Clearing the Cognition: A Study of the Impact of High-Frequency Trading on Equity Market Volatility

Samuel D. Evans*

University of Warwick

Department of Economics

Economics BSc Dissertation

Abstract

This paper investigates the response of equity market volatility to an increase in the proportion of total equity trades conducted by High-Frequency Traders. Subject to this, the paper seeks to explain the results via an emphasis on the substitution of cognitive error by algorithmic rationality whilst evaluating the additional implications that high and low levels of market uncertainty have for the influence of High-Frequency Trading (HFT) on volatility. Utilising a unique dataset; primarily consisting of the firm-specific characteristics of the S&P100 constituents between 1995q1 and 2009q2, this study implements three regressional techniques – Pooled Ordinary-Least-Squares, Fixed Effects and Arellano-Bover/Blundell-Bond Dynamic Panel estimation – accompanied with an instrumental variable approach to extend the existing literature. This investigation finds that the net impact of HFT is to increase equity market volatility. Interestingly, a rise in HFT, that is, algorithmic rationality inherent in the market, is empirically determined to have a reductive influence on the positive effect that cognitive error has on volatility. Moreover, in line with proposed theory, this study reveals that the positive impact of HFT on equity price instability is diminished during periods of low market uncertainty, however, strongly accentuated at high levels of uncertainty.

Word Count: 5297

*I would like to thank Dr. Lei Zhang and Dr. Jeremy Smith of the University of Warwick for their helpful comments and support. I am grateful to Asst. Prof. Arie Gozluklu of Warwick Business School for our invaluable discussions of the institutional idiosyncrasies of HFT. I also thank Dr. Frank X. Zhang of the Yale School of Management for providing me with key data for this study. Special thanks are given to Dr. Gianna Boero for nominating this paper for the Carroll Round. Any errors are my own.

Contents

I. Introduction

There is currently no universally accepted definition of "High-Frequency Trading" (HFT), however, the Securities and Exchange Commission (SEC) describe it as "the use of extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders", SEC (2010). The advent of these sophisticated computer programs has changed the very nature by which equity markets[1] operate and now "the lexicon of financial markets is dominated by talk of [this development]"[2]. The well documented real impact of equity markets on the well-being of a population (see Bencivenga et al., 1995; Deveraux and Smith, 1994; Levine, 1991; and Obstfeld, 1994) would infer that any such developments affecting the stability of markets are of great economic significance. As a direct result, HFT is now receiving a growing amount of attention in the academic literature and regulatory space.

There are several perceived benefits of an increasingly computerized element inherent in equity markets. Proponents of HFT highlight the improved provision of liquidity and reduction in transaction costs, whilst also claiming that HFT has improved price discovery. Nevertheless, opponents put forward the argument that the millisecond time-horizon of HFT distorts equity prices from fundamental value and, in addition to this, their aggressive momentum strategies stimulate excessive price volatility. If these implications of HFT for volatility are true, investors will in future respond less to fundamental news (Zhang, 2006) and demand a higher equity premium, driving up the cost of capital for firms (Froot, Perald and Stein, 1992). As examined in the review of the existing literature, the evidence for these conflicting opinions is inconclusive; however, as these computerised traders become responsible for an increasingly large proportion of the trading volume in equity markets[3], it is now ever more important to determine the true implications of this revolutionary development in stock markets such that correct regulation and policy may be formulated.

The principle objective of this paper is to determine the response of volatility in equity markets to an increase in the proportion of total equity trades that are made by High-Frequency Traders.

Hypothesis 1: *A rise in the proportion of trades conducted by High-Frequency Traders will increase equity market volatility.*

Furthermore, this study aims to extend the literature via an empirical examination of the substitution of overconfidence by perfect, computerised rationality. Intuition would suggest that as HFT increases, the irrationality of the human cognition inherent in the market is substituted by the perfect rationality

[1] The terms "stock markets" and "equity markets" are used interchangeably.
[2] Haldane, A. (2011). "Race to Zero", speech given to the Bank of England.
[3] TABB group estimate HFT is responsible for 73% of total trading volume in US Equities in 2009.

of electronic algorithms, therefore the positive effect that the cognitive bias of overconfidence has on equity volatility diminishes.

Hypothesis 2: *A rise in the proportion of trades conducted by High-Frequency Traders will diminish the positive impact of traders' overconfidence on equity market volatility.*

In addition to this, the implications that high and low levels of market uncertainty have for HFT's influence on volatility shall be investigated. The author makes the following hypothesis based on related studies that are referred to in the literature review:

Hypothesis 3: *At high (low) levels of uncertainty there is an additional positive (negative) impact of HFT on volatility.*

The results of the aforementioned analysis shall be compared and contrasted against previous findings in the literature.

II. Related Literature

A review of the related literature will be structured as follows: firstly, I shall examine the literature relating directly to HFT and its affect on equity volatility. Secondly, the literature on two opposing possible explanations for HFT's volatility influence - i) algorithmic rationality substituting human cognitive error and ii) the theoretical nature of short-horizon trading and its interaction with longer-horizon investors - shall be investigated.

There has been a substantial escalation in the HFT literature following the "Flash Crash" of May 6th 2010[4]. Empirical work in this area, as noted by Chaboud et al. (2009), has been restricted due largely to the difficulty in acquiring data whereby the volume of trades conducted via HFT or more generally, algorithmic platforms[5], can be directly observed.

Several strands of the recent literature provide evidence that HFT increases volatility within equity markets. Kirilenko et al. (2011) study the aforementioned "Flash Crash" and, taking inspiration from the theoretical literature investigating interaction between heterogeneous traders, attempt an explanation through exploiting the differences in the market participants. The paper uses audit trail, transaction-level data to categorise trade accounts based on their trading characteristics (e.g. average size of trade) and compares the behaviour of each category on the day of the crash with their behaviour three days prior. Due to the observational methodology and lack of econometrics, it appears little consideration is given to bias. The authors support the CFTC-SEC report[6], finding that it was fundamental – not high-frequency – traders that triggered the crash, highlighting a negligible change in behaviour of HFTs over the sample period. However, the paper interestingly concludes that due to the very nature of HFT – executing vast numbers of small trades but having net inventories of approximately zero – HFT can create a "hot-potato" effect [7] that accentuates price volatility. This observational explanation is unique among the literature. Naturally however, the short observation period – three days – and specific nature of the event reduces the significance of the paper's findings with regards to longer term effects of HFT, and thus policy implications.

[4] On May 6, 2010, in the course of about 30 minutes, U.S. stock market indices dropped more than five percent but experienced a rapid rebound immediately afterwards. Many market commentators attributed the blame to HFT strategies that accentuated price momentum swings.

[5] HFT is technically a subset of algorithmic trading as it can only be conducted via computer programmes due to the millisecond holding periods that the strategy relies upon, however, the vast majority of literature covering the HFT topic use the terms algorithmic trading and HFT interchangeably: as will this paper.

[6]The regulatory bodies produced a report on the crash titled, "Preliminary Findings Regarding the Market Events of May 6, 2010".

[7]"During the Flash Crash, High Frequency Traders initially bought contracts from Fundamental Sellers...[then]...HFTs appeared to rapidly buy and sell contracts from one another many times, generating a "hot potato" effect before Fundamental Buyers were attracted by the rapidly falling prices to step in and take these contracts off the market." Kirilenko et al. (2011), pp.3.

Zhang (2010) provides alternative, more robust evidence supporting the conclusion of the aforementioned empirical work. Utilising quarterly data over a 9 year period for over 2000 stocks listed on the Nasdaq and NYSE, the paper uses a difference-in-difference-in-difference approach: a firm- and time-fixed effects model and then a comparison of results over two separate periods under the justifiable assumption that no HFT occurred prior to 1995. This model consequently addresses both the structural change bias due to the change in trading behaviour over the long sample period, and measurement error inherent in the author's method for empirically estimating HFT volume. Incorporating the introduction of the NYSE Autoquote[8] as a multiplicative dummy variable into his model as previously done by Hendershott, Jones and Menkveld (2010), Zhang (2010) robustly infers HFT increases volatility. Interestingly, this result is supported by Jovanovic and Menkveld (2010), who undertake a similar difference-in-difference methodology – using the introduction of Chi-X[9] platform that parallels Zhang's use of the Autoquote instrument – to examine HFT's impact in its less active European environment.

Contrary to this, a smaller fraction of the HFT and related literature starkly dispute these findings. Brogaard (2010), aside from confirming positive correlation between volatility and HFT, uses a unique high-frequency dataset constructed from Nasdaq stocks over the period 2008 to 2009 and a difference-in-difference-in-difference approach, akin to Zhang (2010), to test HFT's influence on volatility. The author's methodology entails using the 2008 Short Sale Ban that affected 13 of the 120 sampled stocks as an exogenous shock to HFT; a justifiable proxy given a 22%[10] correlation between short selling activity and HFT activity. The author controls for potential time-varying HFT activity by differencing volatility and HFT for each affected stock with an unaffected stock. The resultant OLS regression yields a significant, negative coefficient for HFT at the 15 and 30 minute intervals, but lacks any longer term evaluation. This result suggests that as HFT increases, equity volatility decreases, contradicting Zhang (2010) among others. A potential explanation for this could be that Brogaard's use of tick-by-tick rather than low-frequency data exaggerates HFT's market-making activities that tend to be more beneficial than their more aggressive trading strategies[11]. Nevertheless, this cannot explain the study's disagreement with Jovanovic and Menkveld (2010) or Kirilenko et al. (2010) who also use high-frequency data and, in addition to this, Brogaard's findings are supported by several related empirical studies (Hasbrouk and Saar, 2010; Chaboud et al., 2009). Consequently, it is clear that further work needs to be done in this field to assist in the development of a consensus.

---

[8] The NYSE Autoquote system is an electronic display book that updates the NYSE's published bid or offer automatically.

[9] Chi-X is an alternative trading system (ATS) that offers fully-automated trading and clearing services that completely bypass Europe's existing exchanges and central counterparty infrastructure.

[10] Value from Brogaard (2010), pp. 23.

[11] See Zhang (2010) pp. 10.

Based on foundations initially laid by Tversky and Kahneman (1974), overconfidence – where one overestimates the precision of one's own information (Wang, 2001) – has gathered the greatest significance out of all cognitive biases in explaining equity market volatility. This naturally has implications for the effects of HFT as the increasing dominance of perfectly rational algorithms will replace this element of human cognition in the market.

The pioneering paper by Chuang and Lee (2006) employs data across all firms listed on NYSE and AMEX from 1963 to 2001 to conduct the only empirical investigation of overconfidence on equity volatility. Strongly supported by the theoretical literature that indicates a positive correlation between previous stock returns and overconfidence (Hirshleifer, 2001) and overconfidence and volume (Benos, 1998), the author proxies for overconfidence using the fitted values from a regression of trading volume on past equity returns. Incorporating this overconfidence proxy into a Garch model, and detrending to account for non-stationarity in weekly data; a unit increase in overconfidence is found to increase short-run volatility by 5.9% and long-run volatility by over 10%. This result appears robust and can be explained by overconfidence stimulating momentum trading strategies as illustrated in Daniel and Titman (1999). Interestingly, no study has yet extended this finding to determine the implications for market volatility if rationality was to replace such overconfidence.

Although replacing the cognitive bias of overconfidence is one explanation for HFT influencing volatility, theoretical study related to HFT – the nature of short-horizon trading and its interaction with longer-horizon investors – posits an alternative explanation.

Morris and Shin (2003), among others, attempt to formally capture Keynes' beauty contest insight into short-horizon trading. Extending the work of Froot, Scharfstein and Stein (1992), the authors account for risk-aversive behaviour whilst using global game techniques to model the interaction between short-term and long-term investors, notably solving for the trigger point of a "liquidity black hole"[12] in the process. Furthermore, this study ensures more robust modelling than previous papers by assuming unknown transaction prices in times of market stress[13]. Involving the idea of feedback rules (De Long et al., 1990), the paper finds short-horizon investors face a downward sloping residual demand curve provided by long-horizon investors. Consequently, the authors propose that when prices fall close to short-term traders' price limits, the irrelevance of equity fundamentals results in a mutually reinforcing sell off of assets: herding. It is thus argued that periods of extreme volatility are attributable to the "strategic complementarities"[14] of short-horizon traders and their interaction with longer-term market participants. This is representative of the theoretical consensus regarding short-horizon speculation.

[12] "V" shaped temporary price shocks where liquidity becomes unavailable (e.g. Flash Crash).
[13] See Kaufman (2000)
[14] Terminology from Admati and Pfleiderer (1988)

This paper contributes to the HFT literature in three ways. Firstly, this study uses a unique dataset and methodology to examine the influence of HFT on equity market volatility. The previous literature provides conflicting suggestions for the direction of this influence; therefore this study further extends the literature towards a more generalised consensus. Secondly, this investigation is the first empirical study of the impact of HFT on the implications of investor overconfidence for volatility. Finally, inspired by the theoretical literature, this paper conducts the first empirical evaluation of the additional impact that HFT has on stock price volatility at extreme levels of market uncertainty.

III. Data and Related Issues

III.a. Outline of Data

This study uses a unique dataset containing quarterly observations of the equity characteristics of the S&P100 constituents, listed on both the Nasdaq and NYSE, from 1995q1 to 2009q2[15]. The dataset is primarily constructed from the *Compustat Monthly*, *Federal Reserve* (FRED) and *Datastream* databases, however, the variable that proxies for HFT's share of total trading volume has been provided by Professor X. Frank Zhang of Yale University. Uncertainty is approximated for by using the standard deviation in analysts' firm-specific earnings forecasts taken from the *IBES* database. Furthermore, the approximation for overconfidence is constructed using the fitted values from a regression of trading volume on past equity returns, as per Chuang and Lee (2006). For the reader's interest, the descriptive details of the firm-specific and macroeconomic control variables[16] are given in Table A2 of the Appendix, whilst descriptive statistics are presented in Table A3 of the Appendix.

Considering the topic of this paper, it may initially appear counterintuitive to use low frequency – quarterly – data for its analysis[17]. However, there is a distinct rationale for this decision. Firstly, lower frequency data avoids the negative bias inherently produced by tick-by-tick data. This is because tick-by-tick data emphasises the more beneficial, volatility-reverting market-making activities of HFT whilst diminishing the influence of the more aggressive HFT strategies (e.g. liquidity detection) that are only reported in the lagged audit trails provided by dark pools. Secondly, lower frequency data enables more interesting policy conclusions to be deduced as the implications of temporary, millisecond price distortions for efficient resource allocation in the real economy are likely to be negligible.

The paper now turns to focus on the key variables of interest.

III.b. Dependent Variable

The dependent variable to be investigated is equity market volatility. In particular, for the purposes of robustness, this paper utilises both an implied and realised measure of the S&P100 index volatility.

The implied volatility measure is Standard & Poor's VXO volatility index, calculated using 30-day at-the-money, S&P100 (OEX) options. The realised equivalent, constructed by the author, is calculated by aggregating the daily price volatility for each S&P100 firm across each quarter over the 14 year sample period[18]. It is apparent from Figures 1(a) and 1(b) that both measures follow a similar time path[19]:

[15] A comprehensive list of the constituent companies is presented in Table A1 of the Appendix.

[16] See Zhang (2010) for a detailed explanation of incentives for including firm-specific controls.

[17] Possible benefits of higher frequency data shall be suggested in Section VI.

[18] It should be noted that Phillip Morris International (Ticker PM) has been omitted from realised vol. index due to lack of observations.

Figure 1(a): S&P100 Implied Volatility Index

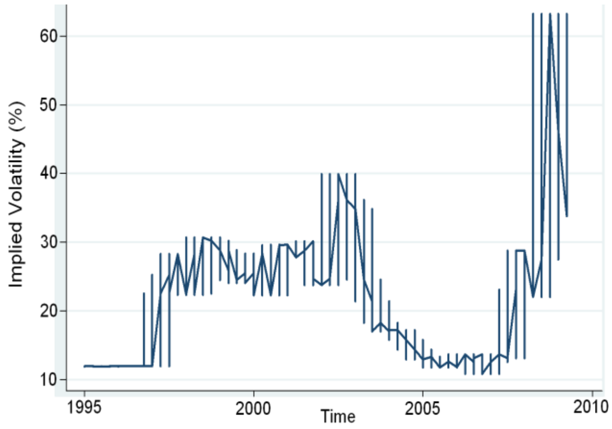
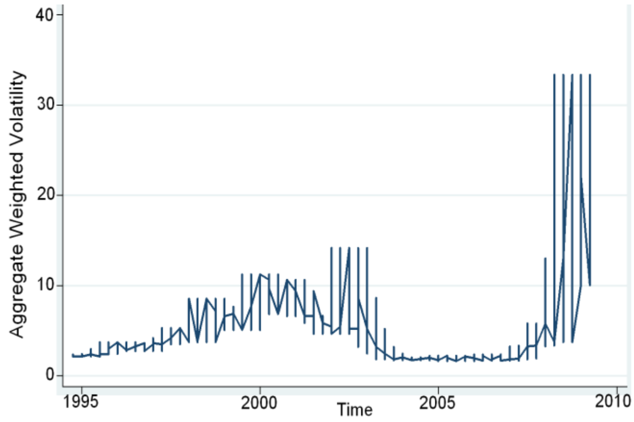


Figure 1(b): S&P100 Realised Volatility Index



It can been observed that VXO almost uniformly exceeded realised volatility with a mean of 22.24% relative to the realised volatility mean of 5.97%. This is consistent with the findings of Andersen and Bondarenko (2007) who attribute this difference to the "presence of a substantial negative variance risk premium" in implied measures. The relative level of each measure is irrelevant for the use of the data in the study as both measures are used in isolation of one another. The element of importance is their comparative time paths which appear extremely similar and suggest the necessity for structural change dummies for both the *Dot Com Bubble* and *Credit Crisis* periods.

Importantly, the Credit Crisis period – observable given the post-2008 explosion in volatility – should not be excluded as anomaly as it encompasses key data for the analysis of the impact of HFT. This is due to HFT being criticised by Kirilenko et al. (2011) for stimulating momentum trading strategies during such highly volatile periods, exacerbating volatility through algorithmic, short-horizