Complexity and Information Content of Financial Disclosures: Evidence from Evolution of Uncertainty Following 10-K Filings *

Frederico Belo † Jun Li‡ Xiaoji Lin§ Xiaofei Zhao¶

October 2016

Abstract

We document two novel findings on the evolution of uncertainty following 10-K filings. First, we find a hump-shaped volatility dynamics for an average firm: following 10-K filings, its volatility increases by 0.36% in the first two to four weeks followed by a 2.55% decrease in the subsequent six weeks. Second and more importantly, this hump-shaped dynamics is more pronounced for firms with larger 10-K file sizes – a recent measure for complexity of financial disclosures. The economic impact of our findings is nontrivial: an options strategy based on this volatility pattern delivers up to 17.3% cross-sectional difference in annualized returns between firms with large and small 10-K file sizes. Our findings therefore highlight two opposing effects of information disclosures on the evolution of uncertainty along the time dimension: While a more complex disclosure is associated with a higher level of uncertainty in the short horizon, its information content can result in an eventual resolution of uncertainty once the information is digested by investors.

Keywords: Complexity; Information content; Financial disclosure; Volatility dynamics
JEL Classification: G14; D82; D83; M41

*We are very grateful to Daniel Cohen, Bill Cready, Zhonglan Dai, Ted Day, Cesare Fracassi, Paolo Fulghieri, Gerard Hoberg, Xu Jiang, Bin Li, Tim Loughran, Jordan Moore, Vikram Nanda, Oleg Nusinzon, Yihui Pan, Xiaoxia Peng, Michael Rebello, Alessio Saretto, Baolian Wang, Tracy Wang, Kelsey Wei, Tony S. Wirjanto, Han Xia, Yexiao Xu, Celim Yildizhan, Jianfeng Yu, Harold H. Zhang, Xiao-Jun Zhang, Feng Zhao, Jane Zhao, and seminar participants at 2015 Lone Star Finance Conference, 2016 MFA Annual Meetings, 2016 Mid-Atlantic Research Conference (MARC) in Finance, 2016 CICF, 2016 FMA Annual Meetings, PanAgora Asset Management and the University of Waterloo for valuable comments. We also thank Tim Loughran and Bill McDonald for making their data on the size of 10-K filings available, and Feng Li for making the Fog index available. All errors are our own.

†Department of Finance, University of Minnesota, fbelo@umn.edu
‡Jindal School of Management, University of Texas at Dallas, jun.li3@utdallas.edu
§Department of Finance, Fisher College of Business, The Ohio State University, lin.1376@osu.edu
¶Jindal School of Management, University of Texas at Dallas, xiaofei.zhao@utdallas.edu

1 Introduction

Investor learning over *time* about certain corporate attributes has gained growing interests in recent academic studies. For example, Pan, Wang, and Weisbach (2014, 2015) examine how investor learning about managerial ability affects the dynamics of asset volatility and cost of borrowing. Giglio and Shue (2014) study the impact of investor learning over time on M&A outcomes. Our paper adds to this emerging line of research by studying how investor learning over time about the annual financial reports affects the evolution of investors' perception of uncertainty.

The fundamental goal of annual reports (10-Ks) is to provide information to investors.[1] Therefore, in principle, annual reports should lead to a resolution of uncertainty. However, how investors react to and digest the information is a more complicated process, especially when the disclosed information is large in amount and complex in nature. Recent empirical evidence suggests this is the case for large disclosure such as 10-Ks. Indeed, Loughran and McDonald (2014) find that a 10-K filing with larger size is associated with higher, instead of lower, volatility within one month after the filing; hence, this finding poses a challenge to the fundamental goal for 10-K disclosures. Our paper is the first attempt to reconcile their seemly contradictory finding and uncover the information role of 10-K disclosures.

Motivated by the idea that investors take time to learn about important corporate attributes (e.g., Pan, Wang, and Weisbach (2014); Pan, Wang, and Weisbach (2015); Giglio and Shue (2014)), we conduct a comprehensive analysis of the volatility dynamics in the event window following 10-K filings. This allows us to uncover the information role of a larger 10-K filing in a relatively longer horizon; this is the main contribution of our paper. We find that a larger 10-K filing in two months following the filing is associated with a bigger reduction in volatility, in addition to a higher volatility in the one month after the filing. Our central findings suggest that consistent with the fundamental goal of annual reports, by and large, a larger 10-K filing carries more information content, although it takes time for investors to learn about it. Coupled with the finding by Loughran and McDonald (2014), our results reveal the dual roles of 10-K disclosures: complexity and information roles of the disclosures are revealed as investors learn through *time*.

We use option implied volatility as a measure for investors' perception of uncertainty. Our first finding indicates a hump-shaped volatility dynamics following 10-K filings as in Figure 1. Specifically, an average firm experiences a 0.36% increase in volatility in the first two to four weeks after 10-K filings, followed by a 2.55% decrease in volatility in the subsequent six weeks, leading to a net reduction of 2.19% relative to the volatility level around the filing

[1] http://www.sec.gov/answers/form10k.htm.

date.[2] More importantly, our second and main finding suggests that the strength of this hump-shaped pattern varies across firms with different disclosure characteristics.

Our main measure of a 10-K disclosure is its file size (Loughran and McDonald (2014)). As an easy-to-construct yet effective proxy for the linguistic complexity of 10-K disclosure, Loughran and McDonald (2014) document that 10-K file size is positively associated with high return volatility in a one-month period following 10-K filings. The file size of 10-K can also be a good proxy for the information content of financial disclosures due to its aggregate nature.[3] While the complexity of a larger 10-K makes it more difficult to digest in the short run, the information carried in the larger disclosure may help resolve uncertainty in the long run. The empirical analysis supports our conjecture (Figure 3). First, consistent with Loughran and McDonald (2014), we find that firms with a larger 10-K experience a greater increase in volatility than firms with a smaller 10-K in the first few weeks after the 10-K filing. Second, complementary to Loughran and McDonald (2014), we find that firms with a larger 10-K also experience a sharper decrease in volatility in the subsequent six weeks.

We carry out our analysis more comprehensively in several settings. First, we use panel analysis to compare the average volatility changes in the short horizon (two to four weeks) with the average volatility changes in the relatively long horizon (eight to ten weeks) and find results consistent with Figure 1 and Figure 3. The short-run volatility change is positively (and significantly) associated with 10-K file size. The relatively long-run volatility change is negatively (and significantly) associated with 10-K file size. All else being equal, a one standard deviation increase in log 10-K file size is associated with 0.37% *increase* in the implied volatility in the short horizon, but 0.59% *net decrease* in the implied volatility in the subsequent six weeks, both relative to the pre-filing volatility level. To alleviate the impact of subsequent quarterly earnings announcements, we study a subsample of firms for which the 10-K filing dates and the subsequent earnings announcement dates are at least seven weeks apart. For this subsample, we find even stronger results: The marginal impact of file size on the relatively long-run volatility doubles while the marginal impact on the short-run volatility remains similar to the full sample. We also simulate a sample

[2]For an average firm, the volatility starts to decrease in about two to three weeks; this reflects the (time) cost of processing the 10-K reports. This is broadly consistent with Lehavy, Li, and Merkley (2011), who use the amount of time it takes analysts to issue reports following 10-K filings (or "analyst report duration") to measure the cost or effort that analysts bear in following firms; they document that the average analyst report duration is about 18 days.

[3]In fact, the dual roles of information disclosure have also been hinted in a survey paper by Li (2010) on page 147 "...The amount of disclosure is relatively easy to measure as it typically involves the length or the size of the file... However, these papers that analyze the length of a document (or a section of the document) often treat it as a measure of complexity or transparency of the disclosure rather than the amount of disclosure. The truth is perhaps somewhere in between: the length of disclosures is likely to capture the level or the amount of disclosure as well as the complexity of disclosure..."

of counterfactual 10-K filings around which the volatility dynamics are entirely driven by earnings announcements. We find that the impact of earnings announcements on our results is likely to be minimal. In addition, although the implied volatility may be a better proxy for investor's perceived uncertainty than the volatility measures based on realized stock returns, our results are robust to these alternative measures (for example, realized volatility calculated using intra-day returns and daily returns).

As a placebo test, we repeat our analysis by replacing 10-K file size (an aggregate measure) with other *normalized* measures of linguistic complexity (for example, Fog index, Flesch index, and Kincaid index, as in Li (2008)). Intuitively, the normalized measures may capture linguistic complexity but should not capture the aggregate information content of 10-K filings. Consistent with this intuition, we find that these normalized measures are positively associated with an increase in short-run volatility but their predictive power for the change in the long-run volatility is very weak.

We also directly model the volatility dynamics as a function of the time passed since a 10-K filing date (in the spirit of a term structure) at the firm-level. To capture the hump-shaped dynamics, we use one of the simplest functional forms, a quadratic function: $\sigma_{i,t} = a_i \times t + b_i \times t^2 + \epsilon_{it}$. One way of interpreting this function is that a_i can be related to the initial learning speed upon receiving a piece of new information and b_i can be viewed as the acceleration of learning. A positive a_i suggests that the uncertainty increases right after the information disclosure. A negative b_i suggests that the uncertainty decreases over time after the peak level of uncertainty. For each firm, we estimate this volatility dynamics using volatility data following the 10-K filing dates to obtain a pair of (a_i, b_i). We find that on average a_i is positive and b_i is negative. Moreover, a_i is positively correlated with 10-K file size and b_i is negatively correlated with 10-K file size. Using estimated parameter values, we are able to generate similar volatility dynamics for firms with different 10-K file sizes (i.e., Figure 4 reproduces volatility patterns similar to those in Figure 3).

To formally gauge the economic impact of our findings, we implement an options strategy based on the hump-shaped volatility dynamics following 10-K filings. The strategy involves shorting straddles during two to four weeks following the 10-K filings and closing the positions in weeks eight to ten. We find that for firms with large 10-Ks, this strategy delivers 17% higher annualized returns than for firms with small 10-Ks. As a comparison, the shorting straddle strategy is unable to deliver the aforementioned cross-sectional difference for randomly selected 6-week windows in the 1-year period around the actual 10-K filing date. This analysis shows that our main findings are robust and economically meaningful. Moreover, we find that information asymmetry decreases significantly in the long run for firms with larger 10-Ks; this finding reinforces the information role of 10-K disclosures.

Our findings suggest that, information disclosure in general, and 10-K disclosure in particular, may have two important opposing effects on the evolution of uncertainty: More information in a disclosure could make it more difficult for investors to digest and lead to a higher level of uncertainty in the short run – this is the complexity aspect of information disclosure. Once digested, more information could eventually result in more resolution of uncertainty – this is the information content aspect of information disclosure.

Our paper is related to several lines of research. First, our study contributes to the emerging line of research on how investor learning over time about important corporate attributes affects the perceived riskiness of the corporate securities. For example, Pan, Wang, and Weisbach (2014) study the impact of learning about managerial ability on volatility dynamics. Our firm-level empirical learning-speed model is similar to theirs. The main differences are the objective and horizon of investor learning. In our paper the learning objective is 10-K information and in theirs the objective is managerial ability. The difference in the learning objective also renders the difference in the horizon of learning (our 8-10 weeks versus their 3-year learning horizon). Pan, Wang, and Weisbach (2015) study the impact of learning about management risk on cost of borrowing. Giglio and Shue (2014) study the impact of investor learning over time on M&A outcomes. We contribute to this line of research by demonstrating the novel impact of learning on volatility dynamics (dual roles of information disclosure) in another important setting.[4]

Second, our paper contributes to the large literature in finance and accounting on the informativeness aspect and on the complexity aspect of corporate information disclosures.[5] Most studies on information disclosures do not jointly consider both aspects in the same context. Our main contribution to this literature is to complement both lines of research by linking the dual roles of financial disclosures through the time dimension; this is novel and has theoretical foundations (e.g., Epstein and Schneider (2007), and Easley and O'Hara (2009)). A recent paper by Bushee, Gow, and Taylor (2013) examines the quarterly earnings conference call information at a fixed time point (conference call) but through different

[4]Our paper is also related to papers that examine the impact of learning on stock volatilities in general. See, for example, Timmermann (1993), Brennan and Xia (2001), Pastor and Veronesi (2003); interested readers should refer to Pastor and Veronesi (2009) for a complete review of how learning affects various aspects of financial markets. Peng and Xiong (2003) study the effect of analysts' cost of effort on long-run volatility persistence.

