



BANK OF CANADA
BANQUE DU CANADA

CELEBRATING 75 YEARS
CÉLÉBRONS 75 ANS

Working Paper/Document de travail
2010-31

Text Mining and the Information Content of Bank of Canada Communications

by Scott Hendry and Alison Madeley

Bank of Canada Working Paper 2010-31

November 2010

Text Mining and the Information Content of Bank of Canada Communications

by

Scott Hendry and Alison Madeley

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
shendry@bankofcanada.ca

Bank of Canada working papers are theoretical or empirical works-in-progress on subjects in economics and finance. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada.

Acknowledgements

We gratefully acknowledge comments provided by Christine Fay, Toni Gravelle, Wei Wang, and seminar participants at the Bank of Canada and the 2010 Northern Finance Association Meetings.

Abstract

This paper uses Latent Semantic Analysis to extract information from Bank of Canada communication statements and investigates what type of information affects returns and volatility in short-term as well as long-term interest rate markets over the 2002-2008 period. Discussions about geopolitical risk and other external shocks, major domestic shocks (SARS and BSE), the balance of risks to the economic projection, and various forward looking statements are found to significantly affect market returns and volatility, especially for short-term markets. This effect is over and above that from the information contained in any policy interest rate surprise.

JEL classification: G14, E58

Bank classification: Financial markets; Monetary policy implementation

Résumé

Les auteurs utilisent l'analyse sémantique latente pour extraire l'information que renferment les communications de la Banque du Canada et ensuite examiner quel type d'information a influencé les rendements et la volatilité sur les marchés de taux d'intérêt à court et à long terme durant la période 2002-2008. Ils constatent que les déclarations prospectives de la Banque et ses analyses du risque géopolitique et d'autres chocs extérieurs, d'importants chocs intérieurs (p. ex., l'épidémie de SRAS et la découverte d'un cas de maladie de la vache folle) et de la résultante des risques entourant sa projection économique ont une incidence significative sur les rendements et la volatilité, surtout dans le cas des marchés à court terme. Cet effet est distinct de celui lié à la diffusion de toute information inattendue concernant le taux directeur.

Classification JEL : G14, E58

Classification de la Banque : Marchés financiers; Mise en œuvre de la politique monétaire

1.0 Introduction

The Bank of Canada takes great care when crafting its monetary policy communications with the public. But what specific focus or topics of these communications are perceived by financial markets to be the most significant? The goal of this paper is to use advanced text mining techniques to extract the important messages from Bank of Canada communications and investigate whether these messages have a significant impact on Canadian interest rates over and above the effect of the monetary policy surprise itself. This work further advances the literature investigating the types of information that moves markets as part of the price discovery process.

We find that it is possible to extract, from Bank of Canada communications related to Fixed Announcement Dates (FADs) and Monetary Policy Reports (MPRs), meaningful measures of information that significantly move market returns and volatility. These effects are stronger for short-term rates and there seems to be a greater effect on volatility than on returns. Over the 2002 to 2008 period, Bank of Canada communication focusing on major shocks hitting the Canadian economy, the balance of risks to the economic projection, and certain forward looking statements had the most significant market impact. This is in addition to any effects on markets coming from target interest rate changes that are actually made.

Modern text mining techniques offer a valuable tool for the extraction of information from text documents. These techniques are designed to yield numerical representations of text documents for several potential uses. Information retrieval, information extraction, document classification, and document comparison are some of the more important uses of text mining. Internet search engines are advanced examples of text mining techniques used for information retrieval. A user supplies a few keywords and the search engine retrieves from a vast collection of documents or web pages those which are most similar.

There is little use of advanced text mining in the economics literature. There is, however, significant use of these techniques by hedge funds and other institutional investors to build algorithmic trading strategies. Reuters, Dow Jones, RTS Realtime Systems Group and other news services now offer real time streaming of their information stories so that algorithmic traders can build text mining models to process the data and design trading strategies. These services are quite expensive illustrating the great value that traders place on text mining techniques to extract meaningful information from text news stories. Some companies (see for example Infonic with its product Sentiment which powers the Reuters NewsScope Sentiment Engine, its market sentiment product) even offer a text mining services to provide real time classification of news stories into various sentiment categories.

Price discovery is the process through which relevant information is incorporated into an asset's price. Most research on this process only examines the information content of macro news announcements. Early literature in this field typically only allowed for the fact that there was or was not a macro news announcement on a particular day. Further research began to measure the information flow with the announcement surprise, i.e. the difference between the actual outcome and the anticipated outcome for a particular announcement. This still leaves a large gap in our understanding of what types of information moves markets. The surprise in a macro announcement is obviously very important but it is generally not the full measure of the information contained in the data release. There is often valuable information in the underlying details and the interpretation of these details that is important to understanding the market reaction but which is not encapsulated in the overall surprise measure.

The few examples in economics that use text mining focus on interpreting the context of monetary policy communication statements, primarily in the U.S. One paper by Boukus and Rosenberg (2006) uses text mining to look at the information that can be extracted from U.S. FOMC minutes and investigates whether these information flows significantly affect market yields. They find that the volatilities of three month, two year, and ten year treasury market yields

react significantly to discussions of current and future financial market and economic conditions. Longer term rates are found to respond more significantly to the information content of the FOMC minutes than do short-term rates. Across all maturities, FOMC minutes contain significant information during periods of increased uncertainty. Our study builds on this work to examine both market return and its volatility in the Canadian case in response to the information flow from Bank of Canada monetary policy communications to determine whether these documents provide significant information to the marketplace.

Our work differs from Boukus and Rosenberg (2006) in that it examines returns as well as the volatility measure. In addition, the Canadian example offers an interesting alternative case in that the monetary policy regime and communication framework has some important differences. Canada has had an explicit and credible inflation targeting regime in place since the early 1990s which implies there is much less uncertainty surrounding the details of the longer term goals of the Bank of Canada. This should lead to different effects on the term structure of interest rates from central bank communication. The Bank of Canada has also made more frequent use of forward looking statements in a wider variety of circumstances to provide near term guidance to markets on the likely path of interest rates. It is important to determine if this guidance has a significant market impact or, possibly more importantly, if other information still has a market impact even in the presence of a forward looking statement.

Some other recent work to use automatic text mining techniques to study central bank communications include Lucca and Trebbi (2009), Bligh and Hess (2009), and Rosa (2007). Lucca and Trebbi (2009) use a technique called content analysis to classify Fed statements as hawkish or dovish and find that short-term Treasury yields respond to changes in policy rates but that long-term government bond rates mainly react to changes in FOMC policy statements. Rosa (2007) uses a different classification algorithm and finds that the variance of short-term Euribor rates responds to ECB statements regarding M3 growth and real economic activity but not to discussions of price stability. Bligh and Hess (2009) use content analysis to classify

communication from Chairman Greenspan into quantitative scores of certainty, pessimism, and macroeconomics. Use of language indicating more certainty is found to raise shorter-term interest rate futures while pessimistic language tended to lower the value of the dollar, the stock market, and near-term Federal Funds futures. Market volatility is significantly reduced by Greenspan's speeches and testimonies emphasizing macro issues.

There is other research that manually or subjectively classifies documents or news stories into positive, negative, or neutral categories depending on the occurrence of certain words.¹ These more subjective evaluations are found to be important for understanding the differing effect on markets across news stories.

Fay and Gravelle (2008) also examine Bank of Canada communication statements (FAD press releases, speeches, interviews, and news conferences) to determine which, if any, have had significant effects on market interest rates. Their analysis, however, simply looks at the occurrence of different types of Bank communications and does not classify the language used in the documents. They find that FAD press releases and, to a lesser extent, speeches significantly affect near-term interest rate expectations. These effects largely disappear, however, in the latter part of the sample (post July 2004) when the Bank of Canada began to consistently include forward looking statements to provide guidance on the future path of interest rates. They find evidence that market participants have been focusing more on the forward looking statement (FLS) and less on the Bank's discussion of the economic outlook. These FLSs provide policy rate guidance or signals to the public and are similar to the policy inclination statements used elsewhere.

Section 2 describes the methodology used in this study while section 3 outlines the data and documents used. Section 4 lays out the results of the study and the final section offers some conclusions and next steps.

¹ See, for example, work looking at U.S. Fed communications by Gurkaynak, Sack and Swanson (2005), Kohn and Sack (2004), Ehrmann and Fratzscher (2004), Zovodny and Ginther (2005), Balke and Peterson (2002), Fetting, Rolnick and Runckle (1999), and Cook and Hahn (1988).

2.0 Methodology

2.1 Introduction to Latent Semantic Analysis

Latent Semantic Analysis (LSA) is used to extract meaning or information from Bank of Canada communications. This technique emerged in the late 1980s to identify patterns in large sets of textual data. It was developed into a theory of knowledge representation by Landauer and Dumais (1997) and is now extensively used in the fields of text interpretation, information retrieval, and search engine technology but has not seen many applications to economics or finance.²

LSA was first introduced as a mechanism for improving information retrieval and is now widely used in internet search engines, automated learning software, automated essay grading, and as a basis for models of cognitive learning. LSA has undergone extensive testing in various environments and is able to perform many meaning-based cognitive tasks (i.e. tasks for which the meaning of the words must be known such as in essay grading and language tests) as well as humans do.

LSA reflects the way people interpret what they read or have heard and creates an assessment of meaning based on the author's word choice. In analyzing what words are included and excluded in a group of passages, LSA calculates a relationship of statistical occurrence which then translates into an overall interpretation.

The meaning of a passage of text, in the LSA framework, is a function of the representations of the words it contains. In other words, the meaning of a document is the sum of the meanings of its individual words. The meaning of an individual word is not explicitly defined but rather is situated with respect to all the other words used in the collection of documents. LSA

² Landauer, Foltz, and Laham (1998) provide an excellent summary of the technique and early applications. See also the Handbook of Latent Semantic Analysis (2007) for an overview of LSA as well as a description of the methodology and its applications.

ignores word order but is still able to extract meaning from a text using the frequency of occurrence of the individual words. Meaning is a relation among words so words do not have meaning in isolation; words have meaning only by virtue of their relations to other words (see Kintsch, 2007). Meaning is derived from the circumstances in which it is and is not used. The meaning of a word is a kind of average of the meaning of all the passages in which it appears and the meaning of a passage is a kind of average of the meaning of all the words it contains (Landauer, Foltz, and Laham, 1998).

Style, syntax and grammar are ignored by LSA; instead the calculations are based on the detailed patterns of occurrences of each word over a large array of written passages. It is important to keep this in mind because LSA does not look at sentence arrangement to find meaning but looks at the direct relationship between word choice and word choice within a passage.³ The principle critique of LSA is that it ignores word order and relies exclusively on word choice. Proponents of LSA, however, estimate that word choice alone accounts for 80%-90% of the meaning of a passage of text, explaining why LSA performs so well in a wide range of applications.

2.2 Applying LSA

Once all the documents are collected, the first step in implementing LSA is to preprocess the documents by removing formatting such as punctuation, capitalization, numbers, and any special characters.⁴ The next step is the removal of all *stop words*. These are typically extremely common words that occur in all or virtually all documents and add little value to the meaning of the texts. These typically include pronouns, articles, prepositions, conjunctions, auxiliary verbs, days, months, etc. The remaining words are then *stemmed*, which is the process of removing suffixes (e.g. s, es, er, est, ly, etc.) so that words sharing a common etymological root are mapped

³ Landauer, Foltz, and Laham. An Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284. 1998.

⁴ We use the TMG program written in Matlab to perform all our LSA. See D. Zeimpekis and E. Gallopoulos (2005, 2006).

into a single “stem.”⁵ Finally, as part of the preprocessing stage, only words that occurred in two or more documents were included. This removed extremely rare words thereby permitting a clearer focus on the underlying themes that ran through the documents.

The next step is to organize the group of documents into a matrix (X) where each row represents a word (t) and each column is a document (d). Each cell of the matrix is the weighted frequency (x_{ij}) of row t 's word in column d 's document.

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix}$$

The X matrix is an mxn matrix where m is the total number of words and n is the total number of documents. Typically, m is much larger than n but this need not be true. The matrix is also usually sparse with most entries being zero so that each document contains only a small subset of the words from the dictionary.

The next processing stage requires both local and global weights to be calculated for each non-zero element of the X matrix. The purpose of this is to either increase or decrease the importance of each term within documents and across all documents. This helps to distinguish the documents from each other and allows a better representation of the most important words by giving low weight to high-frequency terms (occur in many documents) and high weight to low-frequency terms (occur in some documents). We now have

$$x_{td} = \text{local}(t,d) * \text{global}(t)$$

where $\text{local}(t,d)$ is the local weighting for word t in document d , and $\text{global}(t)$ is the word's global weighting.

⁵ The TMG program did not perform as much stemming as desired so a pre-processing step was added to do additional stemming. For instance, “strong” and “strength” were stemmed to the same root.

The within-document relative frequency calculation is the most common example of a local term weighting scheme⁶ where the more often a term appears within a particular document, the more it contributes to the content of that document.

$$local(t, d) = \frac{f_{t,d}}{\max_t \{f_{t,d}\}}$$

$local(t, d)$ for a particular term t in a particular document d is calculated by dividing the frequency $f_{t,d}$ of term t in document d by the highest frequency across all the words in document d .

The most common global term weighting function is the Inverse Document Frequency (IDF), whereby the more documents in which a term occurs, the less discriminating it is.⁷

$$global(t) = \log\left(\frac{n}{n_t}\right)$$

This global term frequency is the logarithm of the inverse of the document frequency, the number of documents in which word t occurred (n_t) divided by the number of all documents in the collection (n).

Finally, to ensure that longer documents do not dominate the results, the documents are normalized to adjust for differences in length. There are different normalizations used but the most common is cosine normalization. In the end, we have the following for each element of the X matrix.

$$x_{td} = \frac{local(t, d) * global(t)}{\sqrt{\sum_{t=1}^T (local(t, d) * global(t))}}$$

2.3 Singular Value Decomposition

⁶ Alternative local weighting schemes include binary frequency (1 if the word is in the document, 0 otherwise) and log of word frequency plus one.

⁷ Alternative global weighting schemes include normal, log entropy, and gfidf.

SVD is vital to the success of LSA since it is what creates or maps the relationships between the words across the documents taking account of not only where words are used but where they are not used. SVD is a generalized form of principal component analysis in which any arbitrary rectangular matrix can be exactly decomposed into the product of three characteristic matrices: $X=USV^T$ (see Figure A). The columns of U ($m \times n$) and V ($n \times n$), referred to as the left and right singular vectors of X , represent the orthonormal eigenvectors of XX^T and $X^T X$, respectively. S ($n \times n$) is a diagonal matrix of singular values where the diagonal entries are the non-negative square roots of the eigenvalues of XX^T and $X^T X$, ordered by decreasing size. SVD yields a unique decomposition of X up to the sign of the singular vectors in U and V .

The U matrix describes words (m rows) as vectors of the derived orthogonal factor values (n columns) while the V matrix describes the documents (n rows) as vectors of the same factors (n columns). These factors may be thought of as underlying concepts or themes that run through the documents. The meaning of each word or document can then be characterized by a vector of weights indicating the importance of each of these underlying themes. The i,j^{th} element of U is the contribution or importance of the i^{th} word to the meaning of the j^{th} theme. Similarly, the i,j^{th} element of V represents the importance of the j^{th} theme in the i^{th} document. The S matrix represents the importance of each theme for explaining the variance of meaning across the documents. With the elements of S ordered by decreasing magnitude, the first theme is thus the most important underlying concept spanning the set of documents. Each successive theme accounts for less of the subject or meaning variation across the documents. If the documents are ordered chronologically, then a column of V represents the importance of each theme in a given document while a row of V defines a time series of the importance of a given theme. It is these rows of V that define the time series of information measures that will be investigated for their importance in moving financial markets.

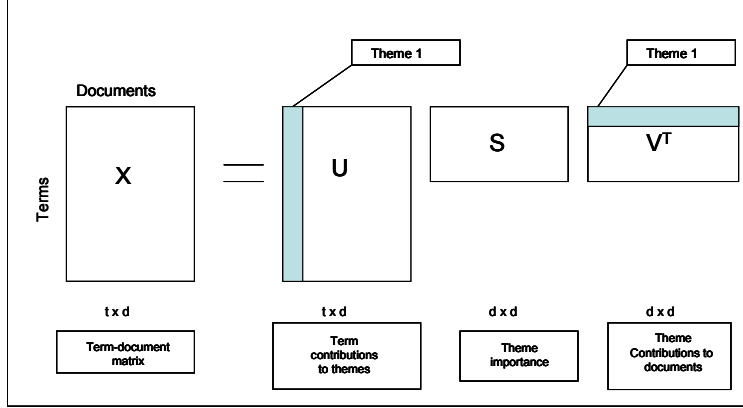
It should be noted that most applications of LSA do not attempt, or even require, interpretation of the themes. LSA typically is used to create numerical representations of

documents (vectors) that are used to find similarities between documents (e.g. to find all documents on a similar topic, to find all existing documents closest to the “document” represented by a few keywords typed into an internet search engine) although this is a rather simplistic view of what a very powerful tool actually achieves. Our study investigates whether extracted themes move markets and then attempts to offer some interpretation of the ideas those themes may represent by examining the most important words in each theme.

