How Long Do Financial Markets Need to Fully Respond to FOMC Announcements?

Paul L. Tran*
University of Texas at Austin

This version: August 26, 2025
Click here for the latest version

Abstract

This paper shows that financial markets need more time to fully react to Federal Open Market Committee (FOMC) policy announcements than typically assumed. Using finance literature techniques and neural network methods for text analysis, I systematically estimate the event window lengths that best reflect market fundamental reactions to only the information content of FOMC statements. On average, markets fully react within an event window *ending at least 30 minutes* after release. This optimal window increases with asset maturity, reaching 50–60 minutes for maturities at least two quarters out, showing that a single window length does not fit all asset types.

Keywords: Event window studies, FOMC statements, neural networks, natural language processing, text analysis

JEL Codes: C45, E52, E58, G14

*Department of Economics, The University of Texas at Austin. 2225 Speedway, BRB 2.128, C3100, Austin, TX 78712, USA. Email: pltran@utexas.edu. Website: https://paulletran.com/. I am indebted to my committee—Oli Coibion, Chris Boehm, Saroj Bhattarai, and Amy Handlan—for their invaluable guidance and support. For their helpful comments and suggestions, I also wish to thank Andi Mueller and Andrew Chang. Finally, I thank Trieu Tran for her research assistance.

1 Introduction

> To establish the right price for a stock the market must have adequate information, but it by no means follows that if the market has this information it will thereupon establish the right price. The market's evaluation of the same data can vary over a wide range, dependent on bullish enthusiasm, concentrated speculative interest and similar influence, or bearish disillusionment. Knowledge is only one ingredient on arriving at a stock's proper price. The other ingredient, fully as important is sound judgement.
>
> Graham (1974)

How much time do financial markets need to fully respond to news? The answer might not be obvious. Consider the following example of the S&P 500 Index price movement shortly after the release of the August 1999 Federal Open Market Committee (FOMC) monetary policy announcement. As shown in Figure 1, measuring the price change up to 20, 50, or 60 minutes after the announcement would lead to the conclusion that the stock market index had a positive, negative, or zero reaction respectively. In other words, the *time window* chosen to measure financial market price reactions to news can result in drastically different answers to the original question.

Because financial prices can serve as proxy for market expectations, meaning changes in these prices arguably represent changes in expectations, tools such as event window studies have become integral for empirical macroeconomics. The method allows econometricians to measure financial market responses to news and capture the predicted effects of events on macroeconomic outcomes. However, news releases can be dense and difficult to interpret, especially when the information content deals with longer time horizons. As a result, whilst markets may need more than a few minutes to react to complex events or policy announcements, longer windows increase the likelihood that the measurement will be contaminated as additional news is revealed, making it difficult to assign market price responses to a particular event. Balancing these two forces is what allows the field to pin down the event window length that should be used to determine the response of financial markets to news releases. Therefore, determining the optimal window length is an empirical question and not a parameter whose value we should simply assume.

A relevant and important application of event window studies is to construct monetary policy surprises, which are changes in interest rates around the Federal Open Market Committee (FOMC)'s policy announcements. Monetary policy surprises can be used directly (e.g., Kuttner, 2001; Gürkaynak, Sack, et al., 2005) or used as an instrument in a structural vector autoregressive or local projections framework (e.g., Gertler and Karadi, 2015; Handlan, 2022b) to estimate the effects of monetary policy on asset prices and the macroeconomy.

The choice of event window size for monetary policy surprises falls into two categories: Short or long window lengths. Proponents of narrow windows around FOMC announcements (e.g., 30 minutes) argue that this high-frequency identification (HFI) makes plausible the assumption that observed price changes are caused solely by the policy announcement. This approach has become the popular choice and de facto standard in the literature (Gürkaynak, Sack, et al., 2005; Gertler and Karadi, 2015; Nakamura and Steinsson, 2018; Gürkaynak, Kisacikoğlu, et al., 2020; Handlan, 2022b; Swanson and Jayawickrema, 2023; Bauer and Swanson, 2023). An implicit assumption from this argument is that the initial price reaction of financial markets within small event windows is equal to the "fundamental" or full reaction. However, this implicit assumption could be violated due to various issues potentially facing financial markets, such as over-reactions (e.g., Boguth et al., 2023; Bianchi et al., 2024), under-reactions (e.g., Boguth et al., 2023; Bianchi et al., 2024), the presence of noise trading (e.g., Hervé et al., 2019; Ben-David et al., 2022), and algorithmic trading (e.g., Bazzana and Collini, 2020; Ben Ammar and Hellara, 2022). In addition, Ramey (2016) finds that the macroeconomic effects of monetary policy surprises constructed with samples beginning after 1984 declined in precision likely because of how monetary policy was conducted more systematically after this time period. As a result, Ramey (2016) argues that monetary policy shocks now consist mostly of information effects from the FOMC and *noise*. The combination of all these effects could result in monetary policy surprises to lack relevance and precision.

Conversely, those who utilise longer event windows (e.g., 1-hour to 1-day) contend that this approach has a higher probability of capturing the market's full, fundamental reaction to FOMC announcements, resulting in shock measures constructed from the monetary policy surprises that have higher explanatory power (Kuttner, 2001; Rigobon and Sack, 2004; Gürkaynak, Sack, et al., 2005; Gürkaynak, Kisacikoğlu, et al., 2020; Bauer, Lakdawala, et al., 2022; Swanson and Jayawickrema, 2023; Boehm and Kroner, 2023). Unfortunately, the trade-off with this alternative size is that longer event window lengths increase the chance that measurements are contaminated with unrelated news, causing a decline in statistical significance and precision of estimates (Gürkaynak, Sack, et al., 2005; Nakamura and Steinsson, 2018).

Ultimately, the choice of event window length has remained an "ad-hoc" decision. Relatedly, the literature has implicitly assumed that this choice applies to all assets, an assumption that may not hold. For example, markets for assets with short maturities might price in FOMC policy announcements more easily, whilst markets for longer-maturity assets might need more time.[1] Therefore, one might expect that longer event window lengths are required to capture the full reaction of the markets for assets such as longer-term bonds and equities.

The observed trade-offs with different event window sizes bring up a crucial question: how does one choose the appropriate event window length for their study in monetary policy? For a methodology as widespread as event window studies in empirical macroeconomics, there are surprisingly very few papers that directly investigate the appropriate length of time over which to measure price reactions. One of the earliest papers looking into this question was Hillmer and Yu (1979), who developed a theory around the assumption that there exist two price reaction processes: One during a period of adjustment to the event of interest and one outside the reaction period. Tests of their model show that as long as markets are still reacting to the new information, the variance of the price reactions within the event window was higher than in periods outside of the window. Using intraday data, the paper finds that

[1]Throughout this paper, I differentiate between "asset/underlying maturity" and "futures contract maturity". The former typically refers to the maturity of the underlying asset for futures contracts (e.g., 2-year Treasury notes). The latter refers to the expiration date of the futures contract themselves, (e.g., front-month 2-year Treasury futures).

the event window should last several hours after the initial event of interest. Chang and Chen (1989) apply the methodology of Hillmer and Yu (1979) on daily data, finding that event windows should span several days due to the market continuing to react to the initial event. Krivin et al. (2003) propose various proxy rules for determining the length when looking at a limited number of observations, such as those based on significant abnormal returns. The authors find that rules based on continuing price movements result in event window sizes that correlate with the "size" of the news. A recent paper by Das and King (2021) attempts to provide some systematic structure by measuring the informativeness of earnings announcements using abnormal R^2 and other statistics, discovering that asymmetric, narrower one-day event windows exhibit greater information content.

My paper builds upon this literature from an alternative angle. Whilst the aforementioned studies infer the optimal window length using only observed price dynamics, my approach combines price dynamics with a direct, text-based signal of the market's fundamental price response to further pin down the optimal event window length. In particular, I use neural network methods for text analysis to construct this signal solely from the information content of FOMC statements, representing the market's direct reaction to the news, word-for-word. My method provides an explicit and systematic approach for setting this basic step in event window studies for the literature.

This method yields two key findings. First, in line with the insights from Hillmer and Yu (1979), my results show that financial markets take more time to process news than is typically assumed in modern HFT studies. Specifically, I find that regardless of contract maturity, financial markets for futures and equities fundamentally react to the information content of FOMC statements within an event window starting 10 minutes before and *ending at least 30 minutes* after release. In other words, I show that the time financial markets need to fully react to FOMC policy announcements is longer than what is popularly assumed in the literature. Second, and in contrast to those earlier papers, I find the optimal window lengths are ultimately shorter than several hours and, importantly, vary systematically across asset classes and maturities.

Figure 2 offers a stylised summary of the main results. It displays the systematically estimated event window lengths from my method for all considered underlying asset maturities, such as 2-year Treasury futures (i.e., two years out). As mentioned above, markets for assets with the shortest underlying maturities are estimated to fundamentally react within a 40-minute event window. As the maturity increases, the window length rises accordingly, levelling off at an event window beginning 10 minutes before and ending 40–50 minutes after statement release. In other words, my paper documents that the optimal event window length varies across asset types and underlying maturities in a systematic way.

When it comes to general applications of text analysis methods in economics, most empirical studies have employed the techniques of clustering words into topics or using word counts to construct sentiment lists. Many such examples exist in the monetary policy literature that studies central bank communication, such as Acosta (2023) in their measures of expansionary versus contractionary monetary policy sentiment and Husted et al. (2020) in their creation of a monetary policy uncertainty index. These techniques are popular because they allow researchers to choose the interpretation and emphasis of individual words. However, a common disadvantage amongst these methods is that more complex features of text, such as context and word interdependencies, are not captured. In contrast, neural networks for text analysis do not overlook these characteristics. Until recently, applying these neural networks to economics has been more or less impossible because training a network for language processing typically requires millions of text documents. However, the successful adaptation of neural networks on smaller text samples from the computer science literature now makes it possible to study the communication of institutions, such as central banks, with these methods.

The most similar examples to the methodology of my paper are Handlan (2022b) and Doh et al. (2023). The latter study uses an artificial neural network for text analysis called the Universal Sentence Encoder to measure differences in hawkish-dovish sentiment between official and alternative FOMC statements and create a stance metric over time. The former paper utilises the state-of-the-art neural network, XLNet of Yang et al. (2019), to create monetary policy text shocks from Federal Open Market Committee (FOMC) official and alternative statements. Using XLNet and these statements, she is able to remove the "Fed Information Effect" (Miranda-Agrippino and Ricco, 2021) from the forward guidance component. Impulse responses of macroeconomic variables from her text shock follow the same trends and amplitudes as those predicted by theory. My paper contributes to this growing literature by utilising the same neural network as in Handlan (2022b), but for a first step in the process of obtaining monetary policy shocks: Systematic determination of event window lengths around FOMC announcements such that financial market price responses reflect "fundamental" reactions directly and only to the policy communications, resulting in more relevant monetary policy surprises (Acosta, 2023).

2 Conceptual Framework and Simulations

The conceptual framework presented in this section shows how the dynamics of asset market prices in response to news have conflicting factors, applying structure and motivating the methodology behind the systematic estimation of the appropriate event window length such that these factors are minimised.

Consider the price of one asset market at time t, P_t. Assume that the news is released at the beginning of the time period, $t = 0$. As shown in Equation 1, the asset price is the sum of three components:

$$\{P_t = P_t^f + \varepsilon_t^c + \varepsilon_t^n | t \geq 0\}, \tag{1}$$

where P_t^f is the fundamental price component, ε_t^c is cognitive noise, and ε_t^n is unrelated news.

P_t^f represents what asset price "should" be set by the market due to the release of news, meaning it is assumed that $P_t^f = P^f \in \mathbb{R}$.

Cognitive noise can be thought of as the collective "market noise" due to market participants only being able to react to so much information in a given amount of time. For aforementioned reasons such as algorithmic and noise trading and potential liquidity constraints, it is possible that market participants will have incomplete information or noise as well as different beliefs of what the news event means to the price of the asset they are trading in a short amount of time. As time progresses after the news release, more investors will be able to see each other's price reactions and respond accordingly, collectively converging to the fundamental price component. In other words, cognitive noise "decays" over time and converges to zero. The framework reflects this behaviour by representing cognitive noise as the following modified AR(1) process:

$$\varepsilon_t^c = \rho_c \varepsilon_{t-1}^c + e^{-decay t} \nu_t^c,$$

where coefficient $|\rho_c| < 1$, $decay \in \mathbb{R}^+$, and random noise ε_t^c is normally distributed with mean zero and variance σ_c^2. Two necessary assumptions to ensure the cognitive noise process in the framework exhibits the decaying behaviour over time are that $|\frac{\rho_c}{decay}| < 1$ and the variance of cognitive noise at time $t = 0$ is equal to σ_c^2.

The effect of news unrelated to the event of interest on the asset price is modelled as a random walk:

$$\varepsilon_t^n = \varepsilon_{t-1}^n + \nu_t^n,$$

where random noise ν_t^n is normally distributed with mean zero and variance σ_n^2. Similar to cognitive noise, the variance of the unrelated news process at time $t = 0$ is equal to zero.

2.1 The Driving Forces Behind the Variance of Asset Prices

The expression of P_t as Equation 1 allows for the derivation of the variance of the asset price for all time $t \geq 0$ through iterative substitution.

At time $t = 0$ when news is released, one can express the variance of the asset price as

$\text{Var}(P_0) = \text{Var}(\varepsilon_0^c) + \text{Var}(\varepsilon_0^n) = \sigma_c^2$. Similarly, the expression at time $t = 1$ is $\text{Var}(P_1) = \text{Var}(\varepsilon_1^c) + \text{Var}(\varepsilon_1^n) = \sigma_c^2(\rho_c^2 + e^{-2decay}) + \sigma_n^2$. Continuing this iterative process yields the following expression of

$$\begin{aligned} \text{Var}(P_t|t \geq 0) &= \left[\sum_{i=0}^{t} \rho_c^{2(t-i)} e^{-2decayi}\right] \sigma_c^2 + t\sigma_n^2 \\ &= \left[\frac{\rho_c^{2(t+1)} - e^{-2(t+1)decay}}{\rho_c^2 - e^{-2decay}}\right] \sigma_c^2 + t\sigma_n^2, \end{aligned} \tag{2}$$

where $\lim_{t\to\infty} \left[\frac{\rho_c^{2(t+1)} - e^{-2(t+1)decay}}{\rho_c^2 - e^{-2decay}}\right] = 0$ by the aforementioned framework assumptions.

I define t^{one} as the time where the variance of P_t is minimised. Solving for this time horizon yields the following indirect expression that provides important insight into the factors influencing the appropriate event window length:

$$t^{one} : decay\left[e^{-2(t+1)decay}\right] + ln(\rho_c)\rho_c^{2(t+1)} = \left[\frac{(e^{-2decay} - \rho_c^2)}{2}\right] \frac{\sigma_n^2}{\sigma_c^2}. \tag{3}$$

2.2 Estimator Form

Two important findings from Equation 3 are that $\frac{\partial t^{one}}{\partial \sigma_n^2} < 0$ and $\frac{\partial t^{one}}{\partial \sigma_c^2} > 0$. In other words, an increased presence of unrelated news (cognitive noise) results in the time where the variance of the asset price is minimised to decrease (increase).[2] The dynamics caused by these two components give formal insight into why strictly using narrow event windows isn't appropriate.

The dynamics described by Equation 3 provide insight into the trade-offs for a *single* news announcement. To apply this framework empirically, this concept must be extended to the *average* reaction across many news releases. Therefore, the goal of this paper is not to find t^{one} for any single event, but to systematically estimate t^*, the time at which the market fundamentally reacts to news announcements on average. Formally, consider N news

[2]Because Equation 3 is unable to have t isolated on one side, I numerically verify the dynamics of the t^{one} for various values of σ_c^2 and σ_n^2 in the indirect expression whilst holding the other parameters constant.

announcements and the price of one asset market. For each news i that is released at time t, the asset price $P_{i,t}$ and its components will respond to the announcement. Therefore, I define t^* as the systematically estimated time that minimises the mean squared error (MSE) between $P_{i,t}$ and $P^f_{i,t}$ for all N news:

$$t^* : \min_t \frac{1}{N} \sum_{i=1}^{N} \left(P_{i,t} - P^f_{i,t} \right)^2 = \min_t \frac{1}{N} \sum_{i=1}^{N} \left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} \right)^2 \tag{4}$$

However, assume that the fundamental component is unobservable to the econometrician. Instead, suppose that the econometrician observes a noisy signal of the fundamental component, $s_i = P^f_i + \xi_i$, where $\xi_i \sim \mathcal{N}(0, \sigma^2_s)$. By making the core assumption that the sum of the asset price noise components, $(\varepsilon^c_t + \varepsilon^n_t)$, and the signal noise are *independent*, I am able to derive the following expression for the minimisation problem of the MSE:

$$\begin{aligned}
t^* : \min_t \frac{1}{N} \sum_{i=1}^{N} (P_{i,t} - s_i)^2 &= \min_t \frac{1}{N} \sum_{i=1}^{N} \left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} - P^f_i - \xi_i \right)^2 \\
&= \min_t \frac{1}{N} \sum_{i=1}^{N} \left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} - \xi_i \right)^2 \\
&= \min_t \frac{1}{N} \sum_{i=1}^{N} \left[\left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} \right)^2 + \xi^2_i - 2\xi_i \left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} \right) \right]^2 \\
&= \min_t \left\{ \mathbb{E} \left[\left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} \right)^2 \right] + \mathbb{E} \left[\xi^2_i \right] - 2\, \mathbb{E} [\xi_i]\, \mathbb{E} \left[\left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} \right) \right] \right\} \\
\implies t^* : \min_t \frac{1}{N} \sum_{i=1}^{N} (P_{i,t} - s_i)^2 &= \min_t \left[\frac{1}{N} \sum_{i=1}^{N} \left(\varepsilon^c_{i,t} + \varepsilon^n_{i,t} \right)^2 + \sigma^2_s \right]
\end{aligned} \tag{5}$$

Because the addition of a constant does not affect the optimisation problem of the MSE with respect to time t, this derivation implies that the econometrician is still able to solve for the optimal time horizon, t^*, by observing *only* the noisy signal of the fundamental, s_i.

2.3 Simulations

This section simulates the asset price process over time for multiple news announcements. The simulation serves two purposes: first, to act as a jumping-off point for the systematic estimation method of this paper, and second, to illustrate the effects of cognitive noise and unrelated news on the optimal market reaction time.

The simulation assumes that news is released at time $t = 0$. The fundamental price component, $P^f_{i,t}$, responds by jumping to and remaining at real value $P^f_i \in [-100, 100]$. Simultaneously, the cognitive noise component jumps to a random value within $[-100, 100]$ whilst unrelated news is equal to zero at $t = 0$. The noisy signal observed by the econometrician, s_i, is assumed to have its error component distributed normally with mean zero and finite variance. To analysis this on average, I simulate the price process and the noisy signal for $N = 10,000$ news announcements up to time $t = 100$, calculating both the true and estimated MSEs using P^f_i and signal s_i, respectively. Finally, the true and estimated times at which the market fundamentally reacts on average, t^* and $\hat{t}$, can be determined according to Equation 4.

2.4 Simulation Results

Three scenarios are considered for the asset market price responding to the release of news: one that exhibits cognitive noise but little unrelated news, one with unrelated news but little cognitive noise, and one with both cognitive noise and unrelated news. The simulation parameters corresponding to these three scenarios can be found in the topmost eight rows of Table 1.

Reading the columns of the bottom two rows of Table 1 from left to right, one can see that an asset market possessing a lot of cognitive noise with little unrelated news will have a longer optimal time horizon. The middle column shows the opposite: when faced with a lot of unrelated news but little cognitive noise, t^* becomes very small. Both these results follow the dynamics predicted by Equation 3. Intuitively, when markets need more time to fundamentally react to news, this means the optimal time horizon should be longer. In contrast, when markets are able to react quickly to the news and there is the presence of unrelated information coming out over time, using a short event window would be more appropriate. The rightmost column displays the most "realistic" scenario for an asset market: both cognitive noise and unrelated news exist. In such an event, the optimal time horizon lies somewhere between the two extreme scenarios.

Importantly, the simulations also present minimal differences between t^* and $\hat{t}$—the true and estimated time horizons, respectively—when using only the signal s_i. Therefore, a "good" signal allows for the possibility of estimating the time horizon that reflects when markets have fundamentally reacted to news on average. However, how does one obtain a "good" signal?