[5]For an incomplete list of papers on the informativeness of financial disclosures, see for example, McNichols and Manegold (1983), Lang and Lundholm (1996), Botosan (1997), Botosan and Plumlee (2002), Kothari, Li, and Short (2009), Leuz and Schrand (2009), Berger, Chen, and Li (2012). In particular, Leuz and Schrand (2009) examine the changes of firm's 10-K reports (for example, the number of pages) in response to Enron's shock to draw an inference on the impact of informative disclosure on the cost of capital. On the complexity aspect, see for example, Li (2008), Bloomfield (2008), You and Zhang (2009), Miller (2010), Lehavy, Li, and Merkley (2011), Dougal, Engelberg, Garca, and Parsons (2012), Lawrence (2013), De Franco, Hope, Vyas, and Zhou (2013), Loughran and McDonald (2014).

sources of information and communication (management versus analysts). In contrast to Bushee, Gow, and Taylor (2013), our focus is on the evolution of investors' perception of uncertainty due to the impact of complexity and information content of 10-K disclosures over the short and long horizons.

Third, our study is closely related to but different from papers that study volatility dynamics around earnings announcements.[6] Although the key financial variables are disclosed in earnings announcements, 10-Ks still carry additional material information that investors care about (e.g., Griffin (2003)), and 10-K reports could typically run over several hundred pages that is much longer than the press release for earnings announcements. The distinction between these two types of corporate disclosures implies a different volatility dynamics following earnings announcements and 10-K filings. As can be seen from Figure 2, earnings announcements are more likely to correspond to the type of disclosure that is less complex and easier to process, which is associated with a reduction of uncertainty in a short horizon. In contrast, 10-K filings are more likely to be the disclosure that is large in amount and complex in nature in which the dual roles of disclosure are more pronounced.

The rest of the paper is organized as follows. We provide more details on data construction in Section 2. Then we present our empirical findings in Section 3. At the end of the paper, we make some concluding remarks in Section 4.

2 Data

Our benchmark measure of firm-level uncertainty is the option implied volatility from OptionMetrics. There is a line of research suggesting that this uncertainty measure may be superior to alternative measures that are based on realized stock returns in capturing investor's perceived uncertainty.[7] Though we use the implied volatility as our main measure of

[6]There is a voluminous literature on the impact of earnings on stock prices. Of particular relevance, some papers mainly focus on volatility dynamics. For example, Rogers, Skinner, and Buskirk (2009) focus on the impact of management earnings forecasts on stock market volatility. Barth, Johnson, and So (2011) analyze the volatility dynamics in the 10-day window around earnings announcement date and focus on the performance of a modified option pricing model. Xing and Zhang (2013) focus on the straddle returns and volatility dynamics in the [-5,+1] day window around the earnings announcement date; Govindaraj, Liu, and Livnat (2012) focus on understanding post-earnings-announcement drift using longer windows. Other related papers that examine option straddle returns include Coval and Shumway (2001), Goyal and Saretto (2009), among others. A related paper by Kelly, Pastor, and Veronesi (2014) examines the pricing of political uncertainty in options market.

[7]For example, this could be due to the forward-looking nature of the implied volatility and that it is less affected by large realized first moment shocks on the underlying equity. For an incomplete list of contributions in this literature, see for example, Day and Lewis (1992), Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Christensen and Prabhala (1998), Fleming (1998), Blair, Poon, and Taylor (2001), Jiang and Tian (2005), Busch, Christensen, and Nielsen (2011), Chang, Christoffersen, Jacobs, and Vainberg (2012). Poon and Granger (2003) also provide a nice review of relevant studies.

uncertainty, we also show that our findings are robust to realized return based uncertainty measures.

In order to cover the horizon of investor learning in the period of 10 weeks after 10-K filings, we only use options with relatively longer maturities. In particular, we only include options with 62 to 213 days (or 1/6 to 7/12 year) to maturity and positive trading volumes for constructing the firm-level implied volatilities. We take the mid-quote value as a fair reflection of the option price. We also require the option price to satisfy the arbitrage bound. For call options, the arbitrage bound is max(0, stock price−strike price)$<$mid quote$<$stock price. For put options, the arbitrage bound is 0$<$mid quote$<$strike price. The moneyness of an option – the ratio between the stock and strike prices – is required to be between 0.8 and 1.2. Once we have screened the option data, we calculate the firm-level daily implied volatility as the simple average of the implied volatilities of all remaining options on that firm on each trading day.[8]

Our main characteristic regarding 10-K filings is the file size from Loughran and M-cDonald (2014) because of its potential capability of capturing the dual roles of disclosure.[9] Loughran and McDonald (2014) suggest that if firms are trying to obscure mandated earnings-relevant information, they are more likely to bury the results in longer documents. They find strong evidence on the positive association between 10-K file size and future uncertainty measures with a focus on the relatively short horizon (1-month) following 10-K disclosures. Another aspect of file size is that larger documents may also contain more information despite its difficulty of comprehension. This is the basis of our study on the potential dual roles of information disclosure. There are other measures of 10-K complexity. Specifically, we also test the effect of three alternative measures of 10-K complexity on the dynamics of volatility. These measures are the Fog index (=0.4 $\times$ (words per sentence + percent of complex words)), the Flesch index (=206.835−1.015 $\times$ words per sentence−84.6 $\times$ syllables per word), and the Kincaid index (=11.8 $\times$ syllables per word+0.39 $\times$ words per sentence−15.59).[10] Unlike file size, which is an aggregate measure by nature, these indices are normalized measures (based on per sentence/word), therefore, we do not necessarily expect them to well capture the information content aspect of 10-K disclosures.

The other firm-level characteristics include the firm's market cap, book-to-market equity ratio, the pre-filing CAPM alpha in the window of [-252,-6] days before the 10-K filing, and

[8]Our findings are robust to alternative maturity, moneyness requirements, and weighting schemes. The results are available upon request.

[9]We thank Tim Loughran and Bill McDonald for making the 10-K filing data available online at http://www3.nd.edu/ mcdonald/Data/LoughranMcDonald_10-X_FileSize.csv.

[10]We are also grateful to Feng Li for making the data on Fog index, Flesch index, and Kincaid index available on his website: http://webuser.bus.umich.edu/feng/.

the absolute value of filing date excess return. These variables are constructed using the stock return data from CRSP and the accounting data from Compustat. We further impose the following restrictions on the underlying stocks in our sample: common shares (SHRCD to be 10 or 11) and stock price $\geq$ \$3.

Table 1 reports the summary statistics and correlation coefficient matrix for the benchmark sample from 1996 to 2012. The average logarithm of 10-K file size (Fsize) is 0.317, and it varies substantially across firm-year observations. For instance, the 10-K file size for the 75th percentile is 2.75 megabytes, more than five times larger than that for the 25th percentile. Consistent with Loughran and McDonald (2014), we find that Fsize has a strong correlation with Fog index, Flesch index, and Kincaid index.[11] The 10-K file size increases with firm size but is negatively correlated with the implied volatility. This is because the larger and more complex business for a big firm would naturally require a long and comprehensive description and discussion in financial disclosure. At the same time, the diversification effect for a large business (e.g., multiple segmentations) also suggests a lower uncertainty for big firms. On the other hand, we find much lower unconditional correlations between other measures of readability and the implied volatility.

Table 1 about here

3 Empirical findings

This section describes the empirical findings. Section 3.1 explores the dynamics of volatility for an average firm following 10-K filings. Section 3.2 studies how different 10-K disclosures affect the dynamics of volatility. In particular, we show that the pattern in the evolution of uncertainty varies substantially across firms with different 10-K file sizes. In Section 3.3, we analyze the profitability of an options investment strategy that takes advantage of the dynamics of volatility following 10-K filings. In Section 3.4, we examine how 10-K disclosures affect the dynamics of information asymmetry.

3.1 Average dynamics of volatility following 10-K filings

We first analyze the average effect of 10-K filings on the evolution of uncertainty. A Form 10-K is an annual report required by the U.S. Securities and Exchange Commission (SEC), that gives a comprehensive summary of a firm's financial and operational performance. Besides

[11] Note that Flesch index is a reading ease index. A higher Flesch index indicates that the text is earlier to read. Therefore, the correlations of Flesch index with Fog index, Kincaid index, and 10-K file size are all negative.

the key numbers such as earnings, dividends, and sales growth that have been disclosed to the public in earnings announcements, 10-K reports typically run for hundreds of pages and also include a lot of detailed information regarding the risk factors, financial conditions, operational results, managerial compensations, as well as projections on the firm's future performance. This large amount of additional information is important and valuable to investors, because it provides a comprehensive description of the company; if it were truly *informative*, it may also help resolve some uncertainty regarding the future prospects of the company. However, due to the large average 10-K file size of 1.37 megabytes, it is unlikely for investors to digest all of the information immediately after the disclosure. Instead, it may take some time (perhaps a couple of weeks or even months) for investors to process the 10-K filing, and then incorporate it into the asset prices. For instance, Lehavy, Li, and Merkley (2011) use the amount of time it takes analysts to issue reports following 10-K filings (or "analyst report duration") to measure the cost or effort that analysts bear in following firms, and document that the average analyst report duration is about 18 days.

The above consideration has important predictions on the evolution of uncertainty following 10-K filings: In the short run, the complexity of 10-K information may increase the uncertainty of a firm's information environment and consequently lead to a higher volatility. In a relatively long horizon, however, as the information is gradually digested by investors, the 10-K disclosure may help reduce uncertainty. To the extent that 10-K disclosure contains new information, it may even reduce the uncertainty to a level lower than the pre-filing uncertainty. This potential information content aspect of 10-K reports is also consistent with the objective of financial disclosures. Note these predictions may also find theoretical foundations corresponding to theoretical settings in which investors are ambiguity averse and do not know which information to process due to the large amount of complex information. For example, in a dynamic model, Epstein and Schneider (2007) show that new complex information may make investors less confident and more ambiguous and perceive higher uncertainty in the short run but as investors become more familiar with the information environment they become less ambiguous and behave more like Bayesian learning agents. Therefore, a 10-K filing may have two important aspects that are related to investors' perception about a firm's uncertainty: complexity and information content.

Table 2 and Figure 1 provide baseline evidence supporting the above predictions. To make the volatility dynamics comparable across firms, we benchmark all the volatility observations following 10-K filings to the pre-filing volatility level and calculate the post-filing volatilities as percentage changes relative to the pre-filing volatility. Figure 1 plots the average percentage change in implied volatility against the number of weeks following all 10-K fillings. Consistent with our predictions, we observe a clear hump shape in the dynamics of implied volatility – a mild increase in weeks 2-4 (with a peak at around week 2), followed by a rapid decline in weeks 8-10. This pattern exists in the full sample between 1996 and 2012 (Panel A), as well as in the early and late subsamples (Panel B and Panel C).[12]

Figure 1 about here

To quantitatively evaluate this effect, in Table 2, we define diffs as the average percentage change in the implied volatility from the 10-K disclosure for the short-run (2-4 weeks) and diffl as the average percentage change in implied volatility for the (relatively) long-run (8-10 weeks).[13] We also define diffsl as the difference between diffs and diffl to capture the evolution of uncertainty from the short-run to the long-run. The choice of the short-run and long-run horizons are motivated by previous studies (e.g., Loughran and McDonald (2014)) and our observations from Figure 1. It is meant to capture roughly the first and second half of the gap between a 10-K filing and its subsequent quarterly earnings announcement, but certainly this choice (especially the short-run) is ad-hoc to some extent. In our firm-level time series analysis in Section 3.2.3 below, we relax this restriction and estimate a quadratic learning model in which the peak level of uncertainty (or short-run horizon) could vary across firms and is determined by the data.

