmeier, Andreas, & Palme, Johannes. (2013). APCluster: An R Package for Affinity Propagation Clustering. *Software Manual, 1.3.1*, 47. http://cran.r-project.org/web/packages/apcluster/vignettes/apcluster.pdf

Bomfim, Antulio N. (2003). Pre-announcement effects, news effects, and volatility: Monetary policy and the stock market. *Journal of Banking & Finance, 27*(1), 133-151.

Borch, Christian. (2006). The Exclusion of the Crowd: The Destiny of a Sociological Figure of the Irrational. *European Journal of Social Theory, 9*(1), 83-102. doi: 10.1177/1368431006060464

Borge-Holthoefer, Javier, & Arenas, Alex. (2010). Semantic Networks: Structure and Dynamics. *entropy, 12*, 39. doi: 10.3390/e12051264

Boukus, Ellyn, & Rosenberg, Joshua. (2006). The information content of FOMC minutes. New York, NY: Federal Reserve Bank of New York.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont.

Brown, Gregory W., & Cliff, Michael T. (2005). Investor Sentiment and Asset Valuation. *The Journal of Business, 78*(2). doi: 10.1086/427633

Burkhart, Lucas, & Fischer, Andreas M. (2008). Communicating Policy Options at the Zero Bound. *Working Paper Series*

Cantador, Iván, & Castells, Pablo. (2006). *Building Emergent Social Networks and Group Pro¯les by Semantic User Preference Clustering*. Paper presented at the Proceedings of the 2nd Workshop on Semantic Network Analysis, Budva, Montenegro.

Carley, Kathleen M. (2002). *Summary of Key Network Measures for Characterizing Organizational Architectures*. Institute for Software Research. Carnegie Mellon University. Retrieved from http://www.casos.cs.cmu.edu/publications/papers/MeasuresInfo.pdf

Carley, Kathleen M. (2010). *Network Science for Deterrence: Sheathing the Sword of the Terrorism/Nuclear Horseman*. Paper presented at the American Physical Society, Portland, OR. http://meetings.aps.org/link/BAPS.2010.MAR.B8.3

Carley, Kathleen M., Columbus, Dave, Bigrigg, Mike, & Kunkel, Frank. (2011). AutoMap User's Guide 2011 (School of Computer Science Institute for Software Research, Trans.). In Kathleen M. Carley (Ed.), *CASOS technical report*. Pittsburgh, PA: Carnegie Mellon University.

Carley, Kathleen M., Reminga, Jeff, Storrick, Jon, & Columbus, Dave. (2011). ORA User's Guide 2011 (Center for the Computational Analysis of Social and Organizational Systems Institute for Software Research, Trans.) *CASOS technical report*. Pittsburgh, PA: Carnegie Mellon University.

Carley, Kathleen M., & Kaufer, David S. (1993). Semantic Connectivity: An Approach for Analyzing Symbols in Semantic Networks. 183-213.

Carley, Kathleen M. (1993). Coding choices for textual analysis : A comparison of content analysis and map analysis. *Sociological Methodology*(23), 75-126.

Carley, Kathleen M., Diesner, Jana, & De Reno, Matt. (2006). AutoMap. Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Institute for Software Research International (ISRI), Center for Computational Analysis of Social and Organizational Systems.

Carley, Kathleen M., Reminga, Jeff, Storrick, Jon, & Columbus, Dave. (2011). ORA User's Guide 2011. Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Institute for Software Research, Center for the Computational Analysis of Social and Organizational Systems.

Carley, Kathleen M., & Skillicorn, David. (2005). Special Issue on Analyzing Large Scale Networks: The Enron Corpus. *Computational & Mathematical Organizational Theory*(11), 179-181.

Carpenter, Bob. (2004). *Phrasal Queries with LingPipe and Lucene: Ad Hoc Genomics Text Retrieval*. Paper presented at the 13th Annual Text Retrieval Conference, Gaithersburg, MD.

Carpenter, Bob. (2006). *Character language models for Chinese word segmentation and named entity recognition*. Paper presented at the Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing.

Carpenter, Bob. (2007). *LingPipe for 99.99% recall of gene mentions*. Paper presented at the Proceedings of the Second BioCreative Challenge Evaluation Workshop.

Chan, Louis K.C., & Lakonishok, Josef. (1994). Institutional trades and intraday stock price behavior. *Journal of Financial Economics, 33*(2), 26. doi: http://dx.doi.org/10.1016/0304-405X(93)90003-T

Chen, Shu-Heng, & Yeh, Chia-Hsuan. (2002). On the emergent properties of artificial stock markets: the efficient market hypothesis and the rational expectations hypothesis. *Journal of Economic Behavior & Organization, 49*(2), 217-239.

Cheng, Hui, Zakharov, Alexy, Dorado, Fernando, & Zhang, Di. (2009). KE-90.5100 Process monitoring. https://noppa.aalto.fi/noppa/kurssi/ke-90.5100/luennot/pls.pdf

Chermack, Thomas J. (2004). Improving decision-making with scenario planning. *Futures, 36*(3), 295-309.

Chuang, Wesley T., Tiyyagura, Asok, Yang, Jihoon, & Giuffrida, Giovanni. (2000). A fast algorithm for hierarchical text classification. In Y. Kambayashi, M. Mohania & A M. Tjoa (Eds.), *DaWaK 2000* (Vol. LNCS 1874, pp. 409-418): Springer-Verlag Berlin Heidelberg 2000.

Daimler, Eric. (2009). *Explorations in Semantic Networks* Paper presented at the CASOS Network Seminar, CASOS Network Seminar.

Danker, D, & Luecke, M. (2005). *Background on FOMC meeting minutes*. Paper presented at the Fed. Res. Bull. http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/fedred91&section=25

De Bondt, Werner F.M., & Thaler, Richard. (1984). Does the Stock Market Overreact? *The Journal of Finance, 40*(3), 13.

Deerwester, Scott, Dumais, Susan T., Furnas, George W., Landauer, Thomas K., & Harshman, Richard. (1988). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407. doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Dekker, David, Krackhardt, David, & Snijders, Tom A.B. (2005). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions.

Diesner, Jana, Frantz, Terrill L. , & Carley, Kathleen M. (2005). Communication Networks from the Enron Email Corpus “It's Always About the People. Enron is no Different”.

*Computational & Mathematical Organization Theory 11*(3), 27. doi: 10.1007/s10588-005-5377-0

Diesner, Jana, & Carley, Kathleen M. (2005). *Exploration of communication networks from the Enron email corpus*. Paper presented at the Proc. of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=15739570544496220308related:lBjmtQQwbtoJ

Doerfel, Marya L. (1998). What Constitutes Semantic Network Analysis? A Comparison of Research and Methodologie. *INSNA Connections, 21*(2), 6-26.

Effken, Judith A., Carley, Kathleen M., Gephart, Shelia, Verran, Joyce A., Bianchi, Denise, Reminga, Jeff, & Brewer, Barbara B. (2011). Using ORA to explore the relationship of nursing unit communication to patient safety and quality outcomes. *International Journal of Medical Informatics, 80*(7), 10. doi: doi:10.1016/j.ijmedinf.2011.03.015

Eichengreen, Barry. (2004). Financial Instability. In Bjørn Lomborg (Ed.), *Global Crises, Global Solutions* (pp. 648). Cambridge: Cambridge University press.

Eisenhardt, Kathleen M., & Zbaracki, Mark J. (1992). Strategic decision making. *Strategic Management Journal, 13*(S2), 17-37. doi: 10.1002/smj.4250130904

Esuli, Andrea, & Sebastiani, Fabrizio. (2006). *Sentiwordnet: A publicly available lexical resource for opinion mining*. Paper presented at the Proceedings of LREC.

Fallucchi, Francesca, & Zanzotto, Fabio Massimo. (2011). *Transitivity in Semantic Relation Learning*. Department of Computer Science. Università Telematica "G. Marconi" Department of Computer Science, University of Rome "Tor Vergata". Retrieved from http://www.sensocomune.it/portale/images/papers/zanzotto_transitivity_sc_2011.pdf

Fama, Eugene F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics, 49*(3), 283-306.

Feghali, Antoine Assaad. (1991). *A study of engineering college students' use of computer-based semantic networks in a computer programming language class*. (Ph.D. Dissertation), Purdue University, ProQuest. Retrieved from http://proquest.umi.com/pqdlink?did=746494641&Fmt=7&clientId =79356&RQT=309&VName=PQD (DAI-A 53/03)

Fisher, Kenneth L., & Statman, Meir. (2000). Investor Sentiment and Stock Returns. *Financial Analysts Journal, 56*(2), 16-23.

Fleming, M, & Remolona, E. (1997). *What moves the bond market?* Paper presented at the Economic Policy Review. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1016154

Fleming, M, & Remolona, E. (1999a). *Price formation and liquidity in the US Treasury market: The response to public information*. Paper presented at the The Journal of Finance. http://www3.interscience.wiley.com/journal/119065673/abstract

Fleming, M, & Remolona, E. (1999b). *What moves bond prices?* Paper presented at the The Journal of Portfolio Management. http://www.iijournals.com/doi/abs/10.3905/jpm.1999.319756

Fleming, M, & Remolona, E. (2001). *The term structure of announcement effects*. Paper presented at the EFA 2001 Barcelona. http://ssrn.com/abstract=166430 or doi:10.2139/ssrn.166430

Foner, Leonard Newton. (1986). *Political Artifacts and Personal Privacy: The Yenta Multi-Agent Distributed Matchmaking System*. (Doctor of Philosophy), Massachusetts Institute of Technology, Cambridge, MA.

Fox, Lynn S, Alvarez, Scott G, Braunstein, Sandra, Emerson, Marianne M, Johnson, Jennifer J, Johnson, Karen H, . . . Stockton, David J. (2005). *The Federal Reserve System Purposes & Functions*.

Frambach, Ruud T., & Schillewaert, Niels. (1999). Organizational Innovation Adoption: A Multi-Level Framework of Determinants and Opportunities for Future Research. University Park, PA: Institute for the Study of Business Markets, Penn State University.

Frantz, T.L., & Carley, K.M. (2009). *Information assurances and threat identification in networked organizations*. Paper presented at the Computational Intelligence for Security and Defense Applications, 2009, Ottawa, ON.

Freeman, Linton C. (1978). Centrality in Social Networks I: Conceptual Clarification. *Social Networks, 1*(1), 215-239.

Frendreis, J, & Tatalovich, R. (2000). *Accuracy and bias in macroeconomic forecasting by the administration, the CBO, and the Federal Reserve Board*. Paper presented at the Polity. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed

&cmd=Retrieve&dopt=AbstractPlus&list_uids=3376490972732180056related:WMYDeaa02y4J

Frey, Brendan J., & Dueck, Delbert. (2007). Clustering by passing messages between data points. *Science, 315*(5814), 972-976. doi: 10.1126

Friedkin, Noah E. (1981). The development of structure in random networks: an analysis of the effects of increasing network density on five measures of structure. *Social Networks, 3*(1), 41-52. doi: http://dx.doi.org/10.1016/0378-8733(81)90004-6

Gennheimer, Heinrich. (2002). *Model Risk in Copula Based Default Pricing Models*. Working Paper Series. National Centre of Competence in Research Financial Valuation and Risk Management. Zurich.

Geroski, P.A. (2000). Models of technology diffusion. *Research Policy, 29*, 603-625.

Gershon, Nahum, & Page, Ward. (2001). What storytelling can do for information visualization. *Commun. ACM, 44*(8), 31-37. doi: 10.1145/381641.381653

Gervasio, Melinda. (2004). *Enron email corpus* [email]. Retrieved from: http://www-2.cs.cmu.edu/~enron/.

Godbole, N, Srinivasaiah, M, & Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.

Gong, F, & Remolona, E. (1996). *A three-factor econometric model of the US term structure*. New York: Federal Reserve Retrieved from http://www.nyfedeconomists.org/research/staff_reports/research_papers/9619.pdf.

Green, T. Clifton. (2004). Economic News and the Impact of Trading on Bond Prices. *The Journal of Finance, 59*(3), 1201-1234. doi: 10.1111/j.1540-6261.2004.00660.x

Havrilesky, T, & Schweitzer, R. (1990). A theory of FOMC dissent voting with evidence from the time series. In Thomas Mayer (Ed.), *The Political Economy of American Monetary Policy*. New York: Cambridge University Press.

Heikkilä, Jussi. (2002). From supply to demand chain management: efficiency and customer satisfaction. *Journal of Operations Management, 20*(6), 747-767.

Heiner, Ronald A. (1989). The origin of predictable dynamic behavior. *Journal of Economic Behavior & Organization, 12*(2), 233-257.

Herring, S.C. (2012). Grammar and electronic communication. *The Encyclopedia of Applied Linguistics*.

Hofmann, T. (1999). *Probabilistic latent semantic analysis*. Paper presented at the Proceedings of Uncertainty in Artificial Intelligence, UAI'99, Stockholm, Sweden.