As mentioned earlier, SVD is unique only up to the sign of vectors in U and V . In order to offer interpretation, however, to the extracted themes the sign will need to be fixed. This is achieved using the methodology outlined in Bro, Acar, and Kolda (2007).

LSA does not use the full decomposition of the X matrix. Instead, a k -dimensional approximation is found by using only the first k columns of U and V and the $k \times k$ upper left matrix of S . This approximation, $\hat{X}_k = U_k S_k V_k^T$, of X is found to remove much useless “noise” or extraneous information that is in the document set and focuses on only those factors explaining the important variation in meaning across documents. The matrix \hat{X}_k is the least-squares best fit of X . Performing the SVD and truncating it is what captures the important underlying semantic structure of the words and documents while excluding the noise. Words that have similar meanings are found to be “near” each other in k -dimensional vector space even if they never occur in the same documents. Similarly, documents that have similar conceptual meaning are near each other even if they share no words in common. The selection of k depends on the sample of words and documents being used. Typically a very large number of documents and words are being used and k is set in the 100-300 range. Given the small sample of documents used in this example, however, much smaller values for k ($k=10$ or 20 were used but only 10 themes were used in the end).

FIGURE A



2.4 Interest Rate Model

To identify the components of Bank of Canada communications that move markets, we estimate the following regression model following the methodology used by Anderson et al (2007) to examine dynamic news effects in global stock, bond, and foreign exchange markets.

The first step involves estimating the following return equation using OLS.

$$R_t = \beta_0 + \beta_1 R_{t-1} + \beta_2 F_t + \sum_{k=1}^K \rho_k S_{k,t} + \beta_3 FAD_t + \beta_4 MPR_t + \sum_{i=1}^I \gamma_i^{pr} T_{i,t}^{pr} + \sum_{i=1}^I \gamma_i^{mpr} T_{i,t}^{mpr} + \varepsilon_t \quad (1)$$

where R_t is the five-minute return of either the 90-day BAX or the CGB, F_t is the five-minute order flow, $S_{k,t}$ is the standardized surprise in macro announcement k , FAD_t is a dummy with a value of one for the five-minute interval after a 9:00am FAD press release, MPR_t is a dummy with a value of one for the five-minute interval after the 10:30am release of an MPR, $T_{i,t}^{pr}$ is the i^{th} theme for press releases, $T_{i,t}^{mpr}$ is the i^{th} theme for MPRs, and ε_t is the residual. Themes from press releases and MPRs enter the regressions separately so that we can investigate whether these two communication products have differential effects on interest rates.

While this estimate is consistent, the disturbances are likely to be highly heteroskedastic. To improve the efficiency of the estimates a weighted least squares (WLS) model is used. This involves taking the absolute value of the residuals from the OLS estimate of the conditional mean

equation above and estimating a time-varying volatility function which is then used to perform a WLS estimation of equation (1). The volatility model fitted to the first stage residuals is

$$|\hat{\varepsilon}_t| = \sum_{j=1}^J \alpha_0 C_j + \alpha_1 |\hat{\varepsilon}_{t-1}| + \sum_{k=1}^K \sum_{m=-1}^1 \lambda_k D_{k,t+m} + \sum_{m=-4}^1 \alpha_3 FAD_{t+m} + \alpha_4 MPR_t + \sum_{i=1}^I \phi_i^{pr} T_{i,t}^{pr} + \sum_{i=1}^I \phi_i^{mpr} T_{i,t}^{mpr} + \omega_t \quad (2)$$

where C_j is a time-of-day dummy to capture intraday patterns in volatility⁸, $D_{k,t}$ is a dummy for the five-minute interval in which macro news announcement k occurs, and ω_t is the residual.

The five-minute interval was chosen as the object of focus, not because the Bank of Canada is interested in controlling or affecting asset prices at such a high frequency, but because this will best allow for the proper identification or isolation of the effect of monetary policy communication on interest rates. Lower frequency analysis, e.g. hourly or daily, would leave us less confident that the true effect of communication on interest rates has been identified but it would give a better idea about whether the effects of communication were transitory or persistent. This will be the focus of future work.

3.0 Data

3.1 Bank of Canada Text Documents

The data used in our analysis is composed of 95 Bank of Canada text documents from two different sources; interest rate press releases and Monetary Policy Reports and Updates. Our sample of Bank of Canada interest rate press releases spans the 64 Fixed Announcement Dates (FAD) covering the period from their inception in December 2000 to September 2008. Prior to December 2000, the Bank of Canada announced adjustments to its target interest rate as needed throughout the year via a press release which usually contained a brief explanation as to the reasons behind the decision. Beginning December 2000, the Bank of Canada released a schedule

⁸ Dummies are included for each five-minute interval from 8:25am to 9:00am and for each half-hour interval from 9:00am to 3:00pm. As such, there are $J=19$ subintervals each day.

of eight fixed announcement dates over the calendar year when the Bank would set its target overnight interest rate until the next date. The Bank has also committed to provide a press release to explain its decision whether it be an increase, decrease, or no change. The Bank of Canada still has the option to act between FAD's, but will do so only in extraordinary circumstances. The new FAD process was put in place to provide a more regular, transparent process of public communications on monetary policy.⁹

The interest rate press releases are posted at 9am on the day of the FAD, usually a Tuesday, and contain the decision, an explanation of the decision and, possibly, a forward looking statement giving a broad flavour for the future path of rates. We deleted the beginning of each press release document which simply states the title, the rate action taken, and the city in which the decision took place.

The Monetary Policy Report (MPR) and Update (MPRU) are both published twice a year, providing 31 documents from January 2001 to July 2008. The full reports are published in April and October, while the updates are published in January and July and are shorter in length. The MPR and MPRUs outline a detailed account of the Bank of Canada's policies and strategies. They shed light on the economic climate and its implications for inflation and often discuss press releases, speeches and publications that had been made in the months prior. We removed the bibliography section when processing the documents.

3.2 Pre-Processing Text Documents

As described above, the first pre-processing step is to remove all formatting, numbers, and special characters. Next, certain phrases are combined into a single term so as to maintain their meaning in the face of different representations. For example, "Bank of Canada" and "Bank" were converted to "bankofcanada". Next, all the stop words are removed. This takes out

⁹ Bank of Canada. "Bank of Canada Releases Dates for Announcing Bank Rate Actions". October 2000. www.bankofcanada.ca

such common words as *the, is, a, and*, as well as days of the week and months. We also removed words that added little or no value to the meaning of the documents from a monetary policy context. These included proper names of Bank of Canada staff as well as quarter and year references (e.g. 1990Q1). Finally, the remaining words are stemmed to remove suffixes so that such words as *slow, slower, slowest, slows, slowly*, and *slowing* are all reduced to *slow*. Next, the various possible references to certain entities were stemmed into a single “word”. For instance, all the references to the Bank of Canada (e.g. “Bank of Canada” and “Bank”) were converted to *bankofcanada*. Similarly, all the possible versions of U. S. (e.g. “U. S.”, “U. S. A.”, “USA”, and “United States”) were converted to *unitedstates*.

To help with the interpretation of the extracted themes, the basic set of stop words was aggressively expanded to remove words that did not seem to contribute importantly to the interpretation of the themes. In addition, a version of the documents was generated that particularly identified the most important words in the forward looking statements that have consistently appeared in Bank communication since mid-2004.¹⁰ This permits the LSA methodology to attribute special meaning to these words and phrases different from what it will when they appear elsewhere in the documents. For example, the FAD press release from April 26, 2006 contained the forward looking statement “some modest further increase in the policy interest rate may be required to keep aggregate supply and demand in balance and inflation on target over the medium term”. Typically, the stop list removes the words *may, be, and some* and gives the same meaning to “increase” as it will when it’s used in “increase in real growth“. To ensure that the forward looking statement receives special focus and profile as it does with market participants, the stop list is prevented from removing *may, be, and some* when they appear in the FLS, special characters are added to such words as *modest, further, and increase* to ensure they are treated differently from when they appear elsewhere, and the phrases “some modest”, “may

¹⁰ See Fay and Gravelle (2008) for a fuller discussion of forward looking statements and their use by the Bank of Canada.

be required”, and “medium term” are combined into a single word (as we did with bankofcanada) because it is their use together that gives them special meaning. Themes derived for the base case are denoted as Model 1, while themes derived from documents with special identification of the FLSs are denoted Model 2. Finally, Model 3 derives themes while identifying FLSs, and by separating press releases and MPR(U)s into separate sets of documents before applying LSA. Deriving themes from press releases and MPR(U)s simultaneously in Models 1 and 2 will provide more data to the LSA technique which should better identify common themes running across the two communication products. In contrast, extracting themes separately from press releases and MPR(U)s in Model 3 will potentially suffer more small-sample problems but will better highlight the important themes that may be different across the two products.

3.3 Interest Rate Data

Intraday tick-by-tick quote and transaction data on the 10-year Government of Canada Bond futures (CGB) contract and the three-month Canadian Bankers’ Acceptance futures (BAX) contract were obtained from the Montreal Exchange where they are traded. These two markets represent the most actively traded interest rate markets in Canada and the best intraday data that is available. This data covers the period from January 2002 to June 2008. There were approximately 1114 trades/day for the BAX contract and 1674 trades/day for the CGB contract. Only days with a full intra-day sample were used. The ten-year futures contract has been identified by asset managers as an important method of hedging anticipated interest rate moves and often offers portfolio enhancement opportunities.¹¹ As a benchmark for Canadian short-term interest rates, the BAX futures are used for hedging over-the-counter derivatives and were the first interest rate contracts to be listed on the Montreal Exchange.¹²

¹¹ Montreal Exchange. “CGB (Ten-year Government of Canada Bond Future)”. Reference Manual. Bourse de Montreal Inc. October 2001.

¹² Montreal Exchange. “BAX – Three-Month Canadian Bankers’ Acceptance Futures”. Reference Manual. Bourse de Montreal Inc. December 2007

From both data sets, we extracted the front contract's midpoint of the bid and ask prices for the end of each five minute interval. From this we calculated five-minute returns.¹³ The raw tick data did not identify buys and sells so these were imputed according to the following algorithm:

- if trade price=ask (bid) price that just disappeared and trade size=ask (bid) quote size that just disappeared then trade=buy (sell);
- if trade price=current ask (bid) price and trade size=current reduction in ask (bid) quote size then trade=buy (sell).

All but a very few of the trades could be signed using this algorithm. From this, order flow was calculated as the sum of the number of buys for a five-minute interval less the number of sells.

3.4 Macro News Announcements

The expected value and actual outcome for the following macro news announcements were collected from Bloomberg: Bank of Canada target rate, international securities transactions, Ivey purchasing managers index, and month over month growth in Building Permits, GDP, retail sales, and wholesale sales.¹⁴ The Bank of Canada target overnight rate is released at 9:00am, the Ivey purchasing managers index at 10:00am, and the rest at 8:30am. Standardized surprise values are calculated as $(Act_t - Exp_t) / Var_t$ where Act_t is the actual value, Exp_t is the expected value, and Var_t is the variance of the difference between them over the full sample of news releases. Other news announcements were investigated but were not found to be significant in any specification. The Bank of Canada interest rate surprise was included so that we can separate the information

¹³ Boukus and Rosenberg (2006) analyse the 20 minute change in treasury yields from five minutes before the FOMC minutes release to 15 minutes after. We choose to focus on five-minute returns to better isolate the impact of the policy communications on markets. Focusing on the higher frequency data will better isolate the impact but is less able to determine whether these effects are persistent.

¹⁴ A total of 24 news announcements were investigated initially (including also CPI, GDP, unemployment, housing starts, current account, merchandise shipments, and others) but only these 6 showed any consistent significance across the various specifications examined.

contained in the actual interest rate decision (i.e. the surprise) from the information contained in the rest of the press release as measured by the themes extracted by LSA.

4.0 Results

4.1 Interpreting Themes

In order to offer some interpretation to the various themes, it is necessary to closely examine the U matrix derived by the SVD to determine which are the most important words in each theme. The size of an element of the U matrix represents the importance of that word in defining the particular theme. Figures 1 to 8 plot the top 10 words (i.e. value in the U matrix) for each of the top 10 themes in the three models. Figures 9 to 20 plot the importance of these top 10 themes for each of the documents (i.e. the first 10 rows of the V matrix) in a time series format. They are plotted separately for press releases and MPRs. In addition, Table 1 displays the largest correlations (>0.3 in magnitude) between the themes and the price of oil, the US/CAD exchange rate (i.e. an increase is an appreciation), GDP growth, inflation, and the actual change in the target overnight rate between the current date and four FADs ahead. These correlations will arise, not as a result of the LSA technique, but rather through the Bank of Canada's choice of particular words during certain macroeconomic conditions.

Table 2 shows the percent contribution of each theme to the overall variation of meaning across the documents. By construction, the themes are ordered by their contribution. The first 50 themes (just over half) of Models 1 and 2 account for 81% and 82%, respectively, of the variance. For Model 3, half of the FAD press release themes account for about 72% of the variation while the first half of the MPR(U) themes account for only 55%. The top 10 themes used below in the interest rate regressions account for about 34% of the variation for Models 1, 2, and press release part of Model 3. The top 10 themes of the MPR(U)s in Model 3 account for 53% of the variation of meaning across these documents.

Theme 1 of Model 1 (Figure 1) emphasizes words like *GDP*, *real*, *labour*, *financ(e)*, *credit*, and *busi(ness)*. This theme represents factors affecting GDP growth, primarily those coming from the real side but also including variables such as credit. The theme is correlated with interest rate declines over the subsequent FADs as well as with the exchange rate and the price of oil (see Table 1). This first theme, by construction, accounts for more of the variation of meaning, about 10%, across the documents than any other theme (see Table 2). Relative to the other themes, Theme 1 contributes a much more stable amount to meaning variation across the documents (see Figures 9 (PRs) and 11 (MPRs)).

Theme 2 of Model 1 (Figures 1, 9 (PRs), and 11 (MPRs)) emphasizes the words *project(ion)*, *risk*, *upsid(e)*, *downsid(e)*, *tilt*, and *balance* with positive weight and *geopolitical*, *confid(ence)*, *uncertain*, *terrorist*, and *attack* with negative weight. The positive side represents discussion about the risks to the projection and the overall balance of risks. Note that the theme does not differentiate between the treatment of upside and downside risks given both words are present with nearly equal weight. The negative side refers to geopolitical risk following the September 2001 terrorist attacks which resulted in significant uncertainty and low levels of confidence. The positive words are emphasized in documents for which the theme is positively weighted (see Figure 1) and vice versa. This theme is strongly correlated with the price of oil and the USD/CAD exchange rate and weakly correlated with the growth rate of GDP.

Theme 3 (Figures 1, 9 (PRs), and 11 (MPRs)) highlights the words *credit*, *employ*, *unitedst(ates)*, *slowdown*, and *condition* with positive weight and *oil* and *rais(e)* with negative weight. This theme is correlated with interest rate declines and achieves its highest weight (see Figure 9 (PRs)) during the financial crisis period post-July 2007. This theme represents discussion of a U.S. slowdown, especially in employment, based primarily on deteriorating and tight credit conditions.

Possible explanations can be offered for the remaining themes in a similar fashion. For example, theme 4 (Figures 1, 9 (PRs), and 11 (MPRs)) emphasizes the negative effects of *SARS*,

BSE, and the *appreciation* of the *dollar*. Theme 5 discusses an economy that's at *full* capacity with strong *momentum* to its economic *recovery* and the resulting need to take *action* by *raising* rates.

In Model 2, the significant words in the forward looking statements (FLS) giving guidance on the future direction of policy rates were treated differently so they were not removed by the stop list and were not grouped with instances of the same words outside the FLS. The first two themes for Model 2 (Figures 3, 13 (PRs), and 15 (MPRs)), emphasizing real factors and risks to the projection, were essentially the same as for Model 1. The next four themes, however, emphasized different FLS that the Bank has used over the sample period. Theme 3 focused on the FLS which used wording like: *a reduction of monetary stimulus will be required*. This theme was positively correlated with GDP growth. Theme 4 emphasized the FLS *some modest further increase* in rates *may be* required. In addition, this theme highlights an appreciating Canadian dollar and is positively correlated with GDP growth. The negative weighted words in this theme focused on the FLS saying that *a reduction* in monetary stimulus *will be* required. The positive side of theme 5 emphasized the same FLS but on the negative side focused on *SARS*, *BSE*, and other risks to economic *prospects*, especially those working through the *dollar*. The next five themes (Figures 4, 14 (PRs), and 16 (MPRs)) emphasize other FLS (theme 6), geopolitical risk and uncertainty in the Middle East (theme 7), the effects of the terrorist attacks of 2001 (theme 8), economic recovery and high employment (theme 9), and high energy prices but appropriately accommodative monetary policy (theme 10).