Panel A of Table 2 reports the average diffs, diffl, and diffsl for the full sample from 1996 to 2012. Following the 10-K filing, the implied volatility increases by 0.36% on average. This increase is statistically significant, reflecting the complexity aspect of an average 10-K disclosure. However, as the information is gradually understood by investors, the information content aspect becomes more important and starts to dominate the impact on uncertainty; we observe a rapid decline of 2.55% in implied volatility from weeks 2-4 to weeks 8-10, resulting a 2.19% overall decrease in implied volatility from before the filing. The patterns are robust and similar for the two sub-samples in Panel B (1996–2003) and Panel C (2004-2012) of Table 2.

Table 2 about here

[12] The increase in implied volatility after week 9 can be due to the subsequent quarterly earnings announcement. According to the SEC, historically, Form 10-K had to be filed within 90 days after the end of the company's fiscal year. However, in September 2002, the SEC approved a Final Rule that changed the deadlines to 75 days for Form 10-K for "accelerated filers" that have a public float of at least $75 million. In December 2005, the SEC created a third category of "large accelerated filers", accelerated filers with a public float of over $700 million. Beginning with the fiscal year ending on or after December 15, 2006, the deadline for "large accelerated filers" is 60 days. See SEC's website for more details. The average and median gap between 10-K filing dates and subsequent quarterly earnings announcement dates are both about 9 weeks in the most recent regulation regime since 2006; the 1^{st} and 3^{rd} quartile of this gap are 8 and 11 weeks respectively.

[13] Short-run and long-run could mean completely different horizons in different settings. It should be clear that in our context these two terms are defined relatively within a 3-month period.

Our discussion above has an implication on the dynamics of implied volatility around earnings announcements. The information released during earnings announcements usually only includes key financial variables, a balance sheet, an income statement, and sometimes a statement of cash flows. This information is standard and is not too complex to digest. Therefore, we could expect that only one aspect of the information disclosure during earnings announcements, the information content, would matter the most for the evolution of uncertainty. Indeed, when we compare the volatility dynamics around 10-K filings and earnings announcements (Figure 2), the difference is striking. Unlike the hump-shaped dynamics following 10-K filings (Panel B), the implied volatility falls immediately within the next couple of days after earnings announcements and stays at roughly the same level weeks afterwards (Panel A). Despite that both types of disclosures contain valuable information and lower the uncertainty ultimately, the difference in the complexity aspect of information disclosure gives rise to totally different patterns in volatility dynamics.

Figure 2 about here

3.2 Volatility dynamics in the cross section

This previous section documents an average hump-shaped volatility dynamics following 10-K disclosures, and attributes the different behaviors of implied volatility in the short-run and long-run to two distinctive aspects of 10-K disclosure: complexity and information content. In this section, we provide additional evidence by exploring the data in the cross section. In particular, firms may choose to disclose more or less information depending on the type and complexity of the business, firm's age, industry as well as macroeconomic and overall market conditions. Our major 10-K characteristics for the level of disclosure is the 10-K file size from Loughran and McDonald (2014). On the one hand, 10-K file size proxies for disclosure complexity/readability, partly because a larger file size takes more efforts to understand and digest, and/or because if firms are trying to obscure mandated earnings-relevant information, they are likely to bury the results in longer documents. This is one aspect of 10-K file size that Loughran and McDonald (2014) highlight. On the other hand, as we will see in this section, 10-K file size may also be an important proxy for information content. Despite the difficulty of comprehension due to complexity, a large 10-K disclosure could contain a great amount of information, once digested, that helps to resolve the firm's uncertainty. Therefore, we expect that a larger file size is associated with a stronger hump shape of volatility dynamics following 10-K disclosure.

We test the relation between volatility dynamics and 10-K file size using panel regression in Section 3.2.1. In Section 3.2.2, we replace the 10-K file size with alternative measures of linguistic complexity commonly used in the finance and accounting literature, and compare the panel regression result with that in Section 3.2.1. In Section 3.2.3, we estimate a quadratic function between the volatility dynamics and time to learn for individual firms, and study the cross sectional link between firm-level learning speed (and its components) and 10-K disclosure characteristics.

3.2.1 Panel analysis on volatility dynamics and 10-K file size

We start by plotting the dynamics of the implied volatility following 10-K filings for firms with small and large 10-K reports. In each calendar year, firms are sorted into quintile portfolios based on 10-K file size. The implied volatilities are then averaged within a portfolio and across calendar years, and the resulting volatility dynamics are plotted in Figure 3.

Figure 3 about here

The left plot of Figure 3 shows that both firms with large 10-Ks and firms with small 10-Ks display a hump shape, but the pattern is much stronger for large 10-K firms. Following 10-K filings, the volatility for firms with large (small) 10-Ks increases by about 1.5% (0.6%) between weeks 2 and 4, but decreases to -4.5% (-1.9%) at around week 9. This pattern can also be seen in the right plot of this figure, where the difference in the volatility dynamics between these two quintile portfolios also displays a hump shape.

These patterns are also formally tested in panel regressions. In Table 3, diffs, diffl, and diffsl are regressed on the natural logarithm of 10-K file size with or without control for other firm-level characteristics, including firm size, book-to-market equity ratio, the logarithm of the implied volatility, and the CAPM alpha [-252,-6] days before 10-K-filing, and the absolute value of filing period return. The results without other control variables are reported in columns (1), (3), and (5) and the results with control variables are reported in columns (2), (4), and (6). Overall the results with and without other control variables are similar and statistically significant at the conventional levels. To understand the economic impact of file size, we focus on columns (2), (4), and (6). All else being equal, a one standard deviation increase in log 10-K file size is associated with $0.293 \times 1.265 = 0.371$ percent *increase* in implied volatility in 2-4 weeks, but $0.469 \times 1.265 = 0.593$ percent *decrease* in implied volatility in 8-10 weeks following the 10-K filing.

Table 3 about here

The positive association between 10-K file size and short-run volatility change is consistent with the readability interpretation in Loughran and McDonald (2014). Loughran and McDonald (2014) find that larger 10-Ks are significantly associated with higher return volatility in the 1-month period following 10-K filings, supporting file size as an effective measure for readability and complexity of financial disclosures. Our focus on the *dynamics* of volatility and how it relates to 10-K file size reveals another potentially important role of 10-K file size, as a proxy for information content of 10-K disclosures. Intuitively, a larger 10-K filing, on average, may be more difficult to comprehend, but could also provide more information about the financial and operational prospects of a company, and hence could help resolve more uncertainty ultimately. Our study therefore complements Loughran and McDonald (2014).

We also split the full sample into two subsamples based on the sign of the earnings surprises proceeding the 10-K filings and repeat the same analysis for the subsamples. Following Chordia and Shivakumar (2006), we define an earnings surprise as the standardized unexpected earnings (or SUE) that are defined as the earnings in quarter t less earnings in quarter $t-4$ standardized by the standard deviation of earnings changes over the last eight quarters. Note that we require non-missing data on earnings for this analysis. Therefore, the final sample size in this analysis is slightly smaller than the full sample used in Table 3 due to the availability of earnings data. The results are reported in Table 4. It is clear that the volatility dynamics patterns exist for both positive and negative earnings surprises, although the effect of the file size is slightly stronger for negative earnings surprises.

Table 4 about here

Our research design is in an event study setting. For event studies, confounding information is always an important concern; it could either drive the empirical findings or contaminate the results. Of particular relevance in our setting, the main concern about the benchmark sample is that the subsequent quarterly earnings announcements might be too close to the 10-K filing dates. The anticipation of subsequent earnings announcements may contaminate the 10-K information reflected in the volatility dynamics following 10-K disclosures. We have shown in Figure 2 that uncertainty is high before earnings announcements but drops immediately after their release. Therefore the hump-shaped volatility dynamics may be partly affected by the impact of the subsequent earnings announcement. In addition, if the two aspects of 10-K disclosures are indeed important, one should expect our main result to be even stronger by excluding firms that have too short time interval between the 10-K filing and the subsequent earnings announcement. To alleviate the impact of the confounding information due to quarterly earnings announcements, we repeat our analysis in the cross-section using only firms with no earnings announcements in at least seven weeks after 10-K filings.[14] If our main result in the cross section is driven by subsequent earnings announcements, the coefficients on 10-K file size in the panel regressions using this subsample should be much weaker.

The results are presented in Table 5. The impact of 10-K file size on the short-run volatility does not change much whereas the impact on the long-run volatility becomes much stronger. Ceteris paribus, now a one standard deviation increase in log file size is associated with $0.23 \times 1.144 = 0.263$ percent increase in implied volatility in the short run, but is associated with $1.028 \times 1.144 = 1.176$ percent decrease in implied volatility in the relatively long-run following 10-K filings.[15] The marginal impact of the 10-K file size on the resolution of uncertainty almost doubles in the screened subsample, indicating that subsequent earnings announcements are unlikely to be the major source of the pattern in the implied volatility documented in Table 3.

Table 5 about here

To further ensure that the empirical pattern of volatility dynamics following 10-K filings is not purely driven by the volatility dynamics around earnings announcements, we use simulations to construct a sample of *counterfactual* 10-K filings whose volatility dynamics are purely driven by earnings announcements, and then we analyze the volatility dynamics for this simulated sample.

The simulations are conducted in two steps. First, we simulate the volatility dynamics around earnings announcement dates based on the corresponding empirical distribution. To minimize the impact of the actual 10-K filings, we use only the earnings announcements for fiscal quarters 2, 3, and 4 to obtain the empirical distribution. In the second step, we simulate the counterfactual 10-K filing dates based on the empirical distribution of the distance between the actual 10-K filing dates and the subsequent earnings announcement dates for fiscal quarter 1. Now this counterfactual 10-K filing date is a date prior to the earnings announcement date. We use the simulated volatility dynamics starting from this counterfactual 10-K filing date to ten weeks afterwards as the volatility dynamics following a counterfactual 10-K filing.

We repeat these steps for firms at the bottom 20% 10-K file size quintile, the 20-40% file size quintile, the 40-60% file size quintile, the 60-80% file size quintile, and the 80-100% file size quintile, respectively. We obtain 5000 observations for each of the five quintiles and generate a final sample of 25,000 firm-year observations. The results of analyzing the

[14] In the pre-2006 sample, the median gap between 10-K filings and the subsequent earnings announcements is about seven weeks. Our choice of this cutoff time preserves roughly half of our earlier sample. However, our results are not driven by this particular choice and are robust to using longer cutoff times.

[15] The standard deviation of log file size is slightly different from that in the full sample (1.144 vs. 1.265).

simulated volatility dynamics are presented in Table 6. Panel A reports the average volatility dynamics for the simulated sample. The panel shows that the long-run volatility changes from the counterfactual 10-K filing dates to ten weeks after are positive, and this pattern is different from a decrease in volatility after the actual 10-K filings as in Table 2. Panel B reports the results of regressing the volatility dynamics on file size for the simulated sample. Again, the panel shows that the short-run and long-run volatility changes are different from the actual counterparts reported in Table 3. Specifically, in the counterfactual sample, the increase in short-run volatility of small 10-K firms is higher than for large 10-K firms. In the actual 10-K filing sample, the increase in short-run volatility of small 10-K firms is lower. Moreover, although the decrease in long-run volatility of large 10-K firms is more prominent than small 10-K firms in the counterfactual sample, the magnitude of the coefficient for file size (-0.07) is far smaller than that in the actual 10-K sample (-0.47). Most importantly, the coefficient of file size for short-run to long-run volatility change (*diffsl*) is positive and large in magnitude (0.77) in the actual sample whereas this coefficient is small and negative (-0.05) in the counterfactual sample. Therefore, both the average and cross-sectional volatility dynamics patterns that we find following the 10-K filings are unlikely to be mainly driven by the upcoming earnings announcements.

Table 6 about here

Besides the upcoming earnings announcements, there might be other corporate disclosures triggered by 10-K filings. In this case, a larger 10-K filing might trigger more follow-up disclosures which could generate the volatility patterns in our results. We use 8-K reports to control for these unscheduled disclosures.[16] We do so by including the frequency of 8-Ks in our analysis. Specifically, when explaining the short-run volatility changes, we include the frequency of 8-Ks in weeks two to four (denoted as #8Ks). Similarly, we include the 8-K frequency in weeks eight to ten (#8Kl) when analyzing the long-run volatility changes and we include #8Ksl(=#8Ks−#8Kl) when analyzing diffsl. The results are reported in Panel A of Table 7. The results in this panel indicate that our main findings remain robust to controlling for the 8-K disclosures.[17]

Table 7 about here

A related concern is that 10-K file size may be highly correlated with firm's structural complexity. A firm with complex business structure is likely to file large 10-K reports to

[16] http://www.sec.gov/answers/form8k.htm.