Hoser, Bettina, Hotho, Andreas, Jäschke, Robert, Schmitz, Christoph, & Stumme, Gerd. (2005). *Semantic Network Analysis of Ontologies*. Paper presented at the Proceedings of the European Semantic Web.

Hubert, Lawrence. (1987). Assignment Methods in Combinatorial Data Analysis. In INC. Marcel Dekker (Ed.), *Statistics: Textbooks and Monographs* (Vol. 73). New York and Basel.

Hwang, Soosung, & Salmon, Mark. (2008). *Sentiment and Beta Herding*. Warwick Business School, University of Warwick, Coventry. Coventry.

Iqbal, Zahid, & Shetty, Shekar. (1995). Layoffs, Stock Price, And Financial Condition Of The Firm. *Journal of Applied Business Research, 11*(2).

Issing, O. (2005). *Communication, transparency, accountability: Monetary policy in the twenty-first century*. Retrieved from https://research.stlouisfed.org/publications/review/05/03/part1/Issing.pdf.

Jing, Li-Ping, Huang, Hou-Kuan, & Shi, Hong-Bo. (2003). *Improved feature selection approach TFIDF in text mining*. Paper presented at the International Conference on Machine Learning and Cybernetics, 2002. Proceedings. 2002

Joachims, Thorsten. (1999, May 21). *Transductive Inference for Text Classification using Support Vector Machines*.

Jones, Charles M., Lamont, Owen, & Lumsdaine, Robin. (1998). Macroeconomic news and bond market volatility. *Journal of Financial Economics, 47*(3), 315-337.

Judge, T. (2012). A Revolution in American English: The Transformation of American Spoken Language from the 1940s to the 1970s.

Keila, P.S., & Skillicorn, D.B. (2005). Structure in the Enron Email Dataset. *Computational & Mathematical Organization Theory*(11), 183–199.

Kenett, Yoed N., Kenett, Dror Y., Ben-Jacob, Eshel, & Faust, Miriam. (2011). Global and Local Features of Semantic Networks: Evidence from the Hebrew Mental Lexicon. *PLos ONE, 6*(8), 14. doi: 10.1371/journal.pone.0023912

Kilduff, Martin, & Tsai, Wenpin. (2005). *Social Networks and Organizations*. London, England: Sage Publications.

Klibanoff, Peter, Lamont, Owen, & Wizman, Thierry A. (1998). Investor Reaction to Salient News in Close-End Country Funds. *The Journal of Finance, 53*(2), 27.

Klimt, Bryan, & Yang, Yiming. (2004). *Introducing the Enron Corpus*. Language Technology Institute. Carnegie Mellon University.

Klimt, Bryan, & Yiming, Yang. (2004). *The enron corpus: A new dataset for email classification research*: Springer.

Konchady, Manu. (2008). *Building Search Applications: Lucene, LingPipe, and Gate*: Musée d'art contemporain de Montréal.

Krackhardt, David. (1987). QAP Partialling as a Test of Spuriousness. *Social Networks*(9), 171-186.

Krishnamurthy, Arvind, & Vissing-Jorgensen, Annette. (2011). The effects of quantitative easing on interest rates: channels and implications for policy: National Bureau of Economic Research.

Krueger, Joel T, & Kuttner, Kenneth N. (1996). The fed funds futures rate as a predictor of Federal Reserve policy. *Journal of Futures Markets, 16*(8), 865-879.

Landwehr, Peter. (2012). NUBBI: An introduction and a non-examination of Sudanese Newspapers.

Laurance, Edward J. (1990). Events data and policy analysis. *Policy Sciences, 23*(2), 111-132. doi: 10.1007/bf00175597

Lehmann, Fritz. (1992). Semantic networks. *Computers & Mathematics with Applications, 23*(2-5), 1-50. doi: 10.1016/0898-1221(92)90135-5

LeRoy, Stephen F. (2010). *Efficient Market Hypothesis*: John Wiley & Sons, Ltd.

Li, Deyuan, & Peng, Liang. (2009). Goodness-of-fit test for tail copulas modeled by elliptical copulas. *Statistics & Probability Letters, 79*(8), 1097-1104.

Li, Y.H., & Jain, A.K. (1998). Classification of Text Documents. *The Computer Journal, 41*(8), 537-546.

Liaw, Andy, & Wiener, Matthew. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22.

Lucca, David O, & Trebbi, F. (2009, Dec 31). *Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements*. Paper presented at the NBER working paper.

Luce, Laurent. (2012). Twitter sentiment analysis using Python and NLTK [Describes the implementation of sentiment analysis of tweets using Python and the natural language toolkit NLTK. Also describes the internals of NLTK related to this implementation.]. Retrieved from http://www.laurentluce.com/posts/twitter-sentiment-analysis-using-python-and-nltk/

Luss, Ronny, & d'Aspremont, A. (2008, Nov 15). *Predicting Abnormal Returns From News Using Text Classification*.

Malkiel, Burton G. (2003). The Efficient Market Hypothesis and Its Critics. *The Journal of Economic Perspectives, 17*(1 (Winter, 2003)), 59-82.

Mani, I, & Bloedorn, E. (1999). *Summarizing similarities and differences among related documents*. Paper presented at the Information Retrieval.

Mantel, Nathan. (1967). The Detection of Disease Clustering and a Generalized Regression Approech. *Cancer Research*(27), 209-220.

McKenna, Katelyn, & Seidman, Gwendolyn. (2005). You, me, and we: interpersonal processes in electronic groups. In Yair Amichai-Hamburger (Ed.), *The Social Net* (pp. 27). Oxford: Oxford University Press.

Meade, Ellen. (2002). *Central Bank Decisions: How Does One Good Head Compare with a Rule?* Centre for Economic Performance, London School of Economics. London.

Meltzer, Allan H. (2000). Report on the International Financial Institution Advisory Commission. In Allan H. Meltzer (Ed.), *Global Backlash* (pp. 29-33). Washington, D.C.: Government Printing Office.

Meyer, D., Dimitriadou, Evgenia, Hornik, Kurt, Weingessel, Andreas, & Leisch, Friedrich. (2004). Support Vector Machines: The interface to libsvm in Package e1071. cran.r-project.org. Retrieved from http://cran.r-project.org/web/packages/e1071/index.html

Mikosch, Thomas. (2006). Copulas: Tales and facts. *Extremes, 9*(1), 3-20. doi: 10.1007/s10687-006-0015-x

Monge, P.R., & Eisenberg, E.M. (1987). Emergent communication networks. In F . Jablin, L. L . Putnam, K . Roberts & L . Porter (Eds.), *Handbook of organizational communication*. Newbury Park, CA: Sage.

Moon, Il-Chul. (2008). *Destabilization of Adversarial Organizations with Strategic Interventions*. (Ph.D. PhD Dissertation), Carnegie Mellon University. (CMU-ISR-08-124)

Morgan, J, & Kogan, S. (2010). Securities auctions under moral hazard: An experimental study. *Review of Finance, 14*(3), 477-520. doi: 10.1093/rof/rfp008

Mullainathan, Sendhil, & Thaler, Richard H. (2000). *Behavioral Economics* (1st ed.). International Encyclopedia of SocialSciences: Pergamon Press.

Mullen, Tony, & Malouf, Robert. (2006). *A preliminary investigation into sentiment analysis of informal politcal discourse*. Paper presented at the AAAI 2006 Spring symposium on computational approaches to Analyzing Weblogs.

Muresan, Smaranda, Tzoukermann, Evelyne, & Klavans, Judith L. (2001). *Combining Linguistic and Machine Learning Techniques for Email Summarization*. Journal Article. Columbia University; Lucent Technologies.

Nasukawa, T, & Yi, J. (2003). *Sentiment analysis: Capturing favorability using natural language processing*. Paper presented at the Proceedings of the 2nd international conference on Knowledge capture.

Nimick, Francis. (2011). AutoMap Introduction. *AutoMap Introduction, 1.0*, 9. http://www.westpoint.edu/nsc/siteassets/sitepages/Publications/automap.pdf

Olsen, Robert A. (1996). Implications of Herding Behavior for Earnings Estimation, Risk Assessment, and Stock Returns. *Financial Analysts Journal, 52*(4), 37-41.

Pang, B, Lee, L, & Vaithyanathan, S. (2002). *Thumbs up?: sentiment classification using machine learning techniques*. Paper presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.

Pichl, Lukáš. (2010). Semantic Network Closure Structures in Dual Translation of Stochastic Languages. *Databases in Networked Information Systems, 5999*, 210-224. doi: 10.1007/978-3-642-12038-1_14

Poon, Ser-Huang, Rockinger, Michael, & Tawn, Jonathan. (2004). Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications. *The Review of Financial Studies, The Society of Financial Studies, 17*(2), 581-610.

Priebe, Carey E., Conroy, John M., Marchette, David J., & Park, Youngser. (2005). Scan Statistics on Enron Graphs. *Computational & Mathematical Organization Theory*(11), 229-247.

Purpura, Stephen, & Hillard, Dustin. (2006). *Automated classification of congressional legislation*. Paper presented at the Proceedings of the 2006 international conference on Digital government research, San Diego, California.

Qian, Rong, Zhang, Wei, & Yang, Bingni. (2006). *Detect community structure from the enron email corpus based on link mining*. Paper presented at the Sixth International Conference on Intelligent Systems Design and Applications, ISDA'06. .

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106. doi: 10.1007/bf00116251

Reeves, R, & Sawicki, M. (2007, Dec 31). *Do financial markets react to Bank of England communication?* Paper presented at the European Journal of Political Economy.

Reinhart, V, & Sack, B. (2006, Dec 31). *Grading the Federal Open Market Committee's Communications*. Paper presented at the AEA Annual Meetings in Boston.

Reminga, Jeffrey, & Carley, Kathleen M. (2003). Measures for ORA (Organizational Risk Analyzer) (pp. 19): Computational and Organizational Systems at Carnegie Mellon University.

Rice, R.E., & Danowski, J.A. (1993). Is it really just a like a fancy answering machine? Comparing semantic networks of different types of voicemail users . *Journal of Business Communication*(30), 369-397.

Richmond, Federal Reserve Bank of. (2012). Retrieved May 7, 2012, 2012, from http://www.richmondfed.org/faqs/fomc/

Ripley, Brian, Therneau, Terry, & Atkinson, Beth. (2013). rpart: Recursive Partitioning (Version 4.1-1). cran.r-project.org. Retrieved from http://cran.r-project.org/web/packages/rpart/index.html

Ritter, Jay R., Constantinides, G.M., Harris, M., & Stulz, R. M. (2003). Chapter 5 Investment banking and securities issuance *Handbook of the Economics of Finance* (Vol. Volume 1, Part 1, pp. 255-306): Elsevier.

Robertson, JC, & Thornton, DL. (1997). *Using federal funds futures rates to predict Federal Reserve actions*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=6770061116740629237related:9XICY7cW9F0J.

Rogers, Timothy T. (2006). Semantic Memory: Computational Models *Encyclopedia of Cognitive Science*: John Wiley & Sons, Ltd.

Rosa, C. (2007). *Providing Content to ECB Announcements*. Paper presented at the Rivista Internazionale di Scienze Sociali.

Rosé, Carolyn. (2012). Automatic Conversation Analysis. Retrieved 2013, from http://www.cs.cmu.edu/~cprose/TextMining.html

Rubin, Rebecca B., & Morreale, Sherwyn P. (1996). Setting expectations for speech communication and listening. *New Directions for Higher Education, 1996*(96), 19-29. doi: 10.1002/he.36919969604

Sabater, Carmen Pérez, Turney, Ed, & Fleta, Begoña Montero. (2008). Orality and literacy, formality and informality in email communication. *Ibérica, 15*, 71-88.

Sakellariou, Argiris, Sanoudou, Despina, & Spyrou, George. (2012). Combining multiple hypothesis testing and affinity propagation clustering leads to accurate, robust and sample size independent classification on gene expression data. *BMC Bioinformatics, 13*. doi: 10.1186/1471-2105-13-270

Schank, Roger C. (1990). *Tell Me A Story: Narrative and Intelligence*. Evanston, Illinois: Northwestern University Press.

Shleifer, Andrei, & Summers, Lawrence H. (1990). The Noise Trader Approach to Finance. *The Journal of Economic Perspectives, 4*(2), 19-33.

Singer, Jerome E., & Radloff, Lenore S. . (1963). Renegades, heretics, and Changes in Sentiment. *Sociometry, 26*(2), 178-189.

Soroka, Stuart N. (2006). Good News and Bad News: Asymmetric Responses to Economic Information. *Journal of Politics, 68*(2), 372-385. doi: 10.1111/j.1468-2508.2006.00413.x

Steyvers, Mark, & Tenenbaum, Josh. (2011). *The Large-Scale Structure of Semantic Networks*. Department of Psychology. Stanford University. Retrieved from http://psiexp.ss.uci.edu/research/papers/smallworlds.pdf

Tirilly, Pierre, Claveau, Vincent, & Gros, Patrick. (2008). *Language modeling for bag-of-visual words image categorization*. Paper presented at the Proceedings of the 2008 international conference on Content-based image and video retrieval, Niagara Falls, Canada.