Model 3 extracts themes from the FAD press releases and MPRs separately (Figures 5-8, 17-18 (PRs), and 19-20 (MPRs)). The two sets of themes look quite different despite the MPRs generally being simply an expansion of the main points highlighted by the FAD press releases. For the press releases, theme 1 (Figures 5 and 17) emphasizes the *balance* of *risks* to the *projection* coming primarily from a *high dollar* and the US *hous(ing)* sector (positive weights) as well as low *confidence* levels and high *uncertainty* associated with *geopolitical* risks (negative

weights). This theme is strongly and positively correlated with the price of oil and the CAD exchange rate and has been growing from its low point early in the sample. Theme 2 emphasizes the FLS stating that a *reduction in monetary stimulus will be required* (positive weights). On the negative side, the theme highlights the poor *credit conditions* in the *United States* and the *slowdown* in the *housing sector* that has occurred during the financial crisis starting in 2007. This theme is correlated with interest rate increases and GDP growth. Theme 3 emphasizes high oil prices (positive weights) as well as the FLS calling for a *reduction in monetary stimulus* and the discussion of the *upside* and *downside risks* to the forecast (negative weights). Theme 4 talks about the depreciation of the U.S. *dollar/currency* and its effects on Canadian *prospects* (positive weights) as well as the need to *raise* rates because the economy is operating at *full* capacity. Theme 5 highlights *rising oil* prices and *geopolitical* uncertainty (positive weights) along with low *business confidence* levels and the need for *some modest increase* in rates (negative weights).

Turning to the themes extracted from the MPRs, theme 1 (Figures 7 and 19) emphasizes the *GST*, *CERI*¹⁵, *credit*, and two of the FLSs (*some modest further increase* in rates is required, the current level of the overnight rate is *consistent* with inflation target) with positive weights as well as *SARS*, *BSE*, and *terrorist attacks* (negative weights). This theme was negative in the first half of the sample, positive for the second, and is highly correlated with the exchange rate and the price of oil. The next several themes, despite being orthogonal by design, do seem to emphasize some of the same concepts. *SARS*, *BSE*, and the *terrorist attacks* of 2001 figure prominently, as do discussions of *credit*, the *Bank rate*, and relative price movements associated with strong growth in *insurance premiums*. The similarity across these themes may be an indication of the relatively small sample size involved (31 MPRs and MPRUs). The jagged nature of some of the time series plots of the themes indicate that the methodology may also be having some difficulty in dealing with the substantial difference in size between MPRs and MPR Updates.

¹⁵ CERI stands for Canadian exchange rate index.

This exercise shows that it will only be possible to give a flavour of interpretation to the themes derived from the Bank's communications. There is still a lot of overlap between themes despite them being derived in a manner which ensures that they are orthogonal. This is probably due to the fact that there is only a relatively small sample of documents available and their content is quite consistent.

4.2 Interest Rate Effects of Themes

Moving now to a discussion of the interest rate effects of Bank of Canada communication, the WLS estimation of equation (1) is presented in Tables 3 and 4 using 5-minute return data for the short-term BAX rate and Tables 5 and 6 for the longer-term CGB futures rate. For each of the three models, two specifications were estimated. In the first, the themes were simply demeaned.¹⁶ Due to the fact that there was still some residual correlation between the themes (because of the shorter sample used and, for Models 1 and 2, because the themes were extracted simultaneously from press releases and MPRs but they were included separately in the regressions), a second specification was estimated for each model in which the themes were orthogonalized before being included in the interest rate regressions.¹⁷ This was done by regressing each theme on the orthogonalized themes that came before it in ordering by size. For example, theme 1 is first simply demeaned. Theme 2 is then regressed on a constant and the demeaned theme 1. The residual represents the new orthogonalized theme 2. Theme 3 is then regressed on orthogonalized Themes 1 and 2. This continues until the top ten themes are orthogonalized. Themes for press releases and MPRs are treated separately. Regressions using orthogonalized themes are represented as Model 1-o, 2-o, and 3-o.

¹⁶ By construction, the themes have a zero mean over their full sample from December 2000 to June 2008. However, with intra-day interest rate data only available from 2002 to 2008 the raw theme data may have non-zero mean.

¹⁷ The correlations between themes were not strong but with R-squared values in the 0.15-0.20 range even low collinearity between the themes was sometimes problematic.

Table 3 shows the estimate for the variance equation (2) used in the WLS of (1) using 5-minute BAX return data. We see that volatility increases starting in the five minutes before FAD press releases and continues for 20 minutes after (except for during the period 5-10 minutes after). There was no increased volatility around MPR(U) releases. Many of the top-ten themes extracted from press releases were significant while basically none of the MPR(U) themes were significant. Some substantial differences are evident in the magnitude and significance of some of the themes between the base specifications and those with orthogonalized themes. This highlights the residual correlation that remains between the themes when used in these regressions. As such, we will concentrate our discussions on the three models estimated using orthogonalized themes.

Model 1-o shows that all of the top-ten themes except for 3, 5, and 10 were significant at the 5% level. When a theme is positive and it has a positive coefficient in the volatility regression then higher values raise volatility. But volatility will also increase when a theme becomes more negative and the theme has a negative coefficient in the regression. The reverse of each of these two cases are obviously cases in which volatility will fall. Given this, we can see that volatility falls when press releases emphasize discussions about factors affecting the growth of the economy (Theme 1 becomes more positive), the balance of risks to the projection (Theme 2 becomes more positive), the resumption of growth after a slowdown from external factors (Theme 8 becomes more positive), and somewhat surprisingly, risks tilted to the downside related to global imbalances (Theme 6 becomes more negative) or higher energy prices (Theme 7 becomes more negative).

According to Model 1-o, volatility tends to increase when Bank press releases focus on the terrorist attacks and resulting geopolitical uncertainty leading to low business confidence (Theme 2 or 8 becomes more negative, Theme 9 becomes more positive), SARS, BSE and the negative effects of a CAD appreciation (Theme 4 becomes more positive), high oil prices and geopolitical uncertainty in the Middle East (Theme 6 becomes more positive). Therefore, Bank of Canada communication via FAD press releases tend to lower volatility when discussing growth

and the balance of risks but increase volatility when highlighting special factors (e.g. terrorist attacks, SARS, BSE, geopolitical risk).

Turning to Model 2-o, which highlights the FLS, we see that Themes 1 and 2 have the same volatility effects since they are essentially the same themes as in Model 1. In addition, we find that BAX return volatility is reduced by discussions of the FLS that some modest further increase in interest rates may be required (Theme 5 becomes more positive) or credit conditions and the FLS that further monetary stimulus is likely to be required (Theme 6 becomes more positive). Increases in volatility arise from discussions of the FLS that a reduction of stimulus will be required along with high oil prices and geopolitical risk (Theme 3 becomes more positive), SARS, BSE and their risks to the projection (Theme 5 becomes more negative), geopolitical uncertainty in the Middle East (Theme 7 becomes more negative) and the business confidence effects of the terrorist attacks in 2001 (Theme 8 becomes more positive).

Finally, Model 3-o separates out the FLS as does Model 2-o but also extracts themes separately for the press releases and MPR(U)s. BAX return volatility has tended to decline in association with discussions in press releases about the balance of risks to the projection related to the dollar and housing (Theme 1 becomes more positive), the balance between aggregate demand and supply pressures (Theme 6 becomes more positive), reduced interest rates, momentum, and external pressures (Theme 7 becomes more negative), the slowdown in the US associated with falling credit conditions and problems in the housing sector leading to low Canadian exports (Theme 2 becomes more negative) and rising rates, an economy at full capacity, and the FLS that some modest increase in rates may be required (Theme 4 becomes more negative). Themes 8 and 10 also reduced volatility when they became more negative but the interpretation of these themes is more problematic.

Volatility of BAX returns was increased by discussions about geopolitical uncertainty and its effects on confidence (Theme 1 becomes more negative), the same topics in conjuncture with terrorist attacks (Theme 7 becomes more positive), concerns about confidence, high oil

prices, and exports (Theme 8 becomes more positive), increased interest rates to offset a strong recovery, especially through household spending (Theme 6 becomes more negative), the FLS that a reduction in monetary stimulus will be required (Theme 2 becomes more positive), and the effects of a depreciation of the US dollar (Theme 4 becomes more positive).

Note that even when the themes were extracted only from the MPR(U)s, their themes had no significant impact on BAX return volatility.

The results for WLS estimate of the BAX return equation (1) can be found in Table 4. Overall, we see that order flow has the expected strong positive relationship with returns. Of the macro surprises, the positive information in quarterly GDP growth had a significant positive effect on returns while the Ivey Purchasing Managers index had a significant negative effect. The surprise in the Bank of Canada's overnight target interest rate had a large and significant negative effect on BAX returns in line with expectations.¹⁸

There are fewer significant FAD press release themes that are significant than in the volatility equation but some of the MPR(U) themes are now significant. In Model 1-o, only Themes 4 and 9 had a significant effect on returns at the 5% level, while Themes 5 and 8 had some significance at the 10% level. Higher values to Theme 4 indicate emphasis of SARS, BSE, and the effects of a CAD appreciation which are all negative for growth and consequently associated with interest rate declines. Similarly, higher values of Theme 8 (emphasizing slowing growth especially through external demand) and more negative values of Theme 9 (highlighting employment and credit market concerns) also lower BAX returns.

According to Model 1-o, MPR(U) discussions that focus on Themes 3, 5, and 8 also significantly affect BAX returns. Theme 5 has the same positive effect when in an MPR as in a press release but Theme 8, when emphasized in an MPR, has the opposite (but much more significant) effect as the FAD press release on BAX returns.

¹⁸ Recall that the dependent variable is the change in price or return, not yield. As such, a positive (negative) surprise in the target rate, specified in yield space, should have a negative (positive) effect on return.

Moving to Model 2-o, press release Themes 5, 6, and 8 affect BAX returns at the 1% significance levels while Theme 3 has a significant affect at the 10% level. Themes 6 to 10 are significant when they appear in MPR(U)s. Theme 5, which emphasizes the FLS that some modest further increase in interest rates may be required, has a positive effect on BAX returns when it appears in press releases. Also, when Theme 5 becomes more negative, emphasizing the discussion of SARS, BSE and their risks to the projection, BAX returns are pushed down. Theme 6, with its negative coefficient for both press releases and MPR(U)s, puts downward pressure on interest rates because it emphasizes the FLS that further monetary stimulus is likely to be required as well as the problems associated with deteriorating credit conditions. On the other hand, when Theme 6 becomes more negative, emphasizing the FLS that rates are consistent with achieving the inflation target but that there is strong momentum and slightly higher than expected inflation, BAX returns actually tended to be pushed up. Theme 8 (highlighting terrorist attacks and confidence effects when it becomes more positive) has a strong positive effect through press releases but a weaker and negative effect via MPR(U)s.

In the final model, 3-o, Themes 4, 6, and 9 are significant when expressed in press releases but Themes 3, 7, and 10 are significant when emphasized by MPR(U)s. The effects of the press release themes were much higher than that of the MRP(U) themes. Significant positive effects on BAX returns are found to originate from discussions of rising rates, an economy at full capacity, and the FLS that some modest increase in rates may be required (press release Theme 4 becomes more negative), economic recovery and strong household spending (press release Theme 6 becomes more negative), and high energy prices and solid economic growth (press release Theme 9 becomes more positive). From MPR(U)s, positive effects on returns came from discussions of GST increases, hurricane Katrina, the tilt of risks and the FLS that rates were consistent with the inflation target (Theme 7 becomes more negative), the risk of military action in the Middle East and different FLS implying rates will be increasing (Theme 10 becomes more

negative), and terrorist attacks, geopolitical risk, and the FLS that some modest further increase in rates would be required (Theme 3 becomes more negative).

In contrast, pressure for BAX returns to fall came from discussions of the effects of a depreciation of the US dollar (press release Theme 4 becomes more positive), high energy prices, external pressures, and the balance between aggregate demand and supply (press release Theme 6 becomes more positive), and the need to lower rates and add monetary stimulus in the face of an appreciating CAD and other external demand factors (press release theme 9 becomes more negative). Somewhat counter-intuitively, the upward pressure on rates from the MPR(U)s came from discussions of negative Latin American effects on Canadian growth (Theme 10 becomes more positive), credit conditions and financial turbulence in interbank markets (Theme 7 becomes more positive), and shocks from an exposure to increased international competition (Theme 3 becomes more positive).

In sum, we see that BAX returns and volatility responded most to discussions about major shocks hitting the economy (for example, external shocks such as hurricane Katrina and the 2001 terrorist attacks as well as domestic shocks such as SARS and BSE), the balance of risks to the economic projection, geopolitical uncertainty, and forward looking statements. In particular, BAX volatility was increased when Bank of Canada FAD statements focused on the effects of terrorist attacks, SARS, and BSE for Canada as well as the state of geopolitical uncertainty and confidence. FAD statements were, however, able to reduce volatility through its discussions of growth and the balance of risks to the economic projection. Forward looking statements had significant effects on markets but other themes also continued to have significant effects. Financial markets do appear to still pay attention to content on the macroeconomic outlook in addition to guidance on the near term path for interest rates.

The results using 5-minute data for the longer-term 10-year CGB market are contained in Table 5 for the volatility equation and Table 6 for the WLS estimate of the return equation (1).

There is also increased CGB volatility around FAD release times but, in contrast to the BAX market, CGB volatility does not increase ahead of the FAD nor does it seem to stay elevated for as long. As in the BAX market, CGB volatility does not increase around MPR releases. The themes extracted from the FAD and MPR(U) communications have a less significant effect on the CGB market than the BAX market for both volatility and returns.

Theme 4 in Model 1-o (SARS, BSE, and CAD appreciation emphasized as Theme 4 becomes more positive) had a significant positive but stronger effect on CGB volatility than on BAX volatility. Theme 10 (high employment, gradual recovery in the US, accommodative monetary policy), however, had a significantly positive effect on CGB volatility but was insignificant in the BAX market. Themes 1, 2, and 6-9 were all insignificant in contrast to the results for BAX volatility.

In Model 2-o, Themes 5 and 8 were significantly negative. The discussions in Theme 8 of the business confidence effect of the 2001 terrorist attacks worked to increase BAX volatility but reduce CGB volatility. Discussions of real factors affecting GDP growth, risks to the projection for output and inflation, geopolitical uncertainty, and several FLS had no effect on CGB market volatility and yet a significant effect on the BAX market.

Model 3-o revealed significant negative effects for Press Release Themes 1, 7, and 10. The Theme 1 result was the same as for BAX volatility but Themes 7 and 10 were the opposite. There is no clear reason why the longer-term market should respond any differently to these themes than the short-term market.

Moving to the CGB return equation results in Table 6, note that the FAD interest rate surprise variable now has a positive but insignificant effect in contrast to the strong negative effect in the BAX market. The results show that there was little significance to the press release themes but a broader significance for the MPR(U) themes. Focusing in on Model 3-o, we see that there were no significant effects from the press release themes while Themes 3, 5, 6, and 10 were significant for the MPR(U)s. Themes 3 and 10 had negative coefficients as with BAX returns

signifying that CGB returns also moved up in response to discussions of the risk of military action in the Middle East (Theme 10 becomes more negative) or terrorist attacks, geopolitical risk, and the FLS that some modest further increase in rates would be required (Theme 3 becomes more negative). When Theme 5 became more positive (emphasizing SARS and BSE), CGB returns tended to increase. When Theme 5 became more negative, however, and emphasized words like terrorist attacks and geopolitical risk (as in Theme 3) but also mixed in talk of recession and trouble in the Middle East, CGB returns were pushed down. Finally, Theme 6 was an interesting mix of effects. When it became more negative and emphasized concepts like military action in the Middle East mixed together with the effects of large increases in insurance premiums and negative shocks from Latin America, CGB returns were pushed down. CGB returns were pushed up when it focused on terrorist attacks, SARS, BSE, and the potential effects on tourism. The conflicting messages from the effects of discussions about the terrorist attacks of 2001 highlights the difficulty in isolating an interpretation for the themes.

Overall, we find that the themes are less significant for CGB returns and volatility than for BAX rates. The CGB market was, however, responding to some of the same themes when there was significance so the markets had largely consistent results. FAD press releases tended to have more significant market effects than did MPR(U)s which is line with the Bank of Canada's efforts to maintain consistent messages across its communication vehicles. It would be interesting to refine the analysis further to try to determine whether Bank of Canada communications have the same market impact under different conditions. For instance, do the themes have the same impact in expansions, contractions, and at turning points? The degree of financial market or monetary policy uncertainty could also have a role to play in conditioning the market's response to Bank statements. Some of these issues were analyzed in Boukus and Rosenberg (2006) by interacting themes extracted from FOMC minutes with Eurodollar volatility (monetary policy uncertainty proxy) and the yield curve (macroeconomic outlook proxy). They found that some themes did have different impacts depending on the level of uncertainty or the outlook. This type

of analysis has been left for future work given the still relatively small sample of Bank communications in the different possible conditioning circumstances.

