3

Managing real-time risks and returns: The Thomson Reuters NewsScope Event Indices

Alexander D. Healy and Andrew W. Lo[1]

ABSTRACT

As financial markets grow in size and complexity, risk management protocols must also evolve to address more challenging demands. One of the most difficult of these challenges is managing event risk, the risk posed by unanticipated news that causes major market moves over short time intervals. Often cited but rarely managed, event risk has been relegated to the domain of qualitative judgment and discretion because of its heterogeneity and velocity. In this chapter, we describe one initiative aimed at solving this problem. The Thomson Reuters NewsScope Event Indices Project is an integrated framework for incorporating real-time news from the Thomson Reuters NewsScope subscription service into systematic investment and risk management protocols. The framework consists of a set of real-time event indices—each one taking on numerical values between 0 and 100—designed to capture the occurrence of unusual events of a particular kind. Each index is constructed by applying disciplined pattern recognition algorithms to real-time news feeds, and validated using econometric methods applied to historical data.

3.1 INTRODUCTION

As financial markets grow in size and complexity, risk management protocols must also evolve to address more challenging demands. One of the most difficult of these challenges is managing "event risk", the risk posed by unanticipated news that causes major market moves over short time intervals. Examples include terrorist events like September 11, 2001, contagion effects like the Quant Meltdown of August 7–9, 2007, and system glitches like the "Flash Crash" of May 6, 2010. Often cited but rarely managed, event risk has been relegated to the domain of qualitative judgment and discretion

[1] The views and opinions expressed in this chapter are those of the authors only, and do not necessarily represent the views and opinions of AlphaSimplex Group, MIT, or any of their affiliates and employees. The authors make no representations or warranty, either expressed or implied, as to the accuracy or completeness of the information contained in this chapter, nor are they recommending that this chapter serve as the basis for any investment decision—this chapter is for information purposes only. This research was supported by AlphaSimplex Group, LLC and Thomson Reuters.

The Handbook of News Analytics in Finance Edited by L. Mitra and G. Mitra

because of its heterogeneity and velocity. If we cannot measure it, we cannot manage it, and text-based news is hard to quantify.

In this chapter, we describe one initiative aimed at solving this problem. The Thomson Reuters NewsScope Event Indices Project is an integrated framework for incorporating real-time news from the Thomson Reuters NewsScope subscription service into systematic investment and risk management protocols. The framework consists of a set of real-time event indices—each one taking on numerical values between 0 and 100—designed to capture the occurrence of unusual events of a particular kind. For example, the `Macro` index measures the real-time quantity of macroeconomic news, and the `NatDist` index measures the real-time quantity of natural disaster news. Each index is constructed by applying disciplined pattern recognition algorithms to real-time newsfeeds, and calibrated using econometric methods applied to historical data. In this first release, we construct indices that are calibrated to foreign exchange markets; future releases will focus on other markets.

In this chapter, we describe the procedures for constructing and validating the Thomson Reuters/AlphaSimplex Event Indices. We begin with a brief literature review in Section 3.2, and in Section 3.3 we introduce the historical datasets used to calibrate the indices. Section 3.4 contains the algorithms used to construct the indices. In Section 3.5, we describe the event study methodology for validating the indices, and in Section 3.6 we explore the connection between realized volatility (our metric for market impact) and implied volatility. We conclude in Section 3.8.

3.2 LITERATURE REVIEW

There is a surprisingly rich literature on the relationship between news and financial markets going back to Niederhoffer's (1971) pioneering study of world events and stock prices, where world events were defined as five- to eight-column headlines in the *New York Times* and then organized into categories of meaning. Niederhoffer found that large stock price changes did follow world events more than randomly selected days, but that a particular category into which a world event falls did not add much incremental information about future price movements.

Measuring public information by the number of news releases by Reuter's News Service per unit of time, Berry and Howe (1994) showed that there is a positive, moderate relationship between public information and trading volume. Engle and Ng (1993) defined the "news impact curve" which measures how new information is incorporated into volatility estimates. However, by studying the number of news announcements reported daily by Dow Jones & Co., Mitchell and Mulherin (1994) did not find any strong relation between news and market activity. Hong, Lim, and Stein (2000) confirmed that firm-specific information, especially negative information, diffuses only gradually across the investing public.

On the macroeconomic front, Pearce and Roley (1985) showed that on announcement days surprises related to monetary policy significantly affect stock prices, but only found limited evidence of an impact from inflation surprises and no evidence of an impact from real activity surprises.

More recently, papers by Antweiler and Frank (2004), Das, Martinez-Jerez, and Tufano (2005), Tetlock (2007), and Leinweber and Sisk (this volume, Chapter 6) document interesting connections between news, volatility, and stock returns. Chan (2003) shows that the volume of news can explain the difference between mean reversion and momentum in monthly stock returns. And Tetlock, Saar-Tsechansky, and Macskassy (2008) show that simple quantitative measures of language can be used to predict individual firms' accounting earnings and stock returns.

3.3 DATA

Information needed for real-time investment decisions reaches traders through a multitude of news sources such as Thomson Reuters, Bloomberg, and CNN. The event indices described in this chapter reflect the issuance of market-moving information contained in the Thomson Reuters NewsScope Archive. As a proxy for the universe of news sources available to traders, we chose the English-language news from the Thomson Reuters NewsScope feed and, in particular, we have focused on news "alerts" (i.e., the quick news flashes that are issued "[w]hen a newsworthy event occurs"—according to the *Reuters NewsScope Archive User Guide*, V1.0). The basic empirical properties of this dataset are described in Sections 3.1 and 3.A.2 (see appendix on p. 102).

To calibrate the parameters of our news event indices, we use real-time Thomson Reuters foreign exchange spot data, which consist of interbank quotes for 45 currency pairs from January 1, 2003 through July 31, 2007. The characteristics of this dataset are summarized in Sections 3.2 and 3.A.1 (see appendix on p. 100).

3.3.1 News data

Some examples of Thomson Reuters NewsScope alerts include

```
02 AUG 2007 04:44:26.155
TSUNAMI WARNING ISSUED FOR JAPAN'S WESTERN HOKKAIDO COAST
NHK JP ASIA NEWS DIS LEN RTRS

17 AUG 2007 12:16:31.344
FED SAYS DATA SUGGESTS U.S. ECONOMY HAS CONTINUED TO EXPAND AT
  MODERATE PACE
US WASH MCE FED GVD DBT PLCY STIR INT CEN EU WEU FR FIN BNK FRX MTG
  ECB LEN RTRS

22 AUG 2007 20:26:57.587
MOODY'S DOWNGRADES RATINGS OF 120 SUBPRIME RMBS TRANCHES ISSUED
  IN 2005
MTG ABS FINS DBT AAA USC US LEN RTRS
```

This information-rich choice of news inputs has a number of advantages. In intraday risk management or in trading strategy applications, the event indices may race head to head against human response times. Therefore it is vital that they respond in a timely manner and reflect the most current news. The machine-readable Thomson Reuters NewsScope feed is updated on a subsecond basis, allowing the news indices to reflect timely news. Also, by focusing on news alerts, we help to ensure that the indices reflect the most current news.[2]

Furthermore, the characteristics of Thomson Reuters alerts lend themselves to machine analysis. Their textual content is concise and built from a relatively small vocabulary. As a result, we can use robust, simple algorithms to extract information from the text. Another advantage is that Thomson Reuters data are tagged with machine-readable codes that characterize the alerts' topic areas and other important metadata, a powerful aid in analyzing their content.

A preliminary analysis of the NewsScope historical dataset reveals strong seasonality on intraweekly, intradaily, and intrahourly timescales, as expected. However, to identify those times at which incoming news is especially relevant to the market, it is necessary to distinguish true bursts of information from mere seasonal peaks in volume. We present our solution to this challenge in Section 3.4.

Some examples of the seasonalities are as follows: the median weekday sees 1,500 to 2,000 alerts arrive, while over the entire weekend there are typically only 130. Also, as one might expect, few (English language) alerts arrive at midnight GMT, a time when the workday is over in both Europe and America. On an intrahour timescale, alerts arrive more frequently on the hour or half-hour than at other times due to press release schedules and other planned announcements. See Section 3.A.2 (see p. 102) for a more detailed discussion of the seasonality of arrival of English-language alerts.

3.3.2 Foreign exchange data

Because the event indices' role is to rapidly identify and report the arrival of market-moving information, to validate their quality one needs a metric that indicates whether market movements did, in fact, occur. In this first version, the event indices were to be calibrated against foreign exchange markets; we used Thomson Reuters foreign exchange spot data, which consist of interbank quotes for 45 currency pairs since January 1, 2003.

Following convention (see Dacorogna et al., 2001) we approximated tick-by-tick market prices using the geometric mean of bid and ask quotes:

$$p_t \equiv \sqrt{p_{t,\mathrm{bid}} \cdot p_{t,\mathrm{ask}}} \,. \tag{3.1}$$

The dataset was then homogenized at 5-second intervals to facilitate computation while retaining subminute granularity.[3] However, it makes little sense to quantify news impact by measuring the price level. Instead, we consider the instantaneous change in level (5-second log returns):

$$r_{t,5} \equiv \log p_t - \log p_{t-5} \tag{3.2}$$

and the instantaneous variation in level (squared 5-second log returns): $r_{t,5}^2$. For tick-by-tick measurement of volatility, squared returns are our preferred metric because of their similarity to conventional realized volatility (a trailing measure that characterizes multi-

[2] This is in contrast to the follow-on stories that tend to appear 5 to 20 minutes later which provide further details on the event.
[3] Specifically, every 5 seconds we choose the most recent quote to represent the current price; however, if there have been no quotes in the last 30 seconds, we treat the data as missing rather than use outdated quotes.

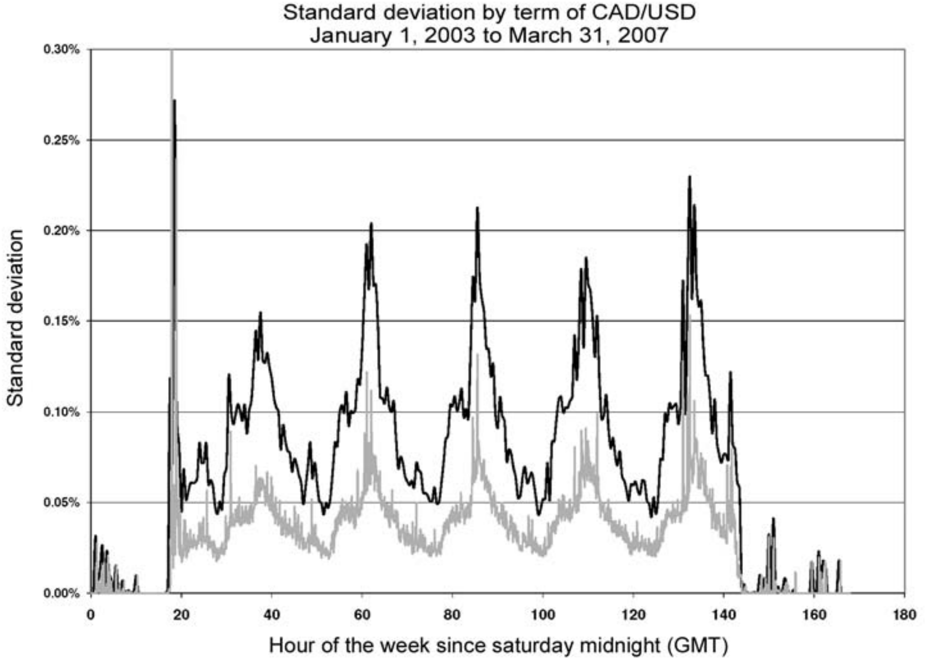

Figure 3.1. Average realized volatility of the CAD/USD exchange rate return over the course of a week, averaged over 5-minute (gray) and 30-minute (black) timescales. Note the strong daily peaks.

tick time periods). We note that the relationship between squared 5-second returns and the realized volatility over the period $[t_1, t_2]$ is as follows:

$$v_{t_1,t_2,5} = \sqrt{\frac{1}{\frac{t_2 - t_1}{5}} \sum_{t_1 \le \tau \le t_2} (r_{\tau,5} - \bar{r})^2} \approx \sqrt{\frac{1}{\frac{t_2 - t_1}{5}} \sum_{t_1 \le \tau \le t_2} r^2_{\tau,5}} \tag{3.3}$$

where $\bar{r}$ is the average return in period $[t_1, t_2]$.

As with NewsScope alerts, volatility exhibits strong seasonalities on intradaily and intraweekly timescales (see Figure 3.1). As one might expect, these seasonalities are only found in the squared returns, not in the returns themselves. This discovery raised the specter of specious results based on the correlation between news alert seasonality and FX volatility seasonality (as measured by squared returns). This potential difficulty is dealt with in Section 3.5, where seasonality is removed. For additional analysis of the properties of the Thomson Reuters FX dataset see Section 3.A.1 on p. 100.

3.4 A FRAMEWORK FOR REAL-TIME NEWS ANALYTICS

The core of our real-time news analysis engine relies on a scoring method that assesses the relative volume/significance of news from a specific category of news. For instance, we wish to identify periods when the volume of news about foreign exchange markets is abnormally high, or when there is a flurry of macroeconomic news announcements.

For a given topic, say foreign exchange news, the scoring procedure has the following parameters:

- A list of keywords/key phrases and real-valued weights: $(W_1, \gamma_1), \ldots, (W_k, \gamma_k)$.
- A rolling window size, ℓ (typically about 5–10 minutes).
- A calibration rolling window size, L (typically about 90 days).

The keywords list and the last ℓ minutes of news are used to create a raw score, and this score is normalized/calibrated using statistics about the news over the last L days (as described below).

3.4.1 Assigning scores to news

The score at a given point in time, t, is assigned as follows: Let $(w_1, \ldots, w_k)$ be the vector of keyword frequencies in the time interval $[t - \ell, t)$ (i.e., w_i is the number of times word/phrase W_i has appeared in the last ℓ minutes). The raw score at time t is then defined to be:

$$s_t \equiv \sum_i \gamma_i w_i \ . \tag{3.4}$$

In this form, the raw score will tend to be high when news volume is high, and so we calibrate/normalize the score using the calibration rolling window: We maintain a record of the scores that have been assigned over the last L days, along with the news volume (measured in words per ℓ minutes) at the time that score was issued. If we denote by $n_{[t-\ell,t)}$ the number of words that have been observed in the time interval $[t - \ell, t)$, then the normalized score is defined by comparing the raw score to the distribution of scores in the calibration window that had the same news volume $n_{[t-\ell,t)}$.

Specifically, the normalized score is equal to the fraction of scores—among scores in the calibration window that had the same news volume—that are less than the current score. Formally:

$$S_t \equiv \frac{\left|\{t' \in [t-L, t) : n_{[t'-\ell,t)} \text{ and } \mathrm{s_{t'}} < \mathrm{s_t}\}\right|}{\left|\{t' \in [t-L, t) : n_{[t'-\ell,t)}\}\right|} . \tag{3.5}$$

Thus, a score of $S_t = 0.92$ can be interpreted as "92% of the time, when the news volume is at the current level, the raw score is less than it currently is."

3.4.2 A natural extension to alerts

The scoring procedure described above is very flexible and, in particular, also has a natural extension to incorporating Thomson Reuters topic codes into the scoring. Specifically, if instead of counting word frequencies we count the fraction of news alerts in the last ℓ minutes that have been tagged with various topic codes, then we can assign scores in exactly the same way, the only difference being that we measure news volume by the number of alerts that appear (rather than the number of words that appear).