3 Natural Language Processing of FOMC Statements

From Subsection 2.4, the question of how one obtains a "good" signal about the effect of FOMC statements on the price reaction of financial markets can be rephrased as follows: How does one approximate the relationship from FOMC statement text to asset price changes? To satisfy this alternative but equivalent question, two requirements must be met. First, the method that approximates this relationship must be able to isolate changes in asset prices to the text of FOMC statements. Second, the method must be able to consider and quantify the "full" text of these statements.

Popular text analysis methods in economics can be summarised as "fitting predictive models on simple counts of text features" (Gentzkow et al., 2019). However, these methods often miss how words within and across sentences relate to each other. For example, the phrases "*employment* went up, but *inflation* did not" and "*inflation* went up, but *employment* did not" would produce the same measures. Alternatively, methods that measure the frequency of neighbouring words, called "n-grams", often miss information spread out across a sentence. For the following phrase, "economic growth slowed, but is expected to pick up pace later this year", a trigram would count "economic growth slowed", but would miss the point of the sentence that economic growth will reverse direction and expand later in the year.

In contrast, neural networks can approximate the complex relationships between words such as context and interdependencies. Using these methods, the words "slowed" and "pick up pace" could both be associated with "economic growth" for prediction, even though these words are not all adjacent within the sentence. Indeed, with other social sciences observing success in applying natural language processing, economics could witness similar benefits when using these methods to study the communication of central banks (Gentzkow et al., 2019).

3.1 XLNet, a Neural Network for Text Analysis

The off-the-shelf neural network for text analysis used in this paper is XLNet from Yang et al. (2019). It is considered a state-of-the-art method for text analysis tasks such as regression. The specific version of XLNet used in this paper, called xlnet-base-cased, is an open-source, pretrained neural network with 12 hidden layers (each of size 768), 12 self-attention heads (each of size 64), and a vocabulary size of 32,000 word tokens.[3] Originally, the authors trained their neural network to predict missing words when given input text that is converted to only contain word tokens.[4] The parameters of the neural network are then updated to more accurately and consistently predict the missing words. This task was only achievable through training XLNet with wide corpora of general English text.

Yang et al. (2019) then took XLNet's structure, trained weights, and embeddings as a starting point to perform other tasks that are considered benchmark exercises in the literature of machine learning for text analysis. These exercises allow for XLNet's parameters to

[3]More information about the neural network can be found on the open-source platform https://huggingface.co/xlnet/xlnet-base-cased.

[4]This conversion reduces the input vocabulary size. For example, {"decreasing", "increasing"} can be broken into the following tokens: {"de", "in", "creas", "ing"}.

undergo further adjustment through a process known as fine-tuning. The authors demonstrated that by using the pre-trained weights, they were able to achieve greater accuracy on these new tasks compared to not using the pre-trained weights and to other text analysis methods. This approach of pre-training parameters on a large and general collection of text and then using them as initial parameters before fine-tuning for a new task is called transfer learning. Most relevant for this paper is how Yang et al. (2019) and others showed that transfer learning decreases the data requirement whilst achieving similar accuracy on new tasks. In particular, the network structure, embeddings, word representations, and generalisability of XLNet allow it to "understand" the direct mapping between FOMC communication about futures price log-differences using the transfer learning approach.

This network is additionally able to satisfy the two conditions for extracting a "good" signal as seen in the beginning of Section 3 for the following reasons. First, Handlan (2022b) finds XLNet to outperform other networks when regressing on small samples of text using a transfer learning approach, which is essential for performing systematic estimation of the appropriate event window lengths with a small, finite sample of FOMC statements. Second, XLNet abandons the literature practices of "corrupting" input text through masking the words of prediction-interest and assuming that these masked words are independent of each other. The former can lead to discrepancies when subsequently training the network on a different set of text without masking. The latter is an assumption that is often violated.[5] Instead, XLNet considers all possible information from all permutations of words around the words of interest in its pre-training phase, learning from bi-directional context without the usage and weaknesses of masking. These advanced capabilities are what allow XLNet to meet the two key requirements for generating a "good" signal, as discussed in the following subsection.

[5]Consider the example phrase, "The Federal Open Market Committee decided to increase its target rate". The words "target" and "rate" are masked. However, these words still share implicit relations to each other.

3.2 How XLNet Satisfies the Signal Requirements

Recall that obtaining a "good" signal requires the neural network to be able to isolate changes in futures prices to the text of FOMC communication. In other words, unrelated news must not be affecting the natural language processing method. XLNet satisfies this condition by the very text analysis task it is being asked to perform. Specifically, XLNet is trained to be "fed" FOMC statements as its sole input and asked to then predict what the futures price log-difference should be using only this information. Therefore, the transfer learning approach by construction does not allow XLNet to use any other sources of unrelated news to make its predictions.

The second condition for obtaining a "good" signal is quantifying the text of FOMC statements in their entirety. The chosen neural network method fulfils this requirement by the initial training data that underpins its state-of-the-art status in the machine learning literature. In particular, Yang et al. (2019) train xlnet-base-cased for its initial task with BookCorpus (11,038 books) and the entire English Wikipedia (over 6.5 million articles). Although the original and larger-scale version of the neural network was initially trained on a corpus that expanded to include Giga5 (9.9 million news articles), ClueWeb12 (733 million websites), and Common Crawl (text from websites totalling over 1,000 terabytes of data), the authors demonstrated xlnet-base-cased outperformed competitive models like BERT by sizeable margins in a variety of tasks and datasets. Overall, XLNet was built on a solid foundation in understanding the general English language. When combined with its ability to learn from bi-directional context from both its permutation- and autoregressive-based training approach, XLNet has the potential to quantify the complex features of the text in FOMC statements and consider them as inputs for predicting the log-difference in futures prices. This ability allows the neural network to be better suited for tasks prioritising understanding context compared to recently popular generative pretrained transformer models (e.g., ChatGPT), which can only process and learn text in a unidirectional manner.[6]

3.3 Addressing Look-ahead Bias

A valid concern when using off-the-shelf neural networks for text analysis is look-ahead bias. As explained in Sarkar and Vafa (2024), look-ahead bias occurs whenever these networks predict values in the past using information in the future. The reason for the existence of this issue comes from how these neural networks are initially trained with a large amount of data from many sources, which increases the probability that information from the future was used in this initial training of the network weights. I argue that the effects of look-ahead bias on the neural network method used for systematic estimation of appropriate event window lengths are mitigated for two reasons. The first reason is from the initial training corpora of XLNet itself and the second pertains to the language composition of the FOMC statements.

From Subsection 3.1, recall that the original and larger-scale version of XLNet was initially trained by Yang et al. (2019) using a large corpus which includes headline news text from sources such as Bloomberg and Reuters. Therefore, it is possible that using the neural network from the original paper brings forth the possibility that XLNet was initially trained with information from FOMC statements indirectly. However, xlnet-base-cased addresses the concern of look-ahead bias because its initial training data was restricted by its authors to *only* BookCorpus and the entire English Wikipedia. This restriction benefits my paper because little to no information about the intersection of FOMC communication and the various futures markets considered is used to initially train the neural network. It should be mentioned that Yang et al. (2019) also showed that the restricted initial training data does not take away from XLNet's state-of-the-art performance when compared to other models.

Look-ahead bias is also mitigated for the application of XLNet in this paper because of information contained in the FOMC statements. After the statements have undergone pre-processing, the fairly consistent language composition of the statements has no references to

[6]Specifically, these neural networks are defined by predicting a word using only those that came before it. This "left-to-right" operation is why such models are commonly labelled as "generative" models.

relative times t and $t+1$. As a result, even if XLNet was trained with FOMC statements that were in the future relative to the statement it is being asked to predict price log-differences, the neural network would not have any information necessary to chronologically order the FOMC statements and subsequently "figure out" that its weights have been trained on future data. Similar to Glasserman and Lin (2023), the temporal anonymisation of the FOMC statements mitigates the effects of look-ahead bias when assessing the predictive qualities of XLNet on performing systematic estimation of appropriate event window lengths.

4 Data

The sample period for my analysis runs from May 1999 through October 2019. Because the FOMC is the principal entity of the Federal Reserve System that determines monetary policy for the U.S., the Committee is closely watched by markets, which react to regularly scheduled announcements about the its policy actions and forecasts. Of particular importance are the released FOMC statements, which are considered as the initial announcement of monetary policy. Because financial markets have been known to dissect and react to these statements word-by-word,[7] understanding when financial markets fully react to the information contained in these headline text is important for event study techniques. This direct relationship between financial market reactions and FOMC communication motivates the use of neural network methods for text analysis. Before discussing the application of these methods for systematically estimating appropriate event window lengths, I provide some descriptive information on the FOMC statements and the financial market asset prices considered in this paper. Table 2 provides an overview of the employed input and outputs data.

[7]E.g., CNBC coverage of the January 2025 FOMC statement text.

4.1 FOMC Statements

The decisions of the FOMC are typically announced in a press release shortly after its meeting concludes. The statements produced within the sample period of this paper were sourced from the Board of Governors of the Federal Reserve System website.[8] I drop statements of unscheduled FOMC meetings from my sample because I want to ensure the change in asset prices within various event window lengths around the statement's publication is driven by the content of the statements themselves and not confounded with the surprise that there was an FOMC meeting. The current version of this paper has a sample size of 165 statements.

Text Pre-processing: To prepare my sample of FOMC statements for analysis done by the neural network, I pre-process all FOMC statements by converting them into plain text format, changing the text coding into standardised UTF-8 format (e.g., change length of "-"), and ensuring the spacing between words is one space. In addition, I remove all URLs, the FOMC member voting record found at the end of each statement, the list of regional bank request approvals, and mentions of release timestamps. Although papers such as C Madeira and J Madeira (2019) have shown that voting records affect stock market reactions, no effect was found for interest-based assets such as Treasuries. As a result, I prioritise removing the voting records such that the main economic discussions within the FOMC statements are entirely considered by the neural network with its hard constraint on the number of words when training. More information about this limitation is discussed in Section 6. However, robustness checks with the voting records included in the text during inference are performed.

Content and Evolution of Statements: A typical FOMC statement begins with a general discussion of current macroeconomic conditions and then communicates the Committee's expectations for the future of the macroeconomy. The statement then mentions the Committee's goal of achieving maximum employment and price stability concludes with the new Federal Funds and discount rates. Following the Great Recession of 2008–2009, these state-

[8] https://www.federalreserve.gov/monetarypolicy/fomc.htm

ments included discussion about unconventional monetary policy, such as quantitative easing programmes. Because this period of time saw an increase in complexity in the language of FOMC communication and the inability of the Committee to use rate changes to influence market expectations about monetary policy, the statements grew rapidly in length up to 2014, as seen in Figure 3. This trend gradually reverses after 2014, but the statements are still longer on average than those before the Great Financial Crisis.

The following example is an excerpt of the FOMC statement released on 01 August, 2018:

> Information received since the Federal Open Market Committee met in June indicates that the labor market has continued to strengthen and that economic activity has been rising at a strong rate. [...]
>
> The Committee expects that further gradual increases in the target range for the federal funds rate will be consistent with sustained expansion of economic activity, strong labor market conditions, and inflation near the Committee's symmetric 2 percent objective over the medium term. Risks to the economic outlook appear roughly balanced.
>
> In view of realized and expected labor market conditions and inflation, the Committee decided to maintain the target range for the federal funds rate at 1-3/4 to 2 percent. The stance of monetary policy remains accommodative, thereby supporting strong labor market conditions and a sustained return to 2 percent inflation.

4.2 Financial Data

The data on asset prices for all considered financial markets come from the Thomson Reuters Tick History database and are obtained from Refinitiv. I use intraday data for all analyses for two reasons. First, prior studies have shown that higher frequency data lowers the risk of contamination by unrelated news and increases precision of estimates. Second, using intraday data allows for the construction of event window lengths popular in the HFI literature (e.g., 30 minutes).

Interest-Rate and Equity Futures: The current version of this paper uses minute-by-minute price data of futures contracts for various financial markets. Federal funds futures and Eurodollar futures contracts are considered because many high-frequency monetary policy surprises are constructed based on the change in expected interest rates at the time horizons measured in these asset types (e.g., Gürkaynak, Sack, et al., 2005; Gertler and Karadi, 2015; Nakamura and Steinsson, 2018; Handlan, 2022b). More specifically, the change in expected interest rate paths can be measured via the variation in prices of interest rate futures in event windows around FOMC announcements. The market for federal funds futures is liquid enough to measure expectations of monetary policy roughly three months out (i.e., $FF1, FF2, FF3, FF4$). Eurodollar futures can be used as instruments to measure expectations beyond three months. As an example in Acosta et al. (2024), a market participant trading on date t of quarter q based on their expectations on interest rates two quarters from now would trade the second-outstanding or 2-quarter Eurodollar future, with the settlement towards the end of quarter $q+1$. This paper considers Eurodollar futures capturing market expectations up to four quarters out (i.e., $EDcm2, EDcm3, EDcm4$). For longer asset maturities, systematic estimation of appropriate event window length is performed on financial markets for Treasury futures with underlying asset maturities of two, five, and ten years (i.e., $TUc1, TUc2; FVc1, FVc2; TYc1$) (Gürkaynak, Kisacikoğlu, et al., 2020). I also focus on equity markets, particularly the S&P 500 Index and its E-mini futures (i.e., $SPX, ESc1, ESc2$), for two reasons. First, aside from interest rates, equities are one of the most studied asset class in the empirical monetary policy literature. Second, stock index futures are traded outside of regular trading hours, providing sufficient data quality for me to consider a wide variety of event window lengths.

Dependent Variable Construction: Price levels for these futures contracts are collected at 10-minute intervals, starting from 10 minutes before an FOMC statement release through 18 hours after publication. The primary output of interest for the neural network is the log-price differences of each futures contract with the "anchor" point being the price 10 minutes before the FOMC announcement,

$$DP_{t+n} = ln\left(\frac{P_{t+n}}{P_{t-10}}\right), \tag{6}$$

where time $t+n$ is n minutes after the publication of the FOMC statement. The neural network of the systematic estimation currently considers price log-differences within a 70-minute event window, where the window ends 60 minutes after statement publication.[9]

5 Systematic Estimation of Event Window Lengths

The core idea behind the estimation procedure of this paper is to test which event window length allows the text of FOMC statements to most accurately predict the corresponding asset price changes. Specifically, I train a separate neural network for each potential window and evaluate its predictive accuracy on a held-out sample of statements. The event window associated with the model that demonstrates the highest average predictive accuracy is then identified as the optimal length.

I begin the systematic estimation by splitting my sample into training and testing sub samples such that 20 per cent of FOMC statements belong in the testing sample through a process called "stratified sampling k-fold cross validation for $k=5$" in the machine learning literature. Splitting the statement observations is specifically conditioned by the type of decision the FOMC made to the target federal funds rate, who the FOMC Chair was, if the date was pre- or post-2007, and by imposing both subsamples to have equal distribution of statements according to word count. This process yields five distinct training-testing sub-sample folds such that no testing subsample shares FOMC statements with one another.

These folds can be thought of as consisting of input-output pairs, where the inputs are the FOMC statements and the outputs are the calculated futures price log-differences. For each 10-minute interval and each futures contract maturity, I train XLNet by refitting its parameters to approximate the relationship between the text of FOMC statements and the price log-differences on the training subsample and then tested on the held-out subsample.

[9] Price log-differences calculated within longer event windows are not considered due to computational and financial constraints, but are still outputs of interest for future versions of this paper. This reason is also why the main text of this paper doesn't show the results for the S&P 500 E-mini futures: both contract maturities share the same underlying maturity of infinity with the stock index and the event window lengths considered by the current version of this paper are all before market close.

This process is repeated five times, once for each fold. The final accuracy for each event window length is the average performance across all five test folds. This methodology ensures a robust measure of out-of-sample performance and accounts for any variation in the network's predictions coming from the folds themselves.[10]

For each fold, I restrict the context window of XLNet (i.e., the number of word tokens the neural network considers as a single input and retains in its memory) to be a sequence length of 512 word tokens for each FOMC statement.[11] Because the average number of words found in FOMC statements is roughly 327 (with a peak of 800 words), I assume that XLNet should still have adequate headroom to understand the full meaning of most statements. Relatedly, a list of the neural network main hyperparameters for the fine-tuning process can be found in Table 3.

5.1 Accuracy Metrics

For each fold, the primary metric considered by the neural network to judge its accuracy during fine-tuning is the out-of-sample R^2 defined in Hawinkel et al. (2024):

$$R^2_{OOS} = 1 - \frac{\widehat{MSE}}{\widehat{MST}} = 1 - \frac{T^{-1}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}{\frac{T+1}{T(T-1)}\sum_{i=1}^{T}(y_i - \overline{y_{IS}})^2}, \tag{7}$$

where $\widehat{MSE}$ is the mean squared error of the neural network over the out-of-sample observations; $\widehat{MST}$ is the mean squared error calculated using the in-sample mean $\overline{y_{IS}}$ as prediction over all out-of-sample observations, and T is the size of the testing sample. Although the explicit objective function of the neural network during fine-tuning is to minimise $\widehat{MSE}$, this simultaneously maximises R^2_{OOS}. Other related metrics that are tracked by XLNet during its fine-tuning process are the out-of-sample Pearson correlation coefficient ρ between the predicted and actual observations, the out-of-sample mean absolute error, and the in-sample

[10] Stratified sampling k-fold cross validation minimises the differences between the population distribution of FOMC announcement characteristics and the subsample distributions of these characteristics, which is necessary for training neural networks on small samples through transfer learning.

[11] For example, {"decreasing", "increasing"} can be broken into: {"de", "in", "creas", "ing"}.

mean squared error. The last criterion is important for determining if the network is learning the relationship between the FOMC statement text and the price log-differences or not.

I should note that the value of R^2_{OOS} is defined explicitly as a comparison of XLNet to the null model (Anderson-Sprecher, 1994), $\overline{y_{IS}}$. Its value can be interpreted as the proportion of the null model's mean squared error that is explained by the neural network. In contrast, R^2_{OOS} does *not* follow the conventional definition of R^2, which is the proportion of variance in the data that is explained by a given model, for two reasons.[12] First, the conventional definition breaks down when applied to non-linear methods like neural networks because the variance of the observations is no longer comprised of only the model and residual variances. Second, applying the conventional R^2 metric onto out-of-sample observations breaks the spirit of prediction because it would be using the average outcome of the testing sample, which is supposed to be inaccessible to models during fitting.

The decisive criterion for systematic estimation is the out-of-sample R^2 *averaged across sample folds* for each 10-minute interval of each futures contract maturity:

$$\overline{R^2_{OOS}} = \frac{\sum_{i=1}^{K} R^2_{OOS}}{K}, \tag{8}$$

where $K = 5$ is the number of sample folds. In other words, the statistic can be interpreted to measure the average performance of the neural network in the task of approximating the relationship between FOMC statements and the price reactions of markets at each event window length, all relative to the null model.

5.2 Fine-tuning XLNet

In essence, performance of the fine-tuned network is measured by the accuracy of its predicted futures price log-differences for FOMC statements that were not used to train and update XLNet's parameters during transfer learning. Note that the machine learning lit-

[12]Note that the conventional definition is equivalent to the definition comparing a given model to the null model when applied to the in-sample observations.

erature evaluates neural networks primarily under this criterion and places no meaning on the network's parameter estimates because network parameters do not have the same interpretation as those of parametric models (Athey and Imbens, 2019). Therefore, my paper cannot describe the causal effect of one word over another within FOMC communication on market expectations. Instead, the training process behind the neural network allows me to approximate this relationship and use its predictions to attribute what event window size best reflects the markets' fundamental reaction to the FOMC statements.