Broadly speaking our results are consistent with Boukus and Rosenberg (2006) in that LSA is able to extract certain themes from communication statements that are important for financial markets. There are, however, some important differences. Boukus and Rosenberg (2006) interpret the greater significance of the themes for long rates than short rates as implying that there is information in the market's interpretation of the minutes that goes beyond simply a clarification of the near-term policy aims. Our results show relatively less significance of the themes for long rates which probably results from the Bank of Canada's clearly articulated inflation targeting regime which has been in place since the early 1990s. While mentioned in all FAD and MPR(U) statements, the inflation targeting regime is well-known and highly credible and, as such, there is little new information in these statements that leads markets to question monetary policy over the longer term. Markets look to Bank communication primarily for information about the short term outlook for interest rates. Over the sample examined, the factors discussed in Press Releases and MPR(U)s have been those relevant for the short, possibly medium, term but not the longer term. Without an explicit inflation target in the U.S., markets are constantly pouring over Fed minutes for information about future monetary policy in the short as well as long term.

5.0 Concluding Remarks

Overall, the Latent Semantic Analysis methodology is able to highlight the major themes from Bank of Canada FAD press releases and MPR(U)s that affect market returns and their volatility. The strongest effects appeared, not surprisingly, in the shorter-term BAX market. The methodology is able to highlight that discussions of the major shocks hitting the Canadian economy (external shocks such as terrorist attacks in the US and geopolitical risk or domestic

shocks such as SARS and BSE), the balance of risks to the economic projection, and the forward looking statements are the factors that markets focus on and respond to.

There is still much work that can be done to refine this methodology and analysis. The sample of Bank communications should be expanded as possible to investigate different communication methods as well as determine whether there is a small sample problem. Other text mining techniques can also be explored to determine if there are any better, alternative methods of extracting the key messages from Bank of Canada communications. Alternative techniques may be able extract more easily interpretable themes. Methods involving non-negative matrix factorizations may be an interesting alternative to SVD.

The same methodology should also be applied as a robustness check to central bank communications coming from other countries and to isolate their effects on Canada. Communication from the U.S. Federal Reserve, for instance, is likely to have a significant effect on Canadian returns while statements from an inflation targeting country like Sweden, another small open economy but not closely connected to Canada, should not. Finally, subsequent research will investigate the similarity between Bank communication and the market news stories that interpret these statements.

References

- Anderson, Torben, Tim Bollerslev, Francis Diebold, Clara Vega. (2007) "Real-time price discovery in global stock, bond and foreign exchange markets." *Journal of International Economics*. 73. pp. 251-277.
- Balke, Nathan S., and D'Ann Petersen. (2002) "How Well Does the Beige Book Reflect Economic Activity: Evaluating Qualitative Information Quantitatively." *Journal of Money Credit and Banking*. 34. 114-136.
- Bank of Canada. (2000) "Bank of Canada Releases Dates for Announcing Bank Rate Actions". October 2000. www.bankofcanada.ca
- Bligh, Michelle and Gregory Hess. (2009) "A Quantitative Assessment of the Qualitative Aspects of Chairman Greenspan's Communications." Unpublished manuscript.
- Bro, R., E. Acar, and Tamara Kolda. (2007) "Resolving the sign ambiguity in the singular value decomposition." *Journal of Chemometrics*. 22. pp. 135-140.
- Boukus, Ellyn, and Joshua V. Rosenberg. (2006) "The Information Content of FOMC Minutes." Unpublished manuscript.
- Cook, Timothy, and Thomas Hahn. (1988) "The Information Content of Discount Rate Announcements and Their Effect on Market Interest Rates." *Journal of Money, Credit and Banking*. 20. No. 2. pp. 167-180.
- Ehrmann, Michael, and Marcel Fratzscher. (2005) "Central Bank Communication: Different Strategies, Same Effectiveness?" Working Paper Series 488, European Central Bank.

- Fay, C. and T. Gravelle. (2009). "The Market Impact of Forward-Looking Policy Statements: Transparency vs. Predictability" *Bank of Canada Review Winter 2008-09*. pp. 25-34.
- Fettig, David, Arthur J. Rolnick, and David E. Runkle. (1999) "The Federal Reserve's Beige Book: A Better Mirror Than Crystal Ball." *Federal Reserve Bank of Minneapolis The Region*. 10-13. pp. 28-32.
- Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson. (2005) "Do Actions Speak Louder than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking*. 1. pp. 55-93.
- Kohn, Donald and Brian Sack. (2004) "Central Bank Talk: Does It Matter and Why?" *Macroeconomics, Monetary Policy, and Financial Stability*. Ottawa: Bank of Canada. pp. 175-206.
- Kintsch, Walter. (2007). "Meaning in Context." In Thomas Landauer, Danielle McNamara, Simon Dennis, and Walter Kintsch, (eds) *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates. Mahwah, New Jersey. pp. 89-106.
- Landauer, T. and S. T. Dumais. (1997) "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge." *Psychological Review*, 104(2), pp. 211-40.
- Landauer, Thomas, Peter Foltz, and Darrell Laham. (1998) "An Introduction to Latent Semantic Analysis." *Discourse Processes*, 25, pp. 259-284.
- Landauer, Thomas, Danielle McNamara, Simon Dennis, and Walter Kintsch. (2007) *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates. Mahwah, New Jersey.

- Lucca, David and Francesco Trebbi. (2009) “Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements”. Unpublished manuscript.
- Rosa, Carlo. (2007) “Providing Content to ECB Announcements”. *Rivista Internazionale di Scienze Sociali*, CXV (4), pp. 525-546.
- Zeimpekis, D. and E. Gallopoulos, (2006) [TMG: A MATLAB toolbox for generating term document matrices from text collections](#), In “Grouping Multidimensional Data: Recent Advances in Clustering” (J. Kogan, C. Nicholas, and M. Teboulle, eds.), Springer, Berlin, pp. 187-210.
- Zeimpekis, D. and E. Gallopoulos, (2005) ["Design of a MATLAB toolbox for term-document matrix generation"](#). In Proc. [Workshop on Clustering High Dimensional Data and its Applications](#), (held in conjunction with [5th SIAM Int'l Conf. Data Mining](#)), (I.S. Dhillon, J. Kogan and J. Ghosh eds.), pp. 38-48, Newport Beach, California.
- Zavodny, Madeline, and Donna K. Ginther. (2005) “Does the Beige Book Move Financial Markets?” *Southern Economic Journal*. 72, No. 1, 138-151.

Table 1a: Correlations between Model 1 Themes and Selected Macro Variables¹⁹

Macro Variable		Model 1 Themes									
		1	2	3	4	5	6	7	8	9	10
Price of Oil	PR	0.39	0.76	0.31							
	MPR	0.53	0.79				0.46		0.48		0.40
Exchange Rate =US/CAD	PR	0.44	0.85								
	MPR	0.54	0.87				0.44		0.46	0.32	0.45
Y/Y GDP growth	PR		0.34	-0.63							
	MPR		0.39	-0.68			0.40		0.39		0.51
Y/Y Inflation	PR							-0.38			
	MPR							-0.31	0.32	-0.30	
Change in target rate*	PR	-0.56		-0.58	-0.40						
	MPR			-0.53						0.46	0.40

Table 1b: Correlations between Model 2 Themes and Selected Macro Variables

Macro Variable		Model 2 Themes									
		1	2	3	4	5	6	7	8	9	10
Price of Oil	PR		0.76				0.32				
	MPR	0.50	0.79								-0.62
Exchange Rate =US/CAD	PR	0.34	0.86				0.31				
	MPR	0.51	0.87								-0.55
Y/Y GDP growth	PR		0.36	0.39	0.41				-0.31		
	MPR		0.41	0.53	0.55	-0.31	-0.34		-0.38		
Y/Y Inflation	PR								-0.33	-0.32	
	MPR								-0.31		
Change in target rate*	PR	-0.62		0.61							
	MPR			0.48			-0.40				-0.46

Table 1c: Correlations between Model 3 Themes and Selected Macro Variables

Macro Variable		Model 3 Themes									
		1	2	3	4	5	6	7	8	9	10
Price of Oil	PR	0.77									
	MPR	0.85									
Exchange Rate =US/CAD	PR	0.87									
	MPR	0.87									
Y/Y GDP growth	PR	0.37	0.42	0.48							
	MPR		-0.49				-0.55				
Y/Y Inflation	PR					0.31	0.30				
	MPR						-0.53				
Change in target rate*	PR		0.69								
	MPR		-0.50			-0.56					

* Change in target rate from current FAD (MPR) to the one in four FADs.

¹⁹ To facilitate reading the tables, only correlations larger than 0.3 in magnitude are shown.

Table 2: Cumulative Percent Contribution of Themes to Overall Understanding of Documents

Theme	Model 1	Model 2	Model 3: PRs	Model 3: MPRs
1	10.0	10.1	5.5	9.0
2	14.4	14.6	10.3	15.9
3	17.5	18.0	13.9	22.3
4	20.2	20.9	17.4	27.9
5	22.6	23.5	20.7	32.9
6	24.8	25.8	23.8	37.7
7	27.0	28.0	26.7	42.1
8	29.1	30.1	29.4	46.2
9	31.1	32.1	32.0	49.7
10	33.0	34.0	34.6	53.2
20	49.4	50.5	55.7	81.5
30	62.1	63.2	71.7	100.0
40	72.5	73.4	83.8	
50	80.9	81.7	92.6	
60	87.6	88.1	98.8	
70	92.8	93.1		
80	96.6	96.8		
90	99.3	99.3		
95	100.0	100.0		

Note: This table summarizes the amount each theme contributes to the overall understanding of the text corpus based on the matrix of singular values from S . Each squared singular value from the diagonal of S is divided by the sum of all squared singular values to determine the percent of the variance explained by each theme.

Table 3: BAX Market Return Volatility Equation Estimates

Variable	Model 1	Model 2	Model 3	Model 1-o	Model 2-o	Model 3-o
$ \varepsilon_{t-1} $	0.24***	0.24***	0.24***	0.24***	0.24***	0.24***
B.Permits _{t+1}	0.66	0.66	0.67	0.66	0.66	0.67
B.Permits _t	-1.50***	-1.50***	-1.50***	-1.50***	-1.50***	-1.50***
B.Permits _{t-1}	-2.83***	-2.83***	-2.83***	-2.83***	-2.83***	-2.83***
GDP _{t+1}	0.09	0.09	0.09	0.09	0.09	0.09
GDP _t	2.55***	2.55***	2.55***	2.55***	2.55***	2.55***
GDP _{t-1}	6.29***	6.28***	6.28***	6.29***	6.28***	6.28***
I.Sec.Tran. _{t+1}	0.52	0.52	0.52	0.52	0.52	0.52
I.Sec.Tran. _t	0.85	0.85	0.84	0.85	0.85	0.84
I.Sec.Tran. _{t-1}	-1.04**	-1.04**	-1.04**	-1.04**	-1.04**	-1.04**
IveyPurch.Mgr _{t+1}	-0.44	-0.44	-0.44	-0.44	-0.44	-0.44
IveyPurch.Mgr _t	0.31	0.31	0.31	0.31	0.31	0.31
IveyPurch.Mgr _{t-1}	2.00***	2.00***	2.00***	2.00***	2.00***	2.00***
Retail Sales _{t+1}	0.34	0.34	0.34	0.34	0.34	0.34
Retail Sales _t	1.65***	1.65***	1.65***	1.65***	1.65***	1.65***
Retail Sales _{t-1}	7.05***	7.05***	7.05***	7.05***	7.05***	7.05***
WholesaleSales _{t+1}	-0.41	-0.41	-0.41	-0.41	-0.41	-0.41
WholesaleSales _t	-1.01*	-1.01*	-1.01*	-1.01*	-1.01*	-1.01*
WholesaleSales _{t-1}	-0.86	-0.86	-0.86	-0.86	-0.86	-0.86
FAD _{t+1}	1.49**	1.49**	1.49**	1.49**	1.49**	1.49**
FAD _t	18.47***	18.23***	18.52***	18.47***	18.23***	18.52***
FAD _{t-1}	44.08***	44.11***	44.07***	44.08***	44.11***	44.07***
FAD _{t-2}	-4.17***	-4.22***	-4.31***	-4.17***	-4.22***	-4.31***
FAD _{t-3}	2.35***	2.35***	2.34***	2.35***	2.35***	2.34***
FAD _{t-4}	3.79***	3.79***	3.78***	3.79***	3.79***	3.78***
MPR	-0.88	-1.36	-0.55	-0.88	-1.36	-0.55
THEME1 - PR	-267.40***	-36.99	-9.62	-132.02***	-85.52***	-24.08***
THEME2 - PR	21.86**	-3.02	31.85***	-17.70***	-31.05***	16.15***
THEME3 - PR	-6.17	47.30***	-14.02***	0.14	25.02***	-5.90
THEME4 - PR	33.31***	9.39*	19.02***	22.86***	4.00	10.35**
THEME5 - PR	-31.46***	-35.15***	-26.25***	5.11	-19.17***	-6.95
THEME6 - PR	42.47***	3.93	-52.05***	31.78***	-16.38***	-36.08***
THEME7 - PR	38.45***	-61.38***	48.31***	38.66***	-45.03***	32.38***
THEME8 - PR	-76.18***	75.85***	33.83***	-64.30***	78.53***	33.12***
THEME9 - PR	16.44**	25.07***	-8.63*	16.11**	23.97***	-7.16
THEME10 - PR	-9.88	-6.05	15.72***	-9.88	-6.05	15.72***
THEME1 - MPR	-94.76	19.49	3.37	-4.40	-1.92	-3.71
THEME2 - MPR	-1.36	-4.03	-20.13	1.15	-9.75	-3.75
THEME3 - MPR	-6.37	-0.40	2.16	-8.72	0.13	0.89
THEME4 - MPR	39.60	-6.29	-20.96	10.20	15.79	-10.42
THEME5 - MPR	56.91	-24.59	-8.96	8.15	-7.49	0.24
THEME6 - MPR	-7.10	4.40	6.78	-3.02	-6.97	4.72
THEME7 - MPR	-43.02*	1.97	-3.33	-32.69	10.59	-3.60
THEME8 - MPR	32.78	11.48	1.96	28.42	12.67	1.82
THEME9 - MPR	-23.06	-29.87	-0.65	-24.12	-27.15	-0.73
THEME10 - MPR	-4.42	13.01	1.45	-4.42	13.01	1.45
R-Bar**2	0.15	0.15	0.15	0.15	0.15	0.15

Notes: This table presents estimates of equation (2) from the text. The dependent variable is the absolute value of the residual from an OLS estimate of equation (1) for the five-minute return in the BAX market. The macro news, FAD, and MPR variables are dummies with a value of 1 for the five-minute interval after the event and zero otherwise. The Theme variables for Models 1, 2, and 3 are demeaned while the Themes for Models 1-o, 2-o, and 3-o are orthogonalized. Model 1 uses the base set of Bank of Canada texts while Model 2 highlights particular phrases in the forward looking statements. Model 3 extracts themes from FAD press releases and MPR(U)s separately. *** is significant at the 1% level, ** is significant at the 5% level, and * is significant at the 10% level. All coefficients, except the lagged dependent variable, should be multiplied by 10^5 .