[17] In untabulated results, we also find that the short- and long-run 8-K frequency differences are not correlated with 10-K file size. This explains why our main findings are not affected by including the 8-K frequency in our analysis.

describe its operations, and our main result may be driven by this firm complexity. To address this concern, we follow Loughran and McDonald (2014) and repeat our main cross-sectional regressions by controlling for a firm complexity measure – the number of business segments from Compustat Segment Data. As reported in Panel B of Table 7, the magnitude of the coefficient on the 10-K file size slightly weakens in the short-run, but remains almost the same in the long run. Combined with the results from the robustness check on earnings announcements and 8-K disclosures, this finding suggests that other corporate disclosures and firm's structural complexity have only minimum impact on our findings, and the volatility dynamics are mainly driven by the information contained in 10-K reports.

As an additional robustness check, we redo our pooling sample regressions based on two alternative measures of volatility: intra-day and daily volatilities. We calculate the intra-day realized volatility using 5-minute intra-day returns. The daily volatility is calculated as the absolute value of the daily stock returns. Acknowledging the measurement issues with volatilities based on realized stock returns, we are able to significantly extend our sample in the cross-section, as our benchmark sample is limited by the availability of the data on implied volatility. Table 8 shows that our major findings in Table 3 and Table 5 are not specific to the use of the option implied volatility. With both alternative measures of volatility, we confirm that a larger 10-K file size corresponds to both higher short-run and lower long-run uncertainty.

Table 8 about here

3.2.2 Placebo tests: Volatility dynamics and normalized measures of readability

In this section, we study the relation between the volatility dynamics and several normalized measures of the readability and complexity of 10-K filings. We focus on three popular measures in the finance and accounting literature: Fog index, (negative) Flesch index and Kincaid index. All three measures are based on two out of three characteristics of 10-K filings: percentage of complex words, words per sentence, and syllabus per word. These indexes have been used as proxies for the difficulty that investors may have in digesting information contained in 10-K filings and analyst reports.[18] Intuitively, these indices are normalized measures, though they may capture some aspect of complexity, they do not necessarily serve as good proxies for the 10-K's information content. As a result, we predict that a less readable 10-K report (based on higher Fox index, lower Flesch index, or higher

[18]See for example, Li (2008), Bloomfield (2008), Miller (2010), Lehavy, Li, and Merkley (2011), Dougal, Engelberg, Garca, and Parsons (2012), Lawrence (2013), De Franco, Hope, Vyas, and Zhou (2013). Loughran and McDonald (2014) focus on measurement issues of these measures and propose 10-K file size as an alternative measure of 10-K linguistic complexity.

Kincaid index) is associated with an increase in short-run volatility, but does not necessarily have much effect on the long-run volatility. Therefore, this analysis can also be viewed as a placebo test.

In Table 9, we repeat the cross-sectional regressions using Fog index, Flesch index, and Kincaid index, and report the coefficients on these indices. The first two columns show that the coefficients on Fog index (Panel A) and Kincaid index (Panel C) are both positive and significant, and the coefficients on Flesch index (Panel B) are all negative. This finding is consistent with the validity of these indexes in representing the readability and complexity aspect of 10-K reports. However, the results from the middle columns indicate that a lower readability based on these normalized measures is not helpful in resolving uncertainty in the relatively long run. If anything, a higher Kincaid and a lower Flesch predict a higher long-run implied volatility after controlling for other firm characteristics. This placebo test, together with the volatility dynamics in Figure 2, again highlights the two important aspects of 10-K disclosures. The uncertainty is affected by the complexity aspect in the short run, but part of the uncertainty will be ultimately resolved by the information content aspect in the long run. The file size of 10-K reports appears to be a good proxy for both aspects.

Table 9 about here

3.2.3 Firm-level time-series analysis of the volatility dynamics and cross-sectional determinants of learning speed

In this subsection, we focus on the firm-level time-series analysis of the volatility dynamics. There are several reasons that firm-level analysis is useful and provides additional evidence to support our main predictions. First, as discussed earlier, in the panel analysis, we have to somewhat arbitrarily choose the short-run window (two to four weeks). At the firm-level, we are able to relax this restriction. We allow each firm to have a different timing for uncertainty peak level and let the data decide that for the individual firms. Second, in the panel analysis, short- and long-run volatilities are used as dependent variables in separate regressions. In some sense, they are not jointly analyzed. At the firm-level, we are able to study the volatility dynamics jointly in a more strict time-series fashion. Third, the firm-level analysis allows us to estimate the learning speed for each firm and uncover the heterogeneity of the firm-level learning speeds. Fourth, we are able to restrict our sample to better alleviate the confounding information from subsequent earnings announcements at the individual firm level, whereas in the panel analysis, imposing these restrictions results in a quite unbalanced panel for which the statistical inference becomes more problematic and challenging (e.g., Nichols and Schaffer (2007), Petersen (2009)).

Specifically, we directly model the volatility dynamics as a function of time passed since a 10-K filing date (in the spirit of a term structure) at the firm-level. To capture the hump-shaped volatility dynamics that we observe in the data, we use one of the simplest functional forms, a quadratic function:

$$\sigma_{i,t} = a_i \times t + b_i \times t^2 + \epsilon_{it}.$$

We scale $\sigma_{i,t}$ by the pre-filing volatility level so that it has the same scale and is comparable across firms. In other words, $\sigma_{i,t}$ captures the percentage change of volatility from the filing date (i.e., time 0) to time t. This functional form between volatility dynamics and time-to-learn may also be derived from a learning model (for example, Pan, Wang, and Weisbach (2014)). We define the average learning speed as the estimated average volatility change per unit of time from time 0 to a fixed time point $\bar{t}$:

$$\text{Speed}_{i,\bar{t}} = \frac{0 - \sigma_{i,\bar{t}}}{\bar{t}} = \frac{-a_i \times \bar{t} - b_i \times \bar{t}^2}{\bar{t}} = -a_i - b_i \times \bar{t}.$$

One way of interpreting this dynamics is that $-a_i$ can be viewed as the initial learning speed upon receiving a piece of new information and b_i can be viewed as the acceleration of learning. A positive a_i would suggest that uncertainty increases right after the information disclosure and vice versa. A negative b_i would suggest that uncertainty decreases over time after its peak level and vice versa. Note that here we define the average learning speed as the volatility change from time 0 to certain fixed time point, so a *positive* average learning speed reflects a *reduction* in uncertainty.

Empirically, we estimate this volatility dynamics for each individual firm i to obtain a pair of (a_i, b_i). We assume (a_i, b_i) is the same across all of the 10-K filing dates for the same firm. We also treat the observations following each 10-K filing date as independent observations for the same firm. The estimation results are reported in Table 10. To be consistent with our panel analysis, we choose the fixed time point to be eight weeks from the 10-K filing date when reporting the estimation of the average learning speed.[19]

Table 10 about here

Panel A reports the summary statistics on the average learning speed (Speed), (negative) initial learning speed (a), and the acceleration of learning (b), all in percentages. The table shows that, on average, a_i is positive (0.29%) and the acceleration of learning b_i is negative (-0.13%). This indicates that an average firm experiences an increase in uncertainty and

[19]Note that for firm-level analysis, we use bi-week as one time unit to increase the number of non-missing daily observations for calculating the average volatility measure per unit of time and as such we can reduce the noise in the volatility measure.

then the uncertainty starts to decrease over time after the uncertainty reaches a peak level at around 2.2 weeks after the 10-K filing date.[20] This is quite close to the peak week in the data as in Figure 1. This combination of negative initial learning speed and acceleration generates a positive average learning speed of 0.26% over the course of eight weeks following a 10-K filing date, an indication of net reduction in uncertainty. The inference based on median values is similar. The median value of a_i is 0.38% and the median value of b_i is -0.2%, implying a peak week at around 1.9 weeks after the filing date. The median (average) learning speed is 0.425%, indicating a reduction of 1.7% in the volatility in the 8-week period following a 10-K filing date, reasonably close to what we observe in the data.

Next we link the firm-level estimations of the average learning speed and its components (a and b) to disclosure characteristics. Since we estimate one pair of (a_i, b_i) for each company, we also take time-series average of the explanatory variables and use the average explanatory variables as the independent variables. Therefore, the multivariate regressions are purely cross-sectional. The main focus of the explanatory variables is on the 10-K file size. We include Fog, Flesch, and Kincaid indices in the analysis as well. Due to the high correlation among these three indices, we report the results separately in Panels B, C, and D respectively. In general, a_i is positively correlated with 10-K file size and the acceleration b_i is negatively correlated with 10-K file size; the average learning speed is positively correlated with 10-K file size. The results are statistically significant at the conventional levels and quite robust and consistent across different specifications. For example, Panel B reports the results when including both the 10-K file size and Fog index, as well as with and without other control variables such as market cap, book-to-market ratio, pre-filing volatilities, pre-filing CAPM alpha, and the absolute value of filing period excess return. The coefficient associated with file size is 0.398 (statistically significant at 1% level) in column (2) when the dependent variable is the average learning speed, which suggests that firms with larger 10-Ks experience a larger percentage reduction in volatilities over the 8-week period following 10-K filing dates. Column 4 of Panel B where the dependent variable is a_i shows that the coefficient for file size is 0.584 (significant at 5%), indicating larger 10-Ks are associated with more increase in uncertainty shortly after the 10-K releases. When the dependent variable is the acceleration of learning (column 6), the coefficient for file size is -0.257 (significant at 1%). This negative coefficient suggests that larger 10-Ks are associated with more rapid decreases in uncertainty after the uncertainty reaches its peak level and this is the main driver for more reduction in the uncertainty for larger 10-Ks.

[20]The maximum (minimum) point of $\sigma_t = a \times t + b \times t^2$ occurs at $t^* = -\frac{a}{2b}$. For an average firm, $b = -0.13\% < 0$ and this t^* corresponds to a maximum point $t^* = \frac{0.29}{2 \times 0.13} = 1.1$. Since in our empirical implementation we use bi-week as one time unit to reduce the measurement noise, t^* corresponds to 2.2 calendar weeks.

We also generate the estimated volatility dynamics from this quadratic function for firms in the bottom and top quintile of file sizes and plot them as a function of time in Figure 4. If we compare the estimated volatility dynamics in Figure 4 with the observed volatility dynamics in Figure 3, we can see the patterns are very similar. Therefore, this quadratic learning-speed model also does a reasonably good job of capturing the volatility dynamics for firms with different 10-K file sizes. Overall the firm-level estimation results based on a simple quadratic functional form are consistent with our panel analysis.

Figure 4 about here

3.3 An options strategy

So far, we have documented the hump-shaped dynamics of implied volatility following 10-K filings. In this section, we formally evaluate the economic impact of this pattern and explore whether this volatility dynamics provides valuable investment opportunities for an options investor. To capture the timing of the volatility changes, we use a strategy that creates a short position of a straddle in three to four weeks (short-term) after 10-K filings and closes the short straddle in weeks eight to nine (long-term). The return on such a trade for each 10-K filing is equal to:

$$\text{Ret} = \frac{\text{Price(short-term)}}{\text{Price(long-term)}} - 1. \tag{1}$$

A straddle is an options strategy in which the investor holds a position in both a call and a put with the same strike price and expiration date.[21] A long position in a straddle is betting on an increase in future volatility, and a short position is betting on a decrease. Existing literature has documented that writing a straddle unconditionally creates a both economically and statistically significant profit, even after taking into account the transaction costs. For instance, Coval and Shumway (2001) document that a simple strategy of selling equal number of at-the-money calls and puts in each month at the bid prices, purchasing the same number of out-of-the-money puts at the ask, and investing the remaining premium and principal in the index, offers a Sharpe ratio up to twice as high as that obtained by investing in the index alone.[22] Xing and Zhang (2013) find similar patterns in straddles on individual stocks. In addition, they document a large profit from taking a long position in straddles a

[21]For the rest of this study, we only report the results for the simple straddle, that is, a pair of call and put options with matching strike prices and maturity dates, due to its simplicity. We also tried the delta-neutral straddle and the result is very similar.