Uijlings, J. R. R., Smeulders, A. W. M., & Scha, R. J. H. (2009). *Real-time bag of words, approximately*. Paper presented at the Proceedings of the ACM International Conference on Image and Video Retrieval, Santorini, Fira, Greece.

Veronesi, P. (1999). Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *Review of Financial Studies, 12*(5), 975-1007. doi: 10.1093/rfs/12.5.975

Walther, J.B. (2012). Interaction through technological lenses: Computer-mediated communication and language. *Journal of Language and Social Psychology*.

Wang, YC, Joshi, M, & Rosé, CP. (2007). *A feature based approach to leveraging context for classifying newsgroup style discussion segments*. Paper presented at the Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions.

Wasserman, Stanley, & Faust, Katherine. (1997). *Social Network Analysis: Methods and Applications*. Cambridge, United Kingdom: Press Syndicate of the University of Cambridge.

White, Kerr L., Williams, T. Franklin, & Greenberg, Bernard G. (1961). The Ecology of Medical Care. *New England Journal of Medicine, 265*(18), 885-892. doi: doi:10.1056/NEJM196111022651805

Wilks, Yorick, Catizone, Roberta, Worgan, Simon, & Turunen, Markku. (2011). Review: Some background on dialogue management and conversational speech for dialogue systems. *Comput. Speech Lang., 25*(2), 128-139. doi: 10.1016/j.csl.2010.03.001

Wilson, Theresa, Wiebe, Janyce, & Hoffmann, Paul. (2005). *Recognizing contextual polarity in phrase-level sentiment analysis*. Paper presented at the Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada.

Witten, Ian H., & Frank, Eibe. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition* (Second ed.): Elsevier.

Woods, W.A. (1975). *What's in a Link: Foundations for Semantic Networks*. Cambridge, MA: Bolt Beranek and Newman Inc.

# APPENDICES

I. Public Policy Data example: FOMC member speeches 2006-2007
II. Public Policy Data example: Sample Speech, full text
III. Delete list used
IV. Thesaurus used
V. Financial Data example: contract for Fed Funds Futures with December 2007 expiration

# Appendix I: Schedule of Fed member speeches in 2006-2007

2006April03_Kroszner
2006April05_Bernanke
2006April06_Kroszner
2006April10_Bies
2006April10_Olson
2006April13_Kohn
2006April13_Olson
2006April17_Ferguson
2006April20_Bernanke
2006April27_Kohn
2006April28_Bies
2006Aug25_Bernanke
2006Aug31_Bernanke
2006Dec01_Bernanke
2006Dec01_kohn
2006Dec15_Berna
2006Feb02_Bies
2006Feb06_Bernanke
2006Feb23_Ferguson
2006Feb24_Bernanke
2006Feb24_Ferguson
2006Jan18_Bies
2006July04_Bies
2006July06_Kohn
2006July18_Warsh
2006June05_Bernanke
2006June06_Bies
2006June09_Bernanke
2006June12_Bernanke
2006June12_Bies
2006June12_Olson
2006June13_Bernanke
2006June14_Bies
2006June15_Bernanke
2006June16_Kroszner
2006Mar03_Ferguson
2006Mar08_Bernanke

2006Mar10_Ferguson
2006Mar13_Olson
2006Mar16_Kohn
2006Mar20_Bernanke
2006Mar29_Bies
2006Mar31_Bies
2006Mar31_Ferguson
2006May03_Bernanke
2006May04_Bies
2006May11_Kohn
2006May16_Bernanke
2006May16_Bies
2006May16_Olson
2006May18_Bernanke
2006May18_Kohn
2006May24_Kroszner
2006May25_Olson
2006Nov01_Bernanke
2006Nov02_Bies
2006Nov03_Kohn
2006Nov10_Bernanke
2006Nov16_Kroszner
2006Nov21_Warsh
2006Nov28_Bernanke
2006Nov30_Schmidt Bies
2006Oct04_Bernanke
2006Oct04_Kohn
2006Oct11_Bies
2006Oct12_Mishkin
2006Oct16_Bernanke
2006Oct17_Bies
2006Sep01_Bernanke
2006Sep11_Kohn
2006Sep27_Kroszner
2007April10_Mishkin
2007April11_Bernanke
2007April20_Mishkin
2007April25_Bernanke
2007April26_Mishkin
2007Aug01_Kroszner

2007Aug31_Bernanke
2007Feb06_Bernanke
2007Feb21_L. Kohn
2007Feb26_Schmidt Bies
2007Jan05_Bernanke
2007Jan08_L. Kohn
2007Jan11_Schmidt Bies
2007Jan17_S. Mishkin
2007Jan18_Schmidt Bies
2007July10_Bernanke
2007July12_Kroszner
2007June01_Kroszner
2007June05_Warsh
2007June05(2)_Bernanke
2007June14_Kroszner
2007June15_Bernanke
2007June23_Mishkin
2007Mar02_Bernanke
2007Mar05_ Warsh
2007Mar05_Kroszner
2007Mar06_Bernanke
2007Mar09_Kroszner
2007Mar09_L. Kohn
2007Mar12_Kroszner
2007Mar22_Kohn
2007Mar22_Kroszner
2007Mar23_Mishkin
2007Mar30_Bernanke
2007May01_Bernanke
2007May10_Kroszner
2007May15_Kroszner
2007May15(2)_Bernanke
2007May16_Kohn
2007May16(2)_Kroszner
2007May17_Bernanke
2007May22_Bernanke
2007May23_Kroszner
2007May24_Mishkin
2007Nov05_Kroszner
2007Nov05_Mishkin

2007Nov06_Bernanke
2007Nov07_Warsh
2007Nov13_Kroszner
2007Nov14_Bernanke
2007Nov16_Kroszner
2007Nov28_Kohn
2007Nov29_Bernanke
2007Nov29_Mishkin
2007Nov30_Kroszner
2007Oct05_Kohn
2007Oct05_Warsh
2007Oct11_Kroszner
2007Oct12_Bernanke
2007Oct12_Kohn
2007Oct15_Bernanke
2007Oct19_Bernanke
2007Oct20_Mishkin
2007Oct22_Kroszner
2007Oct26_Mishkin
2007Sep01
2007Sep06_Kroszne
2007Sep10_Mishkin
2007Sep11_Bernanke
2007Sep21_Kohn
2007Sep21_Warsh
2007Sep24_Bernanke
2007Sep27_Mishkin
2007Sep28_Mishkin

## Appendix II: Sample Speech full text

Chairman Ben S. Bernanke
At the Greenlining Institute's Thirteenth Annual Economic Development Summit, Los Angeles, California
(via satellite)
April 20, 2006

### By the Numbers: Data and Measurement in Community Economic Development

I would like to thank Greenlining for the opportunity to participate in today's conference. In my time at the Federal Reserve, I have had a number of opportunities to meet with community economic development leaders to discuss issues of mutual concern and learn about the valuable role that community development organizations play in economically distressed areas across the country. I have been particularly impressed, and heartened, by the increasingly high degree of professionalism in the field. In this area, as in social policy generally, good intentions are not enough. Successful community development requires knowledge--knowledge about the particular community in question and about what has worked in similar communities in the past--and community development organizations are working assiduously and with sophisticated tools to help develop that knowledge.

Of course, knowledge bearing on community economic development has both qualitative and quantitative aspects, and it can be gained through diverse channels, from talking to people in a neighborhood to performing a regression analysis. Today, I will focus on the progress that is being made on the quantitative side--in particular, the remarkable strides that have been made in developing and analyzing social and economic data at the community level. The information that can be extracted from detailed data profiles of individual communities supports economic development in several distinct ways. First, by making companies, entrepreneurs, and investors aware of new opportunities and by promoting competition in underserved areas, such information helps put market forces in the service of community development. Second, both government policymakers and community development organizations need the reality check that only hard data can provide. To know whether our policies and programs are delivering the desired results, we need to be able to measure inputs and outcomes, program by program and community by community. Better information increases accountability and promotes good governance in both the public and the nonprofit sectors. Third, the increased availability of community-level data facilitates independent research, which is vital to informing the public policy debate and to developing further community development efforts, both public and private.

Historically, government agencies have been the source of the most-comprehensive social and economic data bearing on community development. An important example is the data collected by the Federal Reserve under the Home Mortgage Disclosure Act (HMDA). The HMDA data set provides extensive information on home mortgage applications to virtually all U.S. lenders, including approval rates, the socioeconomic characteristics of applicants, and most recently, mortgage pricing information. As all good social scientists know, the data never "speak for themselves," and the HMDA information, like any data set, must be interpreted with care and insight. Still, for nearly three decades, the

HMDA data have provided valuable information about mortgage lending patterns, contributed to significant changes in mortgage credit practices, informed regulatory policies, and supported fair-lending enforcement.

Although government agencies continue to be an important source of data on community development, data collection and data analysis in this area is increasingly becoming the province of the private and nonprofit sectors, notably including community development organizations themselves. In recent years, we have seen a series of data-collection initiatives outside the public sector, with objectives that include the improvement of development strategies, the identification of new opportunities, the quantification of risk, and the exertion of influence on the direction of public policy. Many of these efforts have already had significant payoffs.

In the rest of my remarks, I will discuss some specific ways data and quantitative measurement have been used in community development. To be clear, I do not believe that all aspects of economic development can or should be quantified; and, as I have already noted, the data never speak for themselves but must be interpreted with care. Still, improving the measurement of inputs and outcomes is critical to better development policy. In this regard, it is interesting to observe that we have seen some convergence between best practices in community economic development and in economic development policy at the international level. I will conclude by noting a few of those parallels and their implications.

Discovering Market Potential

Good data support community growth and development by helping to identify previously unrecognized market opportunities. Free markets can be a powerful source of economic development, but markets work less effectively when information about potential opportunities is absent or costly for private actors to obtain. Several noteworthy initiatives have helped to provide better information about the economic potential of lower-income and underserved communities. For example, the Local Initiative Support Corporation's (LISC) MetroEdge initiative seeks to demonstrate the market potential of diverse communities through customized data analyses of each community's demographics and buying power. Such analysis can provide investors with a different perspective when they assess a neighborhood's viability for investment. In one instance, a national home-improvement retailer used MetroEdge data as the basis for its decision to establish a store in inner-city Chicago, even though the retailer's own site-selection model presented discouraging indications of profit potential for that neighborhood. With access to new market data, the company could justify its investment in the community, and sales performance was triple what was expected within the first six months of operation. 1

Similarly, Social Compact's Neighborhood Market DrillDown methodology uses a multilayered research process to provide profiles of the market potential of high-density, lower-income communities. This approach focuses on business indicators--buying power, market size, unmet needs, and market risks--rather than on the deficiency statistics typically used to describe inner-city neighborhoods, such as rates of poverty, crime, and overcrowding. Social Compact, a coalition of business leaders, has applied its DrillDown approach to 101 neighborhoods over the past five years, beginning with Chicago neighborhoods and, most recently, in Santa Ana, California. By tapping existing public records and conducting intensive economic and demographic surveys, the DrillDown analyses of these 101 neighborhoods in eight cities have, in the aggregate, revealed additional income and buying power averaging nearly $6,000 per household, which is not captured by traditional sources of community-level data.2 Such information may attract private-sector investors to areas that had once been deemed untenable for investment. For example, following Social Compact's study of neighborhoods in Jacksonville, Florida, a developer announced plans to invest $45 million in a multi-use entertainment complex there. A DrillDown study in inner-city Houston revealed a population that was 25 percent larger than Census estimates, resulting in the redevelopment of a 750,000 square foot retail center that brought 2,000 jobs to a neighborhood that had not had new construction in fifty years. This shopping center is now one of the busiest retail centers in the city. 3

Work to improve the measurement of market potential in inner-city communities is continuing. In one such project, Social Compact and the Brookings Institution's Urban Markets Initiative group are collaborating in reviewing methods for measuring the size and composition of economies in urban areas around the world. The objectives of the review are to develop new tools for measuring economic activity at the local level and to identify areas for future research.

Informing Investors in Community Development

The growth and maturation of community development financial institutions (CDFIs) provide another impetus for data development and analysis at the community level. CDFIs are private-sector financial intermediaries with community development as their primary mission. Like banks and other more-conventional financial intermediaries, CDFIs are in the business of attracting funds and putting those funds to work in productive ways. Also like conventional intermediaries, CDFIs depend heavily on the production of accurate information both to guide investment decisions and to provide a basis for attracting new funding. It is difficult to overstate the importance of adequate and accurate information for attracting capital. Managers of pools of capital have many choices, and they tend to be extremely wary when they cannot fully assess the level of risk presented.

With an appreciation for the need for such information, managers and others with an interest in the CDFI industry have invested substantial effort in designing tools for data collection and analysis that focus on measuring the financial performance--the risks and returns--of CDFI portfolios. An important motivation for these efforts is the need to diversify funding sources for community development, which has relied heretofore largely on grants from government and foundations. To attract more return-oriented investors, including both conventional investors and those with social as well as financial goals, CDFIs must demonstrate financial viability as well as the ability to fulfill the broader development mission.