Table 4: BAX Return Estimates

Variable	Model 1	Model 2	Model 3	Model 1-o	Model 2-o	Model 3-o
Constant	0.08***	0.08***	0.08***	0.08***	0.08***	0.08***
$\Delta R_{BAX(t-1)}$	-0.11***	-0.11***	-0.11***	-0.11***	-0.11***	-0.11***
OFLOW	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***
R_{surp}	-18.34***	-19.30***	-18.19***	-18.34***	-19.30***	-18.19***
B.Permits _{surp}	-0.35	-0.35	-0.35	-0.35	-0.35	-0.35
GDP _{surp}	1.20***	1.20***	1.20***	1.20***	1.20***	1.20***
I.Sec.Tran. _{surp}	-1.19	-1.19	-1.19	-1.19	-1.19	-1.19
IveyPurch.Mgr _{surp}	-2.16**	-2.16**	-2.16**	-2.16**	-2.16**	-2.16**
Retail Sales _{surp}	-1.15	-1.15	-1.15	-1.15	-1.15	-1.15
WholesaleSales _{surp}	0.62	0.62	0.62	0.62	0.62	0.62
FAD	2.57	2.59	2.53	2.57	2.59	2.53
MPR	-1.87***	-1.51**	-1.67**	-1.87***	-1.51**	-1.67**
THEME1 - PR	206.61	451.44**	-40.51	-79.43	-84.50	-49.74
THEME2 - PR	-69.16	-73.23	43.95	-29.73	-29.40	47.58
THEME3 - PR	-51.37	140.87***	-7.31	-43.90	75.33*	-1.72
THEME4 - PR	-119.83***	-23.23	-68.44***	-120.84***	-13.16	-71.94***
THEME5 - PR	30.27	139.14***	7.32	72.39*	150.84***	5.96
THEME6 - PR	14.66	-55.76**	-71.47**	0.66	-77.98***	-81.31***
THEME7 - PR	35.07	-49.78	25.75	41.18	-23.57	58.74
THEME8 - PR	-155.55**	121.61***	-32.45	-106.28*	125.43***	-21.32
THEME9 - PR	60.45**	20.37	62.56**	58.84**	15.95	59.47*
THEME10 - PR	-48.21	-24.36	-32.95	-48.21	-24.36	-32.95
THEME1 - MPR	-88.05	40.57	-3.28	-22.48	-15.39	-0.41
THEME2 - MPR	-73.31**	-88.94***	-1.79	-1.60	8.82	-3.57
THEME3 - MPR	-0.85	-58.74***	-6.69**	-18.30*	-2.21	-6.47**
THEME4 - MPR	63.70***	50.55***	3.38	-3.01	10.89	-2.79
THEME5 - MPR	69.78*	23.06*	2.97	29.56**	11.42	5.70
THEME6 - MPR	-21.77	-33.84***	2.68	-8.02	-19.02**	1.89
THEME7 - MPR	-55.89**	76.45***	-16.99***	-24.04	33.92**	-14.90***
THEME8 - MPR	123.21***	-66.57***	-2.87	123.33***	-48.96*	-2.38
THEME9 - MPR	24.75	-52.58***	-1.24	16.94	-70.39***	-0.65
THEME10 - MPR	-32.70	-85.18***	-10.33***	-32.70	-85.18***	-10.33***
R-Bar**2	0.19	0.19	0.19	0.19	0.19	0.19

Notes: This table presents the weighted least squares estimates of equation (1) from the text. The dependent variable is the five-minute return in the BAX market and the weights are the predicted values from equation (2) in Table 3. The variables with a subscript *surp* are macro news surprises while the FAD and MPR variables are dummies with a value of 1 for the five-minute interval after the event and zero otherwise. OFLOW is order flow estimated as the number of buy orders less sell orders. The Theme variables for Models 1, 2, and 3 are demeaned while the Themes for Models 1-o, 2-o, and 3-o are orthogonalized. Model 1 uses the base set of Bank of Canada texts while Model 2 highlights particular phrases in the forward looking statements. Model 3 extracts themes from FAD press releases and MPR(U)s separately. *** is significant at the 1% level, ** is significant at the 5% level, and * is significant at the 10% level. All coefficients, except the lagged dependent variable, should be multiplied by 10^{-5} .

Table 5: CGB Market Return Volatility Equation Estimates

Variable	Model 1	Model 2	Model 3	Model 1-o	Model 2-o	Model 3-o
$ \varepsilon_{t-1} $	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***
B.Permits _{t+1}	-6.51**	-6.51**	-6.51**	-6.51**	-6.51**	-6.51**
B.Permits _t	-7.85**	-7.85**	-7.85**	-7.85**	-7.85**	-7.85**
B.Permits _{t-1}	-25.89***	-25.89***	-25.89***	-25.89***	-25.89***	-25.89***
GDP _{t+1}	-14.93***	-14.93***	-14.93***	-14.93***	-14.93***	-14.93***
GDP _t	-1.75	-1.75	-1.75	-1.75	-1.75	-1.75
GDP _{t-1}	-6.50*	-6.50*	-6.50*	-6.50*	-6.50*	-6.50*
I.Sec.Tran. _{t+1}	-7.68**	-7.68**	-7.68**	-7.68**	-7.68**	-7.68**
I.Sec.Tran. _t	1.80	1.80	1.80	1.80	1.80	1.80
I.Sec.Tran. _{t-1}	-12.84***	-12.84***	-12.84***	-12.84***	-12.84***	-12.84***
IveyPurch.Mgr _{t+1}	1.92	1.92	1.92	1.92	1.92	1.92
IveyPurch.Mgr _t	-4.04	-4.04	-4.04	-4.04	-4.04	-4.04
IveyPurch.Mgr _{t-1}	3.67	3.67	3.67	3.67	3.67	3.67
Retail Sales _{t+1}	-1.33	-1.33	-1.33	-1.33	-1.33	-1.33
Retail Sales _t	-0.75	-0.75	-0.76	-0.75	-0.75	-0.76
Retail Sales _{t-1}	-8.26**	-8.26**	-8.26**	-8.26**	-8.26**	-8.26**
WholesaleSales _{t+1}	-9.03***	-9.03***	-9.03***	-9.03***	-9.03***	-9.03***
WholesaleSales _t	-7.31**	-7.31**	-7.31**	-7.31**	-7.31**	-7.31**
WholesaleSales _{t-1}	-7.68**	-7.68**	-7.68**	-7.68**	-7.68**	-7.68**
FAD _{t+1}	0.89	0.89	0.89	0.89	0.89	0.89
FAD _t	12.01***	12.79***	13.35***	12.01***	12.79***	13.35***
FAD _{t-1}	58.19***	58.08***	58.01***	58.19***	58.08***	58.01***
FAD _{t-2}	14.30***	14.30***	14.30***	14.30***	14.30***	14.30***
FAD _{t-3}	5.44	5.44	5.44	5.44	5.44	5.44
FAD _{t-4}	7.38*	7.38*	7.37*	7.38*	7.38*	7.37*
MPR	-4.26	-1.92	-0.74	-4.26	-1.92	-0.74
THEME1 - PR	-218.57	-126.74	-117.70***	-98.69	-19.53	-60.53*
THEME2 - PR	25.17	-28.67	-37.97	-37.34	-43.69	-3.37
THEME3 - PR	-10.97	-31.80	37.25	-14.08	35.08	23.74
THEME4 - PR	94.77**	9.31	9.82	71.85**	8.33	43.90
THEME5 - PR	-11.26	-55.82	10.51	-44.30	-69.00*	3.64
THEME6 - PR	22.74	-21.01	85.33**	5.49	-9.28	44.04
THEME7 - PR	-21.24	33.17	-105.91**	-36.75	11.82	-77.51*
THEME8 - PR	120.55*	-100.51**	30.77	86.85	-109.90**	21.28
THEME9 - PR	24.70	-57.79*	-11.09	27.87	-48.33	-20.05
THEME10 - PR	94.80**	52.01	-95.69***	94.80**	52.01	-95.69***
THEME1 - MPR	109.00	116.65	-7.66	24.43	-4.57	3.67
THEME2 - MPR	-45.26	-85.05	20.90	15.04	-28.54	-11.69
THEME3 - MPR	7.86	-60.71	16.10	-33.07	20.32	16.22
THEME4 - MPR	61.46	76.18	40.71	64.05	108.03	-3.71
THEME5 - MPR	31.00	-28.63	33.31	31.39	-53.92	-5.98
THEME6 - MPR	20.16	-57.25	-23.64	-13.03	-39.42	0.12
THEME7 - MPR	86.37	-40.13	-11.29	88.04	-32.44	-9.27
THEME8 - MPR	1.41	20.52	-32.07	-1.28	12.39	-32.85
THEME9 - MPR	-20.31	87.80	-3.52	-18.97	86.55	-4.00
THEME10 - MPR	5.60	-5.97	8.30	5.60	-5.97	8.30
R-Bar**2	0.28	0.28	0.28	0.28	0.28	0.28

Notes: This table presents estimates of equation (2) from the text. The dependent variable is the absolute value of the residual from an OLS estimate of equation (1) for the five-minute return in the CGB market. The macro news, FAD, and MPR variables are dummies with a value of 1 for the five-minute interval after the event and zero otherwise. The Theme variables for Models 1, 2, and 3 are demeaned while the Themes for Models 1-o, 2-o, and 3-o are orthogonalized. Model 1 uses the base set of Bank of Canada texts while Model 2 highlights particular phrases in the forward looking statements. Model 3 extracts themes from FAD press releases and MPR(U)s separately. *** is significant at the 1% level, ** is significant at the 5% level, and * is significant at the 10% level. All coefficients, except the lagged dependent variable, should be multiplied by 10^5 .

Table 6: CGB Return Estimates

Variable	Model 1	Model 2	Model 3	Model 1-o	Model 2-o	Model 3-o
Constant	0.40***	0.40***	0.40***	0.40***	0.40***	0.40***
$\Delta R_{CGB(t-1)}$	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***	-0.02***
OFLOW	1.99***	1.99***	1.99***	1.99***	1.99***	1.99***
R_{surp}	2.35	1.25	4.11	2.35	1.25	4.11
B.Permits _{surp}	-1.08	-1.08	-1.08	-1.08	-1.08	-1.08
GDP _{surp}	4.42***	4.42***	4.42***	4.42***	4.42***	4.42***
I.Sec.Tran. _{surp}	1.76	1.76	1.76	1.76	1.76	1.76
IveyPurch.Mgr _{surp}	-8.35**	-8.35**	-8.35**	-8.35**	-8.35**	-8.35**
Retail Sales _{surp}	9.24	9.24	9.24	9.24	9.24	9.24
WholesaleSales _{surp}	2.33	2.33	2.33	2.33	2.33	2.33
FAD	-4.16	-4.87	-4.47	-4.16	-4.87	-4.47
MPR	-4.97	-5.80	-3.59	-4.97	-5.80	-3.59
THEME1 - PR	338.50	258.06	57.98	67.70	51.46	39.56
THEME2 - PR	-129.35***	-47.07	42.40	-2.08	13.69	18.93
THEME3 - PR	37.62	58.25	-75.23	38.01	29.16	-66.34
THEME4 - PR	-135.04*	-73.54	-7.70	-101.80	-80.07*	-23.15
THEME5 - PR	8.08	53.20	-45.84	-0.35	55.59	-28.16
THEME6 - PR	-97.25**	-35.46	-79.87	-72.19*	-35.97	-56.09
THEME7 - PR	3.15	11.61	25.84	28.90	40.27	-15.49
THEME8 - PR	-122.21	133.78*	44.76	-33.00	138.38*	42.59
THEME9 - PR	48.02	-46.61	-31.25	43.19	-65.20	-25.43
THEME10 - PR	-144.81**	-102.27	62.14	-144.81**	-102.27	62.14
THEME1 - MPR	54.70	-127.64	-21.81	-184.86	-200.94	-24.19
THEME2 - MPR	-421.38**	-231.11	-42.68	40.73	32.45	-26.22
THEME3 - MPR	238.80***	-481.56**	-47.56**	29.39	-187.86*	-40.57**
THEME4 - MPR	363.82***	210.79	-23.55	137.01**	178.11*	7.63
THEME5 - MPR	120.99	-131.00	-23.77	107.92	-9.02	53.77*
THEME6 - MPR	96.62	70.68	73.63**	195.96***	35.04	73.04***
THEME7 - MPR	-203.68	68.60	10.39	-140.71	-29.86	22.07
THEME8 - MPR	286.64*	-207.60	-4.98	320.86**	-180.58	-1.68
THEME9 - MPR	239.33***	-340.88**	-5.78	228.60***	-329.42**	-2.07
THEME10 - MPR	-44.91	54.80	-64.93***	-44.91	54.80	-64.93***
R-Bar**2	0.16	0.16	0.16	0.16	0.16	0.16

Notes: This table presents the weighted least squares estimates of equation (1) from the text. The dependent variable is the five-minute return in the CGB market and the weights are the predicted values from equation (2) in Table 3. The variables with a subscript *surp* are macro news surprises while the FAD and MPR variables are dummies with a value of 1 for the five-minute interval after the event and zero otherwise. OFLOW is order flow estimated as the number of buy orders less sell orders. The Theme variables for Models 1, 2, and 3 are demeaned while the Themes for Models 1-o, 2-o, and 3-o are orthogonalized. Model 1 uses the base set of Bank of Canada texts while Model 2 highlights particular phrases in the forward looking statements. Model 3 extracts themes from FAD press releases and MPR(U)s separately. *** is significant at the 1% level, ** is significant at the 5% level, and * is significant at the 10% level. All coefficients, except the lagged dependent variable, should be multiplied by 10^{-5} .