When fine-tuning the network, overfitting becomes an inevitable issue due to XLNet having 110 million parameters. On a per-fold basis, this issue contributes to the difference between R^2_{OOS} and the equivalent calculated for the training sample. I minimise this contribution by limiting XLNet's training through the number of training iterations used and the learning rate. Deploying too many training iterations causes the neural network and its weights to perfectly match the training data whilst deteriorating its out-of-sample prediction accuracy. In other words, too many training iterations causes XLNet to become less generalisable. In contrast, allowing the neural network with too few training iterations can prevent the neural network from fully learning the relationship between the in-sample FOMC statements and price log-differences. The learning rate is essentially the rate at which a neural network updates its parameters to the training data. Too high of a learning rate will result in the network forgetting information from prior training iterations during its current iteration (i.e., the weights of the network are updated too dramatically). This issue results in the network's out-of-sample predictions either degrading rapidly over fewer training iterations or remaining constant because the associated predictions stop updating. Conversely, too low of a learning rate results in XLNet to never fully learning the mapping between FOMC communication and asset price log-differences. Because I am applying the transfer learning approach, the initial values of XLNet's parameters are already optimised to understanding the general English language. Therefore, I can ensure the weighting of XLNet's parameters remain generalisable to new data by limiting how much the parameters can update both within and across training iterations.

Ultimately, the process of training a neural network is a balancing act: I train XLNet to meaningfully approximate the mapping from FOMC statements to price log-differences, but do so without overfitting to the point that the predictions are non-generalisable. I perform this balancing act by following a two-step process with the limitations on the aforementioned hyperparameters in mind. First, I set the maximum number of considered training iterations to be 2040 steps (120 epochs). I then perform what the literature calls a "hyperparameter sweep" on the learning rate. Specifically, Bayesian optimisation is performed on the set of learning rates XLNet uses when fine-tuning up to 2040 training steps.[13] During the sweep, the accuracy metrics mentioned in Subsection 6.1 are tracked in the training process for each considered learning rate. Training XLNet is stopped at the iteration where in-sample accuracy increases, but out-of-sample accuracy decreases. At the end of the sweep duration of 24-48 hours for each fold, the learning rate and training iteration that yields the highest R^2_{OOS} is selected. Because out-of-sample predictions can be volatile during applications of transfer learning on small sample sizes, I select the version of XLNet from the training iteration that exhibits this simultaneity only if future iterations output a permanently worse out-of-sample prediction accuracy. The platform behind the aforementioned fine-tuning process is Simple Transformers from Rajapakse et al. (2024), which simplifies the process of training, experimenting, and performing inference with Transformer model architectures for natural language processing tasks (e.g., Simple Transformers automatically selects the appropriate tokenizer when selecting XLNet).

5.3 When Do Financial Markets Respond to FOMC Statements?

In short, systematic estimation of event window lengths best reflecting market fundamental reactions to FOMC statements can be summarised as the following regression question: Using only the information content contained in the FOMC statements, how much of the null

[13]The set of considered learning rates are $[1e-5, 9e-5]$ and Bayesian optimisation assumes a uniform distribution.

model's squared error loss when predicting the variation in asset prices within a given event window can XLNet explain and attribute to the policy announcements?

Figures 4–17 present the results of this exercise for all maturities of federal funds, Eurodollar, 2-, 5-, and 10-year Treasury futures. The horizontal axis of each figure depicts the end time of the considered event windows (e.g., $t+20$ represents a 30-minute event window starting at 10 minutes before and 20 minutes after FOMC statement publication) and the vertical axis represents the variation of price changes predicted by the neural network to be solely associated with the text of FOMC statements in the testing subsample. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window length with the largest averaged out-of-sample R^2_{OOS}. For each event window size, box-and-whisker plots are shown surrounding the corresponding averages in order to better present the distribution of predictions from the neural network across sample folds.[14]

The cross points show that regardless of futures maturity and asset type, the event window that corresponds with the largest $\overline{R^2_{OOS}}$ is always at least 40 minutes in length, beginning 10 minutes before and ending 30 minutes after FOMC statement publication. In contrast, a 30-minute event window never has such association. This implication is also robust to the act of averaging the out-of-sample R^2_{OOS} of each fold, as similar conclusions are reached when examining other presented metrics of the box-and-whisker plots, such as either the 25^{th} and 75^{th} percentiles: regardless of asset type and maturity, financial markets fundamentally react to the information content of FOMC statements within a 40-minute event window *at minimum*, implying that the popular choice of 30-minute event windows is not enough time.

An interesting observation is that the event window length where XLNet achieves the highest $\overline{R^2_{OOS}}$ increases with the underlying asset maturity of the considered futures. Specifically, the average systematically estimated event window for the front-month and one-month-ahead Federal Funds futures—assets used to measure market expectations about future mon-

[14] Typical standard errors and confidence bands are not calculated because there are too few data points, i.e., the number of sample folds is $k = 5$. Whilst this could be changed such that the neural network performs systematic estimation using leave-one-out cross validation, the computation and financial costs are too great to make this alternative feasible for the current version of this paper.

etary policy from the upcoming FOMC meeting—is 40 minutes in length, whereas the average window length for the two- and three-month-ahead Federal Fund futures is 45 minutes. For Eurodollar futures, which are used to measure expectations about the path of interest rates between two and four quarters into the future, the neural networks predicts markets fundamentally react to the information content of FOMC statements within a 50-minute event window. 50 minutes is also the systematically estimated window length for both 2- and 5-year Treasury futures (when averaging across the first- and second-month contract maturities), but the trend increases for 10-year Treasury futures, which has a systematically estimated length of 60 minutes. The optimal event window length estimated by the neural network for the S&P 500 Index shortens down to 50 minutes. One possible explanation behind this increasing trend overall is that market participants are trading futures with longer underlying asset maturities based on their expectations about interest rates further into the future, which has higher uncertainty as reflected in the significant and larger risk premia (Piazzesi and Swanson, 2008). With regards to equities, it also could be possible that the systematically estimated event window is longer than the conventional 30-minutes because the asset is less directly impacted by monetary policy statements.

Relatedly, the systematic estimation of appropriate event window lengths for most futures contracts by the neural network shows that $\overline{R^2_{OOS}}$ increases with time (e.g., a 30-minute event window has a smaller $\overline{R^2_{OOS}}$ than that of a 40-minute event window). This trend peaks at 30–50 minutes after statement publication, where the measure begins to decline. Overall, the neural network is able to explain an additional 2–17% of the null model's mean squared loss within the systematically estimated event window compared to the conventional 30-minute window. The conceptual framework factors of cognitive noise and unrelated news could be at play here, where the effects of the former decline with time as markets approach fully reacting to the text of FOMC statements, but eventually the effects of the latter dominate the futures market, causing the proportion of the null model's mean squared error explained by XLNet to shrink.

5.4 Comparing Market Responses in Different Event Windows

With the optimal event window length systematically estimated for every asset type and futures maturity, I compare the market price responses measured within these windows to those calculated within the conventional 30-minute window of the literature. Figure 18 depicts this exercise, where sub-figures (a)–(d) are for federal funds futures, (e)–(g) for Eurodollar futures, (h)–(i) for 2-year Treasury futures, and (j)–(k) for 5-year Treasury futures, (l)–(m) for 10-year Treasury futures, and (n) is for the S&P 500 Index. The horizontal (vertical) axis of each sub-figure represents the price log-difference within the systematically estimated (30-minute) event window. The blue dots are the market reactions on scheduled FOMC meetings in my sample. The 45-degree line is depicted as grey.

For a simple ex post characterisation of whether markets for certain asset types and maturities under- or over-react to FOMC statements on release, I regress the price log-differences within the optimal event window lengths on those calculated within a 30-minute window through OLS. If the slope coefficient is less than one and its difference with one is statistically significant for at most $\alpha = 0.05$, then I state that the market for the futures contract and underlying maturity under-reacts to the information content of FOMC statements on release. Otherwise, the price reactions of the market between these two event windows are not different in general. I visually display this in the sub-figures by colouring the regression lines as red for the former scenario and as grey for the latter.

Financial markets for all futures with interest-rate-based underlying assets appear to under-react to the information content of FOMC statements on release, ex post. Furthermore, the slope of the regressions decreases as the underlying maturity of these assets increases. The movement of these markets could potentially be related to the post-FOMC announcement drift documented in Brooks et al. (2023) and Neuhierl and Weber (2024), where the latter specifically finds that the slow adjustment of mutual fund investors their interest rate expectations play a role in the price dynamics after the policy announcements. A notable example of this characterisation can be seen in the 16 December, 2008 FOMC meeting, where the federal funds rate range was cut down to between 0% and 0.25% in response to the Great Recession. In addition to being the first time that the federal funds rate would be below 1%, the decision was also a surprise to financial markets, who were expecting a reduction to only between 0.25% and 0.5%. When combined with how the FOMC statement itself stated that the committee would "employ all available tools to promote the resumption of sustainable economic growth and to preserve price stability", these characteristics possibly contributed to the markets needing more time than typically assumed to fully react to the information describing the weakening U.S. macroeconomy. The top sub-figure of Figure 19 presents the market price reactions for all considered Eurodollar futures maturities on this FOMC meeting date. One can immediately see that as the event window length expands, the change in price becomes increasingly positive compared to the initial reaction within a 20-minute window.

Financial markets for the S&P 500 Index were also found to systematically under-react to the FOMC statement releases, ex-post, corroborating findings from Neuhierl and Weber (2024) on stock return movements in the Center for Research in Security Prices value-adjusted index and from Golez et al. (2025) about belief under-reaction of individual equities, possibly due to individual equity market participants having difficulty fully reacting to the signal content in FOMC announcements. However, it should be noted that the S&P 500 Index was also the market that saw the most price reversals, particularly from negative to positive log-differences as as time progressed after the release of some FOMC statements. One good example can be seen in the S&P 500 Index market response to the 16 September, 2008 FOMC statement, which is depicted in bottom sub-figure of Figure 19. Within 20 minutes after statement release, equity markets responded negatively to the release. Interestingly, the price log-difference reversed signs and continued to grow until market close. A possible explanation behind the reversal could be that the beliefs of equity markets themselves reversed when reading the FOMC's decision to maintain interest rates. In other words, the equity market could only fully react to the relatively positive information behind the FOMC's decision with more time, allowing cognitive noise to die out and the associated negative "knee-jerk" price reactions to reverse. Indeed, Bordalo et al. (2024) documents that over-reactions are a common occurrence in stock markets and will reverse as market beliefs "correct" themselves with time.

Overall, these figures from the systematic estimation and ex post characterisation support the idea that the optimal event window length is not a parameter whose value we should simply assume. Instead, I document how the event window lengths that best reflect market fundamental reactions to FOMC statements varies across asset types and underlying maturities systematically.

6 "One Signal" Approach

Recall that systematic estimation of the event window lengths that best reflect the fundamental reactions of markets is XLNet regressing the changes in asset prices within different windows on the text of FOMC statements. With regard to the terminology from the conceptual framework, systematic estimation could be thought of as a "joint" estimation of the signal and time horizon that best reflects the fundamental reaction of the market to FOMC statements. However, having the neural network perform this "joint" estimation according to best practices from the machine learning literature (e.g., stratified sampling k-fold cross validation for $k = 5$) is computationally intensive. In particular, the financial and mandatory hardware constraints required by the neural network are why the current version of this paper only considers up to 60 minutes after the FOMC statement release for systematic estimation.

An alternative method that potentially circumvents these constraints is what I call the "one signal" approach. If the core assumption is made that the signal (i.e., predictions) from the neural network for the "jointly" estimated event window is *constant* for all time, event window lengths longer than an hour after FOMC statement release can be considered by following the conceptual framework more closely. Specifically, the "one signal" approach uses the out-of-sample predictions of all $k = 5$ folds from the "jointly" estimated event window and calculates the $\overline{R^2_{OOS}}$ for all event window lengths at least as long as the systematically estimated event window length. One advantage of this "one signal" approach is that it allows me to quickly check if event windows beyond the length of the "jointly" estimated one have a larger $\overline{R^2_{OOS}}$.

Figures 20–23 visually display the results of this exercise for maturities of federal funds, Eurodollar, 2-year Treasury, and the S&P 500 Index respectively. The horizontal axis depicts the event window lengths, starting from the systematically estimated event window length and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the highest $\overline{R^2_{OOS}}$. One can see that for most maturities and assets, the event window length with the largest $\overline{R^2_{OOS}}$ is the length systematically estimated by the "joint" approach. In other words, the results from this exercise can be viewed as a robustness check in support of XLNet providing a "good" signal overall for systematic event window length estimation. Although these results may seem trivial, recall that neural networks such as XLNet are non-linear models whose weights and embeddings are sensitive to changes in both the output and input variables of asset price log-differences and FOMC statements, respectively. When combined with the fact that the signal can change with the event window length used to calculate the price log-differences, it is entirely possible that the systematically estimated event window is different from that which has the largest $\overline{R^2_{OOS}}$ under the "one signal" approach.[15] As an example, consider how the "one signal" approach estimate that the optimal event window length for 1-month-ahead federal fund futures, 3-month-ahead federal fund futures, and front-month 2-year Treasury futures should be 1080, 300, and 130 minutes, respectively. To confirm the validity of these estimated window lengths, I calculate $\overline{R^2_{OOS}}$ under the "joint" approach for these futures. The results can be found in Figures 5, 7, and 11.

[15] A rough analogy is how iterative generalised method of moments does not always not converge.

TEXT DESCRIBING THE RESULTS OF THIS ITERATIVE ROBUSTNESS CHECK WILL BE RECORDED HERE WHEN PERFORMED.

6.1 Heterogeneity Analyses

An additional and useful shortcut that the "one signal" approach provides is the ability to observe how the three sources of noise from the conceptual framework—cognitive noise, unrelated news, and white error noise of the signal—affect the event window length conditioned on characteristics of the FOMC statements. Recall that cognitive noise (unrelated news) results in the event window length best reflecting the market's fundamental reaction to FOMC statements to lengthen (shorten), whilst the white error noise of the signal shouldn't have any effect theoretically. Specifically, I use this shortcut method to observe whether the complexity of the information content in FOMC statements, the degree of similarity between sequential statements, and the mention of dissents within the statements changes the optimal event window lengths. I note that these heterogeneity exercises rely on the core assumptions that the signal (i.e., predictions) from the neural network for the "jointly" estimated event window is *constant* for all time *and* this signal is still relevant when conditioning the FOMC statements into subsamples.[16]

Text Complexity: I first observe how the optimal event window length changes for maturities of assets conditioned on the complexity of the FOMC statement text. I assess this complexity using the standard readability formula of the Flesch-Kincaid Grade Level index. These readability formulas determine the grade level required for a reader to properly read and comprehend a provided text based on sentence structure (i.e., average sentence length), word structure (i.e., percentage of difficult words in a sentence), and word phonology (i.e., average number of syllables per word).[17] Figure 26 displays the evolution of the Flesch-Kincaid

[16] The ideal solution would be to acquire large subsamples of FOMC statements satisfying the text complexity and similarity conditions, then have XLNet systematically estimate the optimal event window lengths for all subsamples.

[17] For a description of U.S. education grade levels, see https://en.wikipedia.org/wiki/Education_in_the_United_States.

Grade Level for the sample period of FOMC statements. In particular, we see that the complexity of the FOMC statements under Greenspan's tenure was fairly volatile, frequently swinging between a readability level of some college and a level of 21.3, the equivalent to a doctoral level of education. Under Chair Bernanke, the grade level required to comprehend FOMC statements steadily climbed from slightly above high school to a peak reading comprehension level equivalent to a doctoral degree. The language complexity declined under both Chairs Yellen and Powell and averages between the comprehensive level of a bachelor's or master's degree level of education. Descriptive statistics for the text complexity measure can be found in Table 4.

Given that the median grade level readability of the FOMC statements is some amount of a master's level of education, I split my FOMC statements depending on whether their grade levels are below or above that threshold (i.e., 16.5). I then calculate the MSEs under the "one signal" approach for each subsample of FOMC statements for every maturity of every futures contract. A summary of this exercise is found in the first two columns of Table 5, which displays the minimised MSEs and associated event window lengths averaged across all asset types and maturities. Immediately, one can see that FOMC statements with a required readability level of at least some of a master's degree have a longer average optimal event window of 72 minutes. In contrast, "simpler" statements with a grade level up to some of a master's degree is associated with a shorter average event window length of 58 minutes. Interestingly, both subsamples of FOMC statements possess a similar MSEs on average. One potential explanation for the differences in optimal event window lengths but similar MSEs is that the "simpler" statements exhibit a relatively larger share of unrelated news, whilst most of the noise in the "complicated" statements is coming from cognitive noise. Indeed, papers such as Smales and Apergis (2017) have shown that the increasing complexity of FOMC statement language over time is associated with increasing volatility across equity, bond, and currency markets. This association is consistent with the idea that more complex information results in differing interpretations and reactions to said information, therefore increasing market volatility and trading volume. In other words, financial markets might need more time to fully react to FOMC statements with more complicated language.

Text Similarity: Using the "one signal" approach, I also analyse how the optimal event window length changes when conditioned on the degree of similarity between the FOMC statements over time. The method I use for this exercise is a bag-of-words model, which represents text as an unordered, weighted-frequency of words. Importantly, this method assumes that documents with common weighted frequencies for words are discussing similar topics. By transforming the FOMC statements into vectors of word frequencies, I can perform vector analysis and create a measure of similarity between any two document vectors.

The weighted frequency of words frequently used in text analysis is called Term Frequency-Inverse Document Frequency (TFIDF). Conversion of my sample of FOMC statements into the TFIDF matrix is done through the Python text analysis library of sklearn. Specifically, each word in a document is multiplied by the number of times it occurs within a particular document over the frequency of documents where the word appears. The purpose behind this ratio is to properly weigh down the importance of frequently occurring words that don't actually provide much meaning with regard to distinguishing FOMC statements from one another. For example, words such as "a" and "the" appear in all documents with high frequency, but don't signal anything about any particular statement. Therefore, the value for these words are weighed down by the high number of occurrences across documents, yielding a *weighted frequency.* Additionally, words like "unemployment", "inflation", and "federal funds rate" will also be weighed down because monetary policy discussed in every FOMC statements has these components. However, it should be noted that these words are weighed down by a smaller degree relatively speaking because they could still convey some importance when discerning the degree of similarity between two statements (e.g., if one statement discusses about inflation relatively more than another statement that focuses more on labour market conditions). Lastly, words that do not appear in every FOMC statement, such as "persists" or "tight", are given a higher weighted frequency because they potentially signal a unique economic environment when the FOMC meeting took place compared to other meeting dates. The differences in these unique and informative words are what allow me to calculate the degree of similarity between FOMC statements.

The calculation of the TFIDF matrix is as follows: Let D be the set of FOMC statements and T be the set of all terms that appear within and across all statements in D. Let a single document and term be indexed by $d \in D, t \in T$, respectively. Therefore, the term frequency of a particular term t in document d can be defined as:

$$tf_{d,t} = \ln\left(\frac{tc_{d,t}}{nt_d}\right) + 1, \tag{9}$$

where $tc_{t,d}$ is the number of times term t appears in document d, nt_d is the total number of words in d, and both the natural logarithm and addition term serve to smooth out the frequency measure. We can then calculate the inverse document frequency as follows:

$$idf_{d,t} = \ln\left(\frac{nd}{df_{d,t} + 1}\right) + 1, \tag{10}$$

where nd is the number of documents in set D and $df_{t,d}$ is the number of documents that term t appears. Multiplying the two terms gives us the weighted frequency of each term t in document d:

$$TFIDF_{d,t} = tf_{d,t} * idf_{d,t}. \tag{11}$$

Calculating $TFIDF_{d,t} \forall d \in D, t \in T$ and combining them yields a $D \times T$ matrix, where each element $TFIDF_{d,t}$ represents the number of times a term appears in a particular document, divided by the number of words in said document, then multiplied with the document frequency (ratio between the number of documents where the term appears and the total number of documents). Essentially, the higher the TFIDF value is for a term, the more informative that word is with regard to distinguishing the information content of those FOMC statements from others.

Although the TFIDF matrix can be produced on the current version of the sample of FOMC statements, the results might not be accurate in representing the similarity of the intended or comprehended information content of the statements. For example, the method could potentially view the terms "Federal Funds Rate" and "federal funds rate" as different. In other words, this bag-of-words model is not as "sophisticated" at discerning context from the FOMC statements in their original grammatical structure compared to the neural network, which could cause the $TFIDF_{d,t}$ values to be inaccurate. To combat this issue, I employ additional preprocessing steps on my sample of FOMC statements. First, I make every word lowercase. Second, I remove common words that convey little semantic meaning, such as articles, pronouns, and conjunctions.[18] The last step involves converting all words into their "base" form (e.g., the words "increases", "increasing", and "increase" are all combined into the base form of "increas"). Calculating the TFIDF matrix on the sample of additionally cleaned FOMC statements from May 1999 through October 2019 yields a matrix with 165 rows (statements) and 966 columns (terms). Table 6 presents a list of 30 base terms that have the highest TFIDF scores for my sample of FOMC statements. Recall that terms with higher TFIDF values are more informative terms with regard to differentiating the FOMC statements with them from statements without.