[22]The average straddle returns include risk premia due to volatility and jump risks. See, for example, Coval and Shumway (2001) and Cremers, Halling, and Weinbaum (2015).

few days around corporate earnings announcements. Different from these studies, we focus on a conditional options strategy following corporate 10-K filings.

Panel A of Table 11 reports the average returns of our strategy for firms sorted into quintiles by 10-K file size. We use the quintile breakpoints of 10-K file sizes from the previous year to form portfolios in the current year, so the strategy is tradable with available information. We use two ways to implement the strategy. The first method averages the straddle options prices for the days within weeks 2 - 4 and for the days within weeks 8 - 10, and the return is calculated as the percentage difference in the average prices. This implementation is equivalent to a price-weighted strategy (PW). The second strategy assumes a holding period of exactly six weeks (42 days). For example, if a straddle is written on the 21st day after a 10-K filing, the position is closed on the 63th day after the filing. We then take the average of the returns across all days within the investment windows. This corresponds to an equal-weighted strategy (EW). As reported in Panel A of Table 11, our strategy generates large profits conditional on 10-K filings. The average 6-week PW return increases from 5.84% for firms with small 10-Ks to 7.84% for firms with large 10-Ks. The difference (2.00%) is statistically significant, and economically, this corresponds to around 17.3% per year. Similar results hold for the EW strategy. The large return spread between firms with small and large 10-Ks provides additional evidence on the dynamics of the implied volatility.

Table 11 about here

One issue with the implementation of this strategy on individual stock options is trading frictions. Santa-Clara and Saretto (2009) provide evidence that limits to arbitrage, represented by transactions costs and margin requirement, have an economically important impact on the execution and the profitability of option strategies that involve writing out-of-the money put options. To alleviate this concern, we explicitly take into account of the bid-ask spread in estimating strategy returns. Specifically, we assume the realized spread to be 50% of the quoted spread, and the result for the after-cost strategy return is reported in Panel B of Table 11.[23] One interesting feature is that the transaction cost is large on average. A round-trip trading of shorting straddle removes the profitability of most strategies across different 10-K file size quintiles, and the average after-cost return of the strategies for first three quitiles now becomes negative. In contrast, the strategy return for firms with large 10-K file sizes remains positive and economically large, with an average annualized profitability

[23]This assumption is conservative, since existing studies show that the ratio of the effective spread to the quoted spread is typically less than 50%. See, for instance, Mayhew (2002) and Fontnouvelle, Fishe, and Harris (2003).

about 15%. Therefore, our strategy based on the hump-shaped volatility dynamics is still profitable for firms with large 10-K file size even controlling for the bid-ask spread.

As an alternative way to capture the effect of the transaction cost, we repeat our investment strategy using a sample with randomized filing dates as a benchmark. The "fake" filing date draws randomly and uniformly from the one year window around the true filing date, so the profitability can also be considered as a measure of the unconditional (short) straddle return with a 6-week holding period. Panel C of Table 11 presents the profitability from using the random filing dates. On average, this randomized straddle strategy also creates a large return of more than 6% per 6-week. However, the return is larger for firms with small 10-Ks than firms with bigger 10-Ks.[24] More importantly, by comparing the returns between Panel A and Panel C of Table 11, we can easily see that the cross-sectional performance of the conditional straddle strategy is completely different from that of the unconditional straddle strategy. For the conditional straddle strategy, the return difference between firms with large 10-Ks and firms with small 10-Ks is about 2% for the PW strategy and highly statistically significant. In contrast, this difference is -0.37% for the unconditional straddle strategy.[25] Therefore, the variations in the implied volatility documented in previous sections are not only statistically significant, but also provide an economically large profit from the perspective of an options investor.

3.4 Information content of 10-Ks and information asymmetry

We also examine another important capital market outcome that can be affected by information disclosures, the dynamics of information asymmetry. Similar to our main analysis of uncertainty measures in Table 3, we test how 10-K file sizes affect the short- and long-run changes in two commonly used information asymmetry measures: Amihud (2002) illiquidity measure and bid-ask spread. The results are presented in Table 12. The results in Panel A of Table 12 indicate that there is a negative association between file sizes and long-run Amihud illiquidity changes (i.e., the percentage changes of illiquidity measure 8-10 weeks following the 10-K disclosures). All else being equal, a one standard deviation increase in log 10-K file size is associated with $1.317 \times 1.265 = 1.666$ percent *decrease* in illiquidity in the long-run. However, the relation between file sizes and short-run illiquidity changes is statistically insignificant (i.e., the percentage changes 2-4 weeks following the 10-K disclosures). When we use bid-ask spread as the measure for information asymmetry, we find similar results (Panel B of Table 12). All else being equal, a one standard deviation increase

[24] This is perhaps not surprising since file size has a 26.5% correlation with firm size, and illiquidity is more of a concern for trading options on small firms than on big firms.

[25] The inferences from the EW strategies are similar.

in log 10-K file size is associated with $0.834 \times 1.265 = 1.055$ percent *decrease* in bid-ask spread in the long-run. The relation between file sizes and the short-run changes in bid-ask spread is again insignificant.

Table 12 about here

The results based on information asymmetry measures lend further support to the main findings of our paper: file size can proxy for the amount of information content in 10-K disclosures, and its effect on the resolution of uncertainty becomes prominent in a relatively long horizon (8-10 weeks following the disclosure). The results are also consistent with the complexity aspect of 10-K disclosures, although the complexity aspect does not change information asymmetry significantly in the short-run. This suggests that even the informed investors do not fully process the 10-K information in a short period of a couple weeks to generate an information advantage.

4 Conclusion

Investor learning about corporate attributes affects the perceived riskiness of the corporate securities (e.g., Pan, Wang, and Weisbach (2014), Pan, Wang, and Weisbach (2015), and Giglio and Shue (2014)). Our paper adds to this emerging line of research by studying how investor learning over time about the annual financial reports affects the evolution of investors' perception of uncertainty. Our comprehensive analysis of the volatility dynamics in the event window following 10-K filings uncovers two novel findings.

First, we find that for an average firm, there exists a hump-shaped volatility dynamics – the volatility initially increases by 0.36% in the first two to four weeks after 10-K filings, but decreases by 2.55% in the subsequent weeks up to the next quarterly earnings announcement. Second and more importantly, this hump shape is much stronger for firms with larger 10-Ks. All else equal, a one-standard deviation increase in 10-K file size is associated with a 0.37 percent increase in volatility in the short run, but 0.59 percent decrease in volatility in the long run. The economic impact of our findings is nontrivial: a strategy of shorting straddles in this time frame for firms across 10-K file size quintiles delivers up to 17% cross-sectional differences in annualized returns. Moreover, we find that firms with larger 10-Ks experience more reduction in information asymmetry in the long horizon, which reinforces the information content aspect of 10-K disclosures.

Our focus on the volatility dynamics uncover the information role of 10-K filings in a relatively longer horizon; this complements the result by Loughran and McDonald (2014). Our central findings suggest that consistent with the fundamental goal of annual reports, by and large, a larger 10-K filing carries more information content, although the complexity aspect of a larger report can lead to a higher uncertainty before the information is fully digested by investors. Our results, therefore, suggest that the dual roles of 10-K disclosures – complexity and information content of the disclosures – can be revealed through investor learning over *time*.

References

Barth, Mary E., Travis L. Johnson, and Eric C. So, 2011, Dynamics of earnings announcement news: Evidence from option prices, Working paper.

Berger, Philip, Jason Chen, and Feng Li, 2012, Firm specific information and the cost of equity capital, Working Paper.

Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor, 2001, Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns, *Journal of Econometrics* 105, 5–26.

Bloomfield, Robert, 2008, Discussion of annual report readability, current earnings, and earnings persistence?, *Journal of Accounting and Economics* 45, 248–252.

Botosan, Christine A., 1997, Disclosure level and the cost of equity capital, *The Accounting Review* 72, 323–349.

——— , and Marlene Plumlee, 2002, A re-examination of disclosure level and the expected cost of equity capital, *Journal of Accounting Research* 40, 21–40.

Brennan, Michael J., and Yihong Xia, 2001, Stock price volatility and equity premium, *Journal of Monetary Economics* 47, 249 – 283.

Busch, Thomas, Bent Jesper Christensen, and Morten Orregaard Nielsen, 2011, The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets, *Journal of Econometrics* 160, 48–57.

Bushee, Brian J., Ian D. Gow, and Daniel J. Taylor, 2013, Linguistic complexity in firm disclosures: Obfuscation or information?, Working Paper.

Canina, Linda, and Stephen Figlewski, 1993, The informational content of implied volatility, *The Review of Financial Studies* 6, 659–681.

Chang, Bo-Young, Peter Christoffersen, Kris Jacobs, and Gregory Vainberg, 2012, Option-implied measures of equity risk, *Review of Finance* 16, 385–428.

Chordia, Tarun, and Lakshmanan Shivakumar, 2006, Earnings and price momentum, *Journal of Financial Economics* 80, 627 – 656.

Christensen, B.J., and N.R. Prabhala, 1998, The relation between implied and realized volatility, *Journal of Financial Economics* 50, 125–150.

Coval, Joshua D., and Tyler Shumway, 2001, Expected option returns, *The Journal of Finance* 56, 983–1009.

Cremers, Martijn, Michael Halling, and David Weinbaum, 2015, Aggregate jump and volatility risk in the cross-section of stock returns, *The Journal of Finance* 70, 577–614.

Day, Theodore E., and Craig M. Lewis, 1992, Stock market volatility and the information content of stock index options, *Journal of Econometrics* 52, 267–287.

De Franco, Gus, Ole-Kristian Hope, Dushyantkumar Vyas, and Yibin Zhou, 2013, Analyst report readability, *Contemporary Accounting Research*, forthcoming.

Dougal, Casey, Joseph Engelberg, Diego Garca, and Christopher A. Parsons, 2012, Journalists and the stock market, *Review of Financial Studies* 25, 639–679.

Easley, David, and Maureen O'Hara, 2009, Ambiguity and nonparticipation: The role of regulation, *Review of Financial Studies* 22, 1817–1843.

Epstein, Larry G., and Martin Schneider, 2007, Learning under ambiguity, *Review of Economic Studies* 74, 1275–1303.

Fleming, Jeff, 1998, The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance* 5, 317–345.

Fontnouvelle, Patrick De, Raymond P. H. Fishe, and Jeffrey H. Harris, 2003, The Behavior of Bid-Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999, *Journal of Finance* 58, 2437–2464.

Giglio, Stefano, and Kelly Shue, 2014, No news is news: Do markets underreact to nothing?, *Review of Financial Studies* 27, 3389–3440.

Govindaraj, Suresh, Sangsang Liu, and Joshua Livnat, 2012, The post earnings announcement drift and option traders, Working paper.

Goyal, Amit, and Alessio Saretto, 2009, Cross-section of option returns and volatility, *Journal of Financial Economics* 94, 310 – 326.

Griffin, Paul A., 2003, Got information? Investor response to form 10-K and form 10-Q EDGAR filings, *Review of Accounting Studies* 8, 433–460.

Jiang, George J., and Yisong S. Tian, 2005, The model-free implied volatility and its information content, *Review of Financial Studies* 18, 1305–1342.

Kelly, Bryan, Lubos Pastor, and Pietro Veronesi, 2014, The price of political uncertainty: Theory and evidence from the option market, Working Paper.

Kothari, S. P., Xu Li, and James E. Short, 2009, The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis, *The Accounting Review* 84, 1639–1670.

Lamoureux, Christopher G., and William D. Lastrapes, 1993, Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities, *The Review of Financial Studies* 6, 293–326.

Lang, Mark H., and Russell J. Lundholm, 1996, Corporate disclosure policy and analyst behavior, *The Accounting Review* 71, 467–492.

Lawrence, Alastair, 2013, Individual investors and financial disclosure, *Journal of Accounting and Economics* 56, 130–147.

Lehavy, Reuven, Feng Li, and Kenneth Merkley, 2011, The effect of annual report readability on analyst following and the properties of their earnings forecasts, *The Accounting Review* 86, 1087–1115.

Leuz, Christian, and Catherine Schrand, 2009, Disclosure and the cost of capital: Evidence from firms' responses to the enron shock, Working Paper.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

——— , 2010, Texutal analysis of corporate disclosures: A survey of the literature, *Journal of Accounting Literature* 29, 143–165.

Loughran, Tim, and Bill McDonald, 2014, Measuring readability in financial disclosures, *Journal of Finance*, forthcoming.

Mayhew, Stewart, 2002, Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets, *Journal of Finance* 57, 931–958.

McNichols, Maureen, and James G. Manegold, 1983, The effect of the information environment on the relationship between financial disclosure and security price variability, *Journal of Accounting and Economics* 5, 49–74.

Miller, Brian P., 2010, The effects of reporting complexity on small and large investor trading, *The Accounting Review* 85, 2107–2143.

Nichols, Austin, and Mark E Schaffer, 2007, Clustered standard errors in Stata, United Kingdom Stata Users' Group Meetings 2007 07 Stata Users Group.

Pan, Yihui, Tracy Yue Wang, and Michael S. Weisbach, 2014, Learning about CEO ability and stock return volatility, *Review of Financial Studies* 28, 1623–1666.

——— , 2015, Management risk and the cost of borrowing, Working Paper.

Pastor, Lubos, and Pietro Veronesi, 2003, Stock valuation and learning about profitability, *The Journal of Finance* 58, pp. 1749–1789.

——— , 2009, Learning in financial markets, *Annual Review of Financial Economics* 1, 361–381.

Peng, Lin, and Wei Xiong, 2003, Time to digest and volatility dynamics, Working Paper.

Petersen, Mitchell A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Poon, Ser-Huang, and Clive W.J. Granger, 2003, Forecasting volatility in financial markets: A review, *Journal of Economic Literature* 41, 478–539.

Rogers, Jonathan L., Douglas J. Skinner, and Andrew Van Buskirk, 2009, Earnings guidance and market uncertainty, *Journal of Accounting and Economics* 48, 90–109.

Santa-Clara, Pedro, and Alessio Saretto, 2009, Option strategies: Good deals and margin calls, *Journal of Financial Markets* 12, 391–417.

Timmermann, Allan G., 1993, How learning in financial markets generates excess volatility and predictability in stock prices, *The Quarterly Journal of Economics* 108, pp. 1135–1145.

Xing, Yuhang, and Xiaoyan Zhang, 2013, Anticipating uncertainty: Straddles around earnings announcements, Working paper.

You, Haifeng, and Xiao-jun Zhang, 2009, Financial reporting complexity and investor underreaction to 10-K information, *Review of Accounting Studies* 14, 559–586.

Table 1: Summary statistics and correlation coefficient matrix for the benchmark sample

This table reports the summary statistics and correlation coefficients for the benchmark sample from 1996 to 2012. The variables include: option implied volatility (IV), the natural logarithm of the 10-K file size from Loughran and McDonald (2014) (Fsize), the Fog (Flesch, Kincaid) index from Li (2008) (Fog, Flesch, Kincaid), the natural logarithm of firm's market cap (Size), the natural logarithm of firm's book-to-market equity ratio (BM), the alpha from a market model using [-252,-6] days before 10-K disclosure (Pre-alpha), and the absolute value of filing date excess return (Abs-filing-ret). Panel A reports mean, standard deviation, bottom and upper quartiles, and median of each variable. Panel B reports the correlation coefficient matrix.

Panel A: Summary statistics

Variable	Mean	Std Dev	Q1	Median	Q3
IV	0.461	0.225	0.302	0.407	0.564
Fsize	0.317	1.265	-0.650	0.294	1.011
Fog	19.594	2.094	18.670	19.518	20.480
Flesch	22.313	5.230	19.720	22.581	25.339
Kincaid	15.639	1.956	14.648	15.479	16.473
Size	7.723	1.563	6.581	7.597	8.746
BM	-1.025	0.869	-1.517	-0.957	-0.458
Pre-alpha	0.041	0.240	-0.083	0.026	0.137
Abs-filing-ret	3.075	3.942	0.801	1.838	3.791

Panel B: Correlations

	IV	Fsize	Fog	Flesch	Kincaid	Size	BM	Pre-alpha
Fsize	-0.290							
Fog	-0.013	0.151						
Flesch	0.062	-0.274	-0.480					
Kincaid	-0.023	0.237	0.968	-0.566				
Size	-0.490	0.265	0.035	-0.077	0.058			
BM	-0.189	0.208	0.032	-0.051	0.032	-0.071		
Pre-alpha	0.154	-0.074	-0.014	0.022	-0.015	0.048	0.015	
Abs-filing-ret	0.434	-0.120	0.000	0.029	-0.007	-0.225	-0.073	0.059

Table 2: Average short-run and long-run changes in implied volatility following 10-K disclosure

This table reports the dynamics of option implied volatility in the (relatively) short and long horizons following 10-K disclosures. All the volatility changes are relative to pre-filing volatility level; diffs is the average percentage change in the implied volatility 2-4 weeks after the 10-K disclosure; diffl is the average percentage change in the implied volatility 8-10 weeks after the 10-K disclosure; diffsl is equal to diffs minus diffl. Each panel presents the mean diffs, diffl, and diffsl, and the robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. Panel A reports the result for the full sample between 1996 and 2012. Panel B reports the result for the early sample between 1996 and 2003, and Panel C reports the result for the late sample between 2004 and 2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Panel A: Full sample 1996-2012

	(1) diffs	(2) diffl	(3) diffsl
Mean	0.358***	-2.193***	2.548***
	(0.06)	(0.10)	(0.08)

Panel B: Early sample 1996-2003

	(1) diffs	(2) diffl	(3) diffsl
Mean	0.371***	-3.508***	3.889***
	(0.10)	(0.16)	(0.14)

Panel C: Late sample 2004-2012

	(1) diffs	(2) diffl	(3) diffsl
Mean	0.349***	-1.351***	1.690***
	(0.08)	(0.12)	(0.10)

Table 3: Effect of 10-K file size on the dynamics of implied volatility following 10-K disclosure (Full sample)

This table reports the effect of 10-K file size (Fsize) on the dynamics of option implied volatility in the (relatively) short and long horizons following 10-K disclosures. The dependent variables include the average percentage change in the implied volatility 2-4 weeks after the 10-K disclosure (diffs), the average percentage change in the implied volatility 8-10 weeks after the 10-K disclosure (diffl), and diffsl which equals diffs minus diffl. The main independent variable of interest is the natural logarithm of the 10-K file size from Loughran and McDonald (2014). Other control variables include the logarithm of firm's market cap (Size), the logarithm of book-to-market equity ratio (BM), the logarithm of the pre-filing implied volatility (Pre-IV), the CAPM alpha [-252,-6] days before 10-K disclosure (Pre-alpha), the absolute value of filing period excess return (Abs-filing-ret), and calendar year dummy variables (not reported). The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1996-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.531***	0.293***	-0.256**	-0.469***	0.790***	0.772***
	(0.08)	(0.09)	(0.13)	(0.15)	(0.11)	(0.13)
Size		0.290***		-0.080		0.365***
		(0.06)		(0.09)		(0.08)
BM		-0.166**		-0.205		0.035
		(0.08)		(0.13)		(0.11)
Pre-IV		-3.102***		-7.924***		4.363***
		(0.52)		(0.82)		(0.70)
Pre-alpha		2.917***		6.863***		-4.121***
		(0.34)		(0.52)		(0.43)
Abs-filing-ret		0.161***		0.091***		0.058**
		(0.03)		(0.03)		(0.03)
N	22,406	18,684	21,836	18,279	20,484	17,488
R-sq	0.083	0.106	0.145	0.171	0.204	0.223

Table 4: Effect of 10-K file size on volatility dynamics for positive and negative earnings surprises

We split the full sample into two subsamples based on the sign of the earnings surprises prior to the 10-K filings, a subsample for positive earnings surprises (P) and a subsample for negative earnings surprises (N). This table reports the effect of 10-K file size (Fsize) on the dynamics of option implied volatility in the (relatively) short and long horizons following 10-K disclosures for these two subsamples. Earnings surprise is defined as standardized unexpected earnings (or SUE) which is defined as the earnings in quarter t less earnings in quarter t-4 standardized by the standard deviation of earnings changes over the last eight quarters. The dependent variables include the average percentage change in the implied volatility 2-4 weeks after the 10-K disclosure (diffs), the average percentage change in the implied volatility 8-10 weeks after the 10-K disclosure (diffl), and diffsl which equals diffs minus diffl. The main independent variable of interest is the natural logarithm of the 10-K file size from Loughran and McDonald (2014). Other control variables include the logarithm of firm's market cap (Size), the logarithm of book-to-market equity ratio (BM), the logarithm of the pre-filing implied volatility (Pre-IV), the CAPM alpha [-252,-6] days before 10-K disclosure (Pre-alpha), the absolute value of filing period excess return (Abs-filing-ret), and calendar year dummy variables (not reported). The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1996-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Positive (P) or Negative (N) earnings surprises						
	P	N	P	N	P	N
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.260**	0.376**	-0.456**	-0.527**	0.720***	0.965***
	(0.11)	(0.16)	(0.18)	(0.26)	(0.16)	(0.22)
Size	0.209***	0.382***	-0.106	-0.094	0.296***	0.467***
	(0.07)	(0.10)	(0.11)	(0.15)	(0.10)	(0.12)
BM	-0.099	-0.257*	-0.161	-0.207	0.007	0.034
	(0.10)	(0.15)	(0.17)	(0.22)	(0.14)	(0.19)
Pre-IV	-4.295***	-2.318***	-10.480***	-5.539***	5.516***	3.361***
	(0.68)	(0.83)	(1.07)	(1.29)	(0.92)	(1.06)
Pre-alpha	3.771***	2.251***	9.405***	4.194***	-5.907***	-1.905***
	(0.46)	(0.55)	(0.72)	(0.84)	(0.60)	(0.73)
Abs-filing-ret	0.113***	0.234***	0.059	0.143***	0.057	0.063
	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)
N	10,956	7,005	10,769	6,858	10,346	6,530
R-sq	0.116	0.105	0.160	0.182	0.198	0.256

Table 5: Effect of 10-K file size on the dynamics of implied volatility following 10-K disclosure (Subsample with no earnings announcements in at least 7 weeks after disclosure)

This table reports the effect of 10-K file size (Fsize) on the dynamics of option implied volatility in the (relatively) short and long horizons following 10-K disclosures for the subsample of observations with no earnings announcement in at least 7 weeks after the disclosure. The dependent variables include the average percentage change in the implied volatility 2-4 weeks after the 10-K disclosure (diffs), the average percentage change in the implied volatility 8-10 weeks after the 10-K disclosure (diffl), and diffsl which equals diffs minus diffl. The main independent variable of interest is the natural logarithm of the 10-K file size from Loughran and McDonald (2014). Other control variables include the logarithm of firm's market cap (Size), the logarithm of book-to-market equity ratio (BM), the logarithm of the pre-filing implied volatility (Pre-IV), the CAPM alpha [-252,-6] days before 10-K disclosure (Pre-alpha), the absolute value of filing period excess return (Abs-filing-ret), and calendar year dummy variables (not reported). The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1996-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.846***	0.230**	-0.546***	-1.028***	1.402***	1.247***
	(0.10)	(0.12)	(0.15)	(0.19)	(0.14)	(0.17)
Size		0.426***		-0.008		0.453***
		(0.08)		(0.12)		(0.10)
BM		0.116		-0.082		0.196
		(0.11)		(0.17)		(0.14)
Pre-IV		-5.068***		-9.663***		4.318***
		(0.74)		(1.12)		(0.94)
Pre-alpha		1.873***		5.566***		-3.760***
		(0.51)		(0.77)		(0.64)
Abs-filing-ret		0.126***		0.082*		0.050
		(0.03)		(0.04)		(0.04)
N	12,593	10,963	12,385	10,806	11,704	10,391
R-sq	0.114	0.144	0.131	0.151	0.215	0.229

Table 6: Analysis of volatility dynamics based on a simulated sample of counterfactual 10-K filings

This table reports the results of analyzing volatility dynamics for a counterfactual sample of 10-K filings. We first simulate the volatility dynamics around earnings announcement dates based on the corresponding empirical distribution. To minimize the impact of actual 10-K filings, we use only the earnings announcements for fiscal quarters 2, 3, and 4 to obtain the empirical distribution of volatility dynamics around earnings announcements. Then we simulate the counterfactual 10-K filing date based on the empirical distribution of the distance between the actual 10-K filing dates and the subsequent fiscal quarter 1 earnings announcement dates. These two steps allow us to obtain a counterfactual volatility dynamics of 10-K filings whose impact is mainly generated by earnings announcements. We repeat these steps for firms in the bottom 20% file size quintile, the 20-40% file size quintile, the 40-60% file size quintile, the 60-80% file size quintile, and the 80-100% file size quintile, respectively. We obtain 5000 observations for each of the five quintiles and generate a final sample of 25,000 firm-year observations. Panel A reports the average volatility dynamics for this simulated sample. Panel B reports the results of regressing the volatility dynamics on file size for this simulated sample. The definitions of *diffs*, *diffl*, and *diffsl* are the same as in previous tables.