Formally, we have the following parameters:

- A list of topic codes and real-valued weights: $(W_1, \gamma_1), \ldots, (W_k, \gamma_k)$.
- A rolling window size, ℓ.
- A calibration rolling window size, L.

The score at a given point in time, t, is assigned in an analogous way. Let $(w_1, \ldots, w_k)$ be the vector of topic code frequencies in the time interval $[t - \ell, t)$ (i.e., w_i is the number of times the topic code W_i has appeared in the last ℓ minutes). The raw score at time t is then defined to be:

$$\sum_i \gamma_i w_i \ . \tag{3.6}$$

Just as before, we calibrate and normalize the score using the calibration rolling window: We maintain a record of the scores that have been assigned over the last L days, along with the news volume (measured in words per ℓ minutes) at the time that score was issued. If we denote by $n_{[t-\ell,t)}$ the number of alerts that have been observed in the time interval $[t - \ell, t)$, then the normalized score is defined by comparing the raw score with the distribution of scores in the calibration window that had the same news volume $n_{[t-\ell,t)}$, again by using formula (3.5). Table 3.1 lists the 45 news indices we have constructed and tested using this approach.

3.4.3 Creating keyword and topic code lists

The scoring mechanism described in Sections 3.4.1 and 3.4.2 relies on a list of keywords/topics, together with real-valued weights. The lists were created by first selecting the major news categories they should capture (foreign exchange, natural disasters, etc.) and then creating, by hand, lists of words/topics that suggested news relevant to these categories. These lists were then honed by examining the news that contained high concentrations of these words and adjusting the lists to remove words that were consistently misrepresenting the meaning of the text, and to add new words/phrases. Because this can be a very arduous task, we developed a tool (see Figure 3.2) that extracts news from the period when our indices assign high scores. The news is then presented, with keywords highlighted, and shows how the score evolves over time. Thus, one can quickly and easily determine whether the keywords that contributed to the high score are legitimate, or whether the keywords (and weights) need to be adjusted.

3.4.4 Algorithmic considerations

Given the vast amounts of data involved in this study, some care is necessary to ensure that the algorithms and data structures that are employed are efficient (both in terms of speed and memory use). In particular, maintaining the large rolling "calibration window", described above, is one case where novel algorithmic ideas are important to implementing our approach.

A naive approach to implementing the large rolling window would simply store all previous scores (for the last 90 days) in an array; however, our scoring procedure requires computing the percentile of a new score every second, and to do this for n unstructured data items would seem to require on the order of n operations. Here, 90 days of scores represents $n = 60 \cdot 60 \cdot 24 \cdot 90 = 7{,}776{,}000$ samples, which might be a feasible number for online scoring once per second (as in the final real-time indices), but is too much for rapidly simulating the scoring on months, or even years, of data. To construct the indices from historical data and to refine them in the future, it is essential to be able to simulate years' worth of scores in a matter of minutes (or at most hours).

Table 3.1. Base indices of the Thomson Reuters NewsScope Event Indices family

Base index	Description
Agricultural (topic)	Agricultural topics (as classified by Reuters) such as cotton/silk, grains, cocoa, etc.
ASIA	Asia (as classified by the Reuters topic code ASIA)
Banking (keyword)	Banks, lending, mortgages, and other areas relevant to banking
Bearish (keyword)	Indicates negative market conditions, low earnings, poor sales, drops in financial indices, etc.
Bonds (topic)	Topics related to bonds (as classified by Reuters)
Bullish (keyword)	Indicates positive market conditions, high earnings, strong sales, surges in financial indices, etc.
Central bank (keyword)	Monetary policy, interest rates, inflation, and other central-bank-related subjects
Central bank (topic)	Central banks (as classified by Reuters)
Corporate (keyword)	Earnings, dividends, and other corporation-related subjects
Credit (topic)	Credit default swaps, mortgages, real estate, bankruptcies, and other credit topics
Economic (topic)	Economic indicators, trade, and other economic topics
Emerging markets (topic)	Emerging markets (as classified by the Reuters topic code EMRG)
Emotional (keyword)	Contains emotional terms and subject matter such as fear, apprehension, relief, and nervousness
Energy (topic)	Energy topics (as classified by Reuters)
EUROPE	Europe (as classified by the Reuters topic code EUROPE)
Finance (keyword)	General finance subjects such as brokerages, underwriting, and financial markets
Foreign exchange (keyword)	Foreign exchange, such as monetary policy, announcements from finance ministers, and specific currencies
Foreign exchange (topic)	Topics related to foreign exchange (as classified by Reuters)
GB	Great Britain (as classified by the Reuters topic code GB)
JP	Japan (as classified by the Reuters topic code JP)
Livestock (topic)	Livestock (as classified by the Reuters topic code LIV)
Macroeconomic (keyword)	Macroeconomic subjects such as housing, inflation, and manufacturing
Macroeconomic (topic)	Macroeconomic topics (as classified by Reuters)
Major news (topic)	News in major news topics (as classified by Reuters)
Markets (topic)	Exchanges, hedge funds, and investing
Mergers (keyword)	Mergers, acquisitions, takeovers, and other merger-related subjects
Metal (topic)	Metals (as classified by the Reuters topic code MET)
Military (keyword)	Intelligence, homeland security, fighting, and other military actions
MX	Mexico (as classified by the Reuters topic code MX)
Natural disaster (keyword)	Hurricanes, earthquakes, tropical storms, mudslides, and other natural disasters
Natural disaster (topic)	Weather and disasters (as classified by Reuters)
Oil (topic)	Oil and oil-producing regions (as classified by Reuters)
Political (keyword)	Political subjects such as elections, legislation, referenda, and diplomacy
Political (topic)	Political topics (as classified by Reuters)
Precious metal (topic)	Precious metals (as classified by Reuters)
Rates (topic)	Interest rates (as classified by Reuters)

Base index	Description
RCH	Broker research (as classified by the Reuters topic code RCH)
Regulation (topic)	Regulation (as classified by Reuters)
Stocks (topic)	Stocks and investment funds (as classified by Reuters)
Terrorism (keyword)	Terrorist actions and related violence
Terrorism (topic)	Topics relevant to terrorism
Urgent news (topic)	Urgent news (as classified by Reuters)
US	The US (as classified by the Reuters topic code US)
VIO	Violence (as classified by the Reuters topic code VIO)
Violence (keyword)	War, fighting, and other violence

```
2003-03-14 13:39:50: 0.0000 (k=0.0, n=500, percentile=0.0000)
CORRECTED-ANDREW CORP<ANDW.O> Q2 MULTEX EPS VIEW PROFIT $0.04 (NOT LOSS $0.04)
ELC USC DBT RES RESF TEL LEN RTRS

2003-03-14 13:40:14: 0.0000 (k=0.0, n=297, percentile=0.0000)
RPT-ANDREW CORP <ANDW.O> SEES Q2 SALES $190-$200 MLN
ELC USC DBT RES RESF TEL LEN RTRS

2003-03-14 13:40:31: 0.0000 (k=0.0, n=284, percentile=0.0000)
RPT-ANDREW CORP <ANDW.O> Q2 MULTEX REVS VIEW $239.4 MLN
ELC USC DBT RES RESF TEL LEN RTRS

2003-03-14 13:40:34: 0.0000 (k=0.0, n=284, percentile=0.0000)
RPT-ANDREW SEES Q2 SHR LOSS $0.03-$0.06 VS PREVIOUS VIEW PROFIT $0.01-$0.04
ELC USC DBT RES RESF TEL LEN RTRS

2003-03-14 13:40:50: 0.0000 (k=0.0, n=295, percentile=0.0000)
RPT-ANDREW CITES CUSTOMER CAP EX LEVELS FALLING AT GREATER THAN EXPECTED RATES
ELC USC DBT RES RESF TEL LEN RTRS

2003-03-14 13:42:57: 0.0000 (k=0.0, n=239, percentile=0.0000)
MEDICINES CO <MDCO.O> SAYS PRICES 4.9 MLN SHR OFFERING AT $17.50 PER SHR
DRU US ISU LEN RTRS

2003-03-14 13:43:52: 0.0000 (k=0.0, n=221, percentile=0.0000)
BUSH TO TRAVEL TO AZORES SUNDAY FOR SUMMIT WITH SPAIN, UK ON IRAQ - WHITE HOUSE
US WASH VIO POL DIP MEAST NEWS IQ OPEC CRU PROD LEN RTRS

2003-03-14 13:45:10: 0.0176 (k=4.0, n=227, percentile=0.9840)
BUSH TO TRAVEL TO AZORES SUNDAY FOR SUMMIT WITH SPAIN, UK ON IRAQ - WHITE HOUSE
PT LEN RTRS

2003-03-14 13:45:54: 0.0176 (k=4.0, n=227, percentile=0.9839)
WHITE HOUSE SAYS SUMMIT PART OF LAST BIT OF IRAQ DIPLOMACY
US WASH VIO POL DIP MEAST NEWS IQ OPEC CRU PROD LEN RTRS

2003-03-14 13:45:55: 0.0312 (k=7.0, n=224, percentile=0.9971)
------------------------------------------------------------
Press ENTER to proceed to next event ('q' to quit)
------------------------------------------------------------
Proceding to next event . . . 2003-03-14 15:03:20
------------------------------------------------------------
2003-03-14 14:59:32: 0.0000 (k=0.0, n=58, percentile=0.0000)
POLISH M3 MONEY SUPPLY UP 1.1 PCT M/M IN FEBRUARY-CENTRAL BANK
PL EEU EMRG EUROPE FRX CEN LEN RTRS

2003-03-14 15:00:03: 0.0000 (k=0.0, n=72, percentile=0.0000)
MEXICO CENTRAL BANK SAYS LEAVES MONETARY POLICY UNCHANGED
MX LATAM EMRG INT GVD CEN FRX STX LEN RTRS

2003-03-14 15:00:20: 0.0000 (k=0.0, n=102, percentile=0.0000)
POLISH INFLATION 0.5 PCT Y/Y FEB VS 0.4 PCT FCAST, 0.5 PCT JAN, +0.1 PCT M/M
PL EEU EMRG EUROPE ECI INT LEN RTRS

2003-03-14 15:00:45: 0.0175 (k=2.0, n=114, percentile=0.9827)
BUSH TIES MIDEAST PEACE PLAN TO CONFIRMATION OF NEW PALESTINIAN PRIME MINISTER
US WASH VIO POL DIP MEAST NEWS IL PS IQ CRU OPEC LEN RTRS

2003-03-14 15:03:19: 0.0357 (k=2.0, n=56, percentile=0.9649)
BUSH-ISRAEL MUST END SETTLEMENT "ACTIVITY" AS PROGRESS IS MADE TOWARD PEACE
US WASH VIO POL DIP MEAST NEWS IL PS IQ CRU OPEC LEN RTRS

2003-03-14 15:03:20: 0.0441 (k=3.0, n=68, percentile=0.9977)
------------------------------------------------------------
Press ENTER to proceed to next event ('q' to quit)
------------------------------------------------------------
```

Figure 3.2. Screenshot of a tool for honing indices.

The approach we take is to build a data structure that allows for efficiently inserting new scores, removing old scores (after they are 90 days old) and computing percentiles. In particular, we achieve all these tasks in time $O(\log n)$, which, for the relevant sample sizes, represents only a few tens of operations, rather than several million (in the naive implementation described above). This data structure is an extension of *randomized treaps* that allows for fast percentile computations. We refer the reader to Cormen et al. (2001) for background on data structures and treaps, and offer a simplified presentation here.

The data are maintained in a *binary search tree*[4] where each node is augmented to contain an additional value that says how many values are stored in the subtree rooted at this node. Then, a straightforward extension to searching the binary tree allows one to compute how many values in the tree are less than a given value in time proportional to the depth of the tree. This is clearly equivalent to computing the percentile of the given value.

The remaining subtlety is that our binary search tree may not be *balanced*—i.e., its depth may be much larger than the optimal $O(\log n)$. If the tree is not balanced then the worst case performance of our searching/percentile computations may be very poor. Thus, it is imperative to maintain a *balanced* tree. An elegant solution to this problem is the random *treap* data structure, which combines a binary tree with a *heap* data structure to guarantee that the tree remains balanced (with high probability). We omit the details of heaps and treaps (which may be found in Cormen et al., 2001), and simply note that all treap operations can be extended to support our efficient percentile computations.

3.5 VALIDATING EVENT INDICES

To establish the empirical significance of our news indices, we use the event study methodology (see Campbell, Lo, and MacKinlay, 1997 for background on event studies). We review the basics of this well-known technique in Section 3.5.1, and provide a few illustrative examples in Section 3.5.2. We present formal statistical tests for the significance of news index events in Sections 3.5.3–3.5.5.

3.5.1 Event analysis

For a given index, event analysis is performed in the following manner. We compute the index over the sample period from January 1, 2003 to March 31, 2007, and declare that an "event" has taken place whenever the score exceeds a certain threshold, typically 0.995. We then remove any event that follows less than 1 hour after another event, which guards against having many events in quick succession that all reflect the same news event. We then analyze the behavior of exchange rates in the periods before and after these events.

In our analysis, we focus on two time-series describing the behavior of exchange rates.

[4] A binary search tree is a collection of nodes where every node (except for a designated "root" node) has exactly one parent and at most two children labeled left and right, with the property that the value in the left child is less than the value of the parent and the value in the right child is greater than the parent. Such a tree allows for efficient searching for a value v by starting at the root and going to the left or right child (depending on whether v is less than or greater than the value of the root node) and continuing down the tree in this way until the value is found. This allows for searching in time proportional to the depth of the tree (which may be as small as $O(\log n)$, where n is the number of nodes in the tree).

The first is the time-series of *log returns*, denoted $\{r_i\}_i$. Since we only have banks' quote data, this series is derived by taking the logarithm of the geometric mean of bid and ask quotes (as described in Section 3.3). The second time-series we consider is that of *de-seasonalized squared log returns*, denoted $\{s_i\}_i$, which is a measure of volatility in exchange rates. Since exchange rate volatilities exhibit strong weekly seasonalities (see Section 3.A.1 on p. 100), this volatility measure considers only excess volatility over typical seasonal volatility. In particular, this series is constructed by first considering the squared log returns $\{r_i^2\}_i$, from which we compute the *weekly seasonality*:

$$\hat{r}_i^2 \quad = \frac{1}{n} \quad \sum_{j=0}^{n} r^2_{(i \bmod W)+j \cdot W} \tag{3.7}$$

where n is the number of weeks in the data (220 in this case), and W is the number of samples in a week ($12 \cdot 60 \cdot 24 \cdot 7$ for 5-second returns). Finally, we define *de-seasonalized volatility* to be:

$$s_i \quad \equiv \quad \{r_i^2 - \hat{r}_i^2\}_i \ . \tag{3.8}$$

Using the events defined above, we test the null hypothesis that the distributions of returns and de-seasonalized squared log returns before events are the same as after the events.