With the TFIDF matrix calculated, I am able to produce a corresponding matrix whose values represent a measure of similarity between any two pairs of FOMC statements. Specifically, this document similarity matrix is created by multiplying the TFIDF matrix with its transpose:

$$\text{Document Similarity Matrix} = TFIDF \cdot TFIDF^T. \tag{12}$$

To understand the document similarity matrix, consider a document d row vector of the TFIDF matrix. For each term $t \in T$, the $TFIDF_{d,t}$ values are only positive if the terms appear in document d, otherwise the value is zero. Additionally, the vectors for every doc-

[18] I choose to retain dates and numbers because the FOMC's targetted federal funds rate or discussions about date-based forward guidance convey important semantic content.

ument all have the same magnitude due to normalisation by the number of documents and terms in the matrix. Because the product between the TFIDF matrix and its transpose is essentially the dot product between every pair of document row vectors in the sample of FOMC statements. As a result, we can define the degree of similarity between any two FOMC statements as the cosine of the angle between these documents, also called the *cosine similarity*:

$$S^{A,B} := \cos\theta = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}, \tag{13}$$

where $\mathbf{A}, \mathbf{B}$ are the vectors for FOMC statements A and B, respectively. The more similar two FOMC statements are, the angle between the two corresponding vectors approaches zero, resulting in the cosine of the angle to go to one. A cosine similarity value of one between two FOMC statements means they are identical, whilst a value of zero means the two documents have no common words. In other words, the more base terms shared between two FOMC statements, the higher cosine similarity value is assigned to the pair.

The Document Similarity Matrix for FOMC statements from May 1999 through October 2019 is represented as a heat map in Figure 27. The rows and columns are FOMC statements ordered by release statement, top-to-bottom and left-to-right, respectively. For any FOMC statement in row d and another statement in column d', the element (d, d') represents the cosine similarity measure between the pair of statements. Elements coded as darker (lighter) shades of blue (green) represent higher (lower) similarity values. The main diagonal of the matrix is a diagonal of ones (i.e., the darkest shade of blue) because the cosine similarity measure is calculated between an FOMC statement with itself. Moving away from the main diagonal represents a comparison between FOMC statements with other statements over time. An interesting pattern seen throughout the heat map is that FOMC statements have become increasingly similar over time, represented by the increasing area of shades of blue when moving from the top-left to the bottom-right corners. The increased persistence and standardisation of FOMC statement wording begins roughly between early-2009 and early-2010, which coincides with the period when the FOMC was in the process of reducing the federal funds rate towards the zero lower bound. As discussed in Handlan (2022a), one explanation behind this occurrence is that the FOMC wanted to avoid market surprises through reducing the variation in statement vocabulary after the financial crisis. Even when the zero lower bound ended with the Board raising the federal funds rate at the end of 2015, the cosine similarity measures remained high through 2019 (represented by the dark blue shadings), possibly due to the efforts of the FOMC to increase the transparency of its monetary policy under Chairs Yellen and Powell.

The series of document-pairs $(d, d-1)$ represents the degree of similarity between sequential FOMC statements, which I call as S^1 for the rest of the paper and is plotted in Figure 28. Similar to the colour trend observed in the heat map, as time progresses, the cosine similarity between sequential FOMC statements increases. Descriptive statistics for S^1 can be found in Table 4.

I split my FOMC statements depending on whether S^1 value is below or above the median (i.e., 0.88), labelling the subsample of statements as being "different" or "similar" relative to their immediately previous document, respectively. Calculating the MSEs for each subsample of FOMC statements for every maturity of every asset, I present the results of this exercise in the third and fourth columns from the left of Table 5. "Different" FOMC statements observe the smallest MSEs within an event window length of 61 minutes when averaged across all asset types and maturities, whilst "similar" statements see minimised MSEs within a 50-minute window on average. Similar to the condition of text complexity, both subsets of statements have similar minimised MSEs on average. These results with respect to cosine similarity could be due to the presence of unrelated news and cognitive noise: the noise affecting financial market reactions to "simpler" statements is comprised relatively more of unrelated news, shortening the amount of time required to fundamentally react to the information context. In contrast, cognitive noise is the larger contributor when reacting to "different" statements. Intuitively, if an FOMC statement contains less similar base terms and associated ideas with its predecessor, financial markets could require more time in order to discern and react to the new information of the current FOMC statement. Conversely, markets would presumably react to statements that share a lot of terms and concepts with past policy announcements in a short amount of time because most of said information has already been incorporated into the price.

Presence of Dissents: The final characteristic of the FOMC statements I consider under the "one signal" approach is whether the presence of dissents in the monetary policy announcements affects the optimal event window lengths of asset types. Roughly 40% of FOMC statements in my sample have recorded dissents, a share similar to that found in prior studies. While this suggests dissents are not rare, they still provide important signals. According to Federal Reserve tradition, dissents are usually recorded only when Board members find the majority's opinion unacceptable (C Madeira and J Madeira, 2019). In other words, the very existence of a recorded dissent provides additional information for markets to process, which in turn could affect their reaction time. To investigate this characteristic, I split my sample of FOMC statements depending on whether the policy votes were unanimous or not.[19] The fifth and sixth columns of Table 5 present the calculated MSEs for each subset of statements. On average, statements with(out) dissents observe the smallest MSE under the "one signal" approach within an event window length of 61 (51) minutes. The longer reaction time for statements with dissents may be because they introduce additional layers of information regarding policy transparency and internal disagreement, requiring more time for markets to process compared to the unanimous announcements.

7 Baseline Application: Monetary Policy Surprises

SoonTM.

[19] I do not record and distinguish between dissent votes for tighter policy, easier, policy, or other reasons.

8 Conclusion

SoonTM.

References

Acosta, Miguel (2023). "The Perceived Causes of Monetary Policy Surprises". Published manuscript.

Acosta, Miguel, Connor M. Brennan, and Margaret M. Jacobson (2024). *Constructing high-frequency monetary policy surprises from SOFR futures*. Tech. rep. Finance and Economics Discussion Series 2024-034. Board of Governors of the Federal Reserve System.

Anderson-Sprecher, Richard (1994). "Model Comparisons and R2". In: *The American Statistician* 48.2, pp. 113–117.

Athey, Susan and Guido Imbens (2019). *Machine Learning Methods Economists Should Know About.*

Bauer, Michael D., Aeimit Lakdawala, and Philippe Mueller (2022). "Market-Based Monetary Policy Uncertainty". In: *The Economic Journal* 132.644, pp. 1290–1308.

Bauer, Michael D. and Eric T. Swanson (2023). "A Reassessment of Monetary Policy Surprises and High-Frequency Identification". In: *NBER Macroeconomics Annual* 37.1, pp. 87–155.

Bazzana, Flavio and Andrea Collini (2020). "How does HFT activity impact market volatility and the bid-ask spread after an exogenous shock? An empirical analysis on S&P 500 ETF". In: *The North American Journal of Economics and Finance* 54, p. 101240.

Ben Ammar, Imen and Slaheddine Hellara (2022). "High-frequency trading, stock volatility, and intraday crashes". In: *The Quarterly Review of Economics and Finance* 84, pp. 337–344.

Ben-David, Itzhak, Francesco Franzoni, Byungwook Kim, and Rabih Moussawi (2022). "Competition for Attention in the ETF Space". In: *The Review of Financial Studies* 36.3, pp. 987–1042.

Bianchi, Francesco, Sydney C Ludvigson, and Sai Ma (2024). *What Hundreds of Economic News Events Say About Belief Overreaction in the Stock Market*. Working Paper 32301. National Bureau of Economic Research.

Boehm, Christoph E and T. Niklas Kroner (2023). *The US, Economic News, and the Global Financial Cycle*. Working Paper 30994. National Bureau of Economic Research.

Boguth, Oliver, Adlai J. Fisher, Vincent Gregoire, and Charles Martineau (2023). "Noisy FOMC Returns? Information, Price Pressure, and Post-Announcement Reversals". SSRN Working Paper No 878091.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer (2024). "Belief Overreaction and Stock Market Puzzles". In: *Journal of Political Economy* 132.5, pp. 1450–1484.

Brooks, Jordan, Michael Katz, and Hanno Lustig (2023). *Post-FOMC Announcement Drifts in U.S. Bond Markets*. Working Paper 25127. National Bureau of Economic Research.

Chang, Son J and Son-Nan Chen (1989). "Stock-price adjustment to earnings and dividend surprises". In: *Quarterly Review of Economics and Business* 29.1, pp. 68–81.

Das, Somnath and Alexander Z. King (2021). "Measuring the informativeness of earnings announcements: The role of event windows". In: *The Quarterly Review of Economics and Finance* 82, pp. 350–367.

Doh, Taeyoung, Dongho Song, and Shu-Kuei Yang (2023). "Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements". Federal Reserve Bank of Kansa City Working Paper.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). "Text as Data". In: *Journal of Economic Literature* 57.3, pp. 535–74.

Gertler, Mark and Peter Karadi (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity". In: *American Economic Journal: Macroeconomics* 1, pp. 44–76.

Glasserman, Paul and Caden Lin (2023). *Assessing Look-Ahead Bias in Stock Return Predictions Generated By GPT Sentiment Analysis.*

Golez, Benjamin, Peter Kelly, and Ben Matthies (2025). "FOMC News and Segmented Markets". In: *Journal of Accounting and Economics* 79.2, p. 101767.

Graham, Benjamin (1974). *The Renaissance of Value: The Proceedings of a Seminar on the Economy, Interest Rates, Portfolio Management, and Bonds Vs Common Stocks, September 18, 1974.* Financial Analysts Research Foundation.

Gürkaynak, Refet S., Burçin Kisacikoğlu, and Jonathan H. Wright (2020). "Missing Events in Event Studies: Identifying the Effects of Partially Measured News Surprises". In: *American Economic Review* 110.12, pp. 3871–3912.

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". In: *International Journal of Central Banking* 1.1.

Handlan, Amy (2022a). "FedSpeak Matters: Statement Similarity and Monetary Policy Expectations". Published manuscript.

Handlan, Amy (2022b). "Text Shocks and Monetary Surprises: Text Analysis of FOMC Statements with Machine Learning". Published manuscript.

Hawinkel, Stijn, Willem Waegeman, and Steven Maere (2024). "Out-of-Sample R2: Estimation and Inference". In: *The American Statistician* 78.1, pp. 15–25.

Hervé, Fabrice, Mohamed Zouaoui, and Bertrand Belvaux (2019). "Noise traders and smart money: Evidence from online searches". In: *Economic Modelling* 83, pp. 141–149.

Hillmer, S.C. and P.L. Yu (1979). "The market speed of adjustment to new information". In: *Journal of Financial Economics* 7.4, pp. 321–345.

Husted, Lucas, John Rogers, and Bo Sun (2020). "Monetary policy uncertainty". In: *Journal of Monetary Economics* 115, pp. 20–36.

Krivin, Dmitry, Robert Patton, Erica Rose, and David Tabak (2003). "Determination of the Appropriate Event Window Length in Individual Stock Event Studies". SSRN Working Paper No 466161.

Kuttner, Kenneth N. (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". In: *Journal of Monetary Economics* 47.3, pp. 523–544.

Madeira, Carlos and João Madeira (2019). "The Effect of FOMC Votes on Financial Markets". In: *The Review of Economics and Statistics* 101.5, pp. 921–932.

Miranda-Agrippino, Silvia and Giovanni Ricco (2021). "The Transmission of Monetary Policy Shocks". In: *American Economic Journal: Macroeconomics* 13.3, pp. 74–107.

Nakamura, Emi and Jón Steinsson (2018). "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect". In: *The Quarterly Journal of Economics* 133.3, pp. 1283–1330.

Neuhierl, Andreas and Michael Weber (2024). *Monetary Momentum.* Chicago Booth Research Paper, Fama-Miller Working Paper, University of Chicago, Becker Friedman Institute for Economics Working Paper, LawFin Working Paper 20-05, 2020-39, 37. Chicago Booth Research Paper Series, Chicago Booth: Fama-Miller Working Paper Series, Becker Friedman Institute for Economics Working Paper Series, DFG Center for Advanced Studies - Foundations of Law and Finance.

Piazzesi, Monika and Eric T. Swanson (2008). "Futures prices as risk-adjusted forecasts of monetary policy". In: *Journal of Monetary Economics* 55.4, pp. 677–691.

Rajapakse, Thilina C., Andrew Yates, and Maarten de Rijke (2024). "Simple Transformers: Open-source for All". In: *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region.* SIGIR-AP 2024. Tokyo, Japan, pp. 209–215.

Ramey, V.A. (2016). "Chapter 2 - Macroeconomic Shocks and Their Propagation". In: *Handbook of Macroeconomics.* Ed. by John B. Taylor and Harald Uhlig. Vol. 2. Elsevier, pp. 71–162.

Rigobon, Roberto and Brian Sack (2004). "The Impact of Monetary Policy on Asset Prices". In: *Journal of Monetary Economics* 51.8, pp. 1553–1575.

Sarkar, Suproteem and Keyon Vafa (2024). "Lookahead Bias in Pretrained Language Models". SSRN Working Paper No 4754678.

Smales, L.A. and N. Apergis (2017). "Does more complex language in FOMC decisions impact financial markets?" In: *Journal of International Financial Markets, Institutions and Money* 51, pp. 171–189.

Swanson, Eric T. and Vishuddhi Jayawickrema (2023). "Speeches by the Fed Chair Are More Important than FOMC Announcenments: An Improved High-Frequency Measure of U.S. Monetary Policy Shocks". Unpublished manuscript.

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le (2019). "XLNet: Generalized Autoregressive Pretraining for Language Understanding". In: *CoRR* abs/1906.08237.

Figures

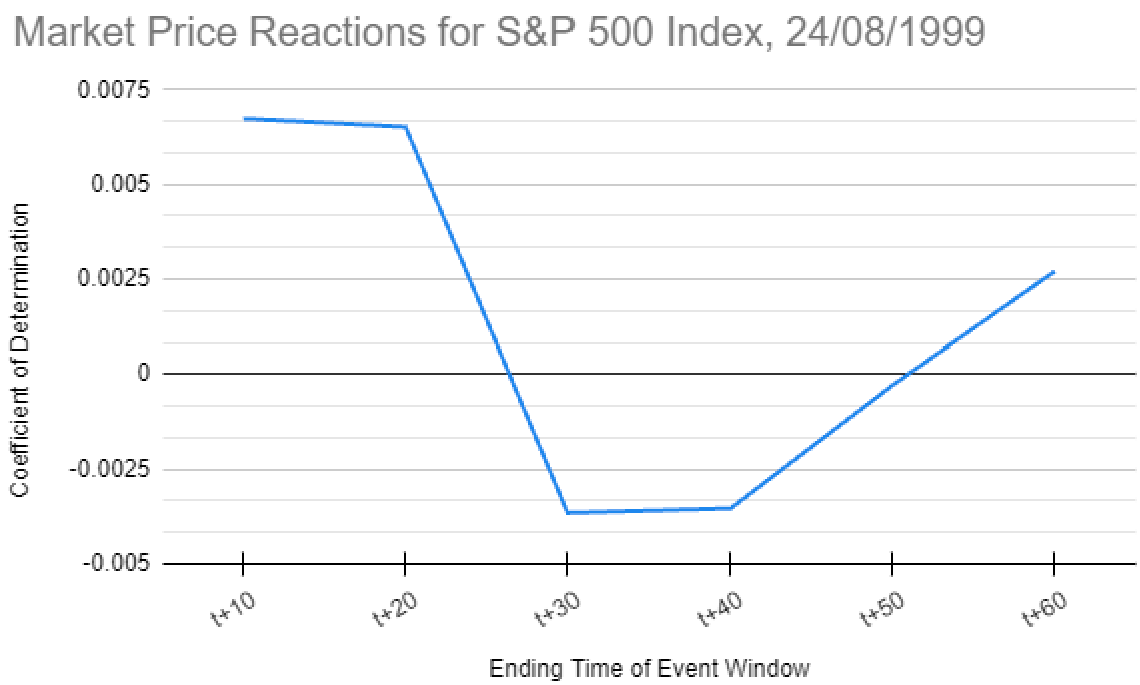


Figure 1: Responses of S&P 500 Index Market to FOMC Statements

The price log-differences of the S&P 500 Index market for different event window lengths after the release of the 24 August, 1999 FOMC statement is depicted. The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the price-log differences.

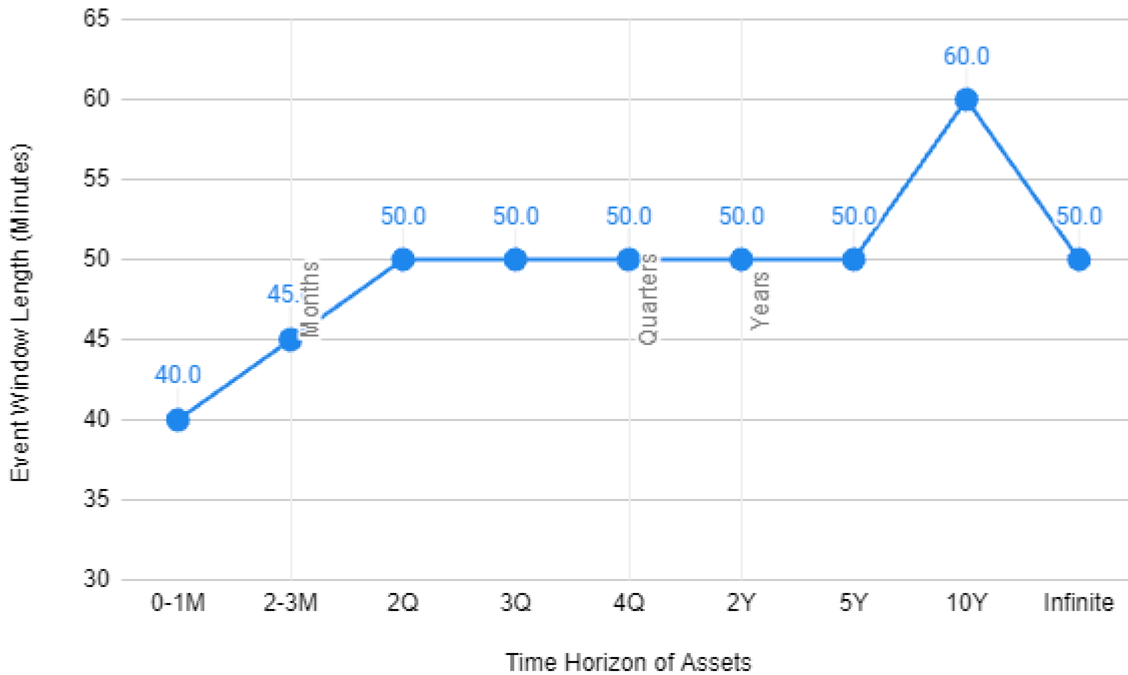


Figure 2: Systematically Estimated Event Window Lengths for Various Asset Maturities

The horizontal axis depicts the horizons of all considered futures contracts and equities. The vertical axis depicts the systematically estimated event window length in minutes, with all windows starting at 10 minutes before FOMC statement release. "0–1M" is the average of the event window lengths for front-month and 1-month-ahead Federal Funds futures. "2–3M" is the average of the event window lengths for 2-month and 3-month-ahead Federal Funds futures. "Infinite" is the maturity of the S&P 500 Index and considered E-mini futures contracts.