Figure 1: Model 1, Themes 1 to 5

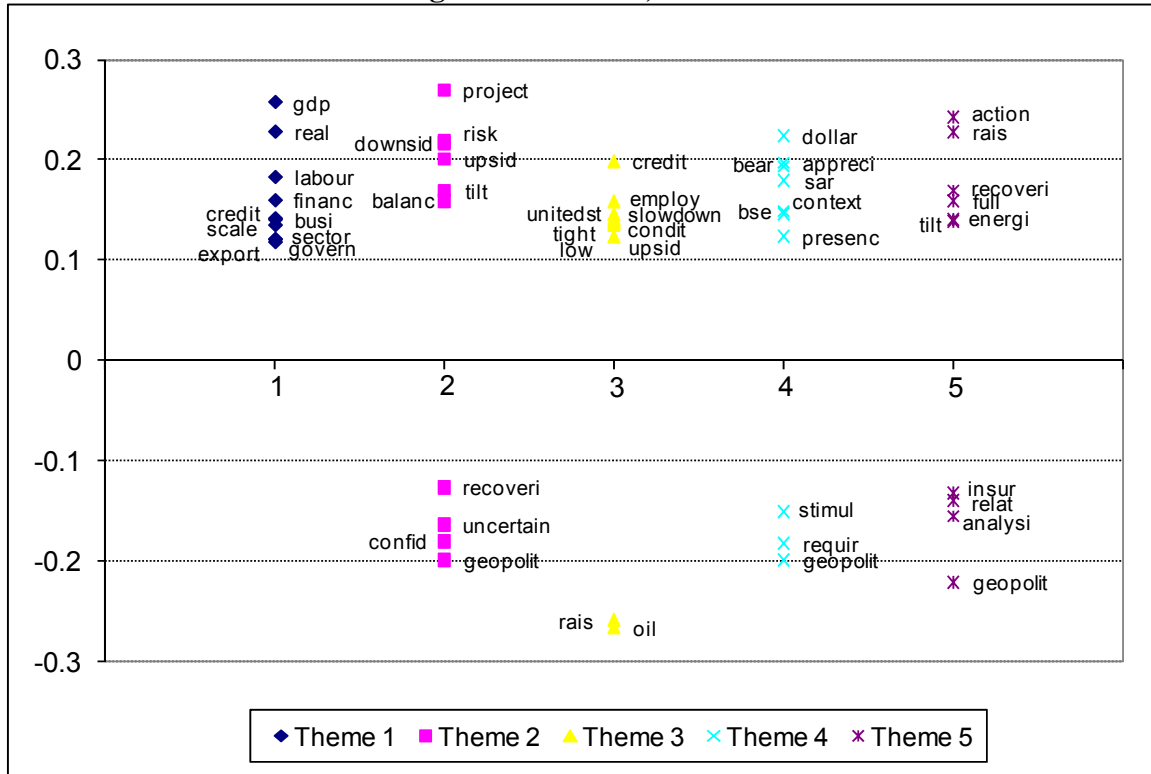


Figure 2: Model 1, Themes 6 to 10

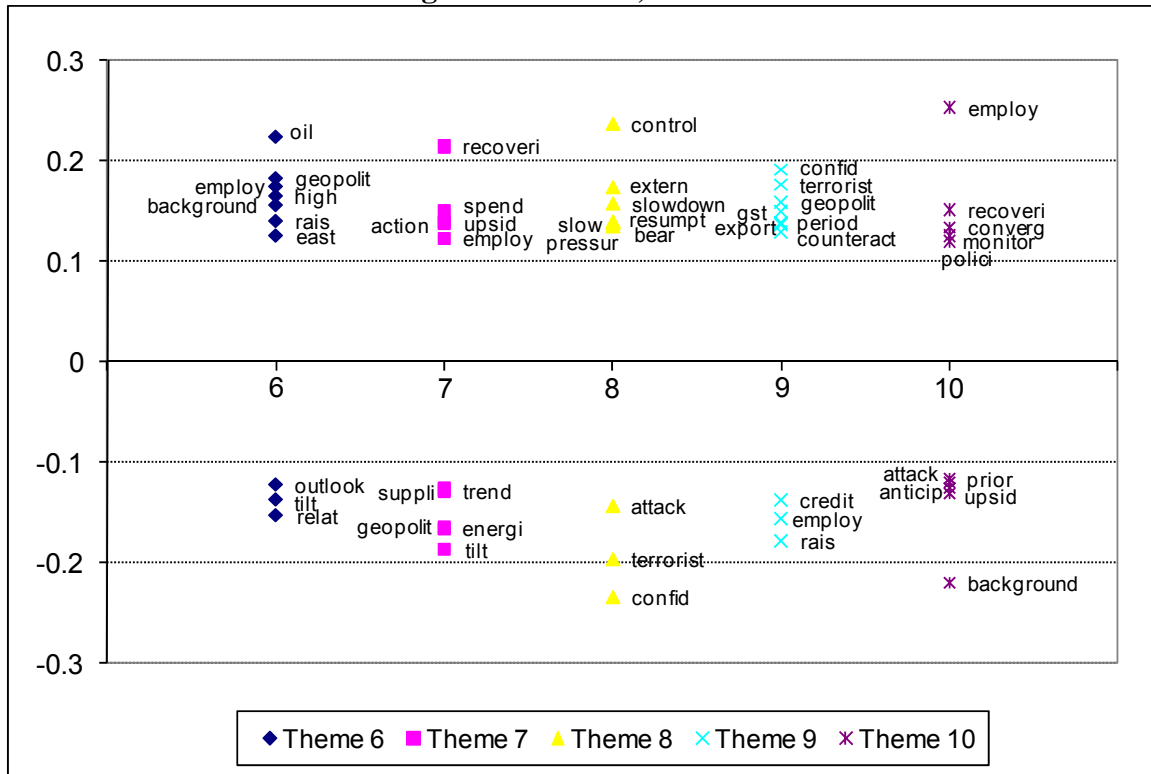


Figure 3: Model 2, Themes 1 to 5

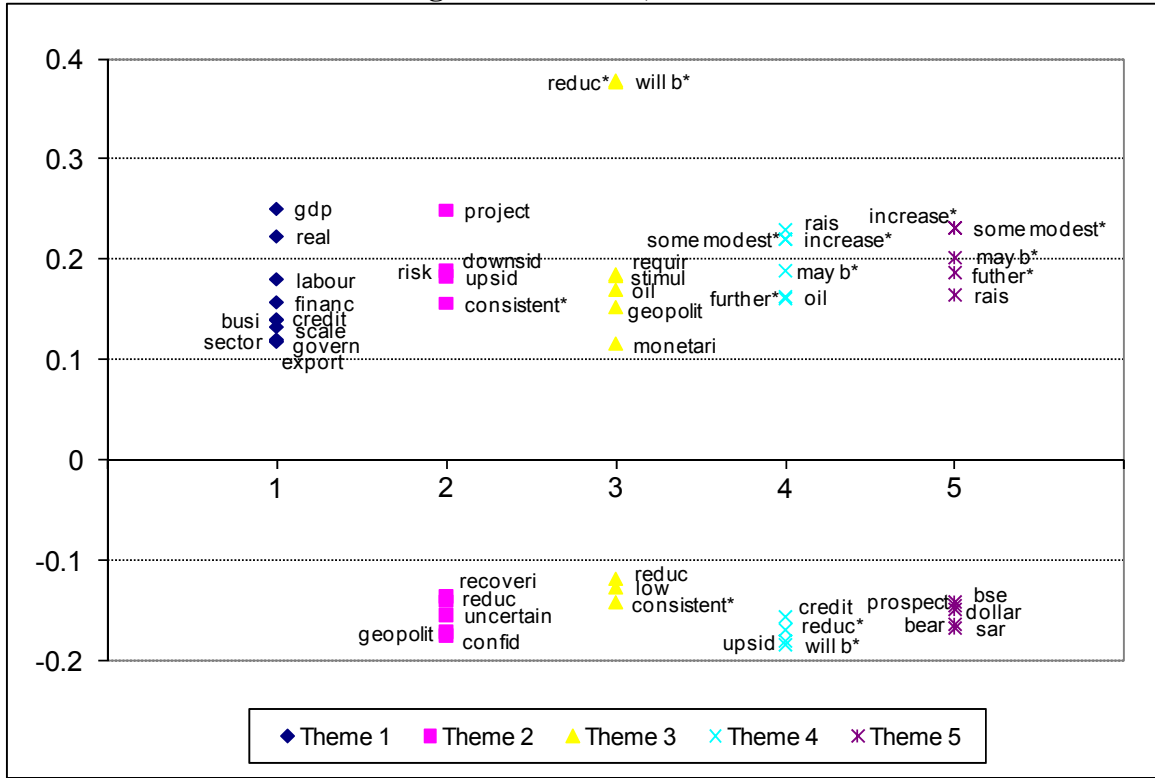
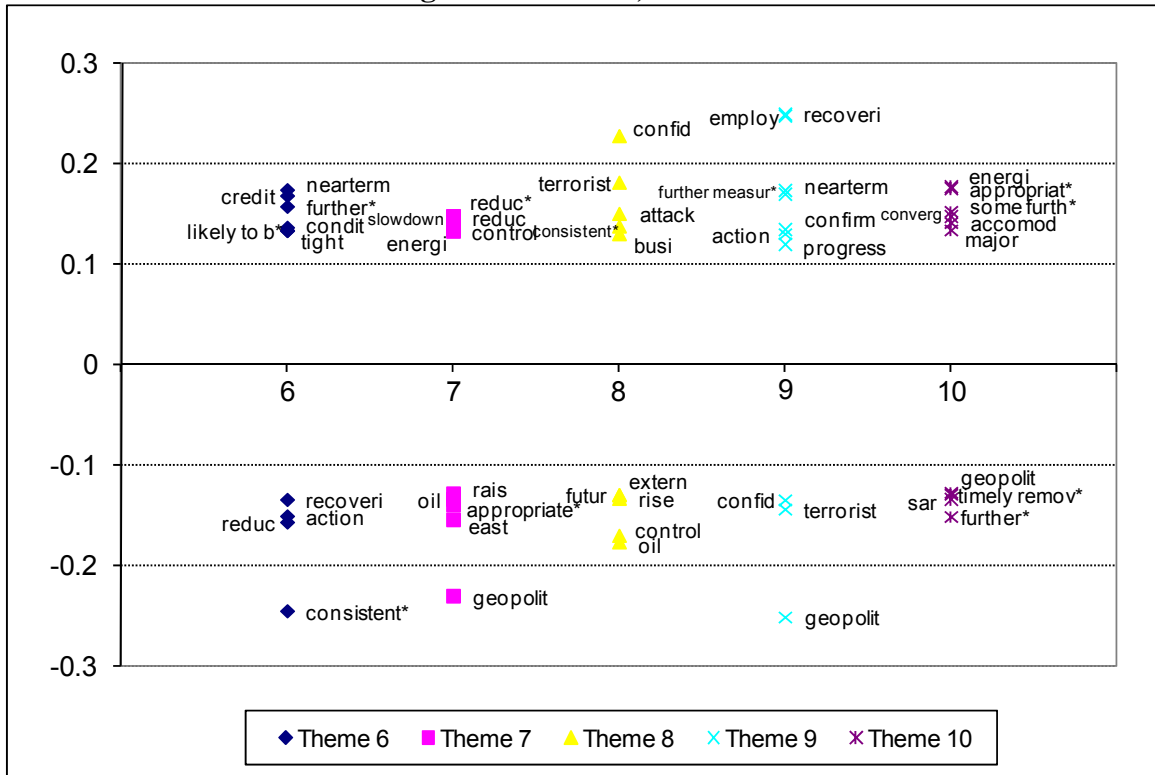


Figure 4: Model 2, Themes 6 to 10



Note: * indicates words appearing in a Forward Looking Statement (FLS).

Figure 5: Model 3, Themes 1 to 5, FAD Press Releases

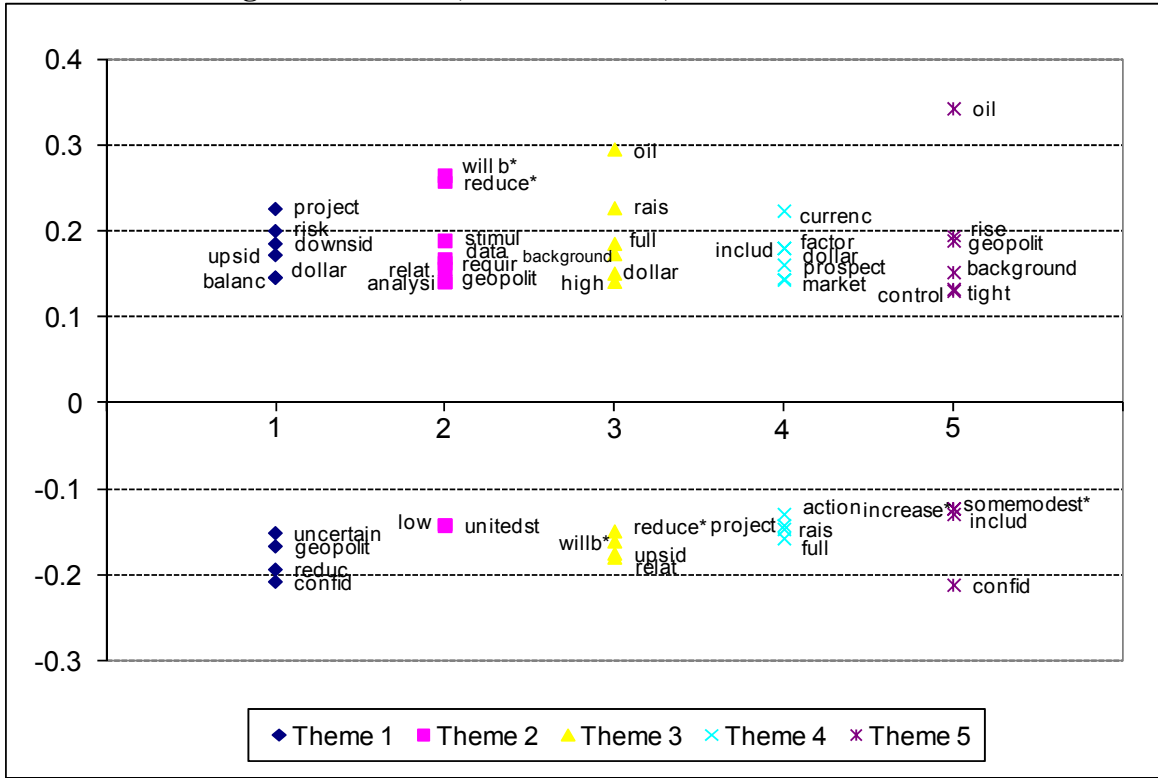
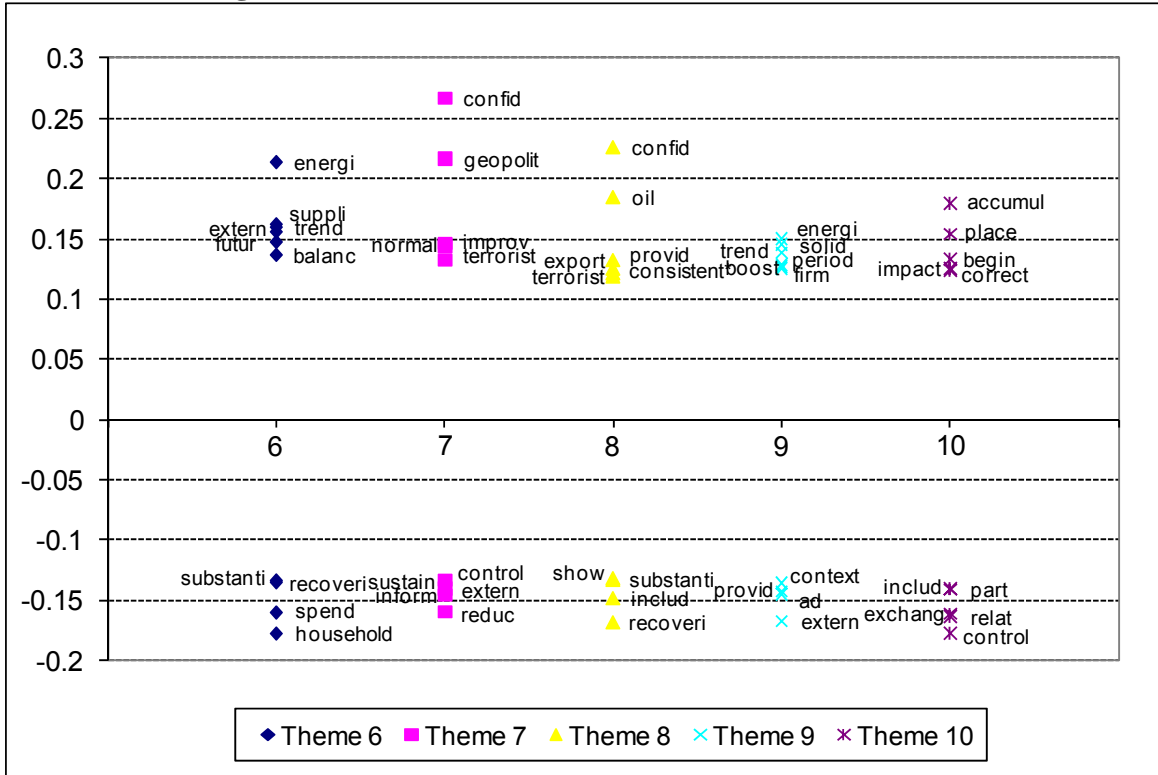


Figure 6: Model 3, Themes 6 to 10, FAD Press Releases



Note: * indicates words appearing in a Forward Looking Statement (FLS).

Figure 7: Model 3, Themes 1 to 5, MPRs

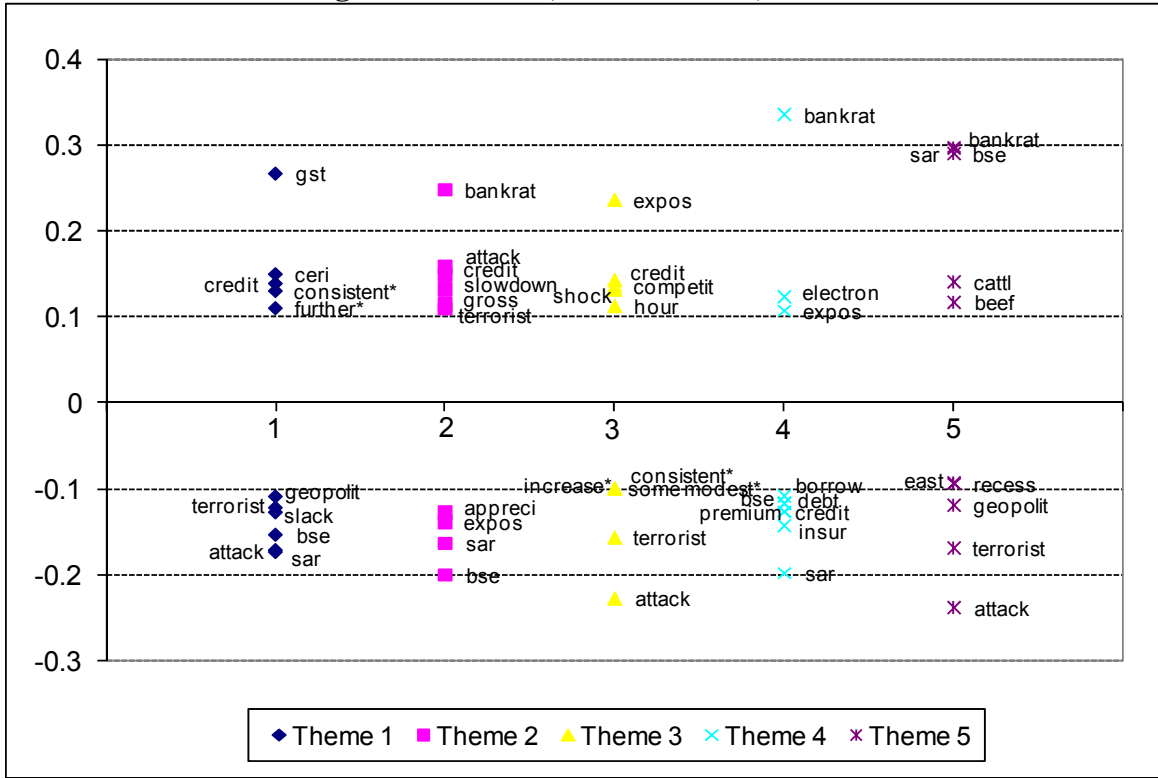
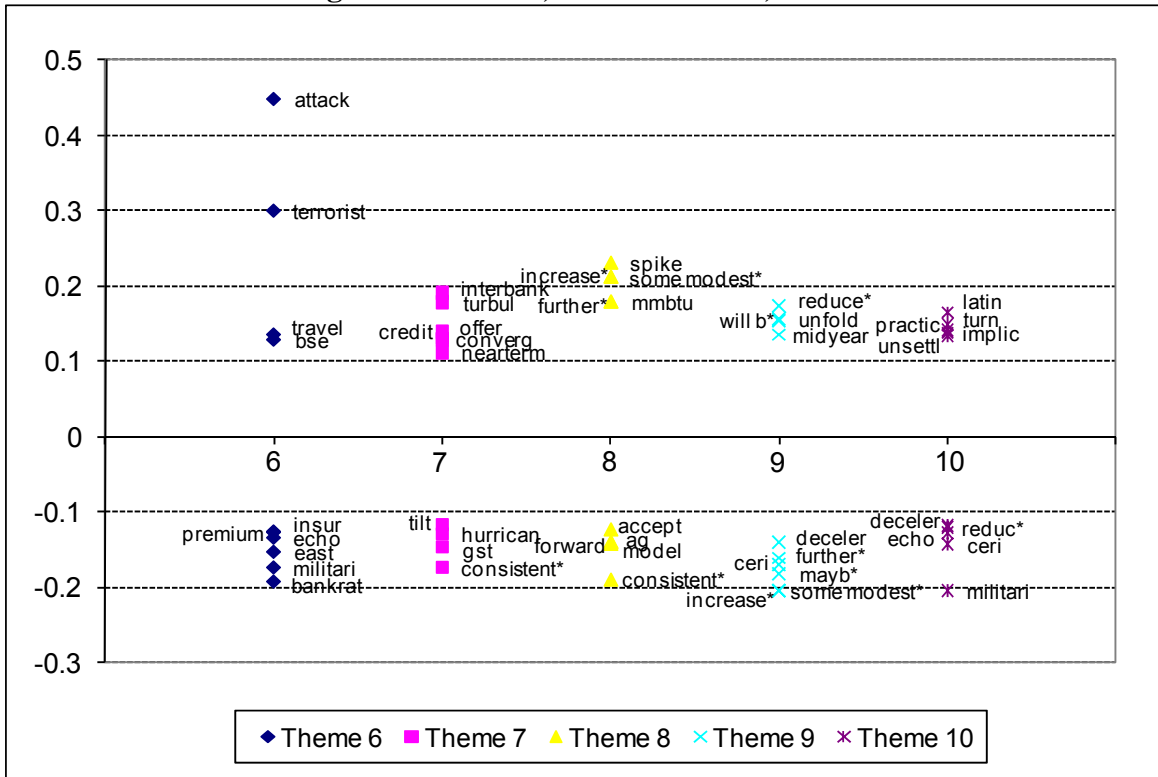


Figure 8: Model 3, Themes 6 to 10, MPRs



Note: * indicates words appearing in a Forward Looking Statement (FLS).