For example, if we begin with the series of volatilities $\{s_i\}_i$, then we denote by $s_i^{(j)}$ the sample from time $i + t_j$, where t_j is the time of event j, and we consider the time-series $\vec{s}$ during a 1-hour window centered at each event; that is,

$$\begin{array}{c} s^{(1)}_{-30}, \ldots, s^{(1)}_0, \ldots, s^{(1)}_{30} \\ \vdots \\ s^{(k)}_{-30}, \ldots, s^{(k)}_0, \ldots, s^{(k)}_{30} \,. \end{array}$$

From these samples we can create an averaged event window:

$$\hat{s}_{-30}, \ldots, \hat{s}_{30} \quad , \quad \hat{s}_i \equiv 1k \sum_{j=1}^{k} s_i^{(j)} \ . \tag{3.9}$$

Then by studying the averaged event window we can assess the impact of the events comprising the event study. Naturally, this analysis can be applied to analyze log returns $\{r_i\}_i$ as well as volatilities $\{s_i\}_i$ (as exemplified above), and we consider both.

3.5.2 Examples of event studies

For concreteness, we present some illustrative examples of event studies that motivate the tests described later in this section. Figure 3.3 shows the graphical interface to our event study engine. The events being studied are surges in our macroeconomic keyword index. The currency pair being considered is EUR/USD and, in particular, we are studying the impact of events on exchange rate volatility (de-seasonalized squared log returns). The large plot at the top shows the averaged event window (i.e., $\hat{s}_i$ in the above notation) with the pre-event samples displayed to the left of the 0-minute mark and the post-event samples displayed to the right of the 0-minute mark. Immediately, we see a peak in the center of the plot, representing a significant increase in

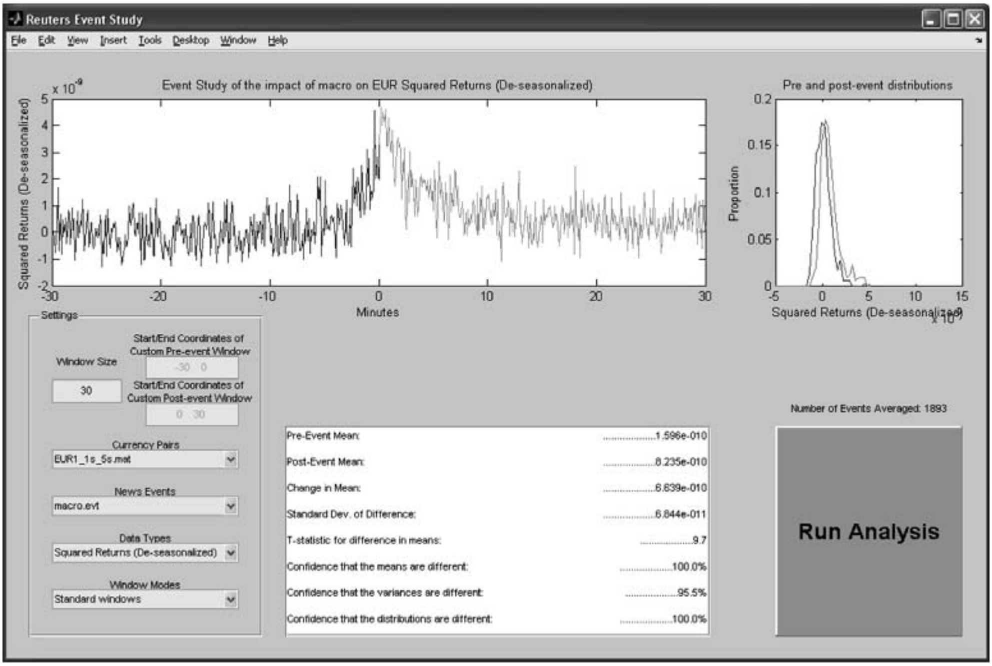

Figure 3.3. Screenshot of event analysis tool GUI, coded in MATLAB.

volatility around the time of the events, which are defined by spikes in our index. Also, average volatility in the post-event window is larger than average volatility in the pre-event window. Indeed, this can also be confirmed by inspecting the statistics reported at the bottom of the window. The second, smaller plot displays the density functions of the pre-event samples and post-event samples; thus, the fact that the pale curve (the post-event density function) is shifted to the right vis-à-vis the dark curve (the pre-event density function) means that there has been an upward shift in volatility, on the average, as a result of the events.

In Figure 3.3 the impact of these events seems clear, but for other indices the impact may be less visually apparent, and thus it is important to measure the impact using rigorous statistical techniques, which we propose in Sections 3.5.3–3.5.5. Nevertheless, it is instructive to consider two more examples.

Figure 3.4 shows an event study for our Agriculture index. This time, the currency pair is AUD/USD and return is the variable of interest. Upon visual inspection, there seems to be no significant change as a result of the events. The density plots to the right confirm this as well, as do the statistical tests described below. This is not surprising, however, since it is not clear that the presence of agriculture news would tend to drive exchange rates in one particular direction. Nonetheless, there is an impact which can again be seen by examining exchange rate volatility (this event study is shown in Figure 3.5).

Figure 3.5 shows an increase in volatility after surges in agriculture news, although it is a more modest effect than the example of macroeconomic news in Figure 3.3. Even so,

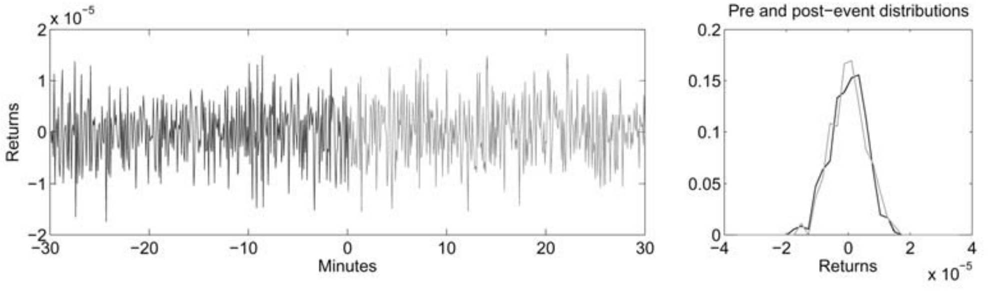

Figure 3.4. Event study of impact of agriculture news on AUD/USD returns.

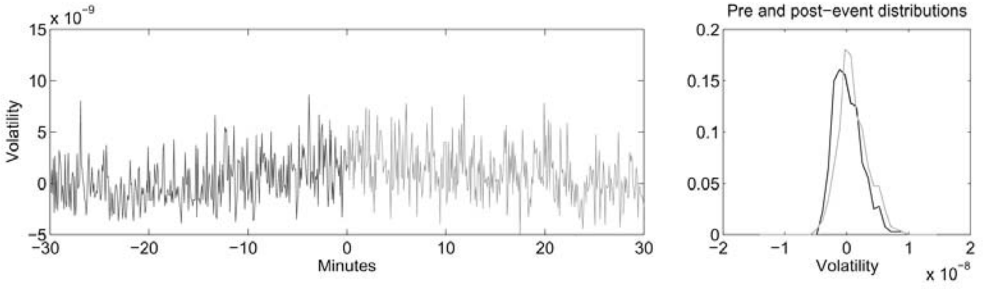

Figure 3.5. Event study of impact of agriculture news on AUD/USD volatility.

a t-test (see Section 3.5.3) establishes that this increase is indeed statistically significant, meaning that our agriculture index is correlated with volatility increases in the Australian dollar exchange rate.

For further examples, we refer the reader to the document *Thomson Reuters NewsScope Event Indices: Event Analysis Results* for the complete set of event studies. Table 3.2 summarizes the t-statistics from all event studies of 30-minute volatility for each of the currency pairs considered in this chapter.

3.5.3 Testing for a change in mean

A natural consequence of market-moving news events would be higher post-event volatility than pre-event volatility. We test for this using a t-test for equality in mean between pre-event and post-event samples $\{\hat{s}_{-w+1}, \ldots, \hat{s}_0, \hat{s}_1, \ldots, \hat{s}_w\}$.

The t-test is formed by computing the t-statistic (denoted t below) in the following manner:

$$\mu_- \equiv \frac{1}{w}\Sigma_{i\leq 0}\hat{s}_i \quad , \quad \mu_+ \equiv \frac{1}{w}\Sigma_{i>0}\hat{s}_i \quad , \quad \delta \equiv \mu_+ - \mu_- \tag{3.10a}$$

$$\sigma^2_{\mu_-} \equiv \frac{1}{w(w-1)}\Sigma_{i\leq 0}(\hat{s}_i - \mu_-)^2 \quad , \quad \sigma^2_{\mu_+} \equiv \frac{1}{w(w-1)}\Sigma_{i>0}(\hat{s}_i - \mu_+)^2 \tag{3.10b}$$

$$\sigma \equiv (\sigma^2_{\mu_-} + \sigma^2_{\mu_+})^{1/2} \tag{3.10c}$$

$$t \equiv \left|\frac{\delta}{\sigma}\right| \tag{3.10d}$$

Table 3.2. *t*-statistics for the significance of each Thomson Reuters NewsScope Event Index with respect to the volatilities of 16 currency pairs

Index	Currency pair															
	AUD USD	CAD USD	CAD JPY	CHF USD	EUR USD	EUR CAD	EUR GBP	EUR JPY	GBP USD	GBP AUD	GBP JPY	JPY USD	MXN USD	NOK USD	NZD USD	NZD EUR
ASIA	2.9	1.1	1.8	0.1	−2.5	0.9	−0.8	−0.5	1.6	2.2	2.2	−0.8	—	0.1	1.7	2.7
Agricultural topics	4.5	−0.8	1.2	2.8	8.7	3.6	2.5	0.7	4.8	0.5	2.3	2.1	−2.3	2.7	2.2	0.9
Banking keywords	−2.2	−2.2	−2.5	3.2	−2.2	−0.7	−2.0	−4.6	2.9	−0.7	−1.2	−4.8	2.2	4.2	−0.8	−1.2
Bearish keywords	1.2	1.1	2.2	1.0	3.2	2.7	1.5	−0.1	1.6	2.4	−0.5	1.2	0.3	1.1	0.2	−0.5
Bonds topics	4.9	2.5	7.0	6.6	6.3	−0.8	6.0	1.8	8.4	1.4	4.1	3.9	1.9	−2.5	3.5	2.6
Bullish keywords	4.1	−0.4	1.4	2.4	0.7	−0.1	−1.4	0.3	3.5	0.4	1.5	3.5	−1.8	−1.3	6.7	0.0
Central bank keywords	7.0	−0.1	3.9	6.5	13.3	4.7	1.2	5.9	5.6	−0.6	4.0	7.3	5.3	−0.1	7.7	2.2
Central bank topics	6.8	6.9	7.8	12.3	12.1	1.8	7.8	9.8	13.4	6.1	8.3	12.7	4.5	7.0	11.1	2.7
Corporate keywords	−1.1	0.1	0.4	1.4	−0.8	0.2	−2.9	1.6	2.4	1.1	1.6	1.0	1.4	1.4	−1.1	1.6
Credit topics	1.8	3.7	3.2	3.8	2.4	−0.7	1.6	1.1	4.9	3.4	5.3	3.4	2.5	−3.5	0.1	0.4
EUROPE	1.7	3.1	1.7	−0.5	−1.8	0.8	−0.9	−0.7	1.4	−1.4	0.9	−3.9	—	2.6	7.2	2.9
Economic topics	8.8	7.0	7.4	12.1	12.9	7.2	2.7	2.7	10.5	2.5	9.4	8.2	−3.4	11.5	6.4	4.0
Emerging market topics	1.4	1.1	1.5	1.8	1.1	2.7	−3.4	1.5	1.2	−0.3	0.2	1.8	0.9	1.2	−1.2	1.9
Emotional keywords	3.2	3.3	0.7	4.5	2.2	0.2	1.0	−0.3	3.1	0.6	2.7	0.7	−3.8	1.8	0.5	0.2
Energy topics	2.5	−3.3	−0.4	−2.5	−2.2	−4.0	0.4	1.1	−2.5	0.0	0.4	0.9	4.1	−7.3	−0.5	−2.1
Finance keywords	−2.3	−1.9	−1.7	1.0	−5.1	0.3	−4.1	2.9	−0.1	2.2	1.8	−1.0	1.9	0.3	0.5	−2.4
Foreign exchange keywords	2.2	4.1	2.5	1.6	0.9	2.0	2.4	2.1	1.0	3.6	3.9	4.0	−1.6	0.0	4.8	7.3
Foreign exchange topics	9.4	8.6	10.7	9.4	7.9	4.0	9.7	6.5	7.0	5.4	8.1	9.2	6.9	5.6	5.6	1.7
Livestock topics	11.6	8.0	11.7	15.0	17.3	7.3	4.7	7.1	15.8	7.5	8.0	12.5	0.6	14.3	10.2	1.0
Macroeconomic keywords	9.9	10.6	10.7	13.8	10.9	3.8	2.9	5.9	8.6	8.1	9.2	12.2	3.6	7.5	6.5	2.2
Macroeconomic topics	13.6	10.7	12.6	19.4	14.6	7.7	4.9	9.7	14.3	6.2	11.2	14.2	0.0	11.7	10.7	6.4

Major news topics	3.5	6.7	5.5	6.8	7.6	1.9	3.9	3.7	6.3	1.0	4.1	6.0	−1.2	5.9	5.1	−1.3
Markets topics	−1.0	−1.3	0.0	−0.9	0.7	−1.4	−0.3	1.4	−2.8	−1.8	−1.1	0.4	−1.8	−0.9	0.0	0.4
Mergers keywords	−4.4	−5.3	−1.4	−4.4	−4.8	2.3	−2.8	−0.9	−2.8	−4.4	−0.5	−2.4	−4.2	−11.6	2.5	3.8
Metal topics	−0.6	−2.3	1.1	−1.5	0.6	0.8	−1.5	−0.6	−2.0	−1.8	0.1	0.6	−1.6	−1.1	−0.6	−0.2
Military keywords	0.7	−2.3	−0.7	−0.7	3.6	0.3	−2.1	1.3	−3.6	0.5	−4.2	−4.9	2.3	3.1	−0.3	0.8
Natural disaster keywords	2.1	−0.1	−1.3	−0.1	−0.2	−0.2	−4.0	−0.4	−1.4	0.7	−0.2	0.7	−0.9	1.5	−4.6	0.0
Natural disaster topics	0.6	−1.4	1.4	−1.7	0.0	1.6	−2.8	−5.8	−4.7	0.6	−0.6	0.1	2.3	−5.6	−2.4	0.4
Oil topics	1.2	−2.1	1.4	−0.2	0.4	−2.2	2.0	−1.6	2.6	−0.8	1.3	−2.1	−4.9	−1.4	2.6	1.2
Political keywords	−0.6	0.9	−2.8	5.7	−1.7	−0.3	−0.7	−1.8	−0.8	−3.7	−3.5	−1.1	−2.3	−3.3	−2.0	−1.7
Political topics	6.0	4.1	3.9	9.9	8.7	2.2	1.6	1.8	10.2	4.3	3.2	7.2	1.7	1.7	5.0	−1.5
Precious metal topics	2.0	−0.3	1.4	1.6	1.9	0.8	0.9	0.2	1.6	−1.0	−0.4	2.2	−0.3	4.8	1.4	0.9
RCH	0.2	−3.3	0.7	−5.9	−5.6	1.7	4.1	1.1	−4.9	−0.9	1.4	−0.6	−2.5	−4.2	−0.7	3.7
Rates topics	8.2	7.0	11.6	12.6	14.5	3.5	6.5	2.8	16.0	5.5	9.9	6.8	−0.1	−2.5	9.3	2.6
Regulation topics	−2.2	1.6	1.4	−1.7	−3.5	−4.0	−2.3	−2.4	−3.3	−1.1	0.0	−3.3	1.8	2.5	2.2	0.4
Stocks topics	6.5	12.6	8.8	6.8	11.6	0.9	0.7	6.0	7.9	3.6	3.1	7.1	3.7	8.6	3.0	1.8
Terrorism keywords	−0.6	−0.3	2.5	0.8	−0.5	−0.9	3.9	1.2	−3.5	−0.4	−2.4	−1.0	3.1	−2.6	−0.3	1.7
Terrorism topics	−0.5	−0.2	2.1	−0.2	0.7	−1.7	−2.2	3.8	−0.8	0.9	−0.2	1.1	−3.0	−1.1	−5.0	−3.2
US new houses ECI	7.8	6.5	8.4	7.4	5.9	4.0	5.6	4.1	10.5	7.0	4.8	6.6	1.6	7.0	3.7	−0.7
US housing starts ECI	9.7	5.0	8.2	9.4	9.6	5.1	0.6	5.8	7.9	7.5	6.4	8.6	0.7	12.5	8.3	3.7
Urgent news topics	10.2	10.7	9.8	15.0	13.4	7.2	7.0	4.9	11.6	4.6	8.9	10.5	0.9	12.5	9.8	8.6
VIO	−1.2	−2.9	2.2	−0.9	−2.8	−0.8	−3.2	−2.6	−0.1	0.8	−0.4	−2.0	−1.8	−1.7	−5.4	0.4
Violence keywords	1.1	−2.5	−0.6	1.9	−1.9	−1.1	1.5	−2.5	1.9	1.1	−0.1	−0.2	−0.1	−1.0	2.5	0.4
All European ECI	3.9	1.8	−0.1	8.2	1.2	1.6	2.5	8.4	2.8	−0.4	3.2	5.4	−1.3	7.2	6.7	0.7
All miscellaneous ECI	13.5	8.6	10.6	13.9	16.8	11.2	9.6	3.4	14.4	8.9	9.4	11.0	−2.8	14.4	8.1	7.1
All US ECI	16.7	12.1	15.0	17.3	15.4	11.6	10.5	11.2	15.5	12.7	12.1	14.9	0.3	19.4	17.3	5.0
Random events	−1.9	−1.4	0.7	−1.2	−2.8	2.1	0.7	0.6	−2.5	−1.1	0.6	−3.1	−0.2	−3.6	−0.1	0.2
Random news events	3.0	0.7	−2.5	−3.0	1.5	0.9	1.3	−2.2	−1.4	−1.6	0.4	−2.3	2.8	−1.3	1.6	−0.8