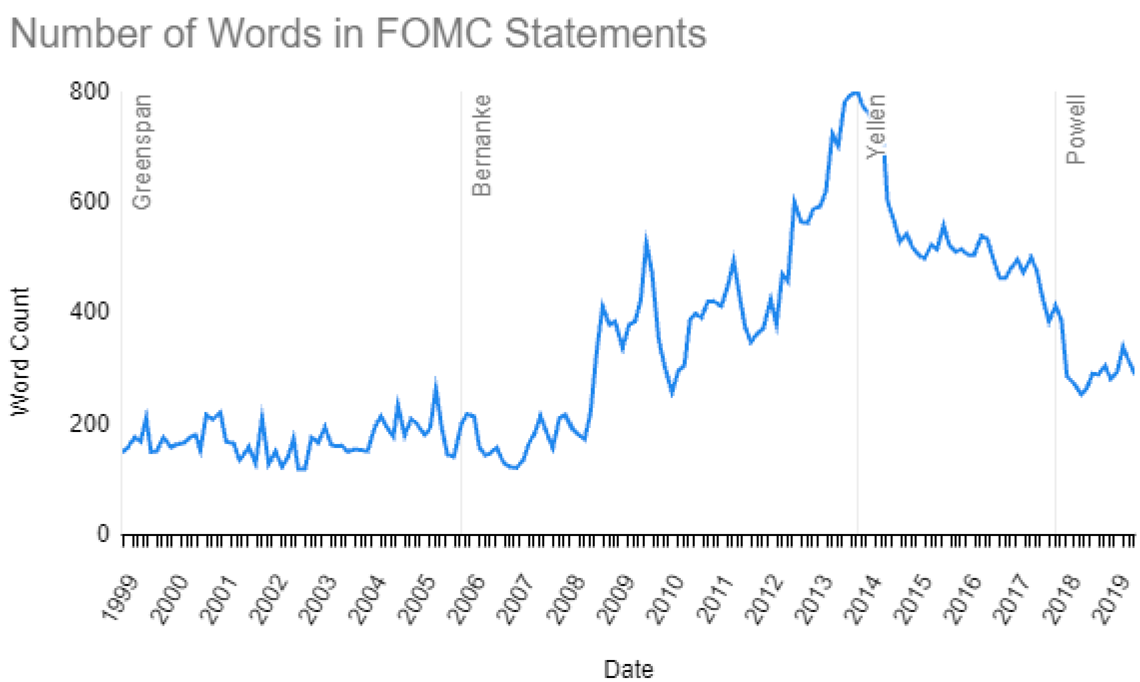


Figure 3: Number of Words in FOMC Statements, May 1999–October 2019

The above counts are for FOMC statements that have underwent pre-processing, which is explained in Subsection 4.1. From left to right, the vertical grey lines indicate the first FOMC meeting with Greenspan, Bernanke, Yellen, and Powell as Fed Chair.

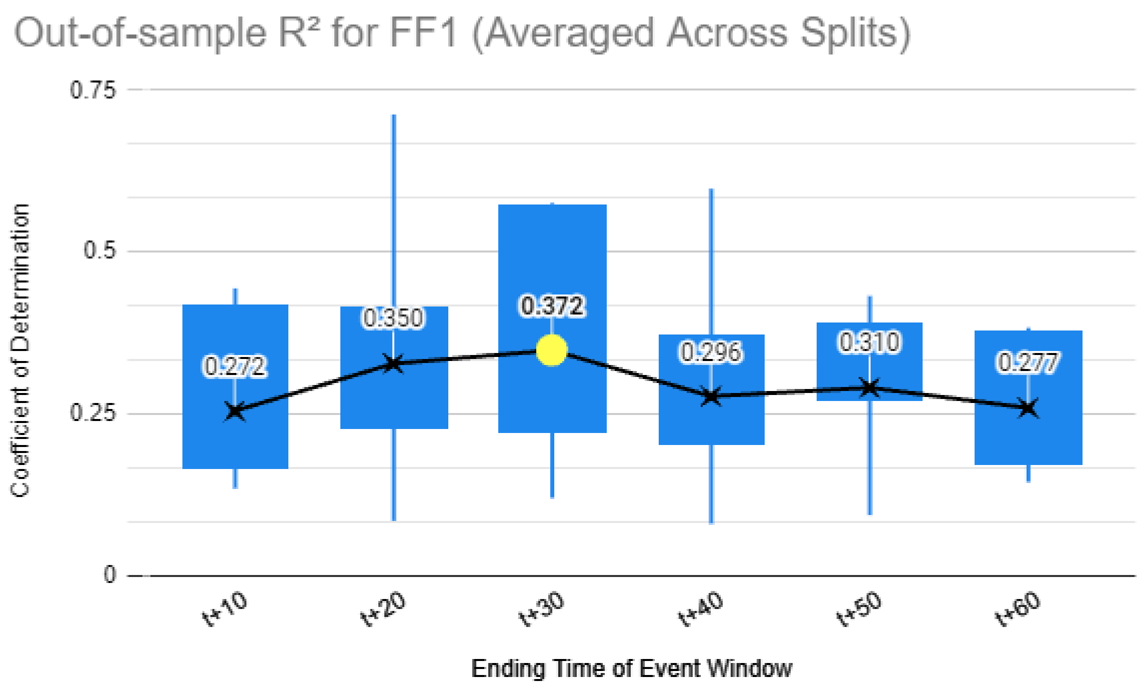


Figure 4: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $FF1$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

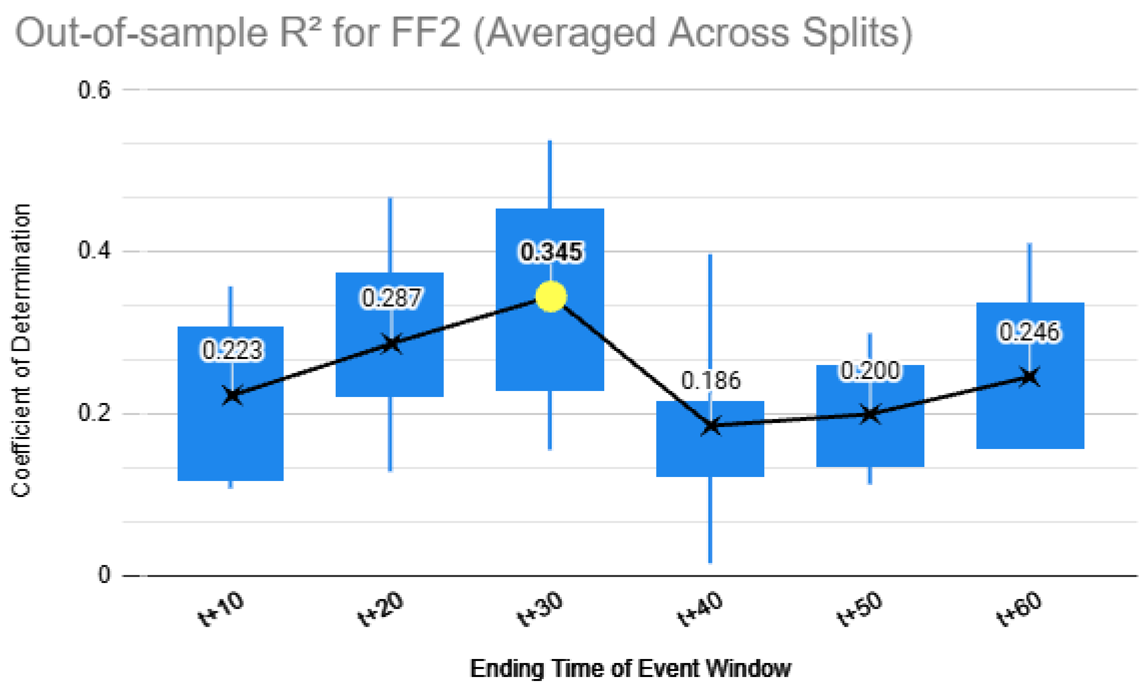


Figure 5: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $FF2$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

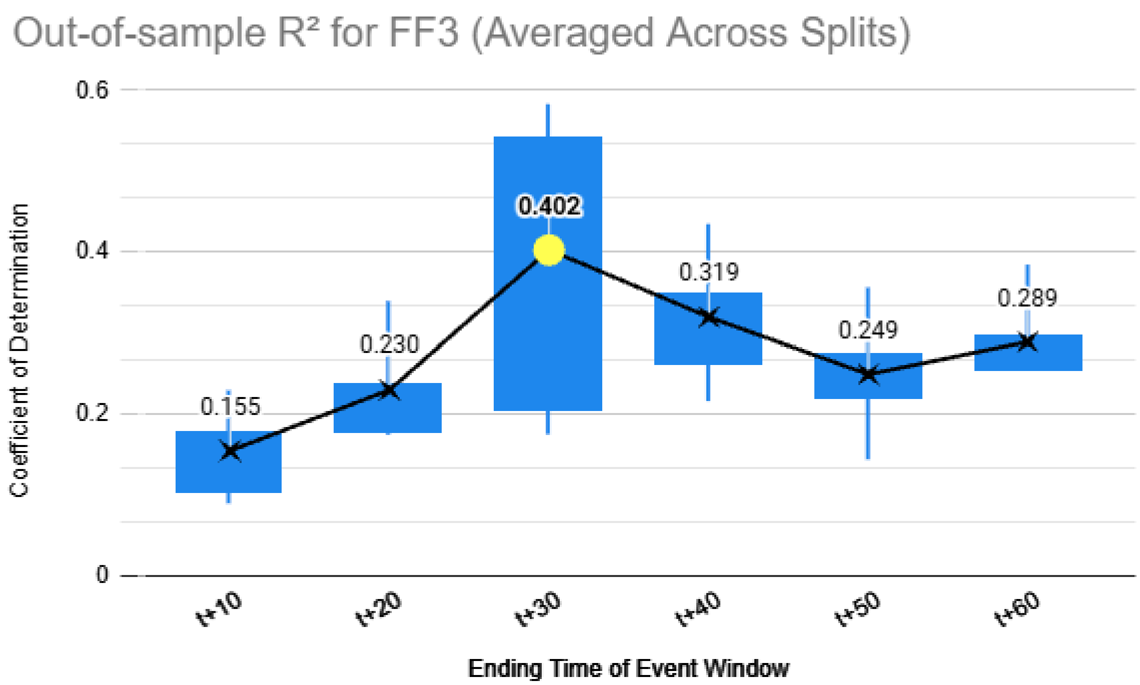


Figure 6: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $FF3$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent t$\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

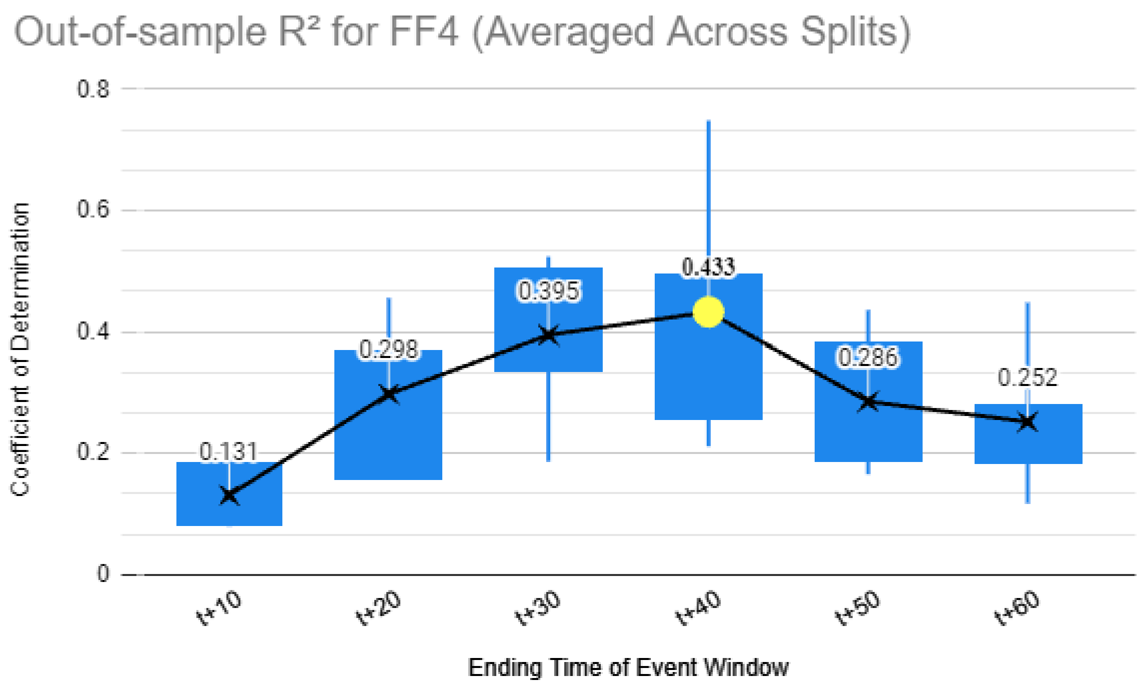


Figure 7: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $FF4$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

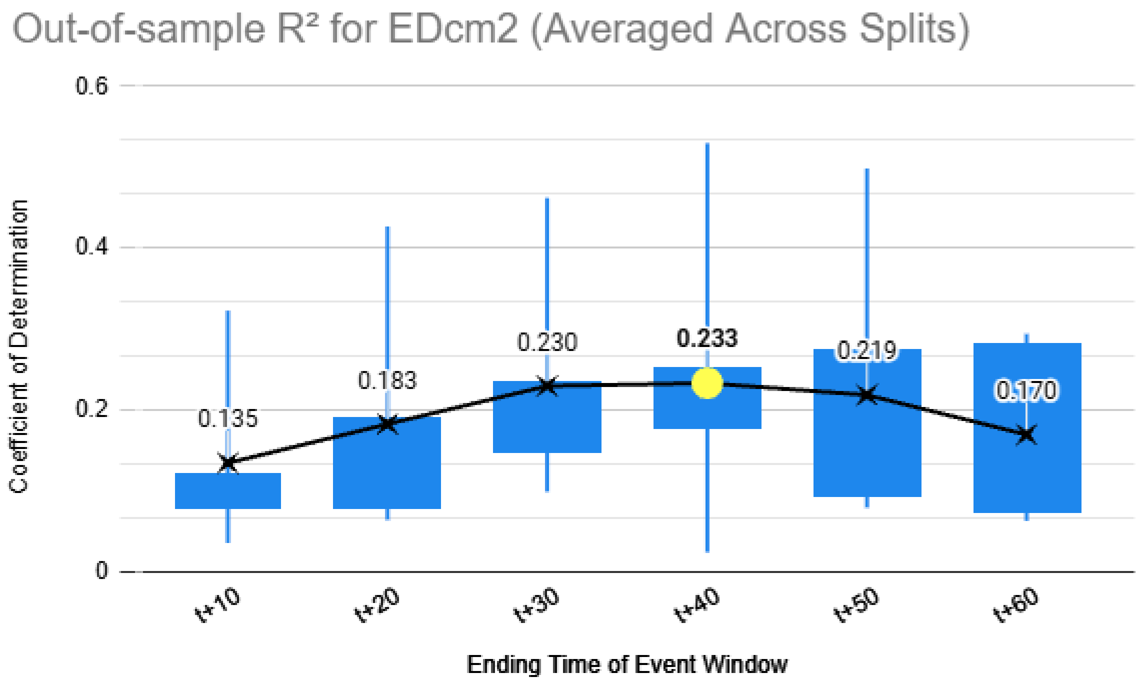


Figure 8: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $EDcm2$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

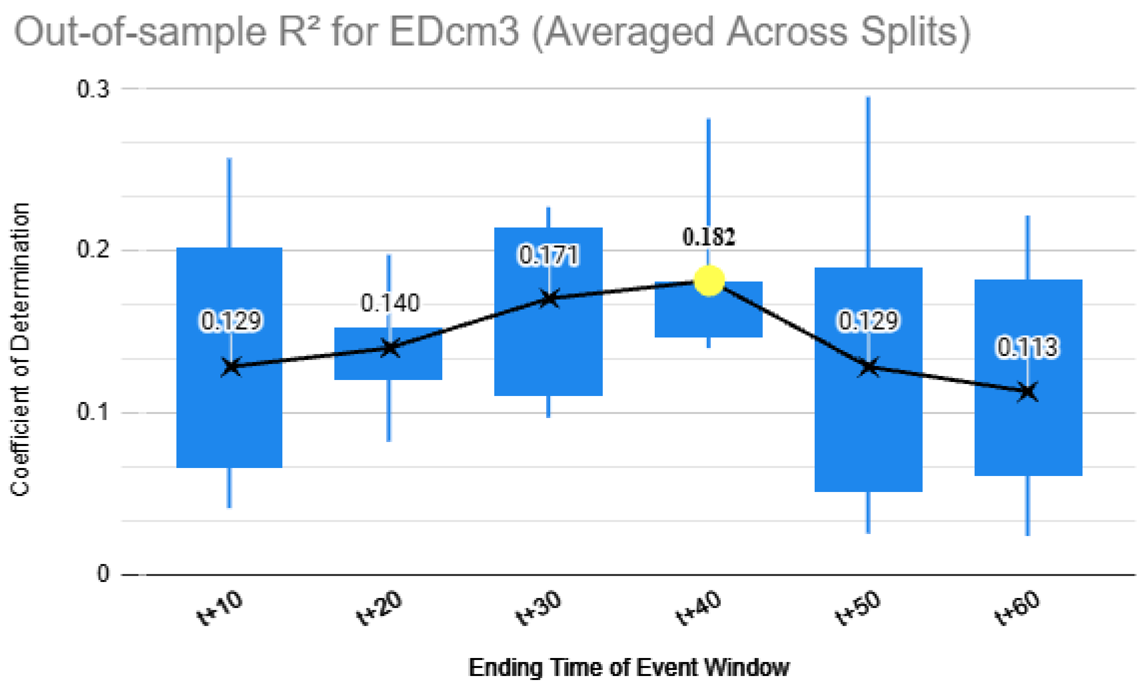


Figure 9: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $EDcm3$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

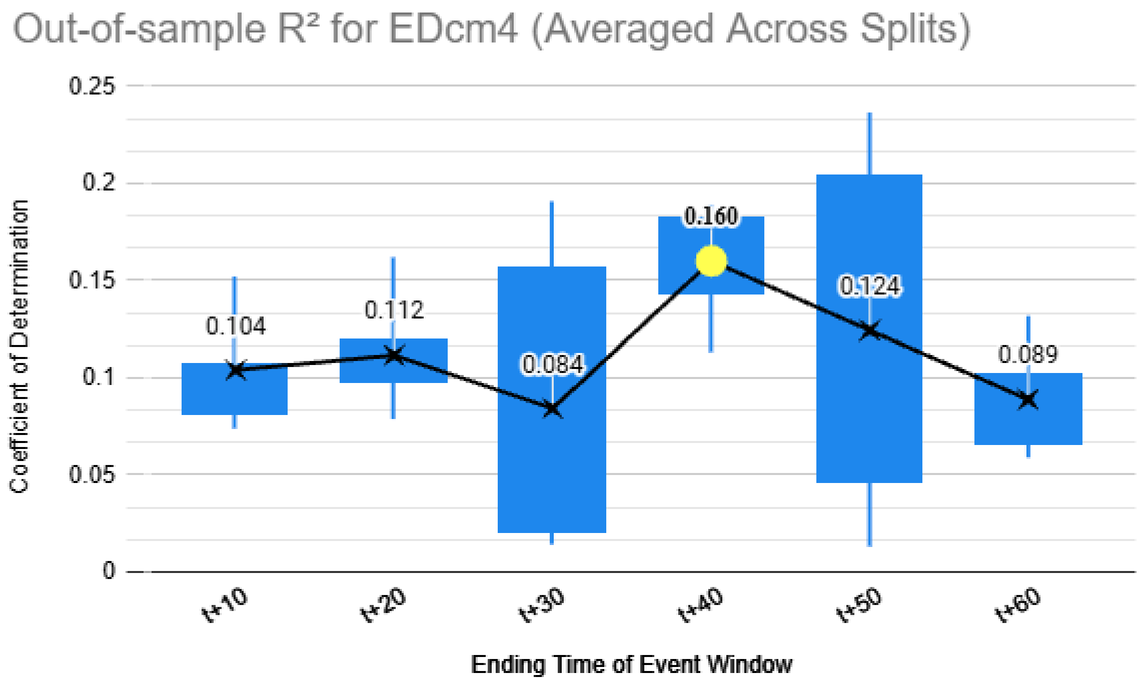


Figure 10: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $EDcm4$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

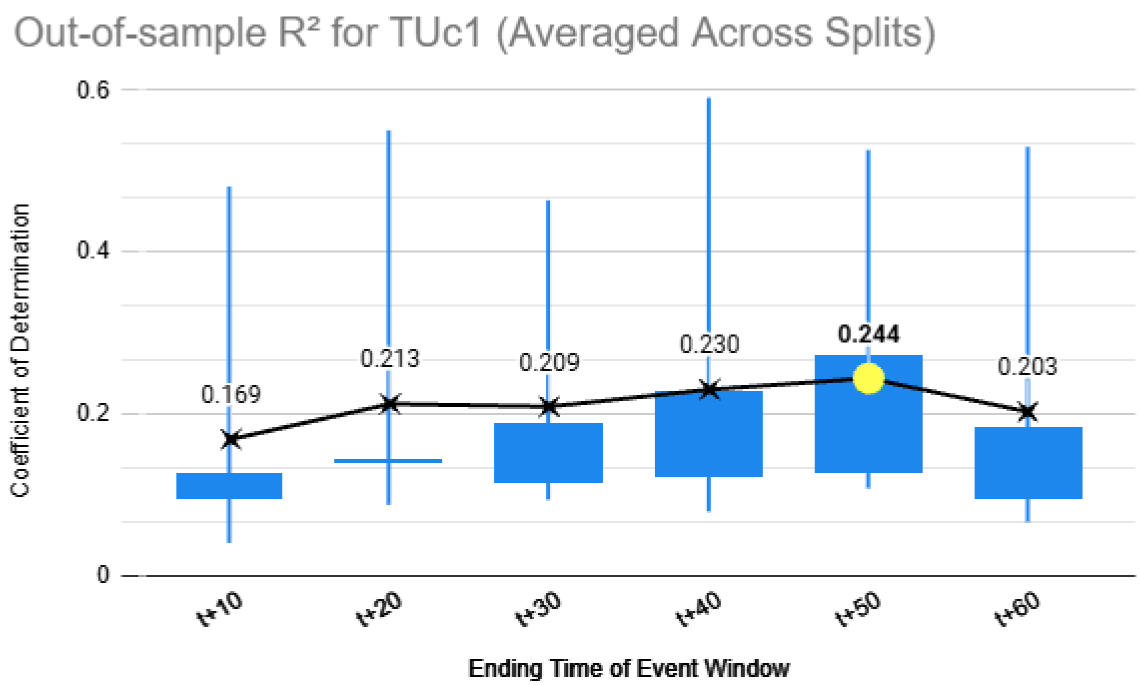


Figure 11: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $TUc1$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

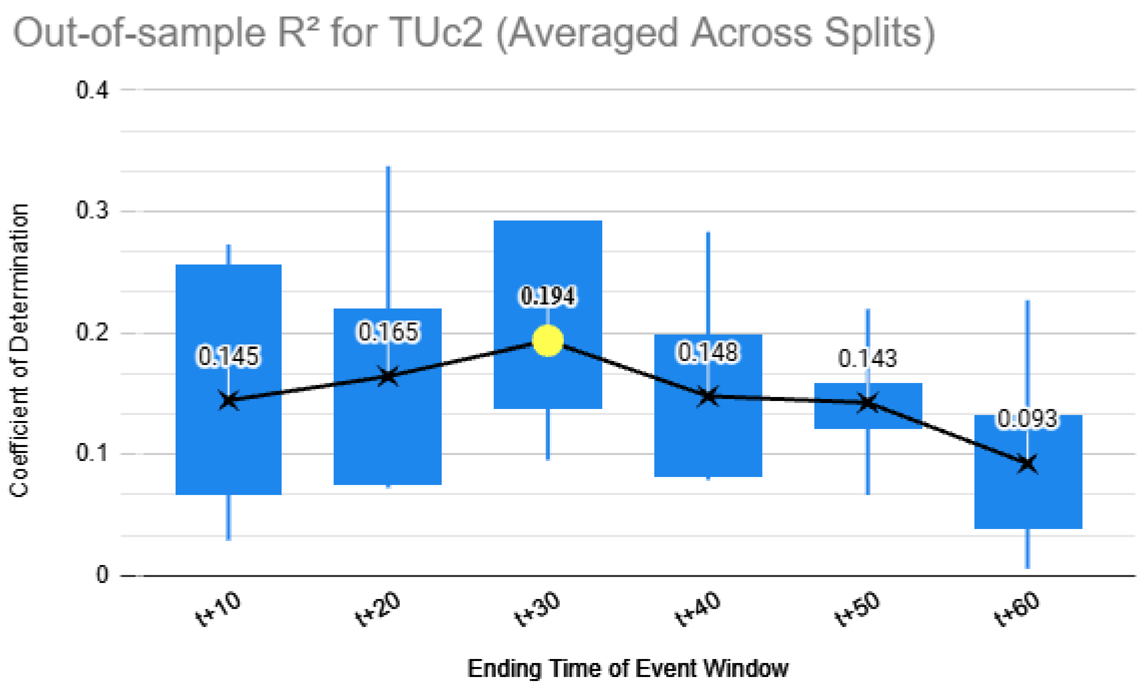


Figure 12: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $TUc2$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

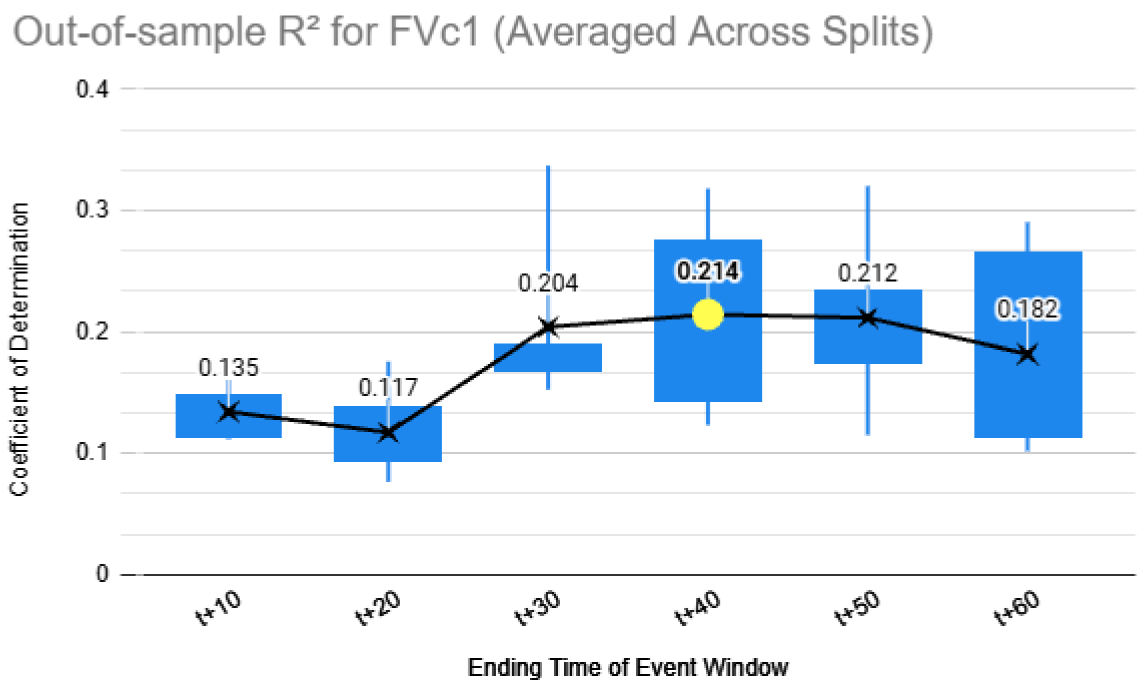


Figure 13: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $FVc1$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

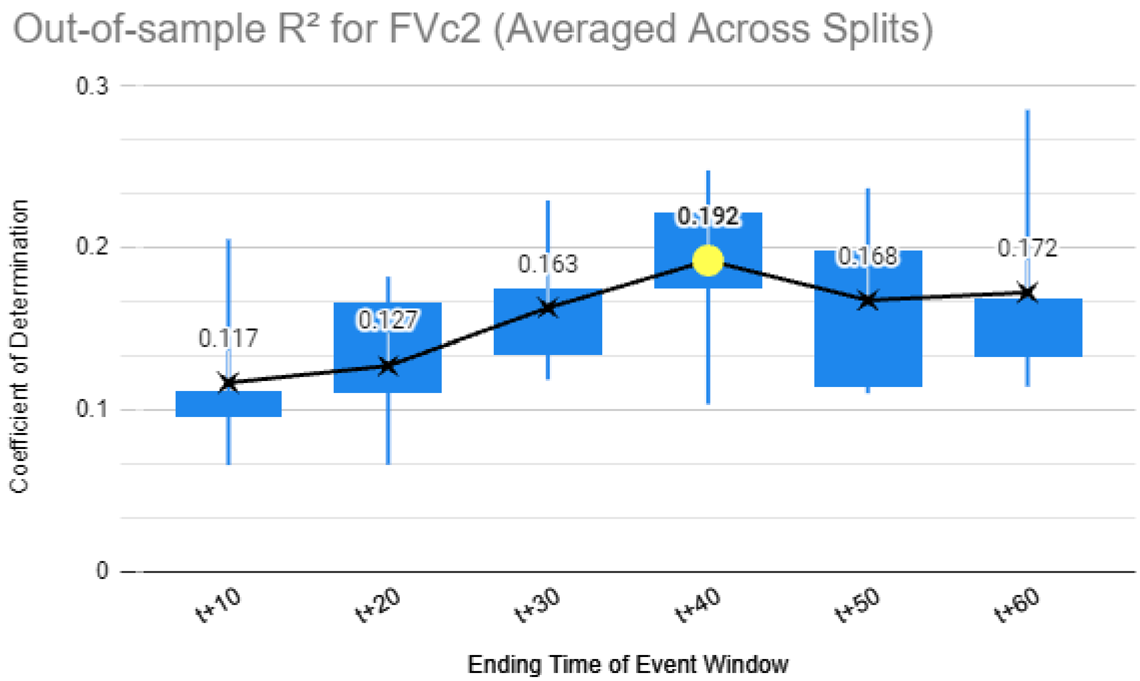


Figure 14: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $FVc2$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

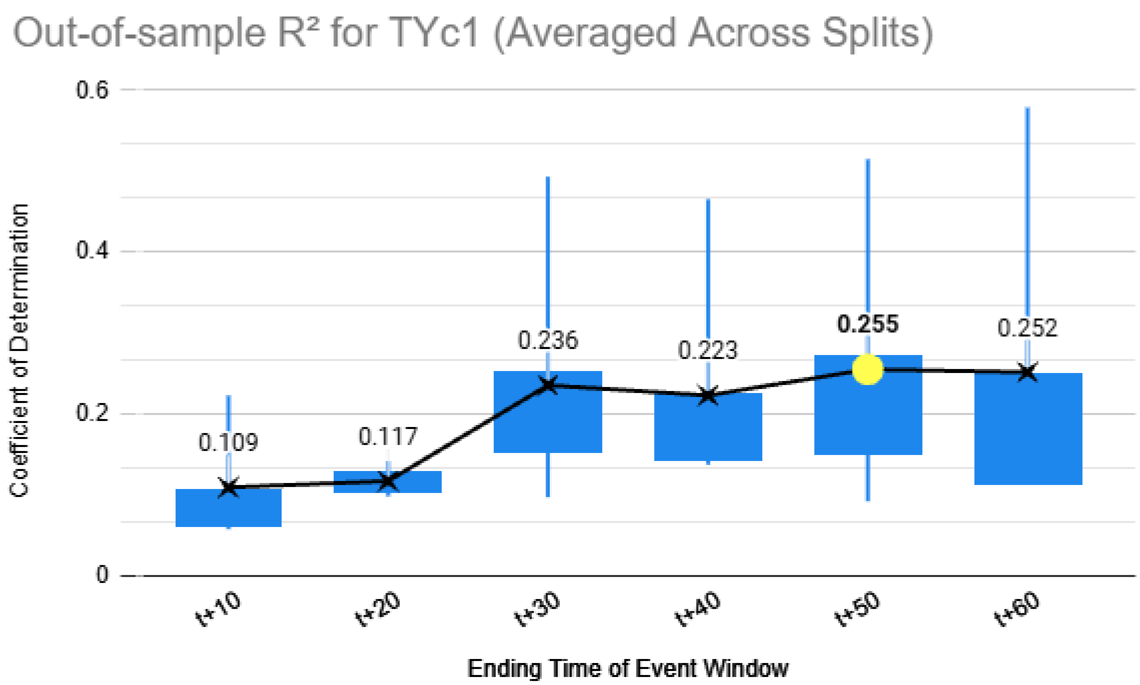


Figure 15: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $TYc1$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

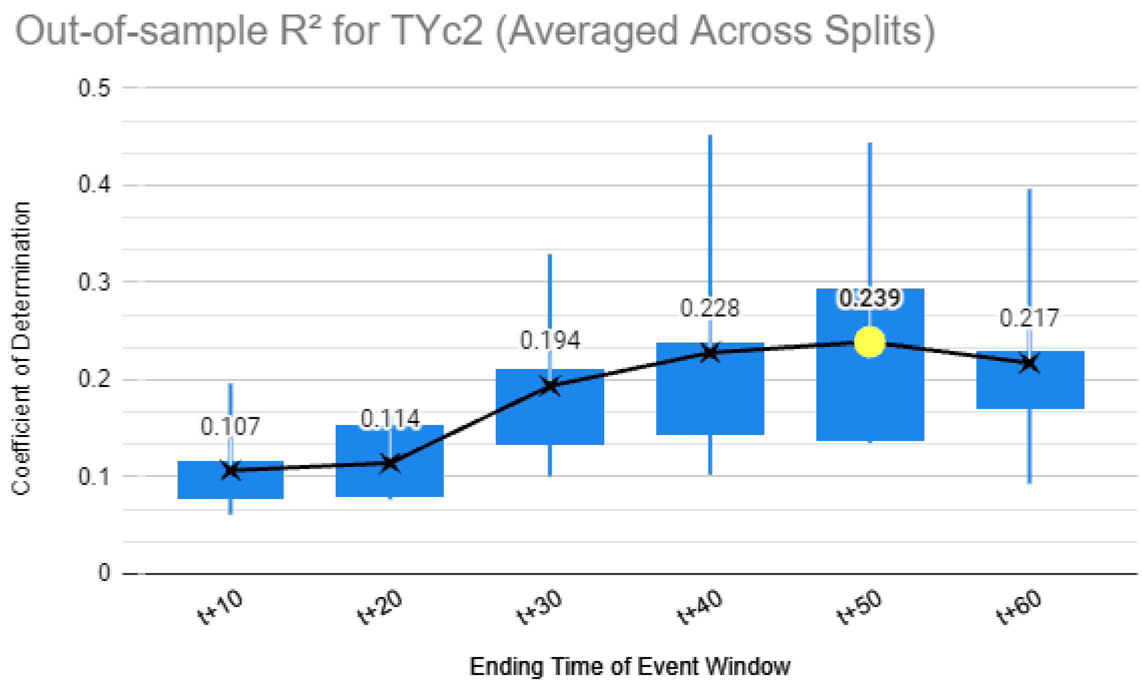


Figure 16: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for $TYc2$

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.

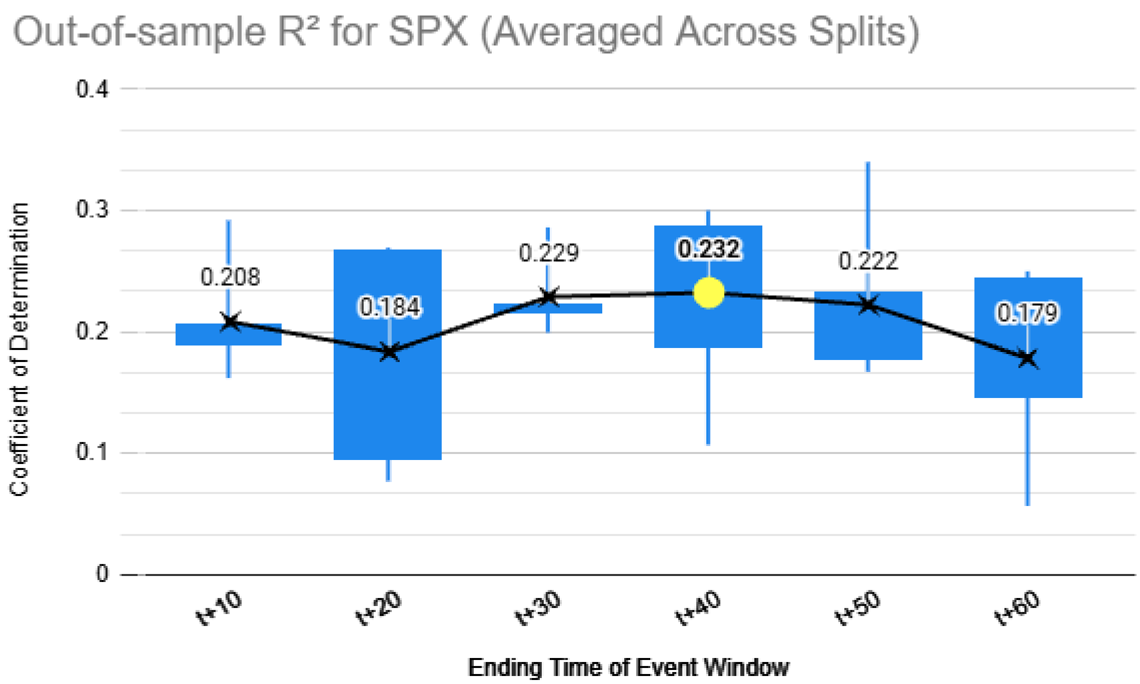
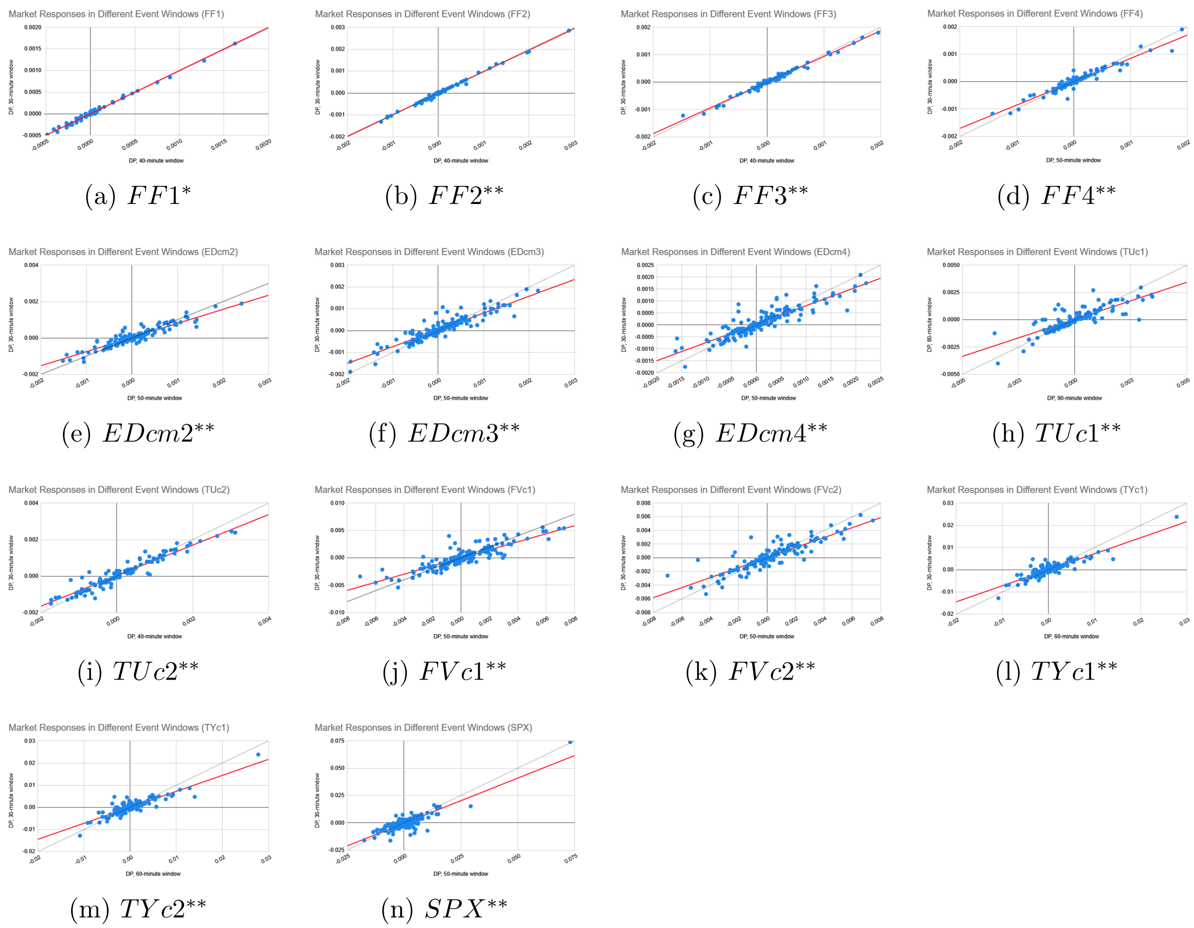


Figure 17: Systematic Estimation of Event Window Size Best Reflecting Market Fundamental Reactions for SPX

The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the percentage points of price changes by the market directly and only due to the FOMC statement text. The cross points represent $\overline{R^2_{OOS}}$ for each event window size, where the solid yellow point represents the event window with the largest $\overline{R^2_{OOS}}$. For each event window, box-and-whisker plots are shown surrounding the corresponding averages. Negative values imply the null model of using the in-sample mean as prediction has better performance than the neural network. $0 \leq \overline{R^2_{OOS}} \leq 1$ represents the proportion of the null model's mean squared error explained by the superior out-of-sample performance of XLNet.