Panel A: Average volatility dynamics for the simulated sample

	(1) diffs	(2) diffl	(3) diffsl
Mean	0.265***	0.227***	0.037***
	(0.005)	(0.004)	(0.004)

Panel B: Volatility dynamics and 10-K file size for the simulated sample

Dependent variable	(1) diffs	(2) diffl	(3) diffsl
Fsize	-0.121***	-0.070***	-0.051***
	(0.004)	(0.003)	(0.003)
N	25,000	25,000	25,000
R-sq	0.042	0.019	0.010

Table 7: Controlling for 8-K filings and business segment

This table reports the effect of 10-K file size (Fsize) on the dynamics of option implied volatility in the (relatively) short and long horizons following 10-K disclosures, after controlling for the number of corporate 8-K filings (Panel A) and the number of business segments (Panel B). The dependent variables include the average percentage change in the implied volatility 2-4 weeks after the 10-K disclosure (diffs), the average percentage change in the implied volatility 8-10 weeks after the 10-K disclosure (diffl), and diffsl which equals diffs minus diffl. The main independent variable of interest are the natural logarithm of the 10-K file size from Loughran and McDonald (2014), the number of 8-K filings 2-4 weeks after 10-K disclosure (#8Ks), the number of 8-K filings 8-10 weeks after the 10-K disclosure (#8Kl), and the difference between #8Ks and #8Kl (#8Ksl) (Panel A), and the number of business segments (Panel B), with or without control for other firm characteristics. All regressions include calendar year dummy variables. See Table 3 for the definitions of other control variables. The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1996-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Panel A: Controlling for number of 8-K filings

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.519***	0.284***	-0.280**	-0.488***	0.789***	0.771***
	(0.08)	(0.09)	(0.13)	(0.15)	(0.11)	(0.13)
#8Ks	0.093	0.116*				
	(0.06)	(0.07)				
#8Kl			0.196**	0.252**		
			(0.09)	(0.10)		
#8Ksl					0.072	0.103
					(0.07)	(0.07)
Control	No	Yes	No	Yes	No	Yes
N	22,406	18,684	21,836	18,279	20,484	17,488
R-sq	0.083	0.106	0.146	0.171	0.205	0.223

Panel B: Controlling for number of business segments

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.411***	0.252**	-0.276*	-0.475***	0.692***	0.762***
	(0.09)	(0.10)	(0.15)	(0.17)	(0.12)	(0.14)
# of Segment	0.057*	0.000	-0.020	-0.116**	0.066*	0.092**
	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)	(0.04)
Control	No	Yes	No	Yes	No	Yes
N	19,681	16,273	19,128	15,894	17,935	15,189
R-sq	0.076	0.098	0.134	0.157	0.189	0.205

Table 8: Effect of 10-K file size on the volatility dynamics following 10-K disclosure: Alternative volatility measures

This table reports the effect of 10-K file size (Fsize) on the dynamics of volatility in the short run and long run following 10-K disclosures using alternative proxies for volatility. Panel A uses measures (diffs, diffl, diffsl) based on intra-day realized volatilities calculated using 5-minute intra-day returns. Panel B uses measures (diffs, diffl, diffsl) based on daily volatilities calculated as the absolute value of daily stock returns. In each specification, the percentage change in volatility following disclosures (diffs, diffl, and diffsl) are regressed on the natural logarithm of the 10-K file size, with or without control for other firm characteristics. All regressions include calendar year dummy variables. See Table 3 for the definitions of other control variables. The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1994-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Panel A: Intra-day volatility measure

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.900**	1.079**	-1.099**	-1.940***	1.778***	2.868***
	(0.42)	(0.45)	(0.46)	(0.49)	(0.38)	(0.41)
Control	No	Yes	No	Yes	No	Yes
N	61,928	55,414	61,300	54,853	61,093	54,654
R-sq	0.027	0.035	0.028	0.038	0.081	0.083

Panel B: Daily volatility measure

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.979***	0.752***	-0.626***	-1.247***	1.637***	2.048***
	(0.25)	(0.26)	(0.22)	(0.23)	(0.26)	(0.28)
Control	No	Yes	No	Yes	No	Yes
N	71,325	64,390	71,335	64,400	71,310	64,376
R-sq	0.030	0.038	0.023	0.035	0.046	0.047

Table 9: Effect of alternative proxies for 10-K complexity on dynamics of implied volatility following 10-K disclosure

This table reports the effect of alternative proxies for 10-K complexity on dynamics of implied volatility following 10-K disclosure. These proxies are: the Fog index from Li (2008) (Panel A), the Flesch index (Panel B), which is calculated as 206.835-(1.015 $\times$ words per sentence)- (84.6 $\times$ syllables per word), and the Kincaid index (Panel C) defined as (11.8 $\times$ syllables per word) + (0.39 $\times$ words per sentence)-15.59. The regressions replace the Fsize by these complexity proxies in the regression of Table 3 and reports regression coefficients on these proxies with or without control variables described in Table 3. All regressions include calendar year dummy variables. The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1996-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Panel A: Fog index

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fog	0.174***	0.179***	0.046	0.120	0.148***	0.068
	(0.04)	(0.05)	(0.07)	(0.07)	(0.06)	(0.07)
Control	No	Yes	No	Yes	No	Yes
N	16,923	14,198	16,564	13,954	15,532	13,351
R-sq	0.079	0.103	0.150	0.176	0.206	0.226

Panel B: Flesch index

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Flesch	-0.071***	-0.069***	-0.037	-0.053**	-0.027	-0.018
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Control	No	Yes	No	Yes	No	Yes
N	16,923	14,198	16,564	13,954	15,532	13,351
R-sq	0.079	0.103	0.150	0.176	0.206	0.225

Panel C: Kincaid index

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Kincaid	0.189***	0.193***	0.102	0.179**	0.117**	0.029
	(0.04)	(0.05)	(0.07)	(0.07)	(0.06)	(0.07)
Control	No	Yes	No	Yes	No	Yes
N	16,923	14,198	16,564	13,954	15,532	13,351
R-sq	0.079	0.103	0.150	0.176	0.206	0.225

Table 10: Estimation results for firm-level learning speed model

This table reports the estimation results from the firm-level learning speed model. For each firm, we estimate a quadratic function for the dynamics of implied volatility following 10-K disclosures as $\sigma_{i,t} = a_i \times t + b_i \times t^2 + \epsilon_{it}$, where t is the number of (bi) weeks from a 10-K filing date and $t = 0$ corresponds to the 10-K filing date; $\sigma_{i,t}$ is the percentage change of volatility from the filing date (i.e., time 0) to time t. We define the learning speed as the estimated average volatility change per unit of time from time 0 to a fixed time point $\bar{t}$: $\text{Speed}_{i,\bar{t}} = \frac{0-\sigma_{i,\bar{t}}}{\bar{t}} = \frac{-a_i \times \bar{t} - b_i \times \bar{t}^2}{\bar{t}} = -a_i - b_i \times \bar{t}$. To be consistent with our panel analysis, we choose the fixed time point to be 8 weeks from the 10-K filing date to report the results associated with the average learning speed. The definitions of the other variables are the same as in previous tables with the only exception that here we take the time series average for the firm-level analysis. Panel A reports the summary statistics on the average learning speed, a, and b. Panel B (C, D) reports the results of regressing average learning speed, a, and b on 10-K file size, Fog (Flesch, Kincaid) index, and other control variables as in previous tables. The robust standard errors are reported in the parentheses below each coefficient estimate. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Panel A: Summary Statistics

	N	Mean	Std Dev	Q1	Median	Q3
Speed	1242	0.261	2.380	-0.810	0.425	1.440
a	1242	0.293	5.184	-2.410	0.380	2.820
b	1242	-0.134	1.259	-0.780	-0.200	0.480

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Speed	Speed	a	a	b	b
Panel B						
Fsize	0.431***	0.398***	0.970***	0.584**	-0.359***	-0.257***
	(0.12)	(0.13)	(0.26)	(0.28)	(0.06)	(0.07)
Fog	0.045	0.047	0.072	0.060	-0.031	-0.029
	(0.06)	(0.06)	(0.13)	(0.13)	(0.03)	(0.03)
Panel C						
Fsize	0.443***	0.408***	0.964***	0.550*	-0.361***	-0.251***
	(0.12)	(0.13)	(0.27)	(0.29)	(0.06)	(0.07)
Flesch	0.002	-0.001	-0.023	-0.044	0.005	0.011
	(0.02)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)
Panel D						
Fsize	0.441***	0.409***	0.957***	0.559**	-0.359***	-0.253***
	(0.12)	(0.13)	(0.26)	(0.28)	(0.06)	(0.07)
Kincaid	0.000	0.001	0.094	0.114	-0.023	-0.029
	(0.06)	(0.06)	(0.14)	(0.14)	(0.03)	(0.03)
Control	No	Yes	No	Yes	No	Yes

Table 11: An options strategy

This table reports the profitability of an option investment strategy. Panel A reports the summary statistics in returns of a strategy with a short position in a straddle during 2 to 4 weeks after the 10-K filing that is closed in weeks 8 to 10. We report the results for both price-weighted and equal-weighted strategies. Panel B reports the summary statistics of the returns of the same strategy, but taking into account of the bid-ask spread. Panel C reports the summary statistics in returns of the same strategy, but using "fake" 10-K filing dates. Specifically, we assign a random date between six months before and six months after the true filing date as if this were the filing date, and then repeat the strategy implementation. P10 and P90 represent the 10th and 90th percentiles for each variable. The difference in variables between the portfolio with large 10-K file size and the portfolio with small 10-K file size as well as the t-statistics from the t-test is reported in the bottom of each panel. The sample is from 1996 to 2012.