In a classical t-test, the t-statistic is distributed according to Student's t-distribution, and thus δ is large enough to be statistically significant if:

$$t \quad > \quad \Phi_t^{-1}(1-\alpha/2, \nu) \tag{3.11}$$

where $(1-\alpha) \cdot 100\%$ is the confidence level, and ν is the number of degrees of freedom, calculated as

$$\nu \quad \equiv \quad \frac{\left(\frac{\sigma_{\mu_+}^2}{w} \frac{\sigma_{\mu_-}^2}{w}\right)^2}{\frac{\sigma_{\mu_-}^4}{w^2(w-1)} + \frac{\sigma_{\mu_+}^4}{w^2(w-1)}}. \tag{3.12}$$

When measuring variables that are not completely independent (such as pre-event returns/volatility), the t-statistic may not follow an exact t-distribution. To ensure that the confidence levels we compute are accurate, we empirically determine the distribution of the t-statistic under the null hypothesis as follows.

We construct random event studies by choosing, say, 500 random points in time and declare these as "events". We then compute the t-statistic of this event study and repeat this process 5,000 times to generate the finite sample null distribution. The resulting samples give a reliable estimate D of the distribution of the variable t for random (insignificant) events. We can then compare the t-statistics obtained from our (non-random) events and measure their significance according to the following formula:

$$\text{sig}(t) \quad = \quad \Pr_{x \leftarrow D}[x \geq t] \ . \tag{3.13}$$

This yields a more robust significance measure for our t-tests. The empirical values of t-statistics are reported in Section 3.A.3 (see appendix on p. 102).

3.5.4 Levene's Test for equality of variance

Another statistical test we apply to averaged event window samples is Levene's Test, which tests for a change in standard deviation before and after the event. This test is most naturally applied to returns since a change in standard deviation would suggest a change in volatility. Thus we begin with the averaged event window samples of log returns $\{\hat{r}_{-w+1}, ..., \hat{r}_0, \hat{r}_1, ..., \hat{r}_w\}$, and then compute the following quantities:

$$\hat{r}_{\text{median}-} \equiv \text{median}\{\hat{r}_{-w+1}, .., \hat{r}_0\} \quad , \quad \hat{r}_{\text{median}+} \quad \equiv \quad \text{median}\{\hat{r}_1, .., \hat{r}_w\} \tag{3.14a}$$

$$Z_{j-} \equiv |\hat{r}_{j-} - \hat{r}_{\text{median}-}| \quad , \quad Z_{j+} \quad \equiv \quad |\hat{r}_{j+} - \hat{r}_{\text{median}+}| \tag{3.14b}$$

$$\mu_{Z-} \equiv \frac{1}{w}\Sigma_{j\leq 0} Z_{j-} \quad , \quad \mu_{Z+} \quad \equiv \quad \frac{1}{w}\Sigma_{j>0} Z_{j+} \quad , \quad \mu_Z \quad \equiv \quad \frac{1}{2}(\mu_{Z-} + \mu_{Z+}) \tag{3.14c}$$

Finally, we compute:

$$Q \equiv w(2w-2)((\mu_{Z-} - \mu_Z)^2 \ + \ (\mu_{Z+} - \mu_Z)^2) \tag{3.15a}$$

$$R \equiv \Sigma_{j\leq 0}(Z_{j-} - \mu_{Z-})^2 \ + \ \Sigma_{j>0}(Z_{j+} - \mu_{Z+})^2 \tag{3.15b}$$

$$W \equiv Q/R \tag{3.15c}$$

Standard deviation changes with $(1-\alpha) \cdot 100\%$ confidence if $W > \Phi_F^{-1}(1-\alpha, 1, w-2)$.

3.5.5 The χ^2 test for goodness of fit

Another test for changes between pre-event returns/volatility and post-event returns/volatility is the χ^2 goodness-of-fit test. Below we describe the test for averaged event window log returns, but the same test could be applied to volatilities as well.

Recall that the χ^2 test consists of the following steps:

- Create histograms of $\{\hat{r}_{-w+1}, ..., \hat{r}_0\}$ and $\{\hat{r}_1, ..., \hat{r}_w\}$ such that they have the same bins and every bin in the first histogram has at least n counts, where $n > 0$.
- Denote the bin frequencies of the pre- and post-event histograms by $\{f_{1-}, ..., f_{k-}\}$ and $\{f_{1+}, ..., f_{k+}\}$, respectively, where $k \le w$.
- Finally, define the χ statistic by

$$\chi \equiv \Sigma_i \frac{(f_{i-} - f_{i+})^2}{f_{i-}} \tag{3.16}$$

The shape of the distribution changes with $(1-\alpha) \cdot 100\%$ confidence if $\chi > \Phi^{-1}_{\chi^2}(1-\alpha, k-1)$.

3.6 NEWS INDICES AND FX IMPLIED VOLATILITY

In Section 3.5, we showed that event indices, on average, have an impact on realized FX volatility. Since FX implied volatility indices also forecast realized volatility (see Pong et al., 2004; Taylor, 2005), this suggests that implied volatility and news indices might be related. On the other hand, there is an important difference between the two: while event indices are calibrated to predict volatility over 30-minute periods, implied volatility indices forecast volatility over much longer periods, typically about 30 days. The event study methodology was employed to determine whether a relationship between the two does, in fact, exist. No evidence to that effect was found; this suggests that implied volatility and event indices may function as complementary sources of information for risk management, each focused on a different time horizon.

3.6.1 Data pre-processing

Bank quotes for implied euro volatility were obtained from Thomson Reuters for 2005 to mid-2007. Preliminary exploration revealed that the major banks quote persistent, yet statistically different implied volatilities (it is not uncommon for different banks to quote implied volatilities that differ by 3 standard deviations or more). This means that one could easily mistake changes in quote provider for genuine changes in implied volatility. To preempt such errors, and to focus on the relationship between implied volatility and news, we select one source of quotes for our analysis, and choose the most frequent provider, Société Générale, which was responsible for 20,691 of the total 53,959 quotes in our sample. Quotes from other banks were ignored.[5] Each tick in the time-series contained both a bid volatility and an ask volatility. We used the arithmetic mean of these two values.

[5] The next-most-frequent providers were BNP Paribas at 12,475, Broker at 4,980, and RBS at 4,341 quotes.

Event indices vs. implied volatilities

As demonstrated in Section 3.5, times with the top event index values tend to forecast increased volatility.

Given that active values of the Thomson Reuters Event Indices typically predict increased realized volatility, it seems plausible that they could also predict an increase in implied volatility. The following event study seeks to disprove the null hypothesis that implied volatility remains the same, on average, before and after a foreign-exchange-related (FRX) event.

The results of this and other, similar, event studies did not provide evidence to suggest that FRX news affects the implied 1-month euro volatility. Several other news indices were studied and similarly could not be shown to impact the implied 1-month euro volatility.

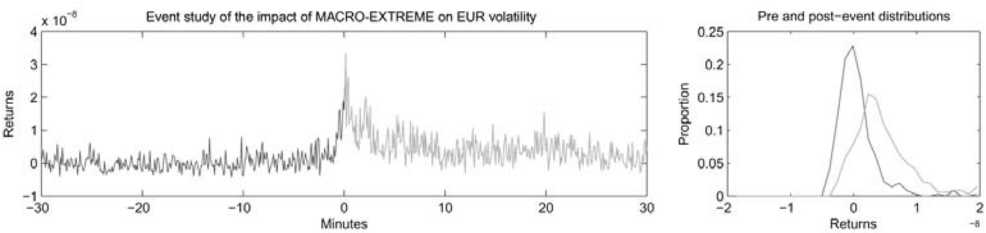

Figure 3.6. EUR realized volatility during 2003–2007 corresponded with the top-161 macro events (99.99% percentile of index).

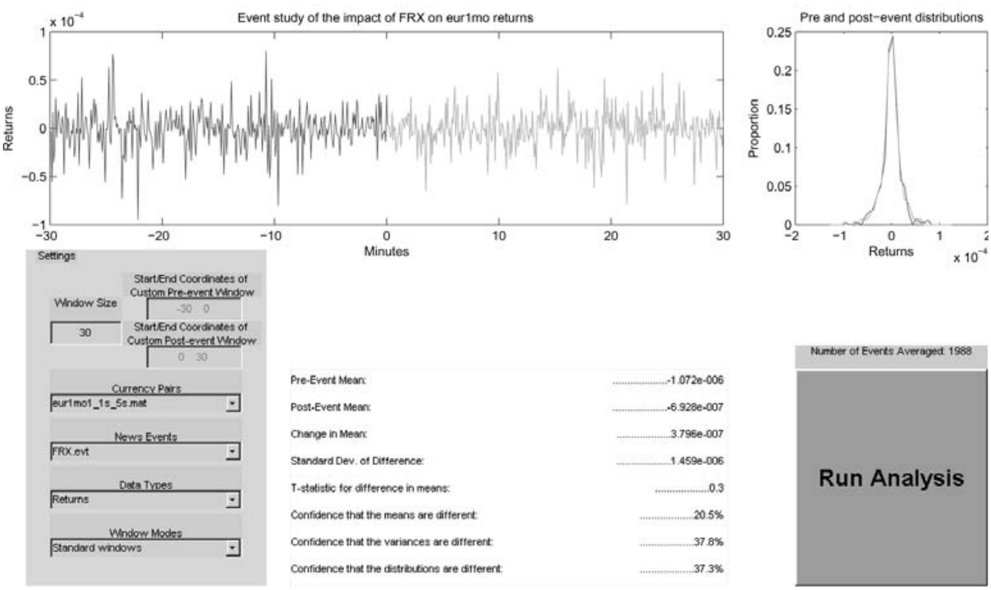

Figure 3.7. FRX events during 2005–2007 did not correspond with statistically significant changes in implied volatility.

In light of the contrasting performance of the event studies for realized volatility and implied volatility, we study the differences between these quantities that would explain this behavior. First, we note that implied volatility is imputed from options prices by inverting the Black–Scholes (or other similar) options-pricing formula. Thus, any forecast power with respect to implied volatility would imply forecast power for options *prices*, and since these options are actively traded we would expect only very modest price inefficiencies with respect to news.

Second, we observe that implied volatility is meant to reflect the expected volatility over a lengthy time horizon (e.g., 1 month), whereas the analysis of Section 3.5 concerns volatility over a much shorter period (e.g., 1 hour). It has been documented that implied volatilities tend to be much better predictors of long-term future price volatility in equity markets, and we find the same effect in foreign exchange data. Indeed, Figure 3.8 plots the correlation between future realized volatility (for horizons from 1 minute to 10 days) and the current implied volatility. For comparison, the grey curve in Figure 3.8 plots the correlation between (past) t-hour EUR realized volatility and the future t-hour realized volatility. We note that the correlation between implied volatility and realized volatility decays as the time horizon gets shorter, and once the time horizon is less than 2 hours, historic volatility outperforms implied volatility as a predictor of future realized volatility. Implied volatilities from other banks, as well as 1-week implied volatilities, showed similar behavior. These findings confirm those of Pong et al. (2004) and Taylor (2005).

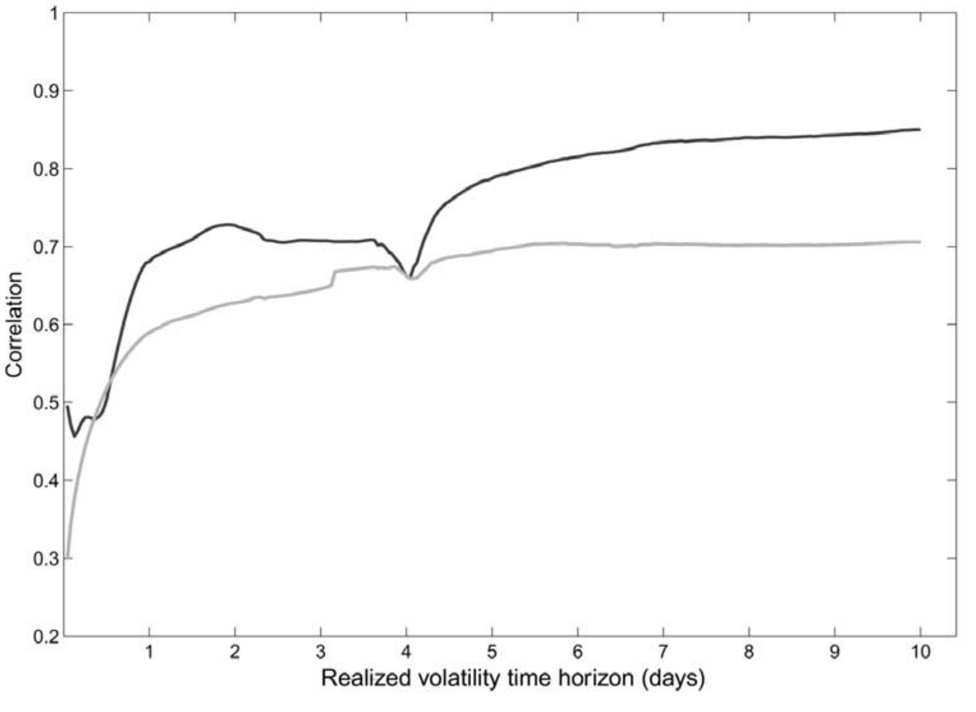

Figure 3.8. Correlation of 1-month Société-Générale-quoted EUR implied volatility with future t-hour realized volatility (black line). Correlation of (historic) t-hour EUR realized volatility with future t-hour realized volatility (grey line). All realized volatilities are de-seasonalized (as described in Section 3.5).

3.6.2 Implied volatility events

Abrupt changes in the 1-month implied volatility reflect a change in the market's beliefs about returns. It seems plausible that such changes might be correlated with contemporaneous changes in realized volatility. To investigate this hypothesis, an event study with 30-minute pre- and post-event windows was conducted.

A rolling window of the last 78 events (about 4 to 16 days' worth of events) was used to calculate the mean and standard deviation of Société Générale's 1-month implied volatility estimates. Deviations of more than 3σ from the rolling mean were taken to be "significant events". An event file of these timestamps was evaluated using the event analysis GUI.

The results of the event study did not provide evidence to support the hypothesis that realized volatility either increases or decreases at times when implied volatility significantly changes.

3.7 EVENT STUDY ANALYSIS THROUGH SEPTEMBER 2008

In the preceding sections, we have focused on event analysis with respect to foreign exchange during the period from January 2003 through July 2007. In this section, we update the results through September 2008, and also consider the impact of Thomson Reuters NewsScope Event Indices on 11 equity indices.

Tables 3.3 and 3.4 report the t-statistics for event analyses of event indices on the returns and volatility of the same 16 currency pairs as before (see Table 3.2), and Tables 3.5 and 3.6 contain t-statistics for the corresponding event analyses for 11 equity indices.

Tables 3.3 and 3.4 show that the event indices have little power to forecast movements in exchange rates, but significant power to forecast exchange rate volatility. However, Tables 3.5 and 3.6 tell a very different story for equity indices—the event indices do seem to have some predictive power for equity index returns, as well as for equity index volatility. One possible explanation for this difference is that equities are not as liquid, hence the impact of news is incorporated into currencies faster than equity indices. While this may suggest potential profit opportunities in equity indices, transactions costs are considerably higher for stock index futures than for currencies given comparable notional exposures. Therefore the magnitude of profits from real-time news-based strategies in equities is an open empirical question.

However, there is no doubt that the event indices have strong predictive power for squared equity index returns (as Table 3.6 illustrates). As in foreign exchange markets, the volatilities of equity indices are greatly affected by real-time news.

3.8 CONCLUSION

The importance of real-time news to the investment process has been well established, but until now there has been no systematic approach that integrates news with investments. The Thomson Reuters NewsScope Event Indices provide a convenient and powerful translation of qualitative information to quantitative signals using Thomson Reuters NewsScope data calibrated to foreign exchange spot data. The significance of the indicated market impact was verified using econometric event studies. Finally, an analysis comparing the volatility-forecasting capabilities of event indices and implied volatility indices suggested that they provide complementary information.

In ongoing research, we plan to construct customized combinations of the 45 base indices to suit a variety of applications (e.g., trading, risk management, and regulatory oversight). Moreover, we are developing a set of adaptive algorithms to automate the process by which new indices are created and old indices are updated to reflect changing market conditions.

Table 3.3. *t*-statistics for the significance of each Thomson Reuters NewsScope Event Index with respect to the returns of 16 currency pairs.

	Currency pair															
Index	AUD USD	CAD USD	CAD JPY	CHF USD	EUR USD	EUR CAD	EUR GBP	EUR JPY	GBP USD	GBP AUD	GBP JPY	JPY USD	MXN USD	NOK USD	NZD USD	NZD EUR
ASIA	0.7	−0.4	0.6	−0.4	0.6	−0.1	0.0	0.5	0.8	−0.9	0.5	0.2	—	−1.0	0.1	0.2
Agricultural topics	0.6	−1.5	1.0	−0.9	0.9	−0.5	−0.2	0.5	1.1	−0.1	0.4	−0.2	−0.1	−0.8	0.2	0.1
Banking keywords	0.0	−0.5	0.5	−0.3	0.1	0.0	0.0	−0.5	−0.3	−0.9	−0.3	−0.7	−1.0	−0.1	0.4	−0.1
Bearish keywords	1.4	−0.3	1.5	0.2	0.0	0.0	−0.5	1.1	0.7	0.0	2.1	1.8	−1.5	0.6	0.3	0.2
Bonds topics	0.5	0.1	0.6	0.0	−0.1	0.0	0.0	0.0	−0.1	−0.7	0.5	0.3	−0.7	0.0	−0.1	0.2
Bullish keywords	0.4	0.0	0.4	0.3	−0.6	0.0	−0.4	−0.2	−0.1	−0.7	0.4	0.2	1.2	0.4	−0.5	−0.1
Central bank keywords	0.0	0.0	0.2	0.4	−0.3	−0.1	−0.3	−0.5	0.3	0.0	−0.4	−0.3	−0.9	0.5	−0.2	−0.3
Central bank topics	0.0	0.2	0.2	0.5	−0.8	−0.2	−0.8	−0.6	0.0	0.7	−0.4	−0.1	0.2	0.5	−0.5	0.5
Corporate keywords	−0.8	0.1	−0.5	0.2	−0.6	−0.1	−0.4	−0.2	0.0	0.6	−0.2	0.1	0.1	0.3	−0.1	−0.1
Credit topics	0.0	0.0	0.8	−0.5	0.0	0.0	−0.1	0.2	−0.4	0.6	0.0	0.4	0.1	0.2	−0.3	−0.1
EUROPE	−0.3	0.5	0.1	0.2	0.0	0.1	−0.3	−0.1	0.2	−0.2	0.2	0.0	−1.3	0.1	0.2	0.0
Economic topics	−0.3	−0.3	−0.1	0.0	−0.3	−0.2	0.5	−0.6	−0.6	0.2	−0.9	−0.6	0.1	0.2	0.0	0.2
Emerging markets topics	−0.1	0.0	0.6	−0.1	0.3	−0.1	−0.1	−0.2	0.7	0.3	−0.6	−0.2	0.0	0.3	0.8	0.1
Emotional keywords	−0.1	−0.5	0.7	0.4	−0.3	0.0	0.0	0.0	−0.2	0.7	−0.1	0.1	0.5	0.3	−0.3	−0.2
Energy topics	0.1	0.6	−0.6	−0.7	0.1	0.3	0.4	−0.3	−0.1	−0.3	−1.0	−0.5	−0.5	−0.2	0.4	0.2
Finance keywords	−0.3	−0.5	−0.4	−0.1	0.1	−0.3	0.1	−0.2	−0.1	−1.3	−0.6	−0.2	−0.2	−0.2	0.1	0.2
Foreign exchange keywords	0.0	−0.3	2.1	1.0	−0.9	−0.4	−0.4	0.5	−0.2	2.0	1.3	1.4	−1.7	0.1	0.4	0.3
Foreign exchange topics	−0.2	1.0	−1.4	1.2	−1.2	0.0	−0.8	−0.3	−0.4	0.9	0.0	0.6	−1.0	0.4	0.0	0.1
GB	−0.1	0.6	−0.9	−0.7	0.8	0.8	0.8	0.1	0.0	0.0	−0.7	−0.5	−0.6	−0.3	0.0	−0.3
JP	0.1	0.4	−0.5	−0.2	0.3	−0.2	0.5	0.1	0.4	−1.2	0.0	0.3	−0.2	−0.3	−0.5	0.2

Livestock topics	0.1	−0.9	−0.1	−0.6	0.4	−0.7	0.0	0.1	0.7	0.5	−0.1	−0.2	−0.8	−0.7	0.0	−0.1
MX	−1.2	1.2	−0.1	0.5	−1.2	0.1	−0.7	−0.4	0.0	0.6	0.3	0.6	−0.5	0.7	−0.6	0.1
Macroeconomic keywords	−0.6	0.3	−0.8	−0.1	0.1	0.1	−0.1	0.1	0.5	−1.1	−0.2	−0.5	−0.6	0.1	0.0	−0.2
Macroeconomic topics	0.2	−0.6	0.3	0.3	−0.3	−0.7	0.0	0.0	−0.2	0.9	0.0	0.5	0.4	0.3	0.0	0.2
Major news topics	0.5	−0.6	1.0	0.8	−0.2	−0.5	0.1	0.0	−0.8	0.7	−0.4	0.5	−0.4	0.2	−0.2	0.2
Markets topics	0.1	−0.6	1.4	1.2	−1.0	−0.8	0.2	0.2	−1.3	−0.8	0.0	1.1	0.0	0.7	0.0	0.2
Mergers keywords	0.8	0.2	0.5	1.1	−0.7	−0.4	−0.8	0.1	0.1	0.4	0.3	0.8	0.4	0.6	0.5	0.2
Metal topics	−0.5	0.1	−1.1	0.3	−0.5	0.3	0.2	−0.4	−0.9	0.2	−1.1	−0.4	−0.1	0.7	−0.2	−0.4
Military keywords	−1.2	−0.1	−0.3	0.0	0.2	−0.2	−0.1	−0.3	0.0	−0.2	0.2	−0.3	0.1	0.3	−0.1	0.0
Natural disaster keywords	−0.6	0.6	−0.1	0.1	−0.2	0.3	0.4	−0.1	−0.1	0.6	−0.3	0.5	1.1	0.9	0.1	−0.1
Natural disaster topics	0.1	−0.2	−0.2	0.2	−0.2	−0.1	−0.1	0.0	−0.3	−0.4	−0.1	0.1	−0.1	0.7	0.1	−0.2
Oil topics	0.1	−0.1	0.5	0.0	0.4	−0.1	0.4	0.3	0.1	0.5	0.1	0.1	−0.4	−0.3	0.1	−0.1
Political keywords	−1.4	0.0	−0.7	−0.1	−0.2	−0.1	0.2	−0.2	−1.3	0.6	−0.5	0.3	0.5	0.5	−1.0	−0.1
Political topics	−0.5	0.0	0.3	0.1	−0.1	−0.1	0.0	−0.1	0.1	0.4	−0.2	0.0	−1.1	0.7	0.1	0.0
Precious metal topics	0.9	−0.5	−0.1	−0.1	0.7	0.1	−0.1	0.3	0.8	−0.7	0.3	0.1	−1.0	−0.2	0.6	−0.4
RCH	0.3	−0.3	−0.8	−1.0	1.0	0.3	0.2	0.3	0.9	−0.1	0.2	−0.7	0.2	−0.3	0.0	0.1
Rates topics	−0.4	0.0	1.8	0.1	−0.6	−0.1	0.0	0.7	−0.5	0.6	0.2	0.7	0.2	0.3	0.5	0.2
Regulation topics	0.5	0.0	−0.6	−0.4	0.5	0.3	0.3	−0.2	0.7	−0.9	0.2	−0.4	−0.8	−0.4	−0.3	−0.3
Stocks topics	0.3	−1.0	1.1	−0.1	0.5	−0.2	0.6	0.2	0.0	0.3	−0.2	−0.2	0.1	−0.4	1.2	−0.1
Terrorism keywords	0.7	−1.3	0.6	−0.2	0.5	−0.5	−0.1	0.8	0.8	0.9	0.7	0.1	0.1	0.1	0.3	0.4
Terrorism topics	−0.2	−0.4	0.6	−0.4	0.3	0.1	−0.1	0.2	0.6	0.6	0.1	−0.2	−0.2	0.1	0.1	0.2
US	0.4	0.1	0.3	0.3	0.1	0.1	0.0	0.5	0.3	−0.3	0.7	0.6	0.2	0.1	0.1	0.4
Urgent news topics	−0.6	−1.1	1.4	0.2	−0.2	−1.0	0.0	−0.2	−0.3	1.2	0.1	0.1	−0.1	0.3	0.3	0.1
VIO	−0.4	0.2	−0.3	−0.2	−0.2	−0.1	−0.4	−0.3	0.3	0.6	−0.4	−0.3	0.0	0.4	0.3	0.1
Violence keywords	−0.8	0.3	−0.2	0.5	−0.6	0.1	0.0	−0.3	−0.6	0.3	−0.3	0.0	0.7	0.4	0.0	0.0

Table 3.4. *t*-statistics for the significance of each Thomson Reuters NewsScope Event Index with respect to the volatilities of 16 currency pairs

Index	Currency pair															
	AUD USD	CAD USD	CAD JPY	CHF USD	EUR USD	EUR CAD	EUR GBP	EUR JPY	GBP USD	GBP AUD	GBP JPY	JPY USD	MXN USD	NOK USD	NZD USD	NZD EUR
ASIA	2.4	0.8	0.4	0.8	−2.3	−0.3	−0.4	−2.3	−0.2	1.0	0.7	0.6	—	−0.3	1.3	1.2
Agricultural topics	3.6	−0.3	1.1	1.8	6.2	2.7	1.2	0.1	5.7	0.7	2.0	−0.9	−0.7	5.6	3.4	1.1
Banking keywords	−3.4	−1.1	−1.9	2.1	−1.3	−0.1	−2.3	−6.7	−0.6	−1.6	−3.6	−3.3	1.4	1.8	−1.3	0.7
Bearish keywords	3.0	2.9	5.1	2.2	4.8	4.1	0.6	4.1	3.1	3.2	3.8	4.6	−1.4	1.6	2.0	−0.6
Bonds topics	5.4	3.9	7.9	8.6	5.8	1.1	5.4	0.0	8.0	2.6	5.1	7.3	1.0	−3.2	3.0	2.2
Bullish keywords	2.7	−3.7	−2.3	3.7	−1.5	−2.3	−2.1	−1.6	2.6	0.1	0.0	−1.3	−3.9	−1.0	1.8	0.0
Central bank keywords	7.8	3.7	4.1	8.0	14.2	3.1	2.4	7.4	7.8	1.7	5.0	8.3	4.9	2.8	9.4	3.9
Central bank topics	11.5	8.6	12.4	13.5	13.8	0.1	8.9	11.1	14.0	8.6	11.4	14.5	4.6	8.9	15.2	5.4
Corporate keywords	−1.0	−0.5	0.7	2.2	−1.1	−0.8	−2.2	−1.2	1.1	0.9	1.4	−1.8	0.2	2.6	−3.7	2.0
Credit topics	4.1	3.5	5.0	4.2	4.3	−0.4	2.5	1.7	5.5	4.4	4.6	4.9	1.4	−2.5	−0.7	0.4
EUROPE	2.1	5.4	2.9	−1.7	0.8	−1.3	−1.0	−1.6	0.0	−0.4	1.1	−2.0	0.2	4.0	4.9	3.1
Economic topics	10.6	8.1	8.0	13.3	12.7	7.2	3.7	3.0	10.7	5.0	9.7	10.1	−1.9	12.4	6.8	6.0
Emerging markets topics	−0.3	2.9	1.5	1.3	2.0	2.9	−3.4	2.5	−0.9	−0.6	0.0	1.9	−0.6	3.1	−1.5	1.5
Emotional keywords	2.9	2.7	1.6	2.8	2.7	−0.2	0.2	1.8	2.9	0.5	2.8	1.9	−1.8	1.2	1.3	−0.7
Energy topics	−0.2	−5.3	−2.8	−2.8	−0.8	−4.1	0.0	−0.3	−3.3	−0.2	0.5	−0.2	2.1	−9.5	−2.6	−2.1
Finance keywords	−0.4	0.4	−1.0	1.8	−5.7	1.5	−2.2	2.2	−1.2	3.0	3.5	−1.2	2.1	−1.6	1.3	−0.8
Foreign exchange keywords	4.7	9.1	6.1	3.5	4.2	1.8	1.2	6.2	3.9	6.7	7.6	5.9	−0.6	1.4	6.3	7.0
Foreign exchange topics	10.9	11.5	11.1	9.8	8.1	4.9	11.4	7.1	8.5	7.6	9.9	9.8	8.1	5.5	8.4	4.9
GB	1.5	−2.3	−5.7	−2.0	4.6	−5.9	1.0	−4.6	3.9	−0.2	−0.6	1.4	1.9	7.4	3.0	0.4
JP	−0.7	0.3	2.1	2.4	1.5	3.7	2.9	2.3	1.3	2.3	2.4	−0.1	−0.9	−1.1	1.0	2.0
Livestock topics	12.1	9.7	11.2	14.2	16.1	6.8	5.5	6.4	16.1	7.4	7.4	11.0	1.4	15.3	9.9	2.5