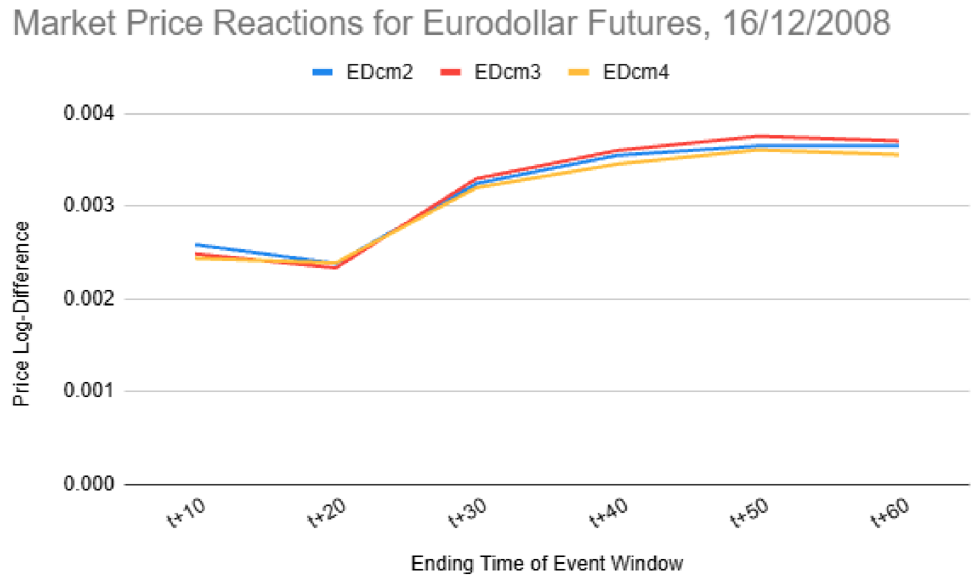
(a) $FF1^{*}$ (b) $FF2^{**}$ (c) $FF3^{**}$ (d) $FF4^{**}$

(e) $EDcm2^{**}$ (f) $EDcm3^{**}$ (g) $EDcm4^{**}$ (h) $TUc1^{**}$

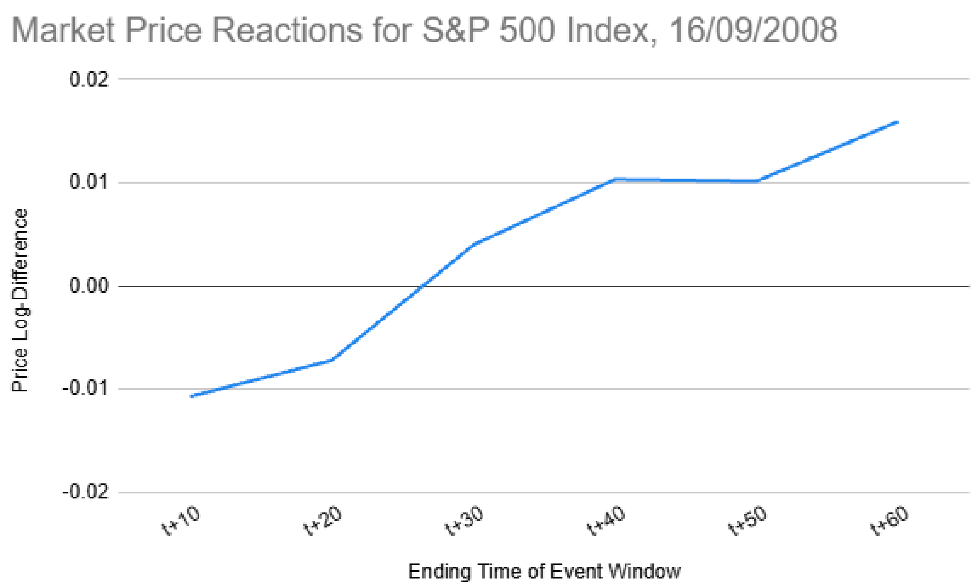
(i) $TUc2^{**}$ (j) $FVc1^{**}$ (k) $FVc2^{**}$ (l) $TYc1^{**}$

(m) $TYc2^{**}$ (n) SPX^{**}

Figure 18: Comparing Market Responses in Different Event Windows
The horizontal axis depicts the price log-difference within the systematically estimated event window length. The vertical axis represents DP_{t+20}. The 45-degree line is depicted as grey. The blue dots are market reactions on scheduled FOMC meetings. The price-log differences calculated within the optimal window lengths are regressed on DP_{t+20} through OLS. If the slope coefficient is greater (less) than one and its difference with one is statistically significant for at most $\alpha = 0.10$, then the financial market under-reacts to the information content of FOMC statements on release, ex post, and is depicted in red. "*" ("**") denotes statistical significance at the 5% (1%) level.



(a) Eurodollar Futures



(b) SPX

Figure 19: Responses of Eurodollar Futures and S&P 500 Markets to FOMC Statements The top (bottom) sub-figure depicts the price log-differences of Eurodollar futures markets (the S&P 500 Index market) for different event window lengths after the release of the 16 December (September), 2008 FOMC statement. The horizontal axis of each figure depicts the end time of the considered event window lengths and the vertical axis represents the price-log differences.

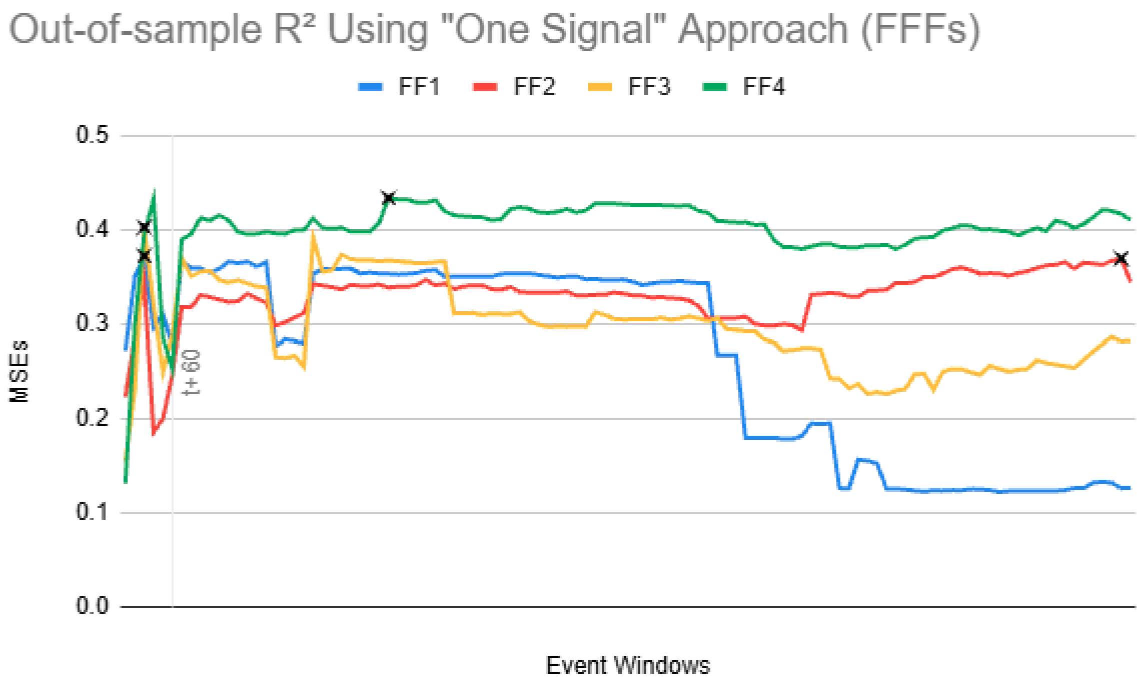


Figure 20: $\overline{R^2_{OOS}}$ Calculated Using the "One Signal" Approach for Federal Funds Futures

The horizontal axis depicts the event window lengths, starting from a 20-minute event window and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the largest $\overline{R^2_{OOS}}$. Estimates to the left of the vertical grey line labelled "$t + 60$" are calculated using the "joint' approach. The 1-month and 3-month-ahead federal funds futures contract are the two exceptions that sees the largest $\overline{R^2_{OOS}}$ within a window length than the systematically estimated length. As a robustness check, $\overline{R^2_{OOS}}$ was calculated with systematic estimation for both futures and considered event window lengths, which can be found in Figures 5 and 7, respectively.

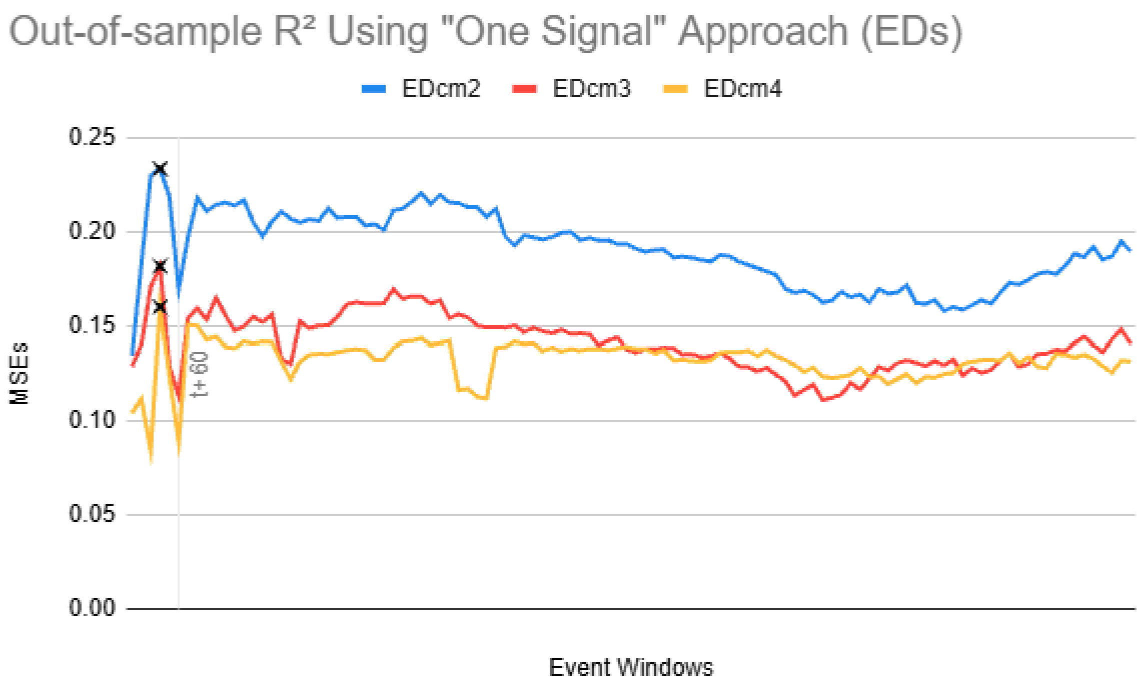


Figure 21: $\overline{R^2_{OOS}}$ Calculated Using the "One Signal" Approach for Eurodollar Futures

The horizontal axis depicts the event window lengths, starting from a 20-minute event window and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the largest $\overline{R^2_{OOS}}$. Estimates to the left of the vertical grey line labelled "$t + 60$" are calculated using the "joint" approach.

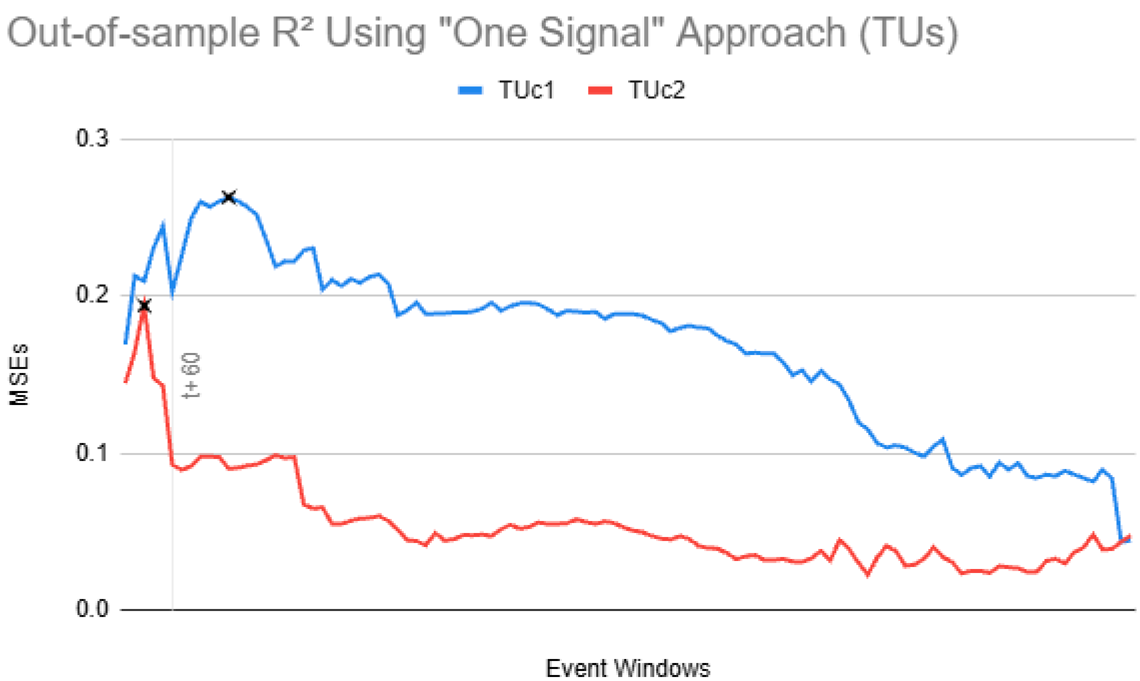


Figure 22: $\overline{R^2_{OOS}}$ Calculated Using the "One Signal" Approach for 2-year Treasury Futures

The horizontal axis depicts the event window lengths, starting from the systematically estimated event window length and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the largest $\overline{R^2_{OOS}}$. Estimates to the left of the vertical grey line labelled "$t + 60$" are calculated using the "joint" approach. The front-month 2-year Treasury futures contract is the one exception that sees the largest $\overline{R^2_{OOS}}$ within a window length than the systematically estimated length. As a robustness check, $\overline{R^2_{OOS}}$ was calculated with systematic estimation for the futures contract and considered event window length, which can be found in Figure 11.

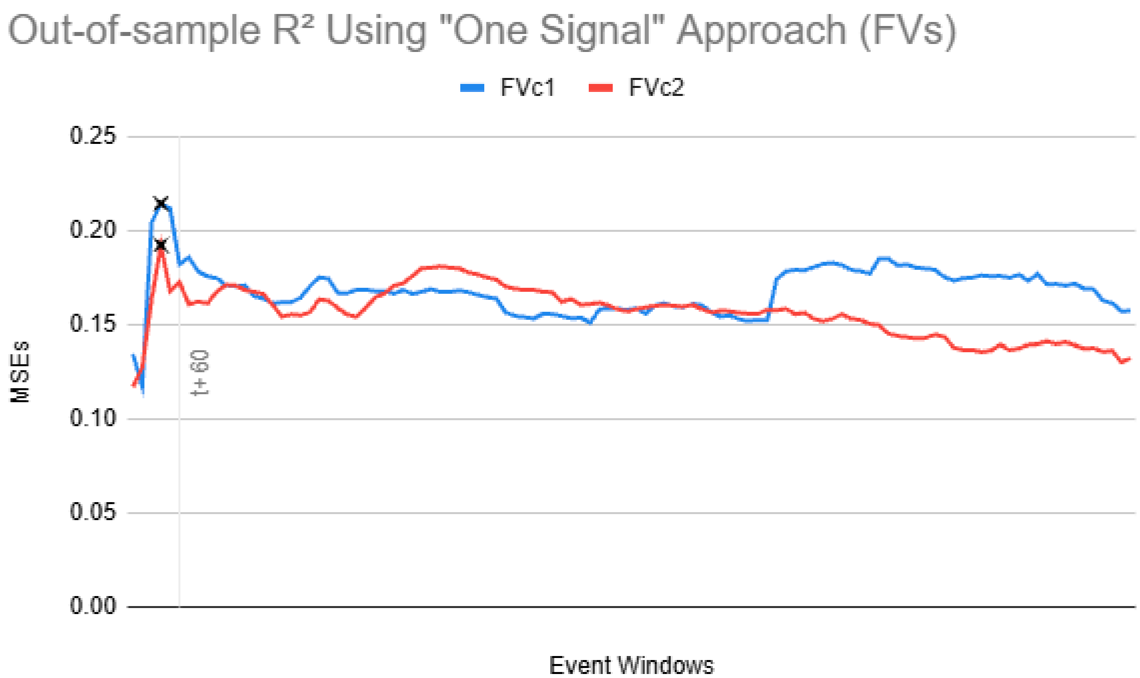


Figure 23: $\overline{R^2_{OOS}}$ Calculated Using the "One Signal" Approach for 5-year Treasury Futures

The horizontal axis depicts the event window lengths, starting from the systematically estimated event window length and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the largest $\overline{R^2_{OOS}}$. Estimates to the left of the vertical grey line labelled "$t+60$" are calculated using the "joint" approach.

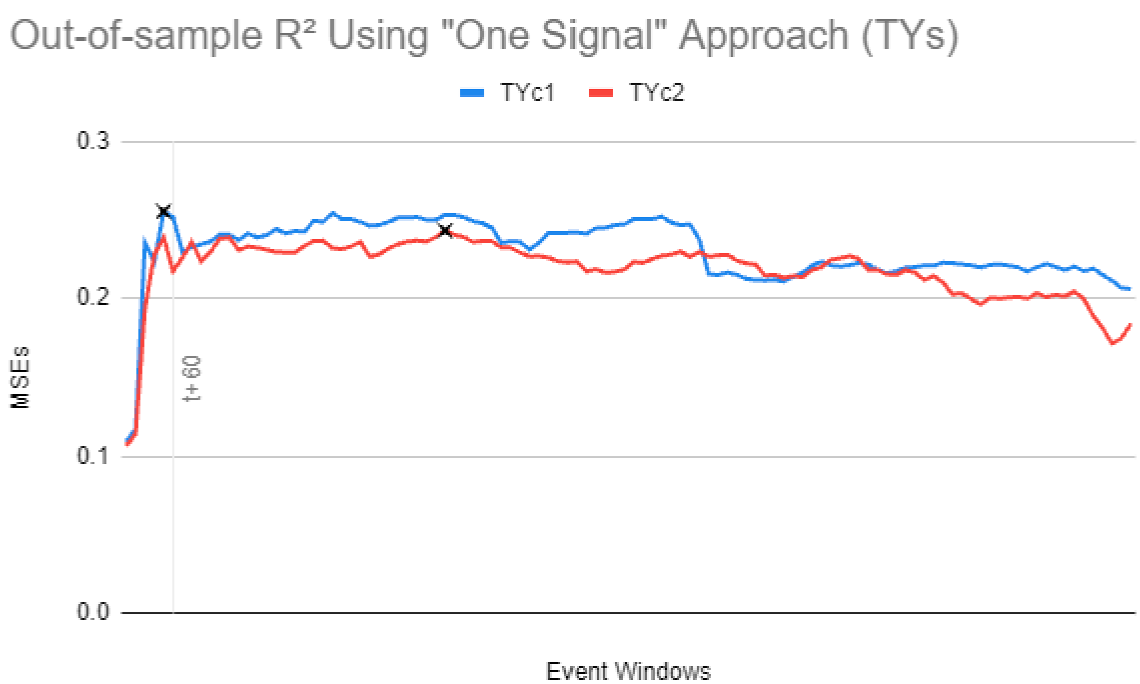


Figure 24: $\overline{R^2_{OOS}}$ Calculated Using the "One Signal" Approach for 10-year Treasury Futures

The horizontal axis depicts the event window lengths, starting from the systematically estimated event window length and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the largest $\overline{R^2_{OOS}}$. Estimates to the left of the vertical grey line labelled "$t + 60$" are calculated using the "joint" approach.