Figure 9: Model 1, Themes 1 to 5, FAD Press Releases

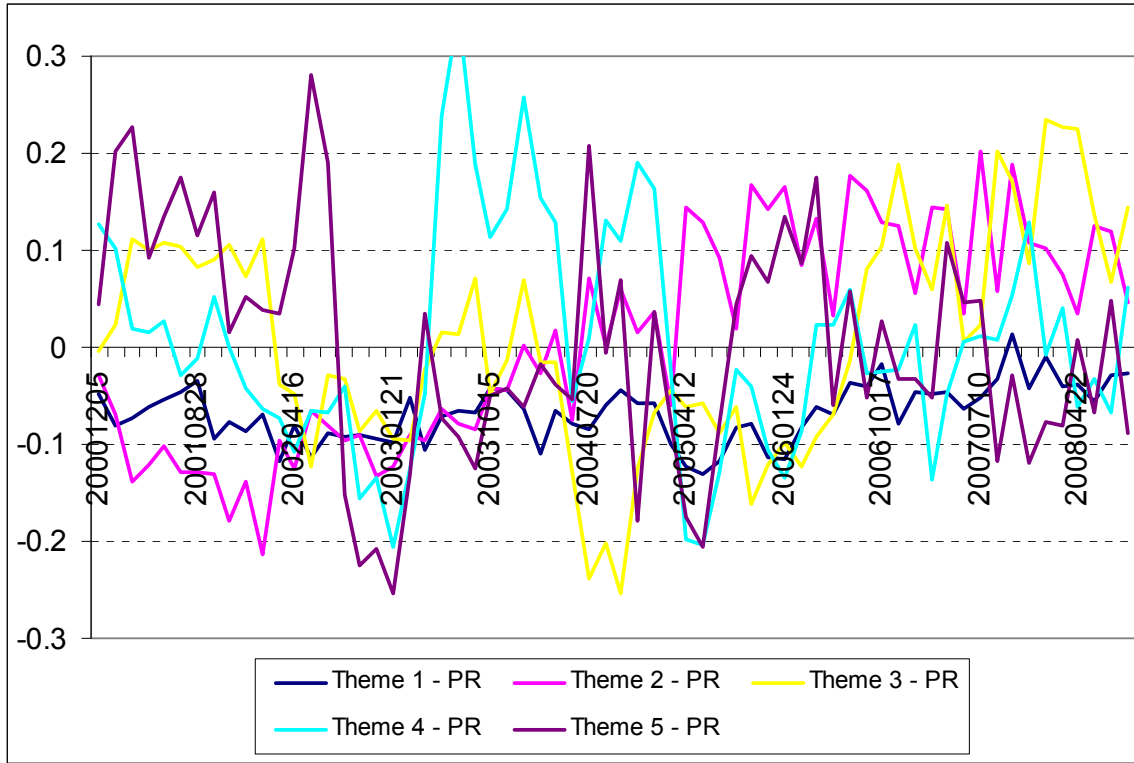


Figure 10: Model 1, Themes 6 to 10, FAD Press Releases

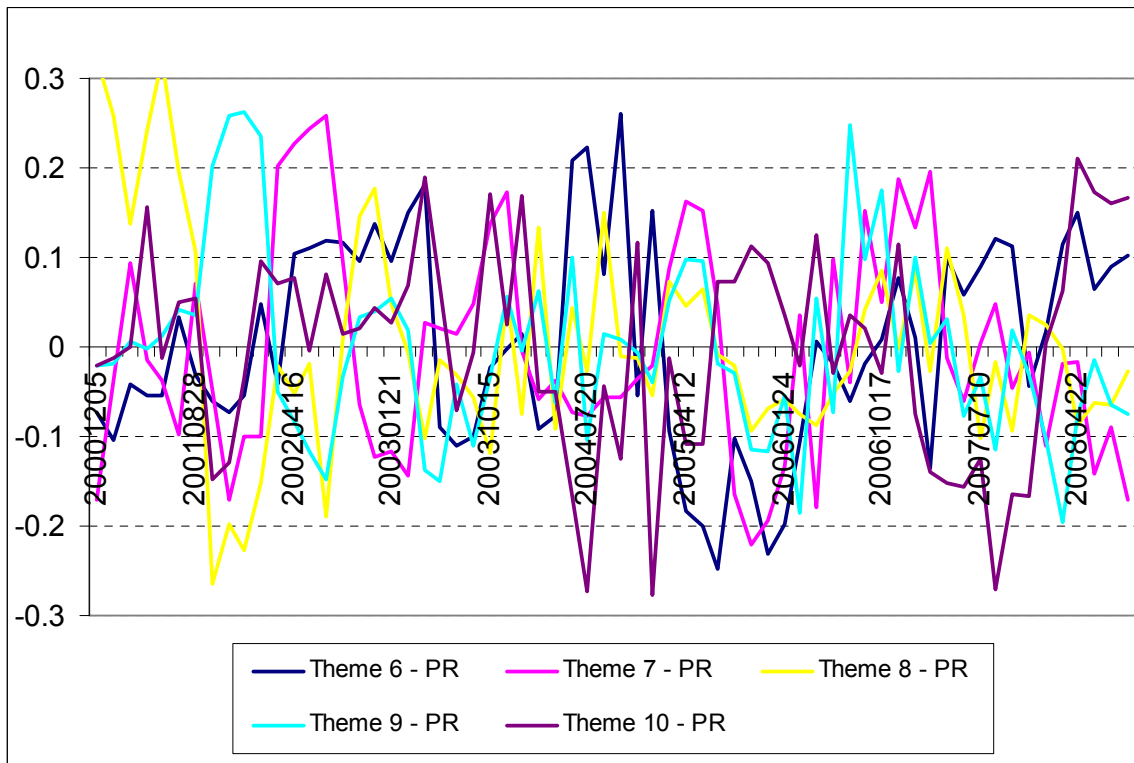


Figure 11: Model 1, Themes 1 to 5, MPRs

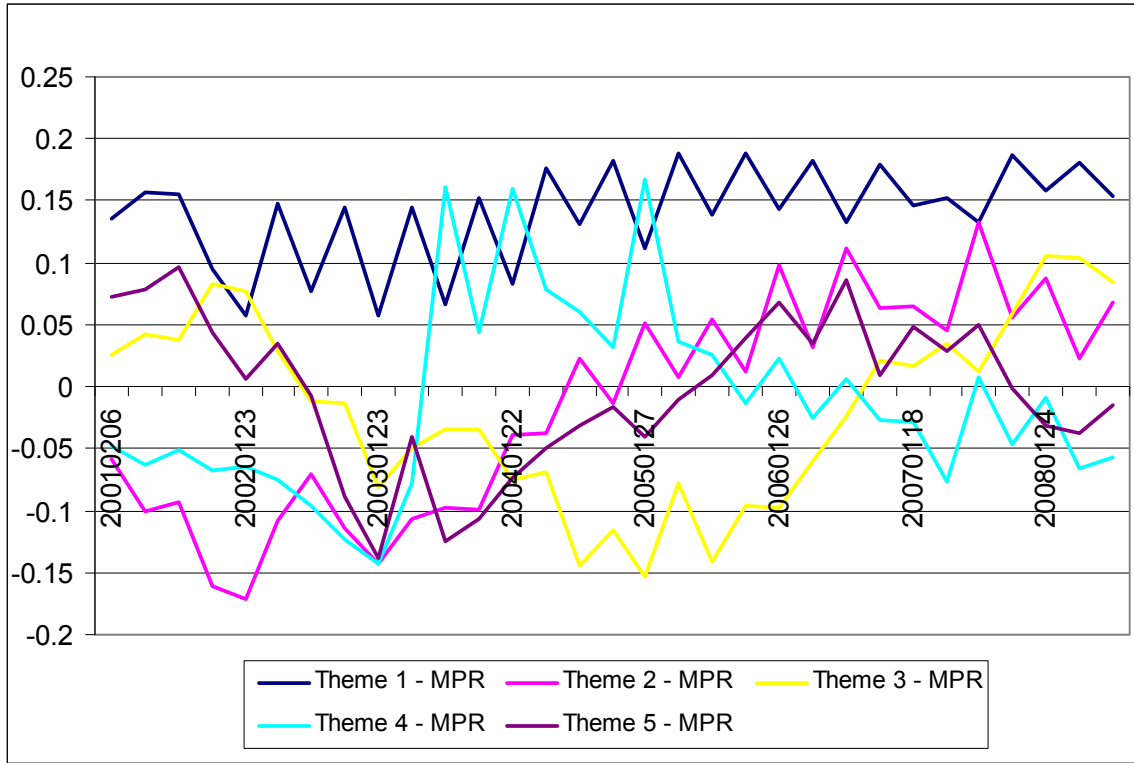


Figure 12: Model 1, Themes 6 to 10, MPRs

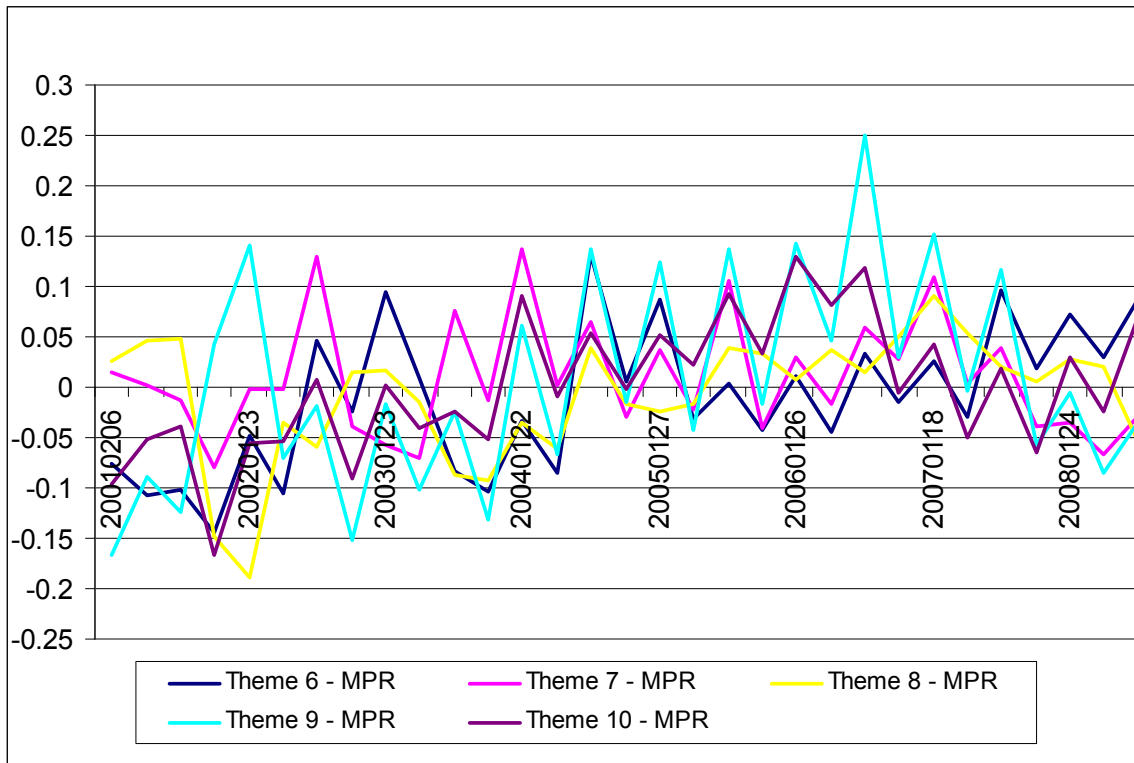


Figure 13: Model 2, Themes 1 to 5, FAD Press Releases

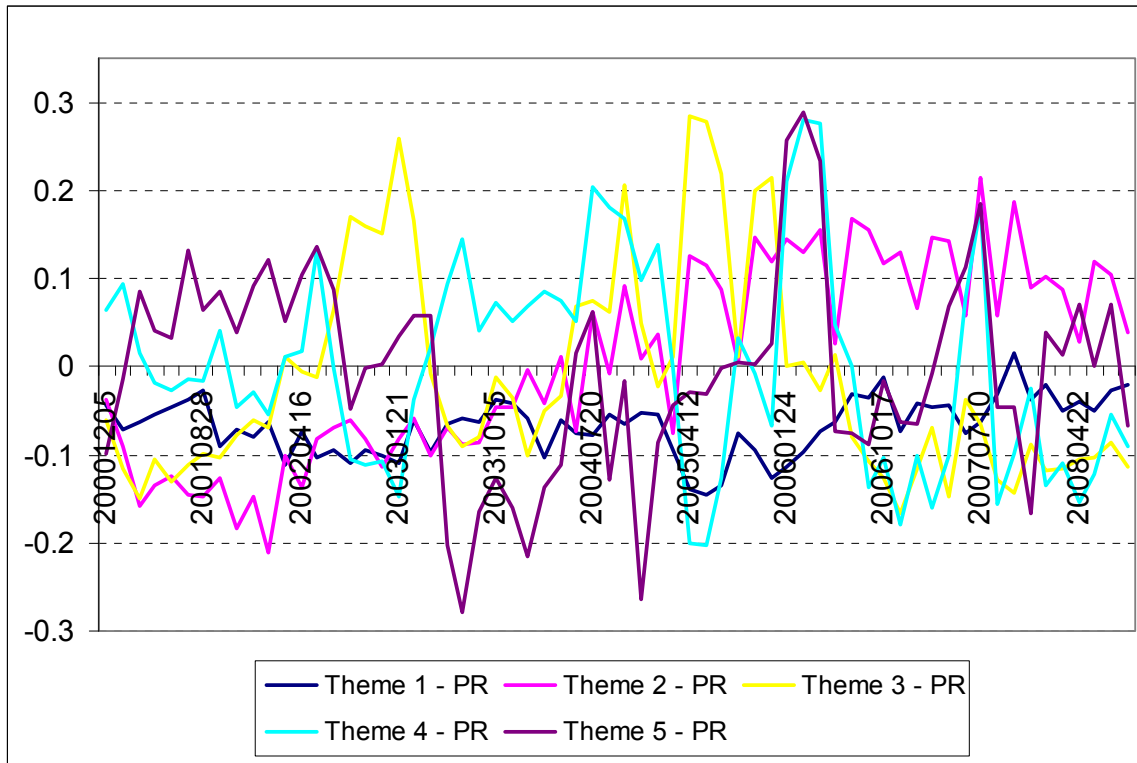


Figure 14: Model 2, Themes 6 to 10, FAD Press Releases