MX	3.7	−1.0	3.9	3.1	5.0	−0.2	0.3	0.3	1.5	0.6	3.0	6.5	3.5	−0.8	2.9	2.4
Macroeconomic keywords	11.1	12.8	14.5	16.1	12.5	6.2	3.4	7.7	10.8	11.6	11.8	15.9	3.4	8.4	9.0	6.6
Macroeconomic topics	12.3	12.1	14.5	19.3	15.9	5.8	5.7	11.2	13.7	8.6	10.5	15.8	0.1	10.8	11.5	6.3
Major news topics	5.4	6.2	6.6	7.7	10.1	0.5	5.0	4.5	6.8	2.4	9.4	8.8	−1.1	7.7	6.0	−0.9
Markets topics	−0.3	0.1	−2.0	−1.7	0.6	0.5	−0.6	0.5	−0.2	−2.1	−1.4	1.6	−2.3	−1.1	1.3	2.4
Mergers keywords	−3.3	−4.9	−4.1	−6.3	−3.7	3.2	−3.9	1.1	−4.0	−3.3	0.0	−4.0	−4.8	−11.4	3.2	2.8
Metal topics	−0.7	−2.0	−0.9	−2.6	−0.4	−0.5	−0.7	−0.3	−1.5	−1.1	1.5	2.5	−1.5	0.7	−1.7	−0.7
Military keywords	−0.7	−2.3	−3.5	−0.5	1.5	−1.5	−1.7	0.3	−3.8	−0.8	−5.4	−3.5	4.3	1.8	0.7	1.9
Natural disaster keywords	0.6	0.7	−1.0	0.2	0.3	−0.1	−3.8	0.9	−0.6	−0.1	−0.8	2.0	−2.1	1.1	−4.2	−0.7
Natural disaster topics	−2.0	−1.7	−3.3	−2.9	−1.1	0.3	−3.1	−5.8	−4.7	−1.2	−0.7	−0.1	1.0	−6.1	−4.2	0.1
Oil topics	1.2	−0.9	−0.9	−2.1	0.1	−1.3	3.6	−1.0	1.8	−0.4	−0.4	−3.5	−4.1	−2.1	1.9	1.8
Political keywords	−1.1	−0.4	−3.0	4.3	−0.9	0.0	−0.1	−5.4	−0.8	−4.4	−4.7	−1.8	−1.6	−3.5	−2.3	−0.7
Political topics	7.4	4.7	2.0	10.6	11.1	−1.1	2.1	0.5	10.7	3.9	1.9	11.0	2.1	3.5	2.9	1.0
Precious metal topics	1.3	−2.4	0.4	1.4	4.5	2.1	2.5	2.5	2.8	−0.6	2.2	6.7	−0.8	6.0	3.0	2.5
RCH	−2.9	−4.0	−2.1	−3.8	−5.0	0.9	3.3	0.5	−2.9	−1.8	0.2	−2.0	−1.6	−4.1	−2.5	2.5
Rates topics	10.2	10.2	14.4	15.7	17.2	4.7	8.1	7.4	16.9	7.5	13.2	10.3	−0.5	−2.2	11.6	4.3
Regulation topics	−3.4	0.4	2.1	−1.5	−4.3	−2.7	−0.6	0.9	−4.6	−0.7	1.3	−1.1	3.0	−0.3	3.0	0.3
Stocks topics	8.3	9.4	2.4	7.1	9.9	1.4	−0.7	5.8	7.5	2.5	−0.4	6.0	3.5	8.1	2.5	1.9
Terrorism keywords	0.0	−0.4	2.3	−0.5	−1.7	−0.1	3.1	1.6	−2.6	−0.5	−1.0	−1.1	2.7	−3.4	−0.5	2.3
Terrorism topics	−2.0	−1.5	−1.2	−0.4	−0.6	−2.0	−0.2	−0.5	−0.7	−0.9	−0.4	0.0	−3.1	−0.3	−6.1	−2.3
US	0.9	2.4	2.7	7.5	4.8	4.1	−3.2	−0.4	3.9	1.6	2.0	5.8	8.4	8.7	3.9	1.2
Urgent news topics	10.1	11.2	9.6	16.3	14.6	6.6	7.9	7.6	10.3	5.9	7.6	12.0	1.9	13.1	11.2	9.2
VIO	−0.8	−2.3	0.1	−1.8	−2.1	0.1	−3.6	−3.5	−1.1	0.3	−2.9	−1.8	−2.6	−2.1	−3.4	−0.4
Violence keywords	0.6	−1.7	−0.5	2.7	−1.5	−0.6	2.2	−3.2	1.8	1.8	0.9	−3.1	1.2	−0.2	3.3	1.0

Table 3.5. *t*-statistics for the significance of each Thomson Reuters NewsScope Event Index with respect to the returns of 11 equity indices

	DJI	FCHI	FTEU3	FTMC	FTSE	GDAXI	HSI	IXIC	N225	SPX	TOPX
ASIA	−1.3	−2.2	0.1	0.3	0.2	−0.5	−0.3	−0.9	−0.6	−1.9	0.2
Agricultural topics	1.6	1.2	0.8	0.7	0.7	0.9	0.5	2.2	−1.2	1.9	0.1
Banking keywords	−1.0	0.0	−0.3	−1.6	−0.4	0.2	0.3	−0.6	0.6	−0.9	−0.1
Bearish keywords	7.0	4.9	6.6	5.0	—	4.0	1.4	7.7	2.0	−0.7	0.2
Bonds topics	−1.3	2.7	3.4	0.7	2.9	3.2	1.3	0.5	−1.8	−1.1	0.2
Bullish keywords	−4.8	−3.6	−6.6	−3.1	−5.2	−5.2	−0.3	−3.0	−0.5	−4.4	−0.1
Central bank keywords	1.6	1.1	1.7	4.4	—	2.4	1.0	1.3	−1.1	1.2	−0.3
Central bank topics	−1.9	−0.7	0.8	0.6	1.1	0.7	−0.7	−2.5	1.1	−2.1	0.5
Corporate keywords	−0.6	−0.9	0.6	0.3	0.5	0.1	0.7	−0.8	0.0	−1.1	−0.1
Credit topics	−0.9	−0.7	−0.7	−1.7	−0.3	−0.3	−0.1	−1.3	0.8	−0.2	−0.1
EUROPE	—	−0.1	0.2	1.0	0.2	−0.7	0.3	—	—	—	0.0
Economic topics	−3.5	−0.5	0.2	1.0	0.5	0.1	0.6	−2.2	−0.8	−2.3	0.2
Emerging markets topics	1.4	0.4	0.0	−2.0	−0.7	0.8	1.1	1.4	−1.5	2.0	0.1
Emotional keywords	−1.1	−0.7	−1.0	1.2	−0.7	0.2	0.5	−0.9	−0.7	−0.9	−0.2
Energy topics	0.7	0.4	1.3	1.4	—	2.0	−0.7	0.9	1.6	0.2	0.2
Finance keywords	−2.4	−0.6	−0.5	0.6	−0.3	−1.3	−0.1	−1.3	1.7	−2.2	0.2
Foreign exchange keywords	2.5	0.0	0.6	−0.6	0.0	0.8	−0.3	2.4	−0.2	2.5	0.3
Foreign exchange topics	0.0	−2.7	−1.8	−1.2	−1.6	−1.8	−0.9	0.3	−1.0	−1.3	0.1
GB	−2.0	0.1	0.3	−2.4	−0.4	0.4	−0.2	−2.7	1.1	−1.5	−0.3
JP	0.5	0.2	−0.9	−2.3	−0.7	0.4	−0.5	0.5	−1.2	0.8	0.2
Livestock topics	−0.5	−0.3	−0.8	1.8	−0.5	−1.1	0.4	−1.1	6.8	−0.8	−0.1
MX	0.5	0.1	0.0	−0.3	0.7	−0.7	0.3	−0.8	0.0	−0.1	0.1
Macroeconomic keywords	−0.1	0.6	0.3	0.5	0.4	0.4	0.6	−1.0	−0.1	0.3	−0.2
Macroeconomic topics	−2.9	0.3	0.1	—	—	0.2	0.5	−3.6	1.2	−2.6	0.2
Major news topics	−0.7	−1.6	−1.0	0.4	−0.6	−0.3	0.3	−1.9	−1.1	−0.8	0.2
Markets topics	1.4	0.4	0.5	2.4	1.4	−0.6	−1.8	0.3	1.4	0.7	0.2
Mergers keywords	−2.2	−0.7	−0.2	—	—	0.7	0.1	0.0	0.5	−1.7	0.2
Metal topics	−1.0	0.0	0.1	0.3	—	−0.1	−1.6	−0.4	−1.6	−1.3	−0.4
Military keywords	0.1	−0.7	0.0	—	—	−0.7	−0.1	−0.7	−0.9	0.2	0.0
Natural disaster keywords	−0.3	2.2	0.7	−1.7	0.7	2.0	0.3	−0.8	−1.2	−0.2	−0.1
Natural disaster topics	−1.3	1.3	1.9	0.1	1.6	2.0	0.1	−1.3	1.5	−1.0	−0.2
Oil topics	1.1	1.3	0.9	0.7	1.5	0.3	−0.1	1.3	0.9	0.7	−0.1
Political keywords	0.3	−1.0	−1.8	0.6	−1.2	−1.5	−0.7	0.3	1.1	0.8	−0.1
Political topics	0.2	−0.2	−0.7	−1.0	−0.9	0.3	0.0	−0.1	0.9	−0.1	0.0
Precious metal topics	1.4	0.5	0.6	1.8	1.2	.2	−1.4	0.8	−1.4	1.4	−0.4
RCH	−2.4	0.8	2.8	—	—	1.3	−0.1	1.2	0.8	−0.4	0.1
Rates topics	−2.9	1.6	2.1	1.8	2.5	2.5	0.2	−1.3	−1.5	−0.3	0.2
Regulation topics	0.2	0.0	0.1	−2.6	−0.3	0.2	−1.2	−1.0	−0.1	−0.9	−0.3
Stocks topics	0.8	−1.1	−4.0	−3.0	−3.5	−2.5	0.7	−0.1	0.7	1.2	−0.1
Terrorism keywords	0.1	0.5	1.3	2.0	1.2	1.0	0.7	0.6	1.9	−0.1	0.4
Terrorism topics	0.1	1.3	2.6	−0.1	2.4	1.6	0.1	−0.5	−0.2	−0.1	0.2
US	1.8	−0.4	−0.1	—	−0.4	1.1	0.4	0.8	−2.0	−1.3	0.4
Urgent news topics	−0.8	0.6	1.5	0.1	1.9	0.5	−0.7	0.0	−0.9	−0.4	0.1
VIO	−0.7	−0.2	0.4	−2.7	−0.1	0.3	0.1	−1.2	0.4	−0.8	0.1
Violence keywords	−1.9	−0.4	−0.9	−2.0	−0.3	−0.9	1.3	−1.4	−0.4	−1.2	0.0

Table 3.6. *t*-statistics for the significance of each Thomson Reuters NewsScope Event Index with respect to the volatilities of 11 equity indices.

	DJI	FCHI	FTEU3	FTMC	FTSE	GDAXI	HSI	IXIC	N225	SPX	TOPX
ASIA	−7.2	0.8	2.3	0.8	−1.2	−0.5	0.2	−3.3	2.3	−3.6	1.2
Agricultural topics	8.0	−3.7	−4.4	−2.2	−4.6	−6.4	−0.8	2.8	−1.4	5.3	1.1
Banking keywords	−3.7	1.4	−0.5	−0.4	0.1	1.9	−2.1	−1.0	−0.5	−1.5	0.7
Bearish keywords	4.5	3.3	2.4	1.2	3.1	0.2	7.5	0.7	−0.7	1.3	−0.6
Bonds topics	3.8	2.7	2.1	0.3	−1.3	6.4	1.4	9.2	−0.5	5.8	2.2
Bullish keywords	−1.4	−2.0	−3.2	−1.9	−1.4	−0.6	0.1	1.0	0.1	0.8	0.0
Central bank keywords	8.3	−0.1	2.0	0.5	1.7	0.1	−2.6	11.9	1.0	10.5	3.9
Central bank topics	16.0	2.9	6.5	2.2	4.8	8.4	−0.6	12.7	0.8	15.1	5.4
Corporate keywords	0.9	−0.2	0.3	2.0	−0.5	0.7	0.9	2.3	0.5	3.7	2.0
Credit topics	1.7	1.1	2.6	0.4	0.9	0.9	−2.0	2.4	0.9	1.9	0.4
EUROPE	—	0.1	1.2	−0.2	0.5	1.7	−3.1	—	—	—	3.1
Economic topics	4.1	4.3	3.8	−2.4	1.1	4.0	1.1	5.5	0.6	3.8	6.0
Emerging markets topics	−1.1	1.6	2.3	0.1	−0.6	0.6	−3.2	0.6	1.2	−0.8	1.5
Emotional keywords	−3.5	3.4	4.5	−0.1	2.6	6.2	−0.6	−1.5	0.8	−0.1	−0.7
Energy topics	3.6	−1.4	−5.4	0.7	4.5	−5.8	0.2	0.0	0.4	3.3	−2.1
Finance keywords	−0.8	1.6	1.8	0.4	1.5	0.3	2.6	−1.3	0.9	−1.2	−0.8
Foreign exchange keywords	2.8	−1.2	−0.9	2.9	−0.1	−1.3	0.3	5.9	−0.3	3.9	7.0
Foreign exchange topics	9.3	0.1	0.7	−0.8	−0.1	1.2	1.4	7.3	0.1	6.1	4.9
GB	−4.8	−1.0	0.3	−1.4	0.9	−0.1	0.3	−3.4	−0.7	−3.6	0.4
JP	−0.5	−0.9	−1.0	−1.8	−2.3	−2.5	−0.3	−0.8	0.4	0.3	2.0
Livestock topics	8.7	1.8	2.0	0.0	−0.7	1.0	0.6	6.8	−0.5	6.6	2.5
MX	6.2	3.6	4.3	0.8	3.1	5.3	1.2	3.3	−0.8	4.2	2.4
Macroeconomic keywords	15.6	1.6	2.3	2.0	3.8	4.9	2.7	18.2	−1.6	17.4	6.6
Macroeconomic topics	9.5	8.5	6.8	1.0	3.2	12.3	−3.2	11.3	−0.6	10.4	6.3
Major news topics	9.1	3.0	4.9	1.5	3.6	9.5	−2.2	8.9	−1.0	8.3	−0.9
Markets topics	0.6	−0.3	−1.4	−1.2	−1.4	−0.6	−0.4	0.5	1.3	−0.5	2.4
Mergers keywords	−9.2	0.9	2.4	−2.8	1.3	3.7	−3.9	−3.4	0.6	−5.1	2.8
Metal topics	2.8	−1.0	−0.3	−1.1	−1.2	−3.1	3.9	3.0	−1.1	1.8	−0.7
Military keywords	2.9	0.6	1.5	0.9	3.3	2.0	−1.4	−1.8	0.3	−1.0	1.9
Natural disaster keywords	−5.6	0.1	0.1	−0.4	−1.6	0.5	1.3	−0.5	0.9	−2.0	−0.7
Natural disaster topics	2.4	−0.3	−2.8	−0.5	−2.3	−1.8	1.8	1.1	1.1	1.5	0.1
Oil topics	−2.2	0.3	−1.0	−0.8	0.7	0.5	−0.3	0.0	1.0	−0.7	1.8
Political keywords	0.7	0.4	0.7	1.6	0.2	−1.0	−1.3	0.0	−0.2	−1.6	−0.7
Political topics	10.8	1.9	−1.2	−1.4	0.3	−0.1	−0.1	7.4	1.1	9.5	1.0
Precious metal topics	3.8	1.4	0.3	0.7	0.8	−0.8	−0.4	1.7	−1.5	1.6	2.5
RCH	1.7	−3.4	−6.6	−4.0	−6.8	−8.6	2.4	−1.0	0.7	−0.5	2.5
Rates topics	11.6	10.5	11.4	1.9	8.4	13.1	0.1	15.1	0.9	12.1	4.3
Regulation topics	4.1	0.1	−2.1	−1.6	−3.5	−0.5	−1.3	0.9	0.4	2.4	0.3
Stocks topics	5.9	2.2	−2.0	−1.2	−0.4	2.8	1.0	8.0	1.0	6.1	1.9
Terrorism keywords	6.0	0.2	−1.4	−2.4	−0.9	0.2	1.4	1.9	−0.1	3.5	2.3
Terrorism topics	5.1	−0.3	0.6	0.4	−0.3	3.2	−1.9	1.5	−1.6	1.9	−2.3
US	6.8	−1.2	−4.5	−1.4	−4.2	−3.6	4.9	−2.7	−1.2	−2.0	1.2
Urgent news topics	11.0	5.8	5.2	−3.1	1.6	5.8	−0.8	11.2	−0.1	12.2	9.2
VIO	2.0	−2.3	−2.5	−0.3	−3.5	−1.0	−4.0	0.4	−1.1	1.6	−0.4
Violence keywords	6.9	0.0	1.4	2.5	0.7	2.4	−4.2	5.2	0.1	5.3	1.0