For example, the Opportunity Finance Network's CDFI Assessment and Rating System (CARS) gathers data to evaluate a CDFI's overall creditworthiness and its effectiveness in using its financial resources to achieve its development objectives. A CDFI is rated for its financial strength and performance in the areas of capital, assets, management, earnings, and liquidity, in a manner broadly analogous to the way a supervisory agency would rate a commercial bank. The financial analysis is supplemented by an evaluation of how well the CDFI is fulfilling its mission, including an assessment of its procedures for tracking the outcomes of its work. To date, more than forty CDFIs have chosen to be evaluated under the CARS, and thirty-one analyses have been completed. Thus far, fifteen potential investors have subscribed to the CARS database, including socially responsible investment funds, brokerage houses, large financial institutions, and national foundations.4 Although still in its early stages, this initiative, if successful, will have the double benefit of attracting more funds into community development and helping to ensure that those funds are effectively used.

More generally, the movement toward quantifying the performance, risk, and community impact of CDFIs is essential to the growth and sustainability of the field, in my view. By demonstrating both financial viability and social impact through hard data, CDFIs are better positioned to obtain the funding necessary to maintain their operations and to respond to emerging needs and opportunities. Indeed, progress has been made in recent years in the rating and securitization of community development portfolios, a development that should provide CDFIs with increased access to the capital markets and to new sources of liquidity. If the new data and evaluation methods of CDFI performance bear scrutiny, investors will gain confidence in using this information for matching their investment choices with their priorities and risk tolerances. In the community development field, to be sure, financial returns and social returns are not necessarily the same, which is why measurement should include both financial and social indicators. Potential investors, including public-sector and foundation sources of funds, will naturally differ on the weights they put on financial and social returns. To attract the widest range of funding, both types of information should be provided.

Evaluating Policy and Practice

Quantitative information plays yet another important role: increasing the effectiveness of policies and programs. The systematic collection and analysis of data on program inputs and outputs is an increasingly important part of learning about what works. For policymakers, data on program results help guide policy development and improve the allocation of scarce public funds. For community development organizations, participation in broad-based data-gathering serves at least two goals. First, in the long run, their analyses of the activities and the associated outcomes in diverse communities will help them achieve the greatest impact for resources expended. Second, such analyses help community development organizations demonstrate their effectiveness to public and private funders.

A number of methods for evaluating community development projects are currently in use, with more in development. The NeighborWorks America'sÆ Success Measures Data System documents the effect of community development programs throughout the country. Using forty-four indicators and a range of data-collection tools, the system quantifies the effects of housing, economic development, and community building programs at the individual, organization, and community levels. By sharing this knowledge, practitioners, funders, and policymakers can identify programs that achieve the best outcomes and gain insights into the reasons they work. Broad access to this information promotes replication of the most effective programs and may diminish the costs associated with trial-and-error learning.5

Another tool available to CDFIs is the Community Investment Impact System developed by the Department of Treasury's CDFI Fund. This system collects detailed information on institutions and transactions, allowing the CDFI Fund to measure community effects and to associate those effects with institutions working in that area. These results can help inform funding decisions, develop programs, establish performance benchmarks, and communicate societal benefits attributable to specific policy. For example, using data from the system, the CDFI Fund found that in a recent year, CDFIs leveraged financial program awards by the fund at a ratio of 20 to 1, using multiple sources of debt and equity financing from banks, local and state governments, private investors, and borrower equity to structure project financing.6

Each of these data-driven initiatives share the goal of increasing understanding of opaque markets to support investment, policy, and research. The need for data and tools is the driving force behind the Brookings Institution's Urban Markets Initiative. In establishing this policy center, Brookings acknowledged that limited access to data that captures the viability of urban communities constrains investment in these markets. The think tank is focusing on initiatives that can demonstrate untapped market potential.7 One such effort is the National Infrastructure for Community Statistics. It will include a central web-based repository that integrates data from federal, state, and local governments and from commercial sources. The ultimate goal of this project, which is under development in collaboration with more than 100 participants from government, nonprofits, and private-sector industries, is to aggregate and to make accessible the data needed to inform decisions about economic development activities.8

Parallels to International Economic Development

The usefulness of microeconomic data in community development raises an interesting parallel to recent analyses of international economic development. Although the U.S. context is obviously different in important respects from that of developing countries, domestic community organizations and providers of international aid both face the challenge of fostering economic development in low-income areas. In the United States, our experience in community development over the past thirty years has resulted in an evolution from a centralized, federal-government-driven approach to a heavy reliance on the involvement of community-based organizations and agencies for project development and implementation. In light of this experience, it is quite interesting that some new thinking on international development has rejected the traditional approach to aid, with its emphasis on large-scale projects and top-down planning, in favor of micro-level, bottom-up approaches that use local information and systematic analyses of inputs and outcomes.

Critics of traditional development aid programs, such as New York University economist William Easterly, argue that such programs have not succeeded because those implementing the programs do not have the information necessary to make effective use of resources.9 For example, a World Bank report describes an irrigation project that was being designed by technical staff for an area of Nepal that was thought to be unirrigated. A delay in the project led to the discovery that, in fact, eighty-five fully functioning farmer-managed irrigation systems existed in the "unirrigated" area. Further, another irrigation program actually reduced productivity because it undermined pre-existing arrangements among farmers.10 Quite obviously, those planning these projects needed local input to make better use of the project resources.

Easterly advocates a more decentralized, grass-roots approach that involves local groups and emphasizes feedback and accountability. Illustrative of this point, a World Bank study of rural water supply projects found that, of those projects with a high level of participation by local beneficiaries, more than two-thirds were successful whereas, among those projects with little local beneficiary participation, only 12 percent were successful.11 Both feedback and accountability depend, of course, on accurate measurement of results. In practice, measuring results is easier at the local level, in part because comparisons can be drawn to other localities that have not received aid. Incentives also matter; and smaller, more-tailored projects for which responsibilities are well defined are likely to provide better incentives to the people who carry them out than those that large, diffuse projects will provide. Follow-up is important as well. Easterly criticizes, for instance, situations in which foreign aid has been used to build highly visible projects, such as new roads, without providing resources or incentives to do the less-glamorous work of maintaining them.

The themes emphasized by Easterly and other analysts of international aid programs are useful, I think, in the context of domestic community development. Although national initiatives have their place, often the most effective programs take place at the level of the individual community, using local information and local participation. Accountability and feedback, facilitated by data development and quantitative analysis as well as by more-qualitative information, are critical for success. Goals should be modest at first; but knowledge is cumulative, and sometimes good results can be replicated at larger scales. Research, both quantitative and qualitative, furthers learning. None of this is easy, particularly since the data have a way of challenging our views about what works and what doesn't. But a great deal is at stake both internationally and domestically and serious empirical analysis has no substitute. The development of more and better data on economically distressed communities, together with sophisticated tools for analyzing those data, is essential for continued progress in community economic development.

---

Footnotes

1. Local Initiatives Support Corporation, "LISC Adds Market Research Initiative to Arsenal of Community Development Tools." MetroEdge, Case Studies, "World's Largest Home Improvement Retailer." Return to text

2. Social Compact. "Social Compact Completes DRILLDOWN in One Hundredth Underserved Neighborhood," DrillDown aggregate statistics provided by Social Compact. Return to text

3. Social Compact, News and Events, "Image Upgrade for Santa Ana's Core," The Orange County Register, February 7, 2006. Social Compact, News and Events, "The Immigrant Dollar: A Driving Force at Gulfgate," The Houston Chronicle, April 9, 2006. Return to text

4. Opportunity Finance Network; National Community Capital Association, CARS, the CDFI Assessment and Rating System. Return to text

5. NeighborWorks America, Success Measures Data System. Return to text

6. United States Department of the Treasury (2005), Community Development Financial Institutions Fund, Impact Data and Reports, " CDFIs Leverage CDFI Program Awards Nearly $20 to $1!" (May). Return to text

7. Brookings Institution, Urban Markets Initiative. Return to text

8. Brookings Institution, National Infrastructure for Community Statistics. Return to text

9. William Easterly (2001), The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics, (Cambridge, Mass.: MIT Press). Return to text

10. World Bank (1998), "Assessing Aid--What Works, What Doesn't, and Why," Policy Research Reports (November). Return to text

11. World Bank (1998), "Assessing Aid. Return to text

# Appendix III: Delete list

.com
a
about
after
ago
all
along
also
among
an
and
another
any
are
area
as
at
be
because
been
before
being
between
but
by
can
cannot
could
date
did
do
during
either
even
ever
five
for
four
from
had
has
have
her
highlights
him
his

how
http
i
if
in
into
is
it
its
kg
more
must
no
nor
not
now
of
on
only
or
other
our
out
over
part
pesos
quickly
s
said
says
she
so
some
soon
such
than
that
the
their
them
then
there
these
this
those
three
through
to
tons

```
under
up
url
us
use
v
very
was
we
well
were
what
when
where
which
while
will
with
without
would
www
you
who
most
```

# Appendix IV: Thesaurus

| Original Word | Replacement |
|---|---|
| qualification requirements | qualification_requirements |
| characteristics preferably | characteristics_preferably |
| quantitative underwriting | quantitative_underwriting |
| sustainable homeownership | sustainable_homeownership |
| financial intermediaries | financial_intermediaries |
| residential construction | residential_construction |
| unnecessary foreclosures | unnecessary_foreclosures |
| preventable foreclosures | preventable_foreclosures |
| consumption expenditures | consumption_expenditures |
| professional forecasters | professional_forecasters |
| investment opportunities | investment_opportunities |
| prescribing quantitative | prescribing_quantitative |
| macroeconomic objectives | macroeconomic_objectives |
| responsible underwriting | responsible_underwriting |
| depository institutions | depository_institutions |
| initial experimentation | initial_experimentation |
| unintended consequences | unintended_consequences |
| disclosure requirements | disclosure_requirements |
| reinvestment coalition | reinvestment_coalition |
| inflation expectations | inflation_expectations |
| financial institutions | financial_institutions |
| communication strategy | communication_strategy |
| underwriting standards | underwriting_standards |
| sustainable employment | sustainable_employment |
| transmission mechanism | transmission_mechanism |
| financial architecture | financial_architecture |
| enhanced communication | enhanced_communication |
| asymmetric information | asymmetric_information |
| inflation compensation | inflation_compensation |
| underwriting practices | underwriting_practices |
| processing information | processing_information |
| particularly important | particularly_important |

| | |
|---|---|
| heightened uncertainty | heightened_uncertainty |
| neoclassical synthesis | neoclassical_synthesis |
| additional information | additional_information |
| inefficient allocation | inefficient_allocation |
| disregarding repayment | disregarding_repayment |
| microeconomic policies | microeconomic_policies |
| contingent liabilities | contingent_liabilities |
| housing administration | housing_administration |
| productive investment | productive_investment |
| information gathering | information_gathering |
| conceptually distinct | conceptually_distinct |
| deceptive advertising | deceptive_advertising |
| financial instability | financial_instability |
| banking organizations | banking_organizations |
| stabilizing inflation | stabilizing_inflation |
| inconsistency problem | inconsistency_problem |
| financial disruptions | financial_disruptions |
| neighborworks america | neighborworks_america |
| rational expectations | rational_expectations |
| sponsored enterprises | sponsored_enterprises |
| unwarranted servicing | unwarranted_servicing |
| standardized approach | standardized_approach |
| experimentation phase | experimentation_phase |
| mitigation techniques | mitigation_techniques |
| further deterioration | further_deterioration |
| microfinance movement | microfinance_movement |
| microfinance programs | microfinance_programs |
| management challenges | management_challenges |
| activities undertaken | activities_undertaken |
| complementary benefit | complementary_benefit |
| conference washington | conference_washington |
| modification programs | modification_programs |
| tailored individually | tailored_individually |
| serious delinquencies | serious_delinquencies |
| accounting standards | accounting_standards |
| incentive structures | incentive_structures |

| | |
|---|---|
| independent mortgage | independent_mortgage |
| conforming mortgages | conforming_mortgages |
| prepayment penalties | prepayment_penalties |
| resource utilization | resource_utilization |
| personal consumption | personal_consumption |
| capital requirements | capital_requirements |
| financial disruption | financial_disruption |
| management practices | management_practices |
| stricter regulations | stricter_regulations |
| residential mortgage | residential_mortgage |
| economic projections | economic_projections |
| economic performance | economic_performance |
| accurate information | accurate_information |
| relationship between | relationship_between |
| respond aggressively | respond_aggressively |
| financial conditions | financial_conditions |
| gathering processing | gathering_processing |
| government sponsored | government_sponsored |
| incoming information | incoming_information |
| affiliate refinances | affiliate_refinances |
| proposed regulations | proposed_regulations |
| greater transparency | greater_transparency |
| mutually reinforcing | mutually_reinforcing |
| inflation objectives | inflation_objectives |
| industrial countries | industrial_countries |
| particular attention | particular_attention |
| carefully considered | carefully_considered |
| enhanced projections | enhanced_projections |
| significant benefits | significant_benefits |
| relevant information | relevant_information |
| national association | national_association |
| workout arrangements | workout_arrangements |
| adverse consequences | adverse_consequences |
| consumer protections | consumer_protections |
| stabilizing economic | stabilizing_economic |
| maintenance expenses | maintenance_expenses |

| | |
|---|---|
| industrial economies | industrial_economies |
| analytical framework | analytical_framework |
| principal writedowns | principal_writedowns |
| becomes sufficiently | becomes_sufficiently |
| specific requirement | specific_requirement |
| became increasingly | became_increasingly |
| assumes incorrectly | assumes_incorrectly |
| via videoconference | via_videoconference |
| refinancing options | refinancing_options |
| considerably higher | considerably_higher |
| market participants | market_participants |
| maximum sustainable | maximum_sustainable |
| inflation objective | inflation_objective |
| consumer protection | consumer_protection |
| financial stability | financial_stability |
| advanced approaches | advanced_approaches |
| manifest themselves | manifest_themselves |
| investment vehicles | investment_vehicles |
| inflation targeting | inflation_targeting |
| structured products | structured_products |
| risk concentrations | risk_concentrations |
| numerical inflation | numerical_inflation |
| risk identification | risk_identification |
| bankers association | bankers_association |
| treasury securities | treasury_securities |
| investor confidence | investor_confidence |
| financial condition | financial_condition |
| legally enforceable | legally_enforceable |
| implementation plan | implementation_plan |
| correlation between | correlation_between |
| collateralized debt | collateralized_debt |
| anticipate vigorous | anticipate_vigorous |
| greater uncertainty | greater_uncertainty |
| market developments | market_developments |
| deceptive practices | deceptive_practices |
| checking timeliness | checking_timeliness |

| | |
|---|---|
| systematic approach | systematic_approach |
| providing liquidity | providing_liquidity |
| conducting monetary | conducting_monetary |
| orderly functioning | orderly_functioning |
| academic economists | academic_economists |
| subprime adjustable | subprime_adjustable |
| remained reasonably | remained_reasonably |
| increasing investor | increasing_investor |
| comprehensive scope | comprehensive_scope |
| productivity growth | productivity_growth |
| institutions should | institutions_should |
| demanded sufficient | demanded_sufficient |
| economic conditions | economic_conditions |
| credible commitment | credible_commitment |
| banking supervision | banking_supervision |
| coercing appraisers | coercing_appraisers |
| these circumstances | these_circumstances |
| counseling agencies | counseling_agencies |
| pitfalls associated | pitfalls_associated |
| between stabilizing | between_stabilizing |
| credit availability | credit_availability |
| leveraged financial | leveraged_financial |
| liquidity providers | liquidity_providers |
| appraisal coercion | appraisal_coercion |
| bank communication | bank_communication |
| headline inflation | headline_inflation |
| subprime mortgages | subprime_mortgages |
| maximum employment | maximum_employment |
| explicit numerical | explicit_numerical |
| market functioning | market_functioning |
| macroeconomic risk | macroeconomic_risk |
| empirical evidence | empirical_evidence |
| mandate consistent | mandate_consistent |
| objective function | objective_function |
| explicit inflation | explicit_inflation |
| complex structured | complex_structured |

| | |
|---|---|
| carnegie rochester | carnegie_rochester |
| troubled borrowers | troubled_borrowers |
| time inconsistency | time_inconsistency |
| regulatory capital | regulatory_capital |
| housing correction | housing_correction |
| financial literacy | financial_literacy |
| originally thought | originally_thought |
| market disruptions | market_disruptions |
| responsible credit | responsible_credit |
| foreclosure starts | foreclosure_starts |
| structured finance | structured_finance |
| rule prescriptions | rule_prescriptions |
| loan modifications | loan_modifications |
| reduce preventable | reduce_preventable |
| thereby increasing | thereby_increasing |
| participants views | participants_views |
| appropriate stance | appropriate_stance |
| price appreciation | price_appreciation |
| these developments | these_developments |
| brokerage services | brokerage_services |
| democratic society | democratic_society |
| interested parties | interested_parties |
| subprime borrowers | subprime_borrowers |
| economic education | economic_education |
| difference between | difference_between |
| american countries | american_countries |
| advanced economies | advanced_economies |
| without conducting | without_conducting |
| financial distress | financial_distress |
| effective consumer | effective_consumer |
| economic downturns | economic_downturns |
| liquidity problems | liquidity_problems |
| sustainable growth | sustainable_growth |
| particular concern | particular_concern |
| compliance reviews | compliance_reviews |
| inventory overhang | inventory_overhang |

| | |
|---|---|
| avoids prescribing | avoids_prescribing |
| business economics | business_economics |
| expansions arising | expansions_arising |
| liquidity provider | liquidity_provider |
| requires examining | requires_examining |
| inflation dynamics | inflation_dynamics |
| under considerable | under_considerable |
| evidence indicates | evidence_indicates |
| reverse causality | reverse_causality |
| consumers receive | consumers_receive |
| financial markets | financial_markets |
| economic activity | economic_activity |
| information about | information_about |
| market operations | market_operations |
| backed securities | backed_securities |
| uncertainty about | uncertainty_about |
| subprime mortgage | subprime_mortgage |
| structured credit | structured_credit |
| unemployment rate | unemployment_rate |
| senior management | senior_management |
| governor frederic | governor_frederic |
| chairman bernanke | chairman_bernanke |
| fomc participants | fomc_participants |
| adverse selection | adverse_selection |
| mentioned earlier | mentioned_earlier |
| percentage points | percentage_points |
| protect consumers | protect_consumers |
| discovery process | discovery_process |
| repayment ability | repayment_ability |
| interbank funding | interbank_funding |
| financial turmoil | financial_turmoil |
| credit conditions | credit_conditions |
| backed commercial | backed_commercial |
| consumer spending | consumer_spending |
| rate depreciation | rate_depreciation |
| strong commitment | strong_commitment |

| | |
|---|---|
| overall inflation | overall_inflation |
| become correlated | become_correlated |
| lending standards | lending_standards |
| loan modification | loan_modification |
| lending practices | lending_practices |
| better understand | better_understand |
| september meeting | september_meeting |
| speculative grade | speculative_grade |
| holding companies | holding_companies |
| vacant properties | vacant_properties |
| keeping inflation | keeping_inflation |
| community affairs | community_affairs |
| delinquency rates | delinquency_rates |
| securities backed | securities_backed |
| provide liquidity | provide_liquidity |
| scenario analysis | scenario_analysis |
| transaction costs | transaction_costs |
| market conditions | market_conditions |
| abusive practices | abusive_practices |
| economic outcomes | economic_outcomes |
| business spending | business_spending |
| evidence suggests | evidence_suggests |
| market turbulence | market_turbulence |
| globalization has | globalization_has |
| recover statutory | recover_statutory |
| remained strained | remained_strained |
| regulations would | regulations_would |
| while maintaining | while_maintaining |
| readily available | readily_available |
| committee members | committee_members |
| equity protection | equity_protection |
| average inflation | average_inflation |
| these instruments | these_instruments |
| leveraged buyouts | leveraged_buyouts |
| sustainable level | sustainable_level |
| downward pressure | downward_pressure |

| | |
|---|---|
| preserve consumer | preserve_consumer |
| standard practice | standard_practice |
| depend critically | depend_critically |
| further weakening | further_weakening |
| systemwide stress | systemwide_stress |
| liquidity support | liquidity_support |
| inflation measure | inflation_measure |
| market discipline | market_discipline |
| determine whether | determine_whether |
| information flows | information_flows |
| community college | community_college |
| their communities | their_communities |
| projection errors | projection_errors |
| detailed analyses | detailed_analyses |
| insurance against | insurance_against |
| capital framework | capital_framework |
| charlotte chamber | charlotte_chamber |
| independent voice | independent_voice |
| market disruption | market_disruption |
| terrorist attacks | terrorist_attacks |
| orleans louisiana | orleans_louisiana |
| discussed earlier | discussed_earlier |
| underlying assets | underlying_assets |
| bank independence | bank_independence |
| our communication | our_communication |
| deposit insurance | deposit_insurance |
| facilitates price | facilitates_price |
| potential source | potential_source |
| right incentives | right_incentives |
| grain inspectors | grain_inspectors |
| called unplanned | called_unplanned |
| commercial paper | commercial_paper |
| percentage point | percentage_point |
| financial system | financial_system |
| financial market | financial_market |
| capital adequacy | capital_adequacy |

| | |
|---|---|
| distribute model | distribute_model |
| market committee | market_committee |
| domestic product | domestic_product |
| governor randall | governor_randall |
| aggregate demand | aggregate_demand |
| economic outlook | economic_outlook |
| policy decisions | policy_decisions |
| adverse feedback | adverse_feedback |
| commodity prices | commodity_prices |
| small businesses | small_businesses |
| auction facility | auction_facility |
| stable inflation | stable_inflation |
| foreign exchange | foreign_exchange |
| highly uncertain | highly_uncertain |
| nominal interest | nominal_interest |
| tradeoff between | tradeoff_between |
| market liquidity | market_liquidity |
| payment defaults | payment_defaults |
| community groups | community_groups |
| adopted explicit | adopted_explicit |
| optimal monetary | optimal_monetary |
| potential output | potential_output |
| expenditures pce | expenditures_pce |
| linkages between | linkages_between |
| underlying trend | underlying_trend |
| assess repayment | assess_repayment |
| adversely affect | adversely_affect |
| banking agencies | banking_agencies |
| debt obligations | debt_obligations |
| risks associated | risks_associated |
| stimulus package | stimulus_package |
| central tendency | central_tendency |
| investment grade | investment_grade |
| master agreement | master_agreement |
| troubles brewing | troubles_brewing |
| mortgage lenders | mortgage_lenders |

| | |
|---|---|
| great moderation | great_moderation |
| overall consumer | overall_consumer |
| credit histories | credit_histories |
| argued elsewhere | argued_elsewhere |
| complex products | complex_products |
| aggregate supply | aggregate_supply |
| loan obligations | loan_obligations |
| forecast horizon | forecast_horizon |
| asset expansions | asset_expansions |
| subprime lending | subprime_lending |
| global financial | global_financial |
| delinquent twice | delinquent_twice |
| mortgage markets | mortgage_markets |
| adverse outcomes | adverse_outcomes |
| household wealth | household_wealth |
| european central | european_central |
| well functioning | well_functioning |
| purchasing power | purchasing_power |
| mortgage brokers | mortgage_brokers |
| borrowers facing | borrowers_facing |
| overall economic | overall_economic |
| anchor inflation | anchor_inflation |
| national council | national_council |
| thereby reducing | thereby_reducing |
| negative effects | negative_effects |
| monthly payments | monthly_payments |
| educated workers | educated_workers |
| recent financial | recent_financial |
| solidly anchored | solidly_anchored |
| abuse unfairness | abuse_unfairness |
| cashing receipts | cashing_receipts |
| insider outsider | insider_outsider |
| consumer testing | consumer_testing |
| channeling funds | channeling_funds |
| mortgage lending | mortgage_lending |
| words scrambling | words_scrambling |

| | |
|---|---|
| risk measurement | risk_measurement |
| finance products | finance_products |
| strong rationale | strong_rationale |
| taylor principle | taylor_principle |
| penalties where | penalties_where |
| hardly insulate | hardly_insulate |
| ongoing process | ongoing_process |
| federal reserve | federal_reserve |
| monetary policy | monetary_policy |
| risk management | risk_management |
| price stability | price_stability |
| discount window | discount_window |
| price discovery | price_discovery |
| economic growth | economic_growth |
| speech governor | speech_governor |
| rating agencies | rating_agencies |
| broader economy | broader_economy |
| mortgage market | mortgage_market |
| funding markets | funding_markets |
| upward pressure | upward_pressure |
| mortgage backed | mortgage_backed |
| loss mitigation | loss_mitigation |
| credit products | credit_products |
| concerned about | concerned_about |
| point objective | point_objective |
| consumer prices | consumer_prices |
| reasonably well | reasonably_well |
| adjustable rate | adjustable_rate |
| enterprise wide | enterprise_wide |
| chairman donald | chairman_donald |
| adverse effects | adverse_effects |
| forward looking | forward_looking |
| officer opinion | officer_opinion |
| central bankers | central_bankers |
| speech chairman | speech_chairman |
| firmly anchored | firmly_anchored |

| | |
|---|---|
| lessons learned | lessons_learned |
| turbulence sets | turbulence_sets |
| consumer choice | consumer_choice |
| moderate growth | moderate_growth |
| bank presidents | bank_presidents |
| firm commitment | firm_commitment |
| weighted median | weighted_median |
| policy strategy | policy_strategy |
| each individual | each_individual |
| fiscal stimulus | fiscal_stimulus |
| leveraged loans | leveraged_loans |
| jumbo mortgages | jumbo_mortgages |
| certain complex | certain_complex |
| coming quarters | coming_quarters |
| subprime market | subprime_market |
| control methods | control_methods |
| judgments about | judgments_about |
| price inflation | price_inflation |
| defined broadly | defined_broadly |
| optimal control | optimal_control |
| better informed | better_informed |
| backstop source | backstop_source |
| flipping scheme | flipping_scheme |
| briefly discuss | briefly_discuss |
| broker steering | broker_steering |
| policy response | policy_response |
| full employment | full_employment |
| quantify latent | quantify_latent |
| timely decisive | timely_decisive |
| different types | different_types |
| days delinquent | days_delinquent |
| recent readings | recent_readings |
| develop prudent | develop_prudent |
| unplanned asset | unplanned_asset |
| senior managers | senior_managers |
| collected under | collected_under |

| | |
|---|---|
| corporate bonds | corporate_bonds |
| good collateral | good_collateral |
| monthly payment | monthly_payment |
| carolinas award | carolinas_award |
| lost confidence | lost_confidence |
| priced mortgage | priced_mortgage |
| likely scenario | likely_scenario |
| spreads widened | spreads_widened |
| uncertain about | uncertain_about |
| else associated | else_associated |
| dual objectives | dual_objectives |
| primary dealers | primary_dealers |
| productive uses | productive_uses |
| special purpose | special_purpose |
| quarter century | quarter_century |
| mortgage broker | mortgage_broker |
| pillowtex site | pillowtex_site |
| session allied | session_allied |
| published last | published_last |
| below baseline | below_baseline |
| loss estimates | loss_estimates |
| adopting basel | adopting_basel |
| interest rates | interest_rates |
| core inflation | core_inflation |
| balance sheets | balance_sheets |
| north carolina | north_carolina |
| policy actions | policy_actions |
| nominal anchor | nominal_anchor |
| gross domestic | gross_domestic |
| downside risks | downside_risks |
| inflation rate | inflation_rate |
| private sector | private_sector |
| concerns about | concerns_about |
| target federal | target_federal |
| committee fomc | committee_fomc |
| liquidity risk | liquidity_risk |

| | |
|---|---|
| latin american | latin_american |
| stress testing | stress_testing |
| reserve system | reserve_system |
| consumer price | consumer_price |
| mortgage loans | mortgage_loans |
| business lines | business_lines |
| tighter credit | tighter_credit |
| housing market | housing_market |
| fourth quarter | fourth_quarter |
| banking system | banking_system |
| market turmoil | market_turmoil |
| unusually high | unusually_high |
| united kingdom | united_kingdom |
| modern science | modern_science |
| exchange rates | exchange_rates |
| risk exposures | risk_exposures |
| credit quality | credit_quality |
| during periods | during_periods |
| housing sector | housing_sector |
| several months | several_months |
| corporate bond | corporate_bond |
| important role | important_role |
| imported goods | imported_goods |
| ban prepayment | ban_prepayment |
| subprime loans | subprime_loans |
| would prohibit | would_prohibit |
| strong nominal | strong_nominal |
| home ownership | home_ownership |
| small business | small_business |
| credit ratings | credit_ratings |
| public comment | public_comment |
| capital ratios | capital_ratios |
| days renewable | days_renewable |
| sheet vehicles | sheet_vehicles |
| closely linked | closely_linked |
| valuation risk | valuation_risk |

| | |
|---|---|
| keep inflation | keep_inflation |
| thinking about | thinking_about |
| september fomc | september_fomc |
| income lending | income_lending |
| help forestall | help_forestall |
| protection act | protection_act |
| relative price | relative_price |
| non depository | non_depository |
| spread between | spread_between |
| phillips curve | phillips_curve |
| credit records | credit_records |
| exchange value | exchange_value |
| modal forecast | modal_forecast |
| their payments | their_payments |
| servicing fees | servicing_fees |
| desired markup | desired_markup |
| state agencies | state_agencies |
| these vehicles | these_vehicles |
| disclosure act | disclosure_act |
| excluding food | excluding_food |
| relatively low | relatively_low |
| local currency | local_currency |
| president bush | president_bush |
| opinion survey | opinion_survey |
| many countries | many_countries |
| proposed rules | proposed_rules |
| sheet capacity | sheet_capacity |
| recent decades | recent_decades |
| data collected | data_collected |
| stigma problem | stigma_problem |
| many borrowers | many_borrowers |
| understand how | understand_how |
| rate mortgages | rate_mortgages |
| help borrowers | help_borrowers |
| term interbank | term_interbank |
| swiss national | swiss_national |

| | |
|---|---|
| which excludes | which_excludes |
| hurt consumers | hurt_consumers |
| refinance into | refinance_into |
| reach troubled | reach_troubled |
| apply stricter | apply_stricter |
| fhasecure plan | fhasecure_plan |
| proposal would | proposal_would |
| minimize moral | minimize_moral |
| rated tranches | rated_tranches |
| take advantage | take_advantage |
| governor kevin | governor_kevin |
| cannot sustain | cannot_sustain |
| excessive risk | excessive_risk |
| toward greater | toward_greater |
| annual meeting | annual_meeting |
| these products | these_products |
| different ways | different_ways |
| estimated pass | estimated_pass |
| property taxes | property_taxes |
| case scenarios | case_scenarios |
| credit markets | credit_markets |
| paper programs | paper_programs |
| lends directly | lends_directly |
| related assets | related_assets |
| haven demands | haven_demands |
| modal outlook | modal_outlook |
| credit scores | credit_scores |
| dinner tables | dinner_tables |
| excludes food | excludes_food |
| address these | address_these |
| woodford 2003 | woodford_2003 |
| delinquencies | delinquencies |
| qualification | qualification |
| accommodation | accommodation |
| dissemination | dissemination |
| international | international |

| | |
|---|---|
| clearinghouse | clearinghouse |
| united states | united_states |
| federal funds | federal_funds |
| balance sheet | balance_sheet |
| central banks | central_banks |
| latin america | latin_america |
| exchange rate | exchange_rate |
| due diligence | due_diligence |
| interest rate | interest_rate |
| higher priced | higher_priced |
| import prices | import_prices |
| energy prices | energy_prices |
| feedback loop | feedback_loop |
| recent events | recent_events |
| reserve board | reserve_board |
| run inflation | run_inflation |
| remarks today | remarks_today |
| third quarter | third_quarter |
| credit rating | credit_rating |
| well anchored | well_anchored |
| incoming data | incoming_data |
| vice chairman | vice_chairman |
| recent months | recent_months |
| borrowers who | borrowers_who |
| real activity | real_activity |
| credit losses | credit_losses |
| stated income | stated_income |
| board members | board_members |
| round effects | round_effects |
| working group | working_group |
| subprime arms | subprime_arms |
| consumers who | consumers_who |
| going forward | going_forward |
| policy makers | policy_makers |
| home purchase | home_purchase |
| twelve months | twelve_months |

| | |
|---|---|
| my colleagues | my_colleagues |
| would require | would_require |
| comfort zones | comfort_zones |
| mishkin 2007b | mishkin_2007b |
| mishkin 2007a | mishkin_2007a |
| risk profiles | risk_profiles |
| these actions | these_actions |
| term interest | term_interest |
| weaker credit | weaker_credit |
| money markets | money_markets |
| already noted | already_noted |
| stable prices | stable_prices |
| policy easing | policy_easing |
| current state | current_state |
| policy making | policy_making |
| coming months | coming_months |
| early payment | early_payment |
| several years | several_years |
| risk managers | risk_managers |
| than expected | than_expected |
| help mitigate | help_mitigate |
| pce inflation | pce_inflation |
| natural rates | natural_rates |
| capital flows | capital_flows |
| policy action | policy_action |
| supply shocks | supply_shocks |
| discount rate | discount_rate |
| check cashing | check_cashing |
| core measures | core_measures |
| noted earlier | noted_earlier |
| fully indexed | fully_indexed |
| business unit | business_unit |
| proposed rule | proposed_rule |
| private label | private_label |
| mishkin 2007c | mishkin_2007c |
| pilot project | pilot_project |

| | |
|---|---|
| sticky prices | sticky_prices |
| outlined here | outlined_here |
| their balance | their_balance |
| ratio exceeds | ratio_exceeds |
| liquid assets | liquid_assets |
| fomc meetings | fomc_meetings |
| during recent | during_recent |
| large numbers | large_numbers |
| best promotes | best_promotes |
| friedman 1968 | friedman_1968 |
| best possible | best_possible |
| timely manner | timely_manner |
| business line | business_line |
| economy faces | economy_faces |
| variable rate | variable_rate |
| demand shocks | demand_shocks |
| these efforts | these_efforts |
| own interests | own_interests |
| during normal | during_normal |
| my discussion | my_discussion |
| two scenarios | two_scenarios |
| steep decline | steep_decline |
| loan workouts | loan_workouts |
| aboveaverage | above_average |
| ii framework | ii_framework |
| rule writing | rule_writing |
| banks should | banks_should |
| central bank | central_bank |
| pass through | pass_through |
| basis points | basis_points |
| dual mandate | dual_mandate |
| house prices | house_prices |
| moral hazard | moral_hazard |
| comfort zone | comfort_zone |
| term funding | term_funding |
| washington d | washington_d |

| | |
|---|---|
| reserve bank | reserve_bank |
| priced loans | priced_loans |
| into account | into_account |
| real economy | real_economy |
| federal open | federal_open |
| because they | because_they |
| simple rules | simple_rules |
| during times | during_times |
| chairman ben | chairman_ben |
| fomc meeting | fomc_meeting |
| labor market | labor_market |
| asset backed | asset_backed |
| now alliance | now_alliance |
| our proposal | our_proposal |
| 2 percentage | 2_percentage |
| past several | past_several |
| we recognize | we_recognize |
| recent years | recent_years |
| recent weeks | recent_weeks |
| latent risks | latent_risks |
| normal times | normal_times |
| part because | part_because |
| even greater | even_greater |
| risk profile | risk_profile |
| less willing | less_willing |
| link between | link_between |
| core measure | core_measure |
| escrow taxes | escrow_taxes |
| unsold homes | unsold_homes |
| our proposed | our_proposed |
| look forward | look_forward |
| yield spread | yield_spread |
| worth noting | worth_noting |
| nominal wage | nominal_wage |
| second round | second_round |
| wide variety | wide_variety |

| | |
|---|---|
| term auction | term_auction |
| take account | take_account |
| loan officer | loan_officer |
| 4 percentage | 4_percentage |
| 3 percentage | 3_percentage |
| double digit | double_digit |
| fomc members | fomc_members |
| upside risks | upside_risks |
| stock market | stock_market |
| shorter term | shorter_term |
| subprime arm | subprime_arm |
| center stage | center_stage |
| cross border | cross_border |
| exactly what | exactly_what |
| risk premium | risk_premium |
| large global | large_global |
| stress tests | stress_tests |
| am delighted | am_delighted |
| run tradeoff | run_tradeoff |
| crucial role | crucial_role |
| asset prices | asset_prices |
| through open | through_open |
| credit cards | credit_cards |
| rate changes | rate_changes |
| natural rate | natural_rate |
| risk factors | risk_factors |
| asset values | asset_values |
| facility taf | facility_taf |
| medium sized | medium_sized |
| value ratios | value_ratios |
| afford their | afford_their |
| highly rated | highly_rated |
| insight into | insight_into |
| under stress | under_stress |
| lq framework | lq_framework |
| fixed amount | fixed_amount |

| | |
|---|---|
| risk spreads | risk_spreads |
| abovemarket | above_market |
| open market | open_market |
| rather than | rather_than |
| longer term | longer_term |
| reserve has | reserve_has |
| most likely | most_likely |
| off balance | off_balance |
| last summer | last_summer |
| taylor rule | taylor_rule |
| help ensure | help_ensure |
| i mentioned | i_mentioned |
| past decade | past_decade |
| medium term | medium_term |
| real estate | real_estate |
| price index | price_index |
| risk taking | risk_taking |
| even though | even_though |
| product gdp | product_gdp |
| lower bound | lower_bound |
| prime jumbo | prime_jumbo |
| john taylor | john_taylor |
| debt growth | debt_growth |
| new zealand | new_zealand |
| home equity | home_equity |
| entire firm | entire_firm |
| most recent | most_recent |
| early 1980s | early_1980s |
| freddie mac | freddie_mac |
| their homes | their_homes |
| pointed out | pointed_out |
| thirty days | thirty_days |
| would apply | would_apply |
| net exports | net_exports |
| anyone else | anyone_else |
| two decades | two_decades |

| | |
|---|---|
| pulled back | pulled_back |
| speech vice | speech_vice |
| write downs | write_downs |
| another way | another_way |
| should help | should_help |
| second half | second_half |
| credit risk | credit_risk |
| less likely | less_likely |
| adage trust | adage_trust |
| 910 billion | 910_billion |
| fha insured | fha_insured |
| kevin warsh | kevin_warsh |
| high degree | high_degree |
| house price | house_price |
| explain why | explain_why |
| percent per | percent_per |
| ease policy | ease_policy |
| target rate | target_rate |
| about their | about_their |
| past couple | past_couple |
| waiting too | waiting_too |
| would allow | would_allow |
| rough patch | rough_patch |
| higher than | higher_than |
| minimum bid | minimum_bid |
| sigma model | sigma_model |
| road tested | road_tested |
| become less | become_less |
| could raise | could_raise |
| hedge funds | hedge_funds |
| senior loan | senior_loan |
| ninety days | ninety_days |
| loans which | loans_which |
| dotted line | dotted_line |
| dashed line | dashed_line |
| larger than | larger_than |