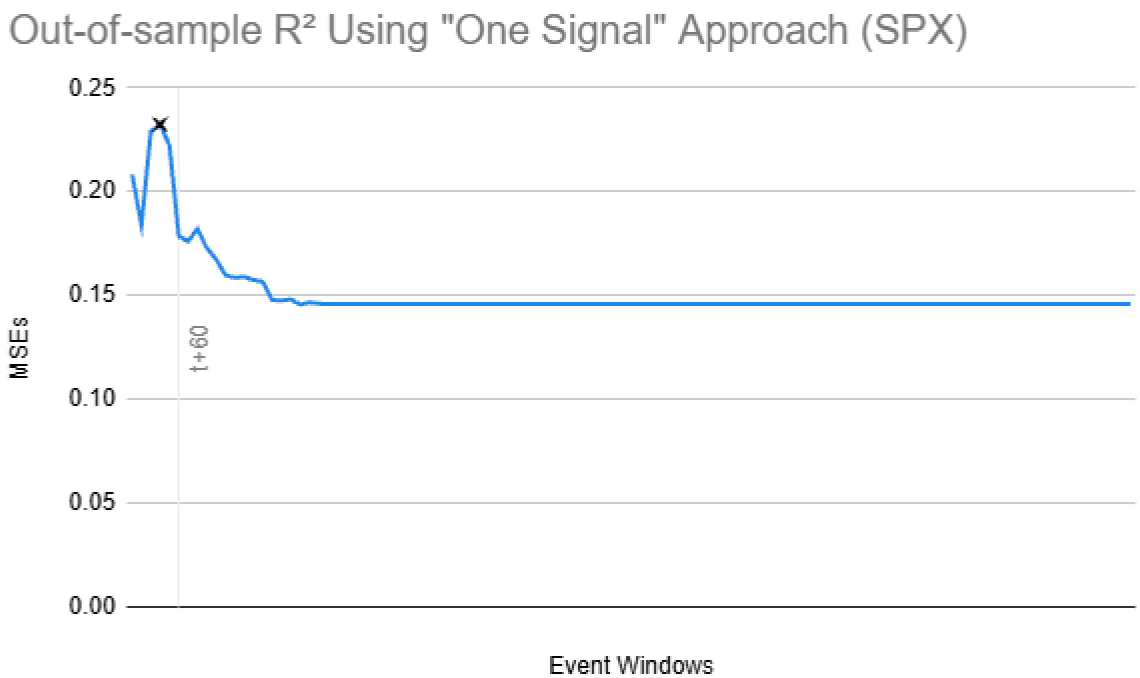


Figure 25: $\overline{R^2_{OOS}}$ Calculated Using the "One Signal" Approach for S&P 500 Index

The horizontal axis depicts the event window lengths, starting from the systematically estimated event window length and ending at an event window starting 10 minutes before and ending 18 hours after FOMC statement release. The cross points represent the event window length associated with the largest $\overline{R^2_{OOS}}$. Estimates to the left of the vertical grey line labelled "$t + 60$" are calculated using the "joint" approach.

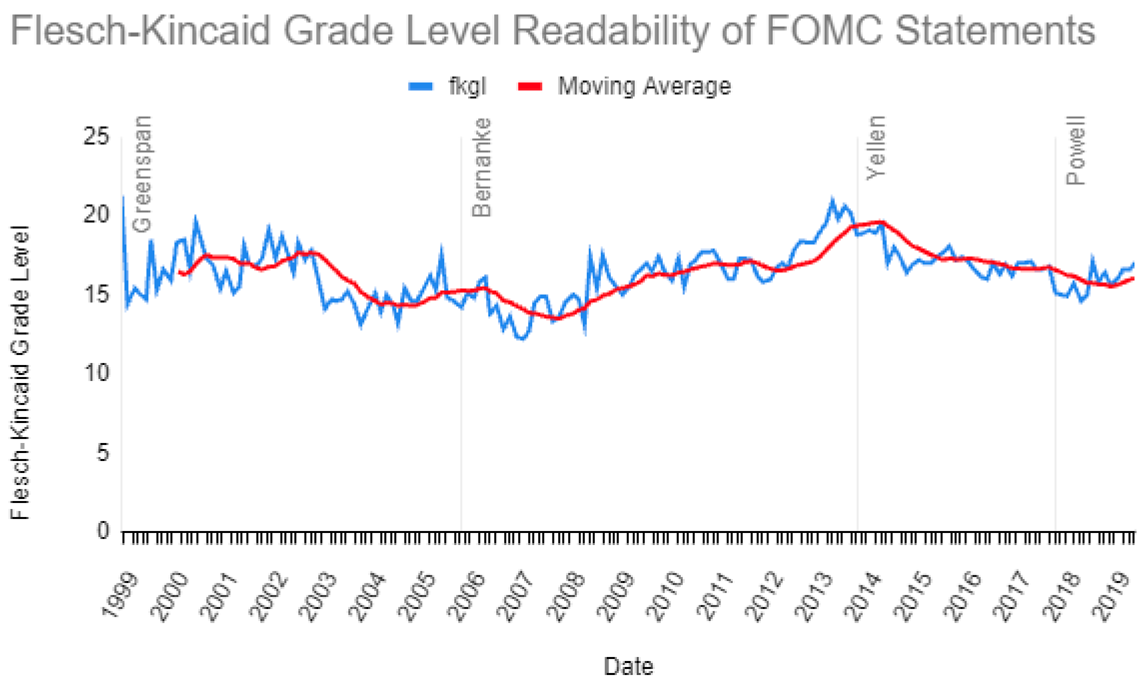


Figure 26: Flesch-Kincaid Grade Level Readability of FOMC Statements

The complexity of FOMC statements is measured by the Flesch-Kincaid Grade Level, defined as: $0.39 \times$ average sentence length $+ 11.8 \times$ average number of syllables per word $- 15.59$. From left to right, the vertical grey lines indicate the first FOMC meeting with Greenspan, Bernanke, Yellen, and Powell as Fed Chair. The moving average is calculated with a period of 10.

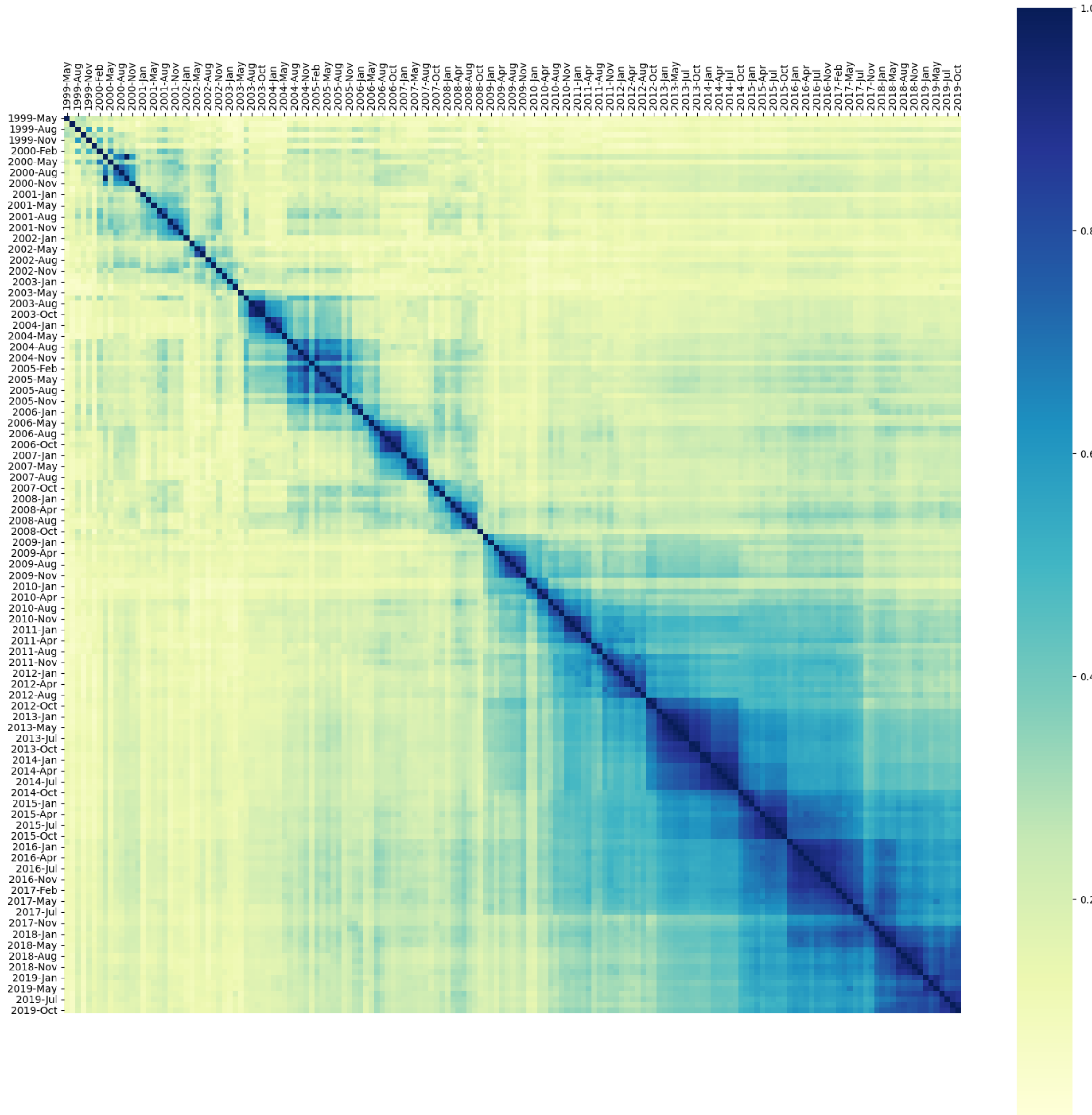


Figure 27: Document Similarity Matrix for FOMC Statements

Each element is the cosine similarity between two FOMC statements of that row and column. The cosine similarity value measures how similar the terms of both statements are. The darker shade of blue represents pairs of documents that have higher similarity measures closer to one, whilst the light shade of green represents pairs of documents that do not have terms in common and similarity values closer to zero. The main diagonal has all ones because the cosine similarity value is calculated between an FOMC statement with itself. All FOMC statements have been preprocessed such that all words are lowercase, all words are reduced to base form, and all stop-words are removed.

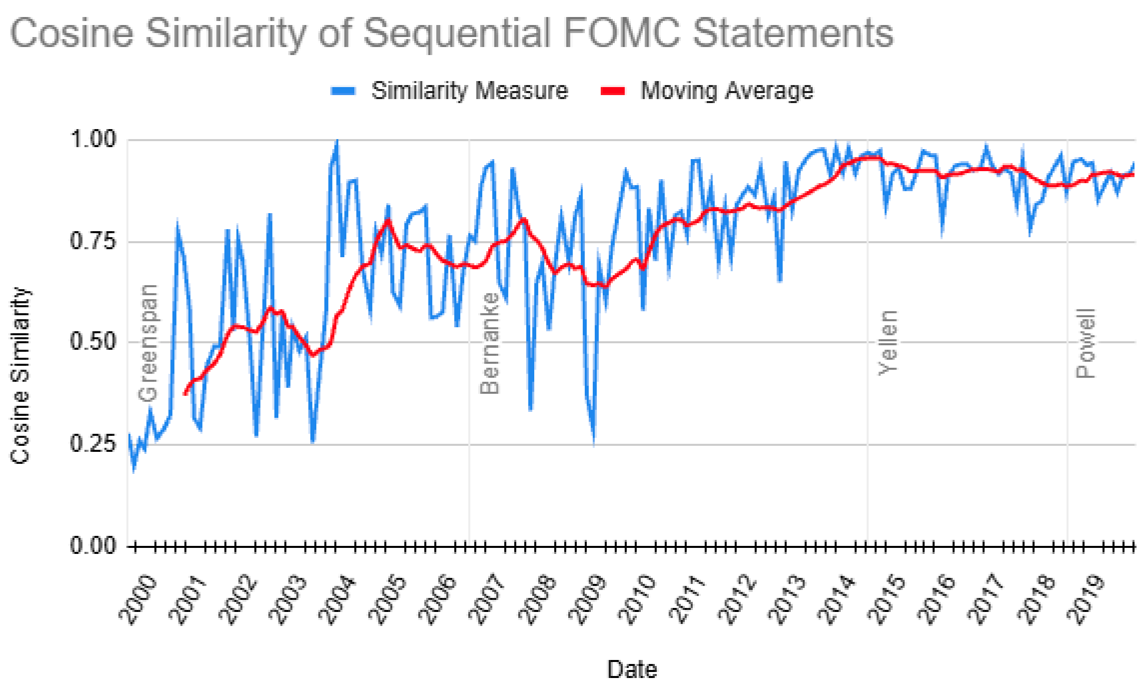


Figure 28: S^1 Cosine Similarity of Sequential FOMC Statements

S^1 is the cosine similarity between the TFIDF value of an FOMC statement and that of the immediately previous FOMC statement. Cosine similarity values closer to one (zero) mean the statements share more (less) common term usage. The moving average is calculated with a period of 10.

Tables

	Scenario 1	Scenario 2	Scenario 3
Framework Simulation Parameters			
P_i^f	$\in [-100, 100]$	$\in [-100, 100]$	$\in [-100, 100]$
$\varepsilon_{i,0}^c$	$\in [-100, 100]$	$\in [-100, 100]$	$\in [-100, 100]$
$\varepsilon_{i,0}^n$	0	0	0
σ_c	10	0.1	5
decay	0.1	1	0.1
σ_n	0.1	10	1
ρ_c	0.47	0.47	0.47
σ_s	$\in \mathbb{R}$	$\in \mathbb{R}$	$\in \mathbb{R}$
Simulation Results			
t^*	37	2	10
$\hat{t}$	37	2	10

Table 1: Framework Simulation Parameters and Results for Different Asset Market Scenarios
Notes: The results are from 10,000 simulations of the asset price framework for the three asset market scenarios considered. t^* and $\hat{t}$ are defined as the time at which the asset market fundamentally reacts to FOMC announcements on average, where the former is calculated using the fundamental price component and the latter with the observed signal according to the price framework.

Name	Maturity	Ticker	Sample	Observations
Inputs				
FOMC Statements	N/A	N/A	1999–2019	165
Outputs				
Federal Funds Rate Futures	Front-month	$FFc1$	1999–2019	165
Federal Funds Rate Futures	1-month-ahead	$FFc2$	1999–2019	165
Federal Funds Rate Futures	2-month-ahead	$FFc3$	1999–2019	165
Federal Funds Rate Futures	3-month-ahead	$FFc4$	1999–2019	165
Eurodollar Futures	2-quarter	$EDcm2$	1999–2019	165
Eurodollar Futures	3-quarter	$EDcm3$	1999–2019	165
Eurodollar Futures	4-quarter	$EDcm4$	1999–2019	165
2-year Treasury Futures	Front-month	$TUc1$	1999–2019	165
2-year Treasury Futures	Second-month	$TUc2$	1999–2019	165
5-year Treasury Futures	Front-month	$FVc1$	1999–2019	165
5-year Treasury Futures	Second-month	$FVc2$	1999–2019	165
10-year Treasury Futures	Front-month	$TYc1$	1999–2019	165
S&P 500 Index	N/A	SPX	1999–2019	165
S&P 500 E-mini Futures	Front-month	$ESc1$	1999–2019	165
S&P 500 E-mini Futures	Second-month	$ESc2$	1999–2019	165

Table 2: Independent and Dependent Variables for Systematic Estimation of Appropriate Event Window Lengths by Neural Networks
Notes: The table shows the FOMC statements and financial market asset prices considered as independent and dependent variables, respectively, in my analysis. The statements are collected from the Board of Governors of the Federal Reserve System website and the asset prices are from the Thomson Reuters Tick History database. All series begin in May 1999 and end in October 2019. *Ticker* refers to the Reuters Instrument Code (RIC).

Hyperparameter	Value
Number of Layers	12
Hidden Size	768
Number of Attention Heads	12
Attention Head Size	64
Hidden Dropout	0.1
Attention Dropout	0.1
Max Sequence Length	512
Vocab Size	32,000
Training Batch Size	8
Evaluation Batch Size	8
Max Num of Steps	2040
Max Num of Epochs	120
Manual Random Seed	47
Learning Rate	*
Warmup Ratio	0.06
Adam Epsilon	1e-8
Learning Rate Decay	Linear

Table 3: XLNet Hyperparameters for Fine-tuning

Notes: The symbol "*" denotes the hyperparameter that undergoes tuning during the fine-tuning process. The random seed is set as a constant to ensure replicability by always initalising random components (e.g., weights and biases) to the same initial values.

Metric	FKGL	S^1
Count	165	164
Mean	16.361	0.751
SD	1.715	0.212
Max	21.3	0.984
75^{th}	17.3	0.920
Median	16.5	0.826
25^{th}	15.1	0.622
Min	12.2	0.200

Table 4: Descriptive Statistics for Heterogeneity Analyses

Notes: The complexity of FOMC statements is measured by the Flesch-Kincaid Grade Level (FKGL), defined as: $0.39 \times \text{average sentence length} + 11.8 \times \text{average number of syllables per word} - 15.59$. S^1 is the cosine similarity measure between sequential FOMC statements.

Metric	Simple	Complicated	Different	Similar	Unity	Dissents
Minimised MSE						
Average	1.12e-5	9.44e-6	9.89e-6	1.02e-5	8.31e-6	1.33e-5
Event Window Length (Minutes)						
Average	58	72	61	50	51	61

Table 5: MSEs and Event Window Lengths Calculated Using the "One Signal" Approach, Conditioned by FOMC Statements Complexity, Similarity, and Presence of Dissents
Notes: The complexity of FOMC statements is measured by the Flesch-Kincaid Grade Level, defined as: $0.39 \times$ average sentence length $+ 11.8 \times$ average number of syllables per word $-$ 15.59, and displayed in the first two columns. Changes in FOMC statements are measured using a pairwise-statement cosine similarity measure and displayed in the third and fourth columns from the left. The event window lengths are displayed in minutes. "Simple" statements have grade levels up to 16.5. "Complicated" statements have grade levels above 16.5. "Different" are sequential statements with a cosine similarity of less than to 0.885. "Similar" are sequential statements with a cosine similarity of more than 0.885. "Unity" statements are those without votes of dissent. "Dissents" are statements with recorded dissent votes. In order to lessen the effects of outliers, the event window length for the 3-month-ahead federal funds future and 3-quarter Eurodollar future under the "one signal" approach are set to equal the median of the sub-set window lengths for the asset type.

domestic	alreadytight	recov	tilt	buildup
alert	alter	foreign	imbal	undermin
direct	excess	quit	favor	perform
strength	fall	trend	eas	concern
background	firm	gener	demand	potenti
core	subdu	cost	longer	gain

Table 6: FOMC Statement Base Terms with Top 30 TFIDF Scores
Notes: TFIDF is a weighted frequency combining word and document counts such that greater weight is given to words that are more informative about the information content of the FOMC statements relative to other statements where the words are not found. Specifically, the weighted term frequency gives higher weight to terms that occur more frequently in a given document. The term frequency is then divided by the number of documents that has this term appear. The more documents that have the word, the less importance and weight will be given to the word as it is less informative for distinguishing documents from one another.