3.A APPENDIX

In this appendix, we present more detailed results regarding the construction of the Thomson Reuters NewsScope Event Indices. In Sections 3.A.1 and 3.A.2, we present some basic empirical properties of foreign exchange quote data and the Thomson Reuters NewsScope Archive, respectively. Section 3.A.3 contains Monte Carlo simulations of the empirical distribution of the *t*-statistic for the event studies of Section 3.5, under the null hypothesis of randomly chosen event times.

3.A.1 Properties of foreign exchange quote data

Our 4-year extract of foreign exchange spot data from the *Thomson Reuters DataScope Tick History* consists of interbank quotes for 45 currency pairs from January 1, 2003 to March 31, 2007. For each quote, the following fields are available: `RIC` (Reuters Identification Code, which specifies the currency pair), `Date`, `Time`, `GMT Offset`, `Type`, `Ex/Cntrb.ID`, `Bid Price`, `Bid Size`, `Ask Price`, and `Ask Size`. There are, in fact, many more fields than these, but we focus only on these pricing fields in our current analysis. A description of the contents of each field is given by the *Reuters DataScope Tick History* document.

For 17 major currency pairs the spot prices were extracted, and we retained only `Date`, `Time Stamp`, `Ask Price`, `Ask Volume`, `Bid Price`, `Bid Volume`, and `Source` (bank). Note that there may be a few missing values in these data, but each line does have, at the very least, an `Ask Price`.

Note that the time stamps for quotes are typically specified in Greenwich Mean Time (GMT), but Thomson Reuters provides the contributor locale of each quote, hence we can convert all GMT (or UTC) times to local times, allowing us to account for daylight savings time in regions that follow this practice.

3.A.1.1 Pre-processing of spot data

Once the data are extracted, we convert them to homogeneous time-series by sampling them at regular intervals. The first entry is the price at 12 : 00 am on January 1, 2003, and each subsequent data point is recorded *n* seconds later (*n* is usually 5), always using the most recent price. This series starts with `NaN`'s[6] until the second after the first price is announced. A quote in this series is considered outdated if it is more than 30 seconds old, at which point `NaN`s are used. The price, p, used in the time-series was defined from the logarithmic middle of the `Bid Price` (p_B) and `Ask Price` (p_A):

$$p \equiv \exp\left(\frac{\log(p_B p_A)}{2}\right). \tag{3.A.1}$$

This is simply the geometric mean of bid and ask quotes, the rationale being that an estimate of the price should be the same whether we look at the quoted rate or the inverse of the quoted rate. Some care needs to be taken to deal properly with this number (see below).

To make sure that this sampling procedure is not discarding too much information,

[6] `NaN` stands for "Not a Number", a quantity that represents an undefined number, in this case a missing data point.

Table 3.A.1. Frequency of quotes within 1-second intervals for six major currencies, and the fraction of quotes discarded using a 1-second sampling interval

Currency	1 Quote	2 Quotes	3+ Quotes	Ignored (%)
AUD	668,882	10,072	88	1.5
CAD	1,311,732	50,619	1,052	3.7
CHF	3,138,848	318,996	18,098	9.3
EUR	4,910,014	1,359,002	255,977	22.6
GBP	3,411,112	496,290	45,355	13.1
JPY	3,847,627	651,332	74,850	15.2

we tabulate the number of quotes that came in within a 1-second interval for each currency. Table 3.A.1 summarizes this with the results from all 52 months.

To reduce memory requirements for handling these time-series (currently in excess of 2 GB in MATLAB), we do not provide explicit timestamps for these data. Each timestamp must be reconstructed by taking its index in the vector and counting forward the appropriate number of seconds.

Given prices (3.A.1), we can construct continuously compounded returns across various intervals by computing log differences:

$$r_t(k) \equiv \log p_t - \log p_{t-k} \tag{3.A.2}$$

More manageable data files can then by constructed by computing non-overlapping returns every 5, 10, 30, 60, 300, 600, 1,200, and 1,800 entries (seconds). As before, indexing in these data is done implicitly from midnight January 1, 2003.

3.A.1.2 Efficiency of quoted prices

Since our analysis uses quotes exclusively, and not transactions prices, an open question is whether such quotes are a reasonable proxy for prices. One method for checking the quality of quotes is to see whether any arbitrage opportunities exist among the quotes of various currency pairs. In particular, we can check whether converting cash through a chain of currency pairs that begins and ends with the same currency yields a profit. In its simplest form—called "triangular" arbitrage—currency A is transformed into currency B which is then exchanged for currency C and then transformed back into currency A, with the hope of ending with more money than we started with. For example, starting with the US dollar, buy the Canadian dollar, then the Euro, then back to the US dollar. Typically we would expect to end with less money than we started with, if for no other reason than the existence of the bid/offer spread.

This intuition is confirmed by the quote data, and after an exhaustive analysis of all possible currency chains of lengths less than seven, starting and ending with the USD and buying AUD, CAD, EUR, GBP, JPY, and NZD in between, using 5-second intervals and bid and offer prices as appropriate (quotes were considered stale after 5 seconds). We find that the longer the chain, the greater the variation in overall return and the lower the mean return. The best chances for arbitrage seem to lie in the following chain:

$$\text{USD} \rightarrow \text{JPY} \rightarrow \text{EUR} \rightarrow \text{USD} \tag{3.A.3}$$

where 1% of the time a profit was found. All chains of length less than seven were analyzed, and several histograms of the value of $1 at the end of specific chains are presented in Figures 3.A.1 and 3.A.2.

3.A.2 Properties of Thomson Reuters NewsScope Data

In this section we provide a summary of the Thomson Reuters NewsScope Archive data and its empirical properties. For the same reason that we expect an increase in price volatility during trading hours, we also expect the news to surge daily and weekly. To measure this seasonality, we perform the same statistical analysis on the news that we performed in Section 3.A.1 for the exchange rate data. Table 3.A.2 contains summary statistics for each type of newsline, using data from January 2003 to April 2007, tallied by day of the week, weekday/weekend, and week. For example, there are an average of 1,565 alerts on Monday, while there are on average only 145 alerts at the weekend (all times are GMT). Note that the volume of news articles is roughly normal, with a small positive skewness and moderate kurtosis. Table 3.A.2 contains finer statistics about the volume per 10-minute interval. Note the large values for the kurtosis and skewness coefficients, indicating that the news volume per 10-minute intervals is much less Gaussian.

Table 3.A.2 shows a clear surge of the news on weekdays, and a lull over the weekends. It is also interesting to note that the bodylines per 10-minute intervals do exhibit large autocorrelations, indicating that news volume is relatively persistent.

3.A.3 Monte Carlo null distributions of the *t*-statistic

In order to assess the significance of the t-statistics computed in our event studies, we have determined the empirical distribution of t-statistics by sampling random event studies. By *random event study*, we mean an event study where each of the events was chosen to be a completely random point in time during the last 4 years. The 90%, 95%, 99%, 99.5%, and 99.9% confidence levels are reported in Table 3.A.3 for 15 currency pairs using random 500-, 1,000- and 2,500-event event studies on both returns and de-seasonalized squared log returns (volatility). We note that the t-statistics of returns seem to have smaller variance than a true t-distribution and the t-statistics of squared log returns seem to have larger variance than a true t-distribution. This suggests that the underlying data (the returns) are not truly independently and identically distributed. Indeed, if they were then the empirical t-statistic would be t-distributed. Thus, the t-statistics obtained in the event studies should be compared with the values in the tables to obtain more robust confidence estimates than applying the inverse cumulative distribution function of the t-distribution.

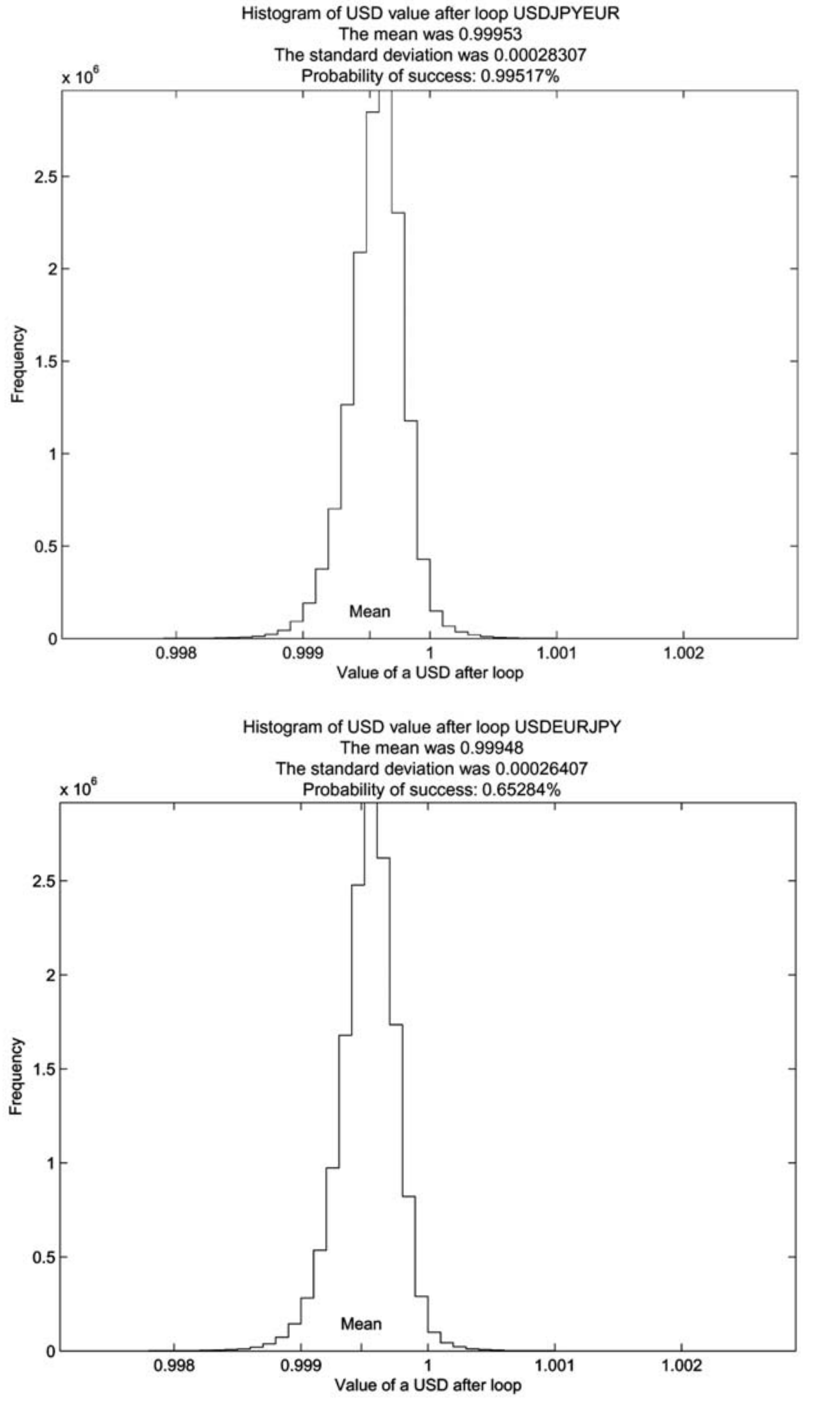

Figure 3.A.1. Histograms of the value of $1 after chains of currency conversions using Thomson Reuters quotes. The histograms in (a) and (b) involve the same currencies but in different directions, giving rise to different distributions.