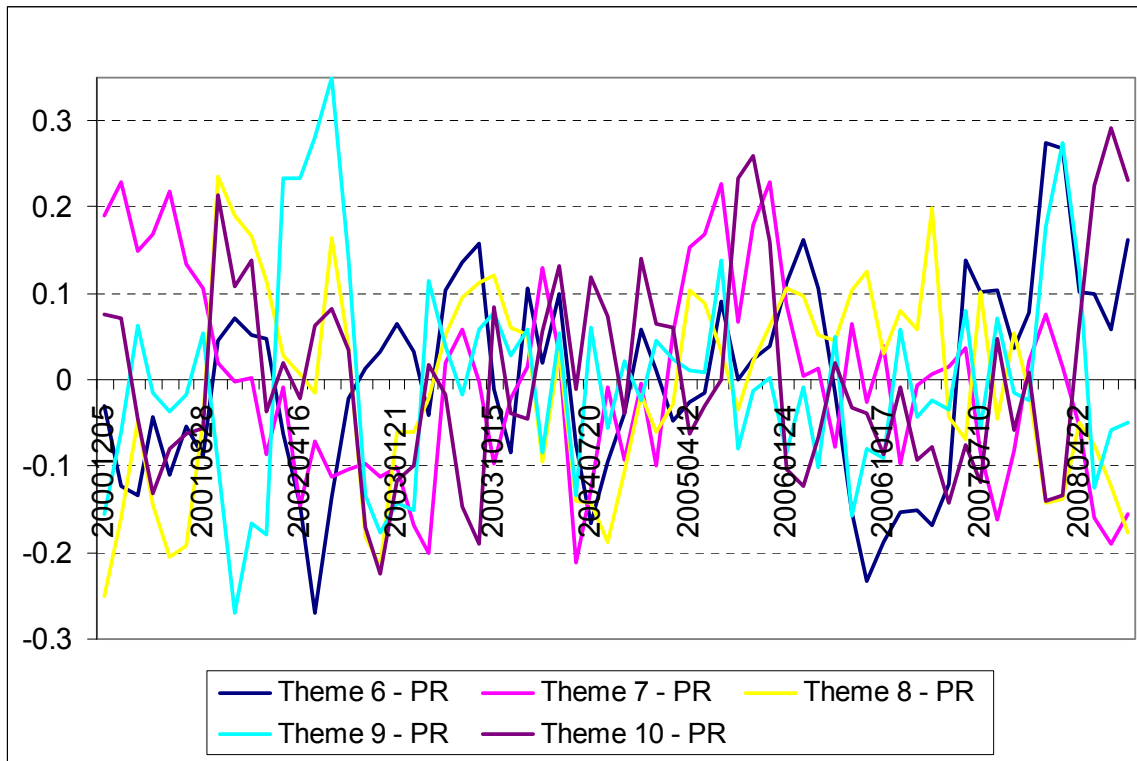


Figure 15: Model 2, Themes 1 to 5, MPRs

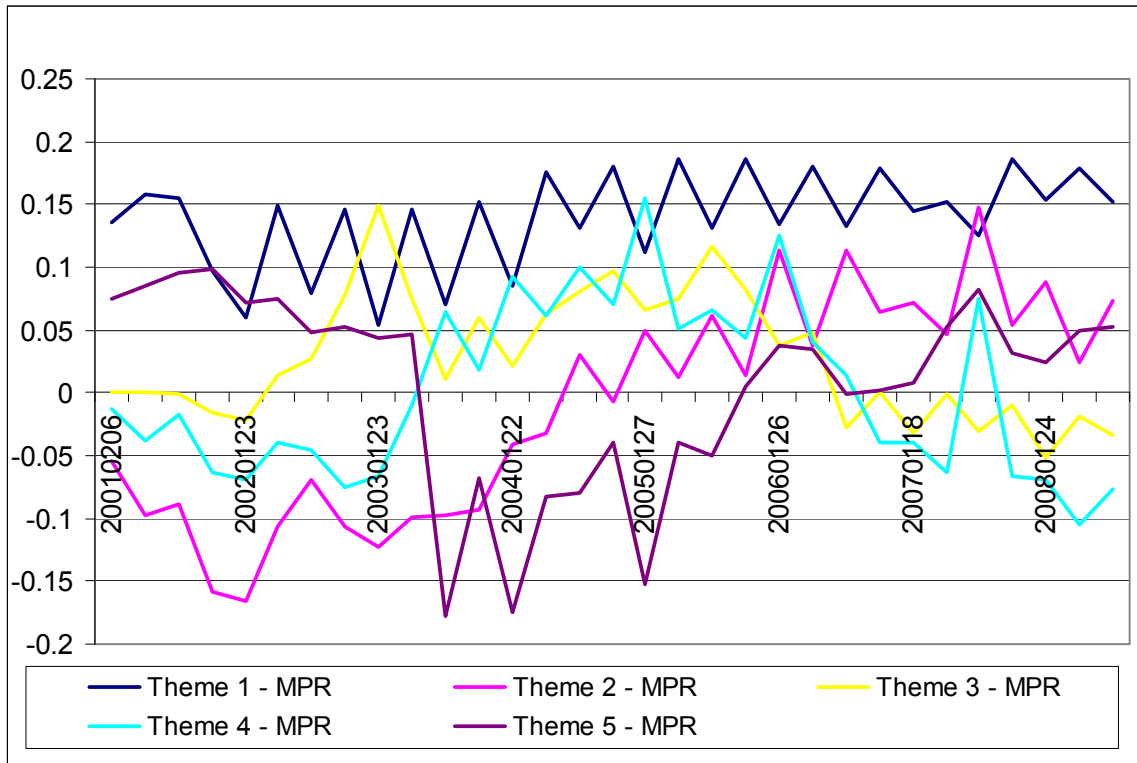


Figure 16: Model 2, Themes 6 to 10, MPRs

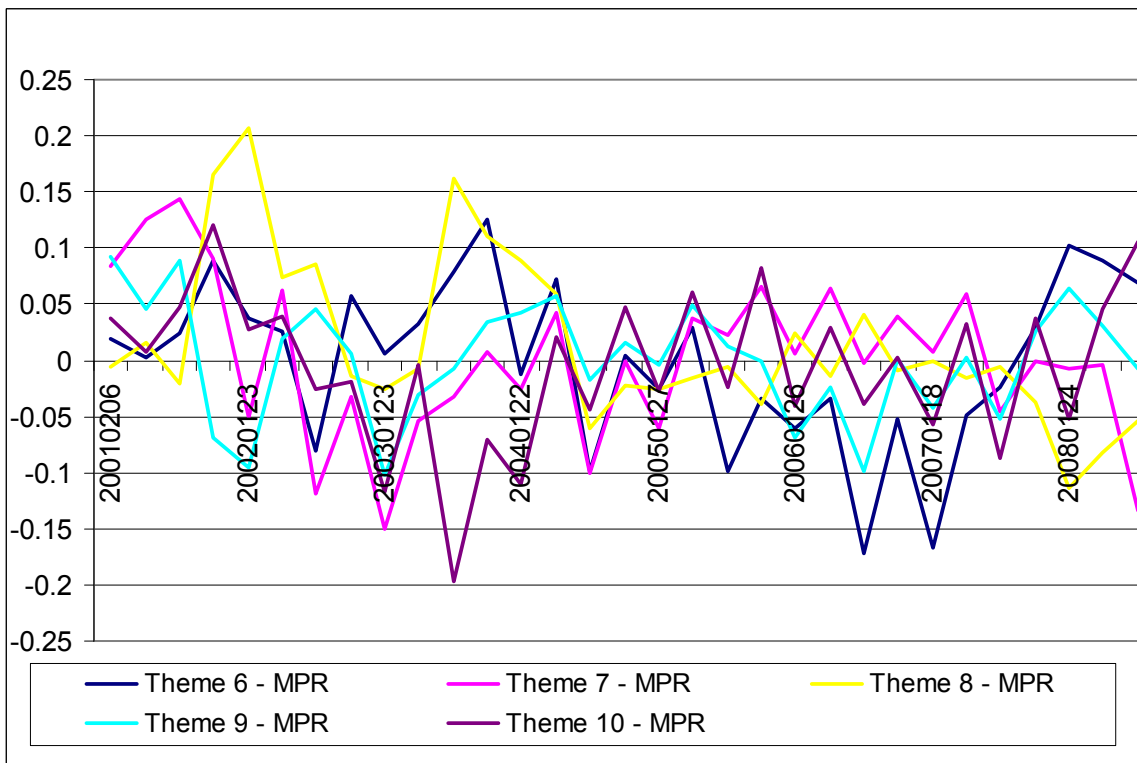


Figure 17: Model 3, Themes 1 to 5, FAD Press Releases

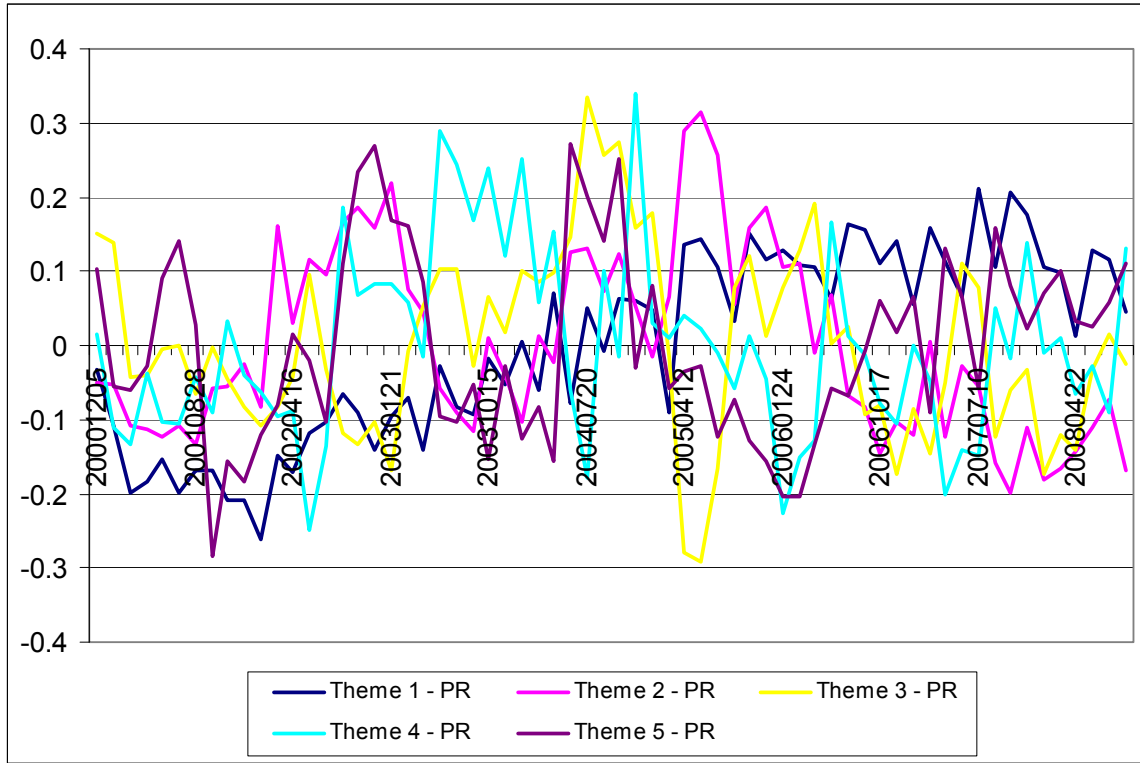


Figure 18: Model 3, Themes 6 to 10, FAD Press Releases

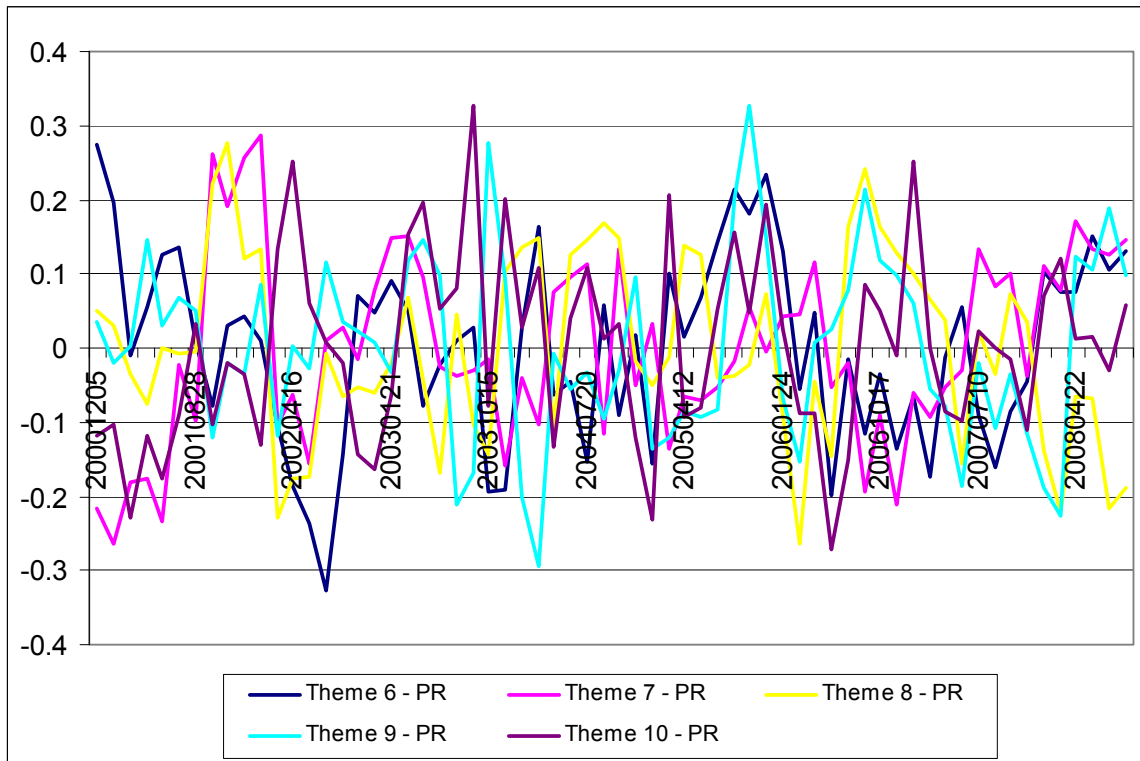


Figure 19: Model 3, Themes 1 to 5, MPRs

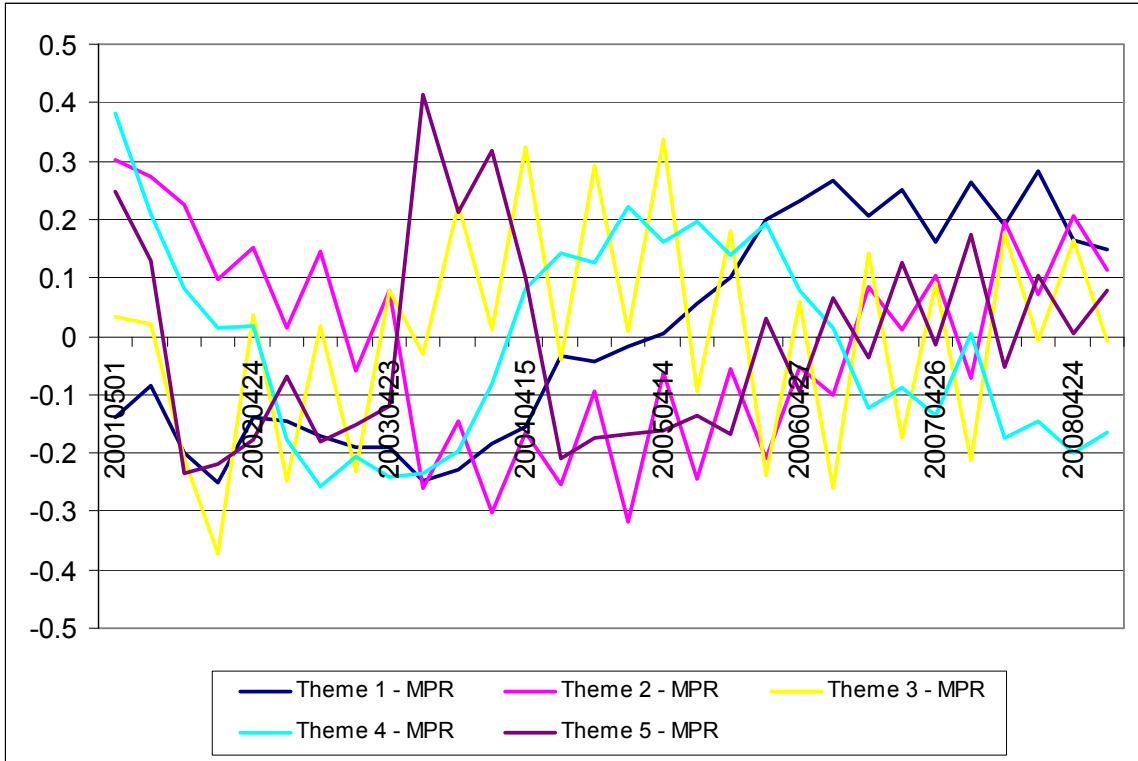


Figure 20: Model 3, Themes 6 to 10, MPRs