| | |
|---|---|
| before they | before_they |
| market desk | market_desk |
| per quarter | per_quarter |
| help offset | help_offset |
| isda master | isda_master |
| rocket ship | rocket_ship |
| food prices | food_prices |
| views about | views_about |
| gap between | gap_between |
| explain how | explain_how |
| early 1990s | early_1990s |
| health care | health_care |
| must always | must_always |
| willen 2007 | willen_2007 |
| canner 2007 | canner_2007 |
| proper due | proper_due |
| else equal | else_equal |
| funds rate | funds_rate |
| short term | short_term |
| my remarks | my_remarks |
| final rule | final_rule |
| would like | would_like |
| longer run | longer_run |
| few months | few_months |
| output gap | output_gap |
| some cases | some_cases |
| gdp growth | gdp_growth |
| wide range | wide_range |
| zero lower | zero_lower |
| sound risk | sound_risk |
| oil prices | oil_prices |
| 40 percent | 40_percent |
| full range | full_range |
| 10 percent | 10_percent |
| fannie mae | fannie_mae |
| write down | write_down |

| | |
|---|---|
| ssg report | ssg_report |
| came under | came_under |
| would also | would_also |
| five years | five_years |
| great deal | great_deal |
| has become | has_become |
| low levels | low_levels |
| real gross | real_gross |
| last month | last_month |
| hard edges | hard_edges |
| put upward | put_upward |
| six months | six_months |
| per barrel | per_barrel |
| lower than | lower_than |
| well known | well_known |
| bill poole | bill_poole |
| 50 percent | 50_percent |
| early 2007 | early_2007 |
| than usual | than_usual |
| rule would | rule_would |
| near prime | near_prime |
| four times | four_times |
| onto their | onto_their |
| very short | very_short |
| safe haven | safe_haven |
| exceeds 50 | exceeds_50 |
| lose sight | lose_sight |
| were often | were_often |
| may prefer | may_prefer |
| paper abcp | paper_abcp |
| ted spread | ted_spread |
| home sales | home_sales |
| key factor | key_factor |
| would help | would_help |
| months ago | months_ago |
| high level | high_level |

| | |
|---|---|
| punch bowl | punch_bowl |
| lucas 1972 | lucas_1972 |
| much wider | much_wider |
| take steps | take_steps |
| rate reset | rate_reset |
| solid pace | solid_pace |
| could lead | could_lead |
| same time | same_time |
| i believe | i_believe |
| so called | so_called |
| near term | near_term |
| 2 percent | 2_percent |
| last year | last_year |
| their own | their_own |
| long term | long_term |
| short run | short_run |
| last week | last_week |
| two years | two_years |
| crude oil | crude_oil |
| much less | much_less |
| after all | after_all |
| 4 percent | 4_percent |
| real time | real_time |
| no longer | no_longer |
| user cost | user_cost |
| must take | must_take |
| thank you | thank_you |
| ten years | ten_years |
| some time | some_time |
| years ago | years_ago |
| 3 percent | 3_percent |
| under way | under_way |
| 100 basis | 100_basis |
| 7 percent | 7_percent |
| late 2005 | late_2005 |
| 5 percent | 5_percent |

| | |
|---|---|
| take into | take_into |
| past year | past_year |
| figure 1b | figure_1b |
| about how | about_how |
| take time | take_time |
| also help | also_help |
| weeks ago | weeks_ago |
| few years | few_years |
| set forth | set_forth |
| bank must | bank_must |
| rate pass | rate_pass |
| late last | late_last |
| we looked | we_looked |
| 9 percent | 9_percent |
| index cpi | index_cpi |
| next year | next_year |
| 6 percent | 6_percent |
| has grown | has_grown |
| tail risk | tail_risk |
| remind us | remind_us |
| act hoepa | act_hoepa |
| plot line | plot_line |
| picked up | picked_up |
| figure 2a | figure_2a |
| they were | they_were |
| hmda data | hmda_data |
| old adage | old_adage |
| too early | too_early |
| less than | less_than |
| one month | one_month |
| has moved | has_moved |
| know what | know_what |
| there was | there_was |
| high cost | high_cost |
| turns out | turns_out |
| last fall | last_fall |

| | |
|---|---|
| new homes | new_homes |
| work outs | work_outs |
| must also | must_also |
| find ways | find_ways |
| http www | http_www |
| basel ii | basel_ii |
| long run | long_run |
| new york | new_york |
| 50 basis | 50_basis |
| past few | past_few |
| you know | you_know |
| doing so | doing_so |
| hope now | hope_now |
| donald l | donald_l |
| york new | york_new |
| low pass | low_pass |
| 25 basis | 25_basis |
| past two | past_two |
| real gdp | real_gdp |
| per year | per_year |
| our dual | our_dual |
| thus far | thus_far |
| i expect | i_expect |
| may well | may_well |
| moved up | moved_up |
| dig deep | dig_deep |
| too much | too_much |
| very low | very_low |
| year end | year_end |
| ten year | ten_year |
| one year | one_year |
| pillar 3 | pillar_3 |
| key role | key_role |
| they had | they_had |
| ll touch | ll_touch |
| dr yunus | dr_yunus |

| | |
|---|---|
| no doubt | no_doubt |
| pillar 2 | pillar_2 |
| wouldn t | wouldn_t |
| st louis | st_louis |
| work out | work_out |
| 1990 91 | 1990_91 |
| ex ante | ex_ante |
| i would | i_would |
| my view | my_view |
| even if | even_if |
| today i | today_i |
| i noted | i_noted |
| n texas | n_texas |
| opt out | opt_out |
| i think | i_think |
| we must | we_must |
| they do | they_do |
| if they | if_they |
| am sure | am_sure |
| basel i | basel_i |
| we also | we_also |
| tell us | tell_us |
| we face | we_face |
| set off | set_off |

# Appendix V: Sample Fed Funds Futures Data, the December 31, 2007 Contract

Fed Fund Future expiring on December 31, 2007

Name FED FUND 30DAY Dec07
Ticker FFZ7 COMB Comdty
Exchange CBT-Chicago Board of Trade
Notional FED FUND 30DAY
Contract Size 5,000,000 USD
Value of 1.0 pt $ 4,167
Tick Size 0.005
Tick Value $ 20.835
Price 0.000 100 - yield
Pt. Val x Price $ 0