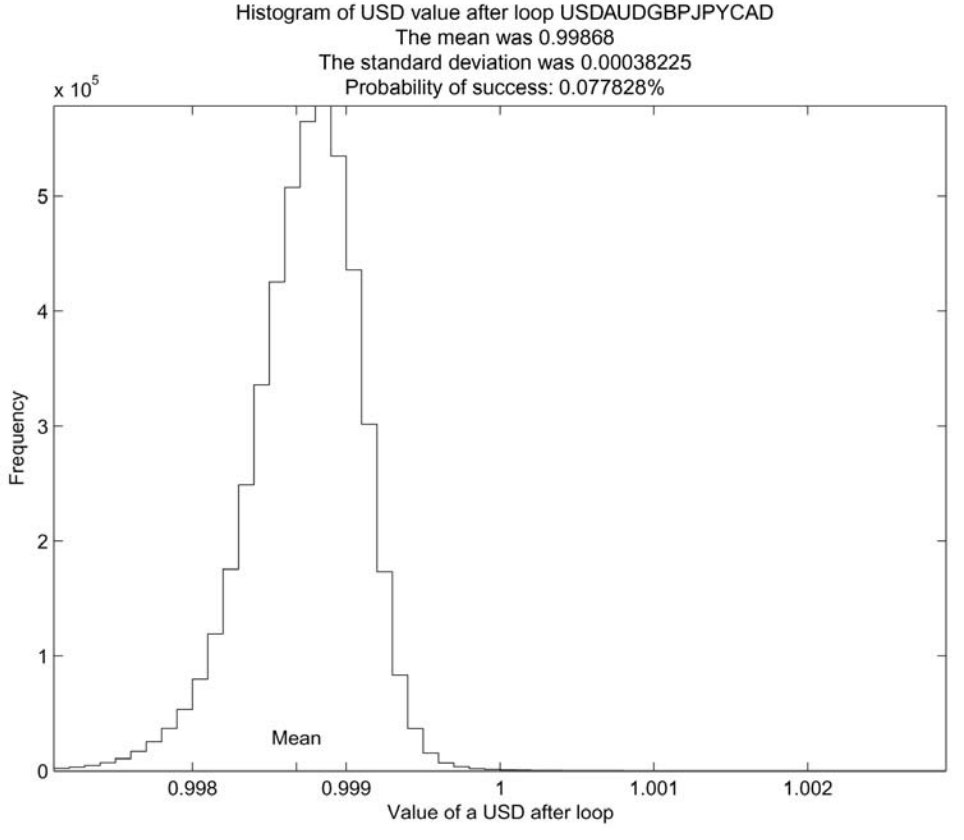

Figure 3.A.2. Histogram of the value of \$1 after a five-currency chain of currency conversions using Thomson Reuters quotes (USD → AUD → GBP → JPY → CAD → USD). This five-currency chain is typical of longer cycles: lower mean and higher variation.

Table 3.A.2. Summary statistics for various types of newslines in the Thomson Reuters NewsScope Archive.

(a) Bodylines

Bodylines	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.	Week-day	Week-end	All
Minimum	6,449	18,206	7,514	0	13,329	3,286	3,805	130,313	9,315	143,577
5%	28,253	34,500	37,223	36,355	22,152	5,319	6,442	189,858	11,851	205,479
Mean	45,321	51,133	52,489	53,515	46,875	7,315	8,516	250,104	15,837	265,872
Median	47,122	52,342	53,605	54,682	48,427	7,299	8,564	250,486	15,830	267,013
95%	56,777	61,741	63,462	65,492	57,801	9,408	10,399	300,506	19,201	317,809
Maximum	66,185	68,761	83,049	73,958	75,880	13,538	14,859	337,051	23,932	355,656
Std. dev.	9,398	8,445	9,023	10,086	9,479	1,358	1,360	35,122	2,253	36,232
Skewness	−1.237	−1.557	−1.458	−1.981	−1.411	0.267	−0.338	−0.546	0.122	−0.589
Kurtosis	5.341	6.868	8.386	9.860	6.111	5.066	6.518	4.084	3.947	4.160

(b) Bodylines per 10-minute intervals

Bodylines	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.	Week-day	Week-end	All
Minimum	0	0	1	0	0	0	0	0	0	0
5%	85	109	113	109	76	5	5	97	5	16
Mean	315	355	364	371	325	51	51	346	55	263
Median	303	345	355	356	306	41	41	333	43	239
95%	603	655	666	708	642	128	128	658	145	617
Maximum	2,418	1,674	2,259	1,789	6,152	630	630	6,152	1,755	6,152
Std. dev.	165	173	177	190	191	44	44	181	51	204
Skewness	0.85	0.71	0.83	0.81	3.54	2.57	2.57	144	367	112
Kurtosis	5.56	4.11	5.28	4.43	73.47	14.38	14.38	21.64	49.21	11.82
ρ_1 (%)	54.92	62.70	60.64	59.50	52.24	57.62	65.82	58.53	62.99	75.95
ρ_2 (%)	49.99	59.58	58.00	56.29	50.58	53.60	63.33	55.55	59.94	74.22
ρ_3 (%)	49.96	57.13	56.29	57.21	47.43	51.58	61.29	54.28	57.92	73.46

(*continued*)

Table 3.A.2 (*cont.*)

(c) Alerts by group

Alerts	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.	Week-day	Week-end	All
Minimum	37	565	50	38	299	1	13	1,992	42	2,093
5%	713	1,179	1,301	1,241	615	20	23	6,032	58	6,029
Mean	1,565	2,084	2,222	2,516	1,513	70	75	9,982	145	10,212
Median	1,466	1,945	2,010	2,274	1,451	55	63	9,195	130	9,505
95%	2,709	3,873	3,920	4,779	2,323	200	153	17,818	309	18,321
Maximum	3,653	5,098	5,400	7,180	3,549	342	356	23,497	479	23,727
Std. dev.	619	791	885	1,108	518	55	49	3,731	74	3,801
Skewness	0.625	1.201	1.197	1.356	0.418	2.007	1.982	1.168	1.387	1.118
Kurtosis	3.550	4.697	4.833	5.671	4.181	7.635	9.364	4.658	5.401	4.520

(d) Headlines by group

Headlines	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Sun.	Week-day	Week-end	All
Minimum	920	4,024	1,054	864	2,643	365	647	19,063	1,362	20,838
5%	5,062	6,835	6,911	6,622	5,120	813	1,030	32,243	1,880	33,864
Mean	7,165	8,074	8,204	8,460	7,466	1,046	1,283	39,473	2,330	41,854
Median	7,287	8,026	8,144	8,425	7,478	1,026	1,272	39,530	2,291	41,937
95%	9,065	9,788	9,814	10,536	9,453	1,395	1,623	47,418	2,888	49,889
Maximum	10,639	11,400	13,265	11,874	11,347	1,599	1,989	53,232	3,534	55,863
Std. dev.	1,304	1,015	1,150	1,355	1,365	182	188	5,259	318	5,492
Skewness	−1.197	−0.319	−0.885	−1.348	−0.999	0.494	0.285	−0.543	0.433	−0.619
Kurtosis	6.790	5.200	11.935	9.602	6.089	4.298	4.972	4.968	4.209	5.074

Table 3.A.3. Empirical *t*-distribution percentiles for 30-minute random event studies using squared currency returns for 15 currency pairs

Currency pair	Returns of *t*-statistic percentiles for random event studies									
	90%	95%	99%	99.50%	99.90%	90%	95%	99%	99.50%	99.90%
	500 events					500 events (de-seasonalized)				
AUD/USD	0.70	0.89	1.25	1.38	1.61	2.71	3.48	5.07	5.38	7.31
CAD/USD	0.63	0.82	1.16	1.25	1.52	2.89	3.85	5.41	5.90	7.59
CAD/JPY	0.87	1.09	1.63	1.79	2.09	2.43	3.26	4.59	5.67	7.03
CHF/USD	0.67	0.89	1.15	1.19	1.44	3.34	4.13	5.79	6.37	6.85
EUR/USD	0.59	0.79	1.09	1.12	1.40	3.49	4.76	6.39	7.18	9.60
EUR/CAD	0.48	0.63	0.93	1.14	1.40	2.23	2.83	4.13	4.91	6.06
EUR/GBP	0.50	0.63	0.90	1.03	1.36	3.06	3.72	6.01	6.34	6.97
EUR/JPY	0.48	0.61	0.82	0.93	1.00	3.19	3.99	5.56	6.52	8.89
GBP/USD	0.70	0.95	1.31	1.41	2.12	3.23	4.13	6.06	6.45	8.67
GBP/AUD	0.80	1.01	1.45	1.57	1.77	2.45	3.11	4.44	5.12	6.69
GBP/JPY	0.66	0.89	1.23	1.35	2.27	2.69	3.49	4.66	6.05	7.20
JPY/USD	0.65	0.81	1.11	1.23	1.46	3.89	4.76	6.88	7.88	9.07
MXN/USD	0.90	1.11	1.54	1.88	2.00	3.24	4.15	6.20	6.88	7.75
NOK/USD	0.66	0.83	1.34	1.46	1.73	3.95	4.92	7.16	8.02	9.83
NZD/USD	0.70	0.88	1.21	1.41	1.65	2.77	3.63	4.86	5.33	7.44
	1,000 events					1,000 events (de-seasonalized)				
AUD/USD	0.61	0.78	1.15	1.33	1.73	2.47	3.33	5.08	6.02	7.23
CAD/USD	0.63	0.81	1.11	1.24	1.51	2.72	3.75	4.86	5.39	6.73
CAD/JPY	0.86	1.10	1.52	1.61	1.99	2.76	3.53	4.91	5.36	6.53
CHF/USD	0.57	0.75	1.04	1.12	1.23	3.15	3.92	5.81	6.30	7.48
EUR/USD	0.63	0.81	1.07	1.11	1.20	3.47	4.44	6.39	6.79	8.66
EUR/CAD	0.49	0.62	0.84	0.97	1.10	2.27	3.03	4.19	4.71	6.87
EUR/GBP	0.46	0.61	0.86	1.03	1.22	3.34	4.36	6.29	6.60	9.03
EUR/JPY	0.46	0.60	0.87	0.96	1.14	3.11	3.99	5.78	6.77	8.24
GBP/USD	0.67	0.84	1.25	1.29	1.71	3.07	3.96	5.58	6.10	7.95
GBP/AUD	0.80	1.03	1.42	1.57	1.67	2.60	3.32	4.54	4.89	5.43
GBP/JPY	0.64	0.82	1.26	1.40	3.00	2.57	3.47	4.75	5.31	8.70
JPY/USD	0.63	0.81	1.15	1.39	1.79	3.76	4.70	6.81	7.28	8.81
MXN/USD	0.82	1.12	1.52	1.71	1.84	2.88	3.80	6.09	6.77	8.01
NOK/USD	0.65	0.82	1.16	1.31	1.54	3.71	5.00	7.21	7.73	8.99
NZD/USD	0.71	0.89	1.26	1.41	1.56	2.69	3.50	5.04	5.34	6.02
	2,500 events					2,500 events (de-seasonalized)				
AUD/USD	0.64	0.88	1.19	1.33	1.76	2.83	3.60	4.70	5.07	7.19
CAD/USD	0.67	0.80	1.16	1.31	1.44	2.81	3.66	5.42	5.68	8.80
CAD/JPY	0.86	1.09	1.57	1.84	2.21	2.57	3.21	4.46	4.84	7.69
CHF/USD	0.56	0.75	1.11	1.15	1.70	3.18	3.92	5.83	6.28	8.46
EUR/USD	0.60	0.76	1.04	1.16	1.40	3.58	4.58	6.28	6.87	8.19
EUR/CAD	0.49	0.61	0.87	1.03	1.31	2.23	2.85	3.87	4.56	6.38
EUR/GBP	0.49	0.65	0.97	1.04	1.28	3.05	4.08	5.33	6.04	6.74
EUR/JPY	0.49	0.62	0.86	0.99	1.10	3.17	4.12	5.67	6.68	8.02
GBP/USD	0.69	0.88	1.35	1.55	2.04	3.35	4.36	6.56	7.04	7.94
GBP/AUD	0.77	1.02	1.44	1.74	1.78	2.34	2.97	4.16	4.64	5.76
GBP/JPY	0.68	0.87	1.16	1.25	1.47	2.31	3.01	4.28	4.92	6.87
JPY/USD	0.60	0.77	1.11	1.15	1.29	3.58	4.64	6.41	7.06	8.66
MXN/USD	0.88	1.09	1.54	1.62	1.80	2.88	3.71	5.39	6.29	9.50
NOK/USD	0.66	0.84	1.18	1.22	1.30	3.90	4.85	6.85	7.80	10.10
NZD/USD	0.67	0.83	1.21	1.54	1.58	2.68	3.47	4.91	5.25	7.89

3.B REFERENCES

Antweiler W.; Frank M. (2004) "Is all that talk just noise? The information content of internet stock message boards," *Journal of Finance*, **59**, 1259–1294.

Bauwens L.; Ben Omrane W.; Giot P. (2005) "News announcements, market activity and volatility in the euro/dollar foreign exchange market," *Journal of International Money and Finance*, **24**, 1108–1125.

Berry T.; Howe K. (1994) "Public information arrival," *Journal of Finance*, **49**, 1331–1346.

Campbell J.; Lo A.; MacKinlay C. (1997) *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Chan W. (2003) "Stock price reaction to news and no-news: Drift and reversal after headlines," *Journal of Financial Economics*, **70**, 223–260.

Cormen T.; Leiserson C.; Rivest R.; Stein C. (2001) *Introduction to Algorithms*, MIT Press/McGraw-Hill Book Company, Cambridge/New York.

Dacorogna M.; Gencay R.; Müller U.; Olsen R.; Pictet O. (2001) *An Introduction to High-Frequency Finance*, Academic Press, San Diego.

Daniel K.; Hirshleifer D.; Subrahmanyam A. (1998) "Investor psychology and security market under- and overreactions," *Journal of Finance*, **53**, 1839–1885.

Das S.; Martinez-Jerez F.; Tufano P. (2005) "eInformation: A clinical study of investor discussion and sentiment," *Financial Management*, **34**, 103–137.

Dominguez K.; Panthaki F. (2006) "What defines 'news' in foreign exchange markets?" *Journal of International Money and Finance*, **25**, 168–198.

Engle R.; Ng V. (1993) "Measuring and testing the impact of news on volatility," *Journal of Finance*, **48**, 1749–1778.

Hirshleifer D.; Shumway T. (2003), "Good day sunshine: Stock returns and the weather," *Journal of Finance*, **58**, 1009–1032.

Hong H.; Lim T.; Stein J. (2000) "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies," *Journal of Finance*, **55**, 265–295.

Johnson G.; Schneeweis T. (1994) "Jump-diffusion processes in the foreign exchange markets and the release of macroeconomic news," *Computational Economics*, **7**, 309–329.

Leinweber D.; Sisk J. (this volume) "Relating news analytics to stock returns" (see Chapter 6).

Liu H.; Lieberman H.; Selker T. (2003) "A model of textual affect sensing using real-world knowledge," paper presented at *Proceedings Seventh International Conference on Intelligent User Interfaces (IUI 2003), Miami, Florida*, pp. 125–132.

Mitchell M.; Mulherin H. (1994) "The impact of public information on the stock market," *Journal of Finance*, **49**, 923–950.

Niederhoffer V. (1971) "The analysis of world events and stock prices," *Journal of Business*, **44**, 193–219.

Oberlechner T.; Hocking S. (2004) "Information sources, news, and rumors in financial markets: Insights into the foreign exchange market," *Journal of Economic Psychology*, **25**, 407–424.

Pearce D.; Roley V. (1985) "Stock prices and economic news,"*Journal of Business*, **58**, 49–67.

Pong S.; Shackleton M.; Taylor S.; Xu X. (2004) "Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models," *Journal of Banking and Finance*, **28**, 2541–2563.

Stone P.; Dunphy D.; Smith M.; Ogilvie D. (1966) *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press, Cambridge, MA.

Taylor S. (2005) *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, Princeton, NJ.

Tetlock P. (2007) "Giving content to investor sentiment: The role of media in the stock market," *Journal of Finance*, **62**, 1139–1168.

Tetlock P.; Saar-Tsechansky P.; Macskassy S. (2008) "More than words: Quantifying language to measure firms' fundamentals," *Journal of Finance*, **63**, 1437–1467.