Panel A: Strategy returns for actual 10-K filings sample (%)

	PW				EW			
Fsize	Mean	Std	P10	P90	Mean	Std	P10	P90
Small	5.84	28.05	-30.31	38.54	6.40	18.98	-17.92	25.58
2	5.83	27.66	-29.89	38.85	6.17	19.23	-18.19	24.87
3	5.64	28.64	-30.61	38.88	6.12	19.34	-18.27	24.47
4	6.48	28.59	-29.38	39.57	6.93	19.17	-16.20	24.42
Large	7.84	28.86	-27.52	41.78	7.85	17.22	-12.78	24.56
Diff	2.00				1.45			
t-stat	(6.72)				(7.71)			

Panel B: Strategy returns accounting for bid-ask spread (%)

	PW				EW			
Fsize	Mean	Std	P10	P90	Mean	Std	P10	P90
Small	-2.58	26.28	-36.67	28.66	-3.20	18.48	-28.11	16.34
2	-1.34	26.27	-35.73	30.23	-2.32	18.36	-26.52	16.98
3	-0.98	26.81	-35.47	30.46	-1.82	17.91	-26.18	16.84
4	0.13	27.01	-34.23	31.93	-0.87	17.73	-24.14	17.36
Large	1.72	27.40	-32.56	34.59	0.47	16.98	-21.05	17.98
Diff	4.30				3.66			
t-stat	(15.25)				(19.73)			

Panel C: Strategy returns for a sample with "fake" randomized 10-K filing dates (%)

	PW				EW			
Fsize	Mean	Std	P10	P90	Mean	Std	P10	P90
Small	6.62	27.02	-28.01	38.66	6.91	17.88	-15.65	24.14
2	5.61	26.95	-29.16	37.15	6.62	18.60	-15.70	24.03
3	5.75	27.56	-29.18	37.88	6.32	18.11	-16.42	23.81
4	6.08	27.64	-28.81	38.33	6.45	17.82	-15.79	23.73
Large	6.25	28.21	-29.17	40.02	6.49	16.89	-14.96	22.97
Diff	-0.37				-0.41			
t-stat	(-1.28)				(2.26)			

Table 12: Effect of 10-K file size on dynamics of information asymmetry following 10-K disclosure

This table reports the effect 10-K file size on dynamics of information asymmetry following 10-K disclosure. We use two measures of information asymmetry from the literature. The first measure is the Amihud illiquidity measure (Panel A), defined as

$$\text{Illiquidity} = \frac{|\text{ret}|}{\text{vol}} \times 10^6,$$

where ret is the daily stock return and vol is the daily trading volume. The second measure is the bid-ask spread (Panel B), defined as

$$\text{Spread} = \frac{(\text{ask} - \text{bid})}{0.5 \times (\text{ask} + \text{bid})} \times 100,$$

where the daily ask and bid prices are measured at closing. In each specification, the percentage change in the information asymmetry measure following disclosures (diffs, diffl, and diffsl) are regressed on the natural logarithm of the 10-K file size, with or without control for other firm characteristics. All regressions include calendar year dummy variables. See Table 3 for the definitions of other control variables. The robust standard errors clustered at the firm level are reported in parentheses below each coefficient estimate. The sample period is 1994-2012. Statistical significance levels of 1%, 5%, and 10% are indicated with ***, **, and * respectively.

Panel A: Amihud illiquidity

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	-0.474	0.055	-2.296***	-1.317**	1.874***	1.462***
	(0.45)	(0.50)	(0.51)	(0.58)	(0.39)	(0.44)
Control	No	Yes	No	Yes	No	Yes
N	69,134	62,261	68,572	61,726	68,121	61,389
R-sq	0.013	0.015	0.036	0.037	0.039	0.039

Panel B: Bid-ask spread

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	diffs	diffs	diffl	diffl	diffsl	diffsl
Fsize	0.003	0.180	-1.406***	-0.834***	1.424***	1.059***
	(0.24)	(0.27)	(0.27)	(0.31)	(0.20)	(0.23)
Control	No	Yes	No	Yes	No	Yes
N	67,877	60,719	67,410	60,283	67,051	59,960
R-sq	0.018	0.019	0.056	0.056	0.040	0.040

Figure 1: Average dynamics of implied volatility following 10-K disclosure

This figure plots the dynamics of the option implied volatility following 10-K disclosures for an average firm. The percentage change of the implied volatility from the filing date is plotted against the number of weeks after the 10-K disclosure. The left panel plots the full sample 1996-2012, the middle panel plots the early sample 1996-2003, and the right panel plots late sample 2004-2012.

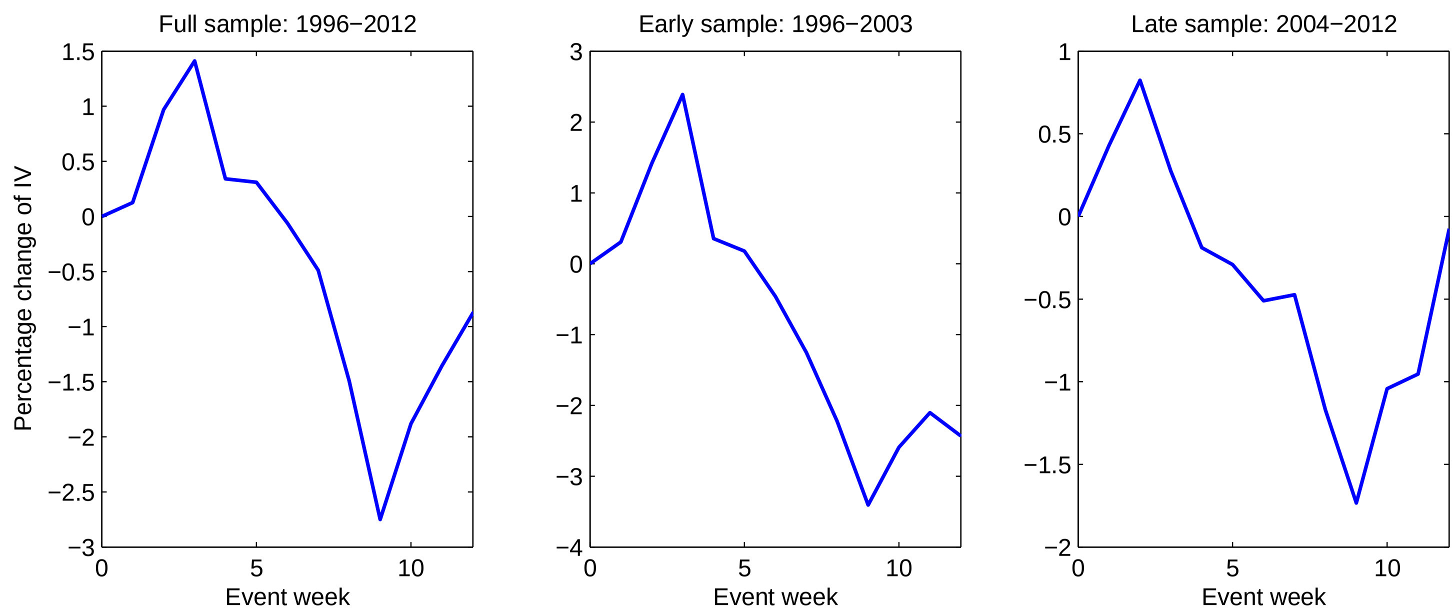

Figure 2: Dynamics of implied volatility around earnings announcements and 10-K filings

This figure compares the dynamics of the option implied volatility around earnings announcement (left panel) and 10-K disclosures (right panel). The option implied volatility is plotted against the number of trading days around these events. The sample is from 1996-2012.

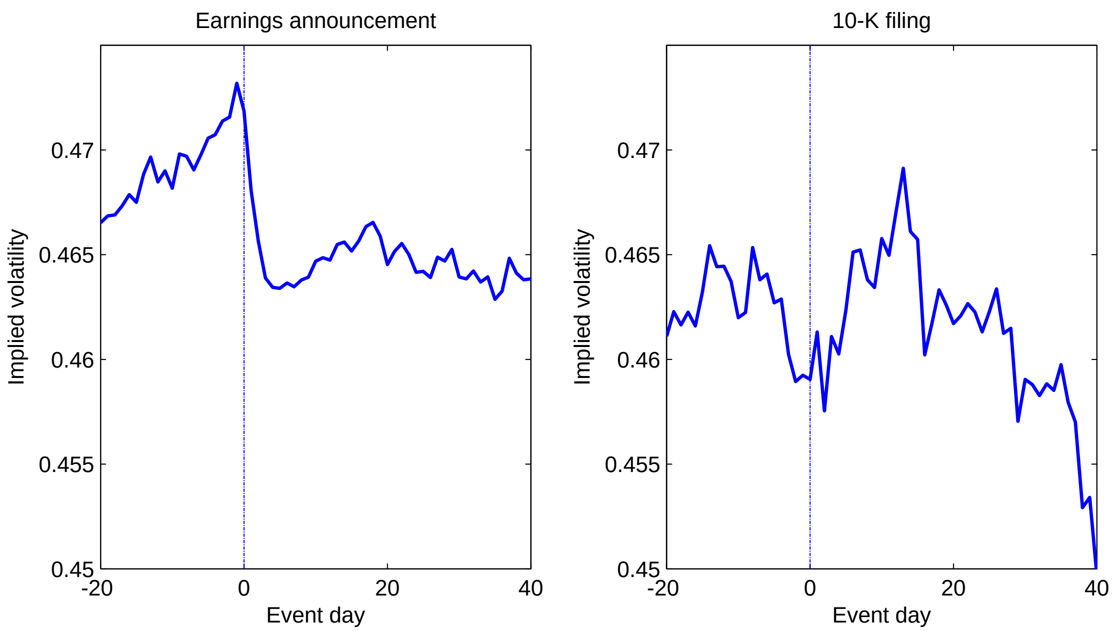

Figure 3: Dynamics of implied volatility for firms sorted by 10-K file size

This figure plots the dynamics of the option implied volatility following 10-K disclosures for firms with small and large 10-K file size. In each year, firms are sorted into quintile portfolios based on their 10-K file size. The left panel plots the percentage change of the implied volatility from the filing date against the number of weeks after the 10-K disclosure for portfolio 1 (small 10-K file size) and portfolio 5 (large 10-K file size), whereas the difference is plotted in the right panel. The sample period is 1996-2012.

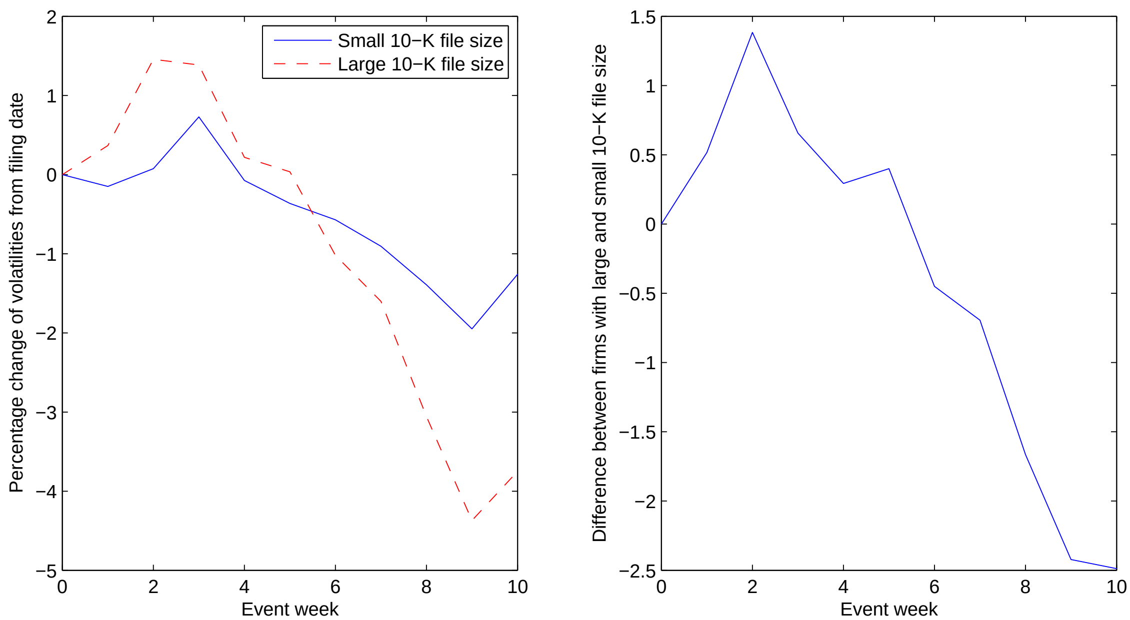

Figure 4: Estimated volatility dynamics for average small and large 10-K firms This figure plots the *estimated* volatility dynamics following 10-K disclosures for firms with small and large 10-K file sizes. The quadratic function between volatility dynamics and time-to-learn, $\sigma_{i,t} = a_i \times t + b_i \times t^2 + \epsilon_{it}$, is estimated for each firm and the estimation results are reported in Table 10. Firms are sorted into quintile portfolios based on their average 10-K file size then the average (a_p, b_p) are calculated for each file size portfolio. The estimated volatility dynamics are calculated based on these portfolio (a_p, b_p) and $\sigma_{p,t} = a_p \times t + b_p \times t^2$. The left panel plots the estimated percentage change of volatility from the filing date against the number of weeks after the 10-K disclosure for firms with small and large 10-K file size, whereas the difference is plotted in the right panel.