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 1/3/06 | 95.265 |
| 1/4/06 | 95.265 |
| 1/5/06 | 95.38 |
| 1/6/06 | 95.35 |
| 1/9/06 | 95.35 |
| 1/10/06 | 95.35 |
| 1/11/06 | 95.35 |
| 1/12/06 | 95.35 |
| 1/13/06 | 95.32 |
| 1/17/06 | 95.32 |
| 1/18/06 | 95.32 |
| 1/19/06 | 95.3 |
| 1/20/06 | 95.3 |
| 1/23/06 | 95.3 |
| 1/24/06 | 95.3 |
| 1/25/06 | 95.25 |
| 1/26/06 | 95.24 |
| 1/27/06 | 95.25 |
| 1/30/06 | 95.25 |
| 1/31/06 | 95.25 |
| 2/1/06 | 95.245 |
| 2/2/06 | 95.175 |
| 2/3/06 | 95.145 |
| 2/6/06 | 95.11 |
| 2/7/06 | 95.12 |
| 2/8/06 | 95.11 |
| 2/9/06 | 95.11 |
| 2/10/06 | 95.08 |
| 2/13/06 | 95.09 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 2/14/06 | 95.075 |
| 2/15/06 | 95.075 |
| 2/16/06 | 95.075 |
| 2/17/06 | 95.095 |
| 2/21/06 | 95.095 |
| 2/22/06 | 95.095 |
| 2/23/06 | 95.05 |
| 2/24/06 | 95.05 |
| 2/27/06 | 95.05 |
| 2/28/06 | 95.05 |
| 3/1/06 | 95.05 |
| 3/2/06 | 95.05 |
| 3/3/06 | 95.055 |
| 3/6/06 | 95.05 |
| 3/7/06 | 95.045 |
| 3/8/06 | 95.045 |
| 3/9/06 | 95.04 |
| 3/10/06 | 95.025 |
| 3/13/06 | 95.02 |
| 3/14/06 | 95.02 |
| 3/15/06 | 95.02 |
| 3/16/06 | 95.22 |
| 3/17/06 | 95.2 |
| 3/20/06 | 95.2 |
| 3/21/06 | 95.195 |
| 3/22/06 | 95.11 |
| 3/23/06 | 95.065 |
| 3/24/06 | 95.135 |
| 3/27/06 | 95.13 |
| 3/28/06 | 95.03 |
| 3/29/06 | 95 |
| 3/30/06 | 94.965 |
| 3/31/06 | 94.965 |
| 4/3/06 | 94.935 |
| 4/4/06 | 94.935 |
| 4/5/06 | 94.985 |
| 4/6/06 | 94.96 |
| 4/7/06 | 94.9 |
| 4/10/06 | 94.885 |
| 4/11/06 | 94.9 |
| 4/12/06 | 94.865 |
| 4/13/06 | 94.88 |
| 4/17/06 | 94.845 |
| 4/18/06 | 94.93 |
| 4/19/06 | 94.91 |
| 4/20/06 | 94.89 |
| 4/21/06 | 94.875 |
| 4/24/06 | 94.885 |
| 4/25/06 | 94.82 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 4/26/06 | 94.78 |
| 4/27/06 | 94.87 |
| 4/28/06 | 94.905 |
| 5/1/06 | 94.85 |
| 5/2/06 | 94.84 |
| 5/3/06 | 94.805 |
| 5/4/06 | 94.765 |
| 5/5/06 | 94.795 |
| 5/8/06 | 94.77 |
| 5/9/06 | 94.78 |
| 5/10/06 | 94.76 |
| 5/11/06 | 94.765 |
| 5/12/06 | 94.78 |
| 5/15/06 | 94.795 |
| 5/16/06 | 94.825 |
| 5/17/06 | 94.77 |
| 5/18/06 | 94.785 |
| 5/19/06 | 94.75 |
| 5/23/06 | 94.77 |
| 5/24/06 | 94.795 |
| 5/25/06 | 94.78 |
| 5/26/06 | 94.8 |
| 5/30/06 | 94.775 |
| 5/31/06 | 94.705 |
| 6/1/06 | 94.705 |
| 6/2/06 | 94.825 |
| 6/5/06 | 94.755 |
| 6/6/06 | 94.73 |
| 6/7/06 | 94.71 |
| 6/8/06 | 94.78 |
| 6/9/06 | 94.77 |
| 6/12/06 | 94.77 |
| 6/13/06 | 94.77 |
| 6/14/06 | 94.665 |
| 6/15/06 | 94.635 |
| 6/16/06 | 94.62 |
| 6/19/06 | 94.61 |
| 6/20/06 | 94.605 |
| 6/21/06 | 94.605 |
| 6/22/06 | 94.555 |
| 6/23/06 | 94.525 |
| 6/26/06 | 94.515 |
| 6/27/06 | 94.535 |
| 6/28/06 | 94.515 |
| 6/29/06 | 94.59 |
| 6/30/06 | 94.61 |
| 7/3/06 | 94.585 |
| 7/5/06 | 94.525 |
| 7/6/06 | 94.54 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 7/7/06 | 94.575 |
| 7/10/06 | 94.57 |
| 7/11/06 | 94.585 |
| 7/12/06 | 94.585 |
| 7/13/06 | 94.635 |
| 7/14/06 | 94.66 |
| 7/17/06 | 94.655 |
| 7/18/06 | 94.585 |
| 7/19/06 | 94.625 |
| 7/20/06 | 94.665 |
| 7/21/06 | 94.665 |
| 7/24/06 | 94.655 |
| 7/26/06 | 94.675 |
| 7/27/06 | 94.68 |
| 7/28/06 | 94.73 |
| 7/31/06 | 94.73 |
| 8/1/06 | 94.72 |
| 8/2/06 | 94.72 |
| 8/3/06 | 94.695 |
| 8/4/06 | 94.77 |
| 8/7/06 | 94.74 |
| 8/8/06 | 94.77 |
| 8/9/06 | 94.76 |
| 8/10/06 | 94.76 |
| 8/11/06 | 94.72 |
| 8/14/06 | 94.76 |
| 8/15/06 | 94.705 |
| 8/16/06 | 94.9 |
| 8/17/06 | 94.875 |
| 8/18/06 | 94.89 |
| 8/21/06 | 94.895 |
| 8/22/06 | 94.89 |
| 8/23/06 | 94.9 |
| 8/24/06 | 94.91 |
| 8/25/06 | 94.92 |
| 8/28/06 | 94.91 |
| 8/29/06 | 94.92 |
| 8/30/06 | 94.95 |
| 8/31/06 | 95.02 |
| 9/1/06 | 95.03 |
| 9/5/06 | 95.02 |
| 9/6/06 | 95 |
| 9/7/06 | 94.995 |
| 9/8/06 | 95 |
| 9/11/06 | 94.975 |
| 9/12/06 | 94.98 |
| 9/13/06 | 94.99 |
| 9/14/06 | 94.975 |
| 9/15/06 | 94.95 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 9/18/06 | 94.935 |
| 9/19/06 | 94.98 |
| 9/20/06 | 94.945 |
| 9/21/06 | 95.135 |
| 9/22/06 | 95.165 |
| 9/25/06 | 95.195 |
| 9/26/06 | 95.17 |
| 9/27/06 | 95.195 |
| 9/28/06 | 95.145 |
| 9/29/06 | 95.28 |
| 10/2/06 | 95.15 |
| 10/3/06 | 95.15 |
| 10/4/06 | 95.19 |
| 10/5/06 | 95.15 |
| 10/6/06 | 95.09 |
| 10/10/06 | 95.04 |
| 10/11/06 | 95.015 |
| 10/12/06 | 95.015 |
| 10/13/06 | 94.94 |
| 10/16/06 | 94.95 |
| 10/17/06 | 94.965 |
| 10/18/06 | 94.97 |
| 10/19/06 | 94.95 |
| 10/20/06 | 94.94 |
| 10/23/06 | 94.9 |
| 10/24/06 | 94.92 |
| 10/25/06 | 94.965 |
| 10/26/06 | 95.04 |
| 10/27/06 | 95.25 |
| 10/30/06 | 95.255 |
| 10/31/06 | 95.335 |
| 11/1/06 | 95.41 |
| 11/2/06 | 95.38 |
| 11/3/06 | 95.165 |
| 11/6/06 | 95.16 |
| 11/7/06 | 95.24 |
| 11/8/06 | 95.275 |
| 11/9/06 | 95.275 |
| 11/10/06 | 95.31 |
| 11/13/06 | 95.285 |
| 11/14/06 | 95.325 |
| 11/15/06 | 95.24 |
| 11/16/06 | 95.22 |
| 11/17/06 | 95.28 |
| 11/20/06 | 95.29 |
| 11/21/06 | 95.29 |
| 11/22/06 | 95.3 |
| 11/24/06 | 95.315 |
| 11/27/06 | 95.32 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 11/28/06 | 95.37 |
| 11/29/06 | 95.35 |
| 11/30/06 | 95.425 |
| 12/1/06 | 95.565 |
| 12/4/06 | 95.55 |
| 12/5/06 | 95.54 |
| 12/6/06 | 95.495 |
| 12/7/06 | 95.475 |
| 12/8/06 | 95.375 |
| 12/11/06 | 95.35 |
| 12/12/06 | 95.375 |
| 12/13/06 | 95.295 |
| 12/14/06 | 95.265 |
| 12/15/06 | 95.28 |
| 12/18/06 | 95.27 |
| 12/19/06 | 95.27 |
| 12/20/06 | 95.26 |
| 12/21/06 | 95.33 |
| 12/22/06 | 95.25 |
| 12/26/06 | 95.25 |
| 12/27/06 | 95.2 |
| 12/28/06 | 95.145 |
| 12/29/06 | 95.12 |
| 1/2/07 | 95.155 |
| 1/3/07 | 95.2 |
| 1/4/07 | 95.26 |
| 1/5/07 | 95.2 |
| 1/8/07 | 95.165 |
| 1/9/07 | 95.155 |
| 1/10/07 | 95.115 |
| 1/11/07 | 95.075 |
| 1/12/07 | 95.045 |
| 1/16/07 | 95.035 |
| 1/17/07 | 94.965 |
| 1/18/07 | 94.98 |
| 1/19/07 | 94.965 |
| 1/22/07 | 94.975 |
| 1/23/07 | 94.95 |
| 1/24/07 | 94.95 |
| 1/25/07 | 94.895 |
| 1/26/07 | 94.885 |
| 1/29/07 | 94.855 |
| 1/30/07 | 94.865 |
| 1/31/07 | 94.91 |
| 2/1/07 | 94.89 |
| 2/2/07 | 94.92 |
| 2/5/07 | 94.945 |
| 2/6/07 | 94.96 |
| 2/7/07 | 94.975 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 2/8/07 | 94.975 |
| 2/9/07 | 94.925 |
| 2/12/07 | 94.895 |
| 2/13/07 | 94.89 |
| 2/14/07 | 94.965 |
| 2/15/07 | 95.005 |
| 2/16/07 | 95.005 |
| 2/20/07 | 95.02 |
| 2/21/07 | 94.99 |
| 2/22/07 | 94.945 |
| 2/23/07 | 94.98 |
| 2/26/07 | 95.025 |
| 2/27/07 | 95.22 |
| 2/28/07 | 95.195 |
| 3/1/07 | 95.195 |
| 3/2/07 | 95.26 |
| 3/5/07 | 95.255 |
| 3/6/07 | 95.255 |
| 3/7/07 | 95.305 |
| 3/8/07 | 95.28 |
| 3/9/07 | 95.15 |
| 3/12/07 | 95.18 |
| 3/13/07 | 95.29 |
| 3/14/07 | 95.255 |
| 3/15/07 | 95.215 |
| 3/16/07 | 95.175 |
| 3/19/07 | 95.14 |
| 3/20/07 | 95.155 |
| 3/21/07 | 95.26 |
| 3/22/07 | 95.195 |
| 3/23/07 | 95.155 |
| 3/26/07 | 95.175 |
| 3/27/07 | 95.18 |
| 3/28/07 | 95.2 |
| 3/29/07 | 95.17 |
| 3/30/07 | 95.16 |
| 4/2/07 | 95.155 |
| 4/3/07 | 95.125 |
| 4/4/07 | 95.13 |
| 4/5/07 | 95.11 |
| 4/6/07 | 94.985 |
| 4/9/07 | 94.97 |
| 4/10/07 | 94.99 |
| 4/11/07 | 94.98 |
| 4/12/07 | 94.975 |
| 4/13/07 | 94.945 |
| 4/16/07 | 94.955 |
| 4/17/07 | 95.005 |
| 4/18/07 | 95.03 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 4/19/07 | 95.03 |
| 4/20/07 | 95.015 |
| 4/23/07 | 95.03 |
| 4/24/07 | 95.055 |
| 4/25/07 | 95.03 |
| 4/26/07 | 94.985 |
| 4/27/07 | 94.975 |
| 4/30/07 | 95.02 |
| 5/1/07 | 94.985 |
| 5/2/07 | 94.98 |
| 5/3/07 | 94.94 |
| 5/4/07 | 94.955 |
| 5/7/07 | 94.96 |
| 5/8/07 | 94.96 |
| 5/9/07 | 94.92 |
| 5/10/07 | 94.945 |
| 5/11/07 | 94.94 |
| 5/14/07 | 94.925 |
| 5/15/07 | 94.925 |
| 5/16/07 | 94.93 |
| 5/17/07 | 94.9 |
| 5/18/07 | 94.87 |
| 5/21/07 | 94.88 |
| 5/22/07 | 94.865 |
| 5/23/07 | 94.855 |
| 5/24/07 | 94.855 |
| 5/25/07 | 94.855 |
| 5/29/07 | 94.835 |
| 5/30/07 | 94.835 |
| 5/31/07 | 94.815 |
| 6/1/07 | 94.785 |
| 6/4/07 | 94.79 |
| 6/5/07 | 94.775 |
| 6/6/07 | 94.79 |
| 6/7/07 | 94.77 |
| 6/8/07 | 94.775 |
| 6/11/07 | 94.775 |
| 6/12/07 | 94.755 |
| 6/13/07 | 94.74 |
| 6/14/07 | 94.74 |
| 6/15/07 | 94.745 |
| 6/18/07 | 94.755 |
| 6/19/07 | 94.775 |
| 6/20/07 | 94.765 |
| 6/21/07 | 94.775 |
| 6/22/07 | 94.805 |
| 6/25/07 | 94.825 |
| 6/26/07 | 94.81 |
| 6/27/07 | 94.815 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 6/28/07 | 94.785 |
| 6/29/07 | 94.81 |
| 7/2/07 | 94.8 |
| 7/3/07 | 94.79 |
| 7/5/07 | 94.77 |
| 7/6/07 | 94.765 |
| 7/9/07 | 94.765 |
| 7/10/07 | 94.79 |
| 7/11/07 | 94.79 |
| 7/12/07 | 94.785 |
| 7/13/07 | 94.785 |
| 7/16/07 | 94.79 |
| 7/17/07 | 94.785 |
| 7/18/07 | 94.795 |
| 7/19/07 | 94.795 |
| 7/20/07 | 94.82 |
| 7/23/07 | 94.805 |
| 7/24/07 | 94.815 |
| 7/25/07 | 94.825 |
| 7/26/07 | 94.94 |
| 7/27/07 | 94.94 |
| 7/30/07 | 94.935 |
| 7/31/07 | 94.935 |
| 8/1/07 | 94.965 |
| 8/2/07 | 94.965 |
| 8/3/07 | 95.045 |
| 8/6/07 | 95.045 |
| 8/7/07 | 94.965 |
| 8/8/07 | 94.94 |
| 8/9/07 | 95.1 |
| 8/10/07 | 95.165 |
| 8/13/07 | 95.135 |
| 8/14/07 | 95.18 |
| 8/15/07 | 95.305 |
| 8/16/07 | 95.47 |
| 8/17/07 | 95.455 |
| 8/20/07 | 95.485 |
| 8/21/07 | 95.505 |
| 8/22/07 | 95.415 |
| 8/23/07 | 95.365 |
| 8/24/07 | 95.29 |
| 8/27/07 | 95.32 |
| 8/28/07 | 95.41 |
| 8/29/07 | 95.41 |
| 8/30/07 | 95.465 |
| 8/31/07 | 95.4 |
| 9/4/07 | 95.435 |
| 9/5/07 | 95.465 |
| 9/6/07 | 95.425 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 9/7/07 | 95.56 |
| 9/10/07 | 95.565 |
| 9/11/07 | 95.5 |
| 9/12/07 | 95.485 |
| 9/13/07 | 95.425 |
| 9/14/07 | 95.405 |
| 9/17/07 | 95.41 |
| 9/18/07 | 95.585 |
| 9/19/07 | 95.59 |
| 9/20/07 | 95.56 |
| 9/21/07 | 95.56 |
| 9/24/07 | 95.555 |
| 9/25/07 | 95.615 |
| 9/26/07 | 95.595 |
| 9/27/07 | 95.595 |
| 9/28/07 | 95.585 |
| 10/1/07 | 95.56 |
| 10/2/07 | 95.56 |
| 10/3/07 | 95.55 |
| 10/4/07 | 95.54 |
| 10/5/07 | 95.47 |
| 10/9/07 | 95.435 |
| 10/10/07 | 95.435 |
| 10/11/07 | 95.435 |
| 10/12/07 | 95.405 |
| 10/15/07 | 95.395 |
| 10/16/07 | 95.425 |
| 10/17/07 | 95.49 |
| 10/18/07 | 95.545 |
| 10/19/07 | 95.605 |
| 10/22/07 | 95.6 |
| 10/23/07 | 95.6 |
| 10/24/07 | 95.66 |
| 10/25/07 | 95.66 |
| 10/26/07 | 95.645 |
| 10/29/07 | 95.61 |
| 10/30/07 | 95.61 |
| 10/31/07 | 95.575 |
| 11/1/07 | 95.605 |
| 11/2/07 | 95.62 |
| 11/5/07 | 95.62 |
| 11/6/07 | 95.625 |
| 11/7/07 | 95.64 |
| 11/8/07 | 95.675 |
| 11/9/07 | 95.69 |
| 11/13/07 | 95.65 |
| 11/14/07 | 95.635 |
| 11/15/07 | 95.685 |
| 11/16/07 | 95.675 |

FFZ7 COMB Comdty

| Date | PX_LAST |
|---|---|
| 11/19/07 | 95.685 |
| 11/20/07 | 95.68 |
| 11/21/07 | 95.705 |
| 11/23/07 | 95.695 |
| 11/26/07 | 95.715 |
| 11/27/07 | 95.7 |
| 11/28/07 | 95.715 |
| 11/29/07 | 95.76 |
| 11/30/07 | 95.78 |
| 12/3/07 | 95.78 |
| 12/4/07 | 95.78 |
| 12/5/07 | 95.755 |
| 12/6/07 | 95.75 |
| 12/7/07 | 95.735 |
| 12/10/07 | 95.745 |
| 12/11/07 | 95.74 |
| 12/12/07 | 95.735 |
| 12/13/07 | 95.73 |
| 12/14/07 | 95.72 |
| 12/17/07 | 95.72 |
| 12/18/07 | 95.725 |
| 12/19/07 | 95.735 |
| 12/20/07 | 95.76 |
| 12/21/07 | 95.745 |
| 12/24/07 | 95.735 |
| 12/26/07 | 95.74 |
| 12/27/07 | 95.74 |
| 12/28/07 | 95.74 |
| 12/31/07 | 95.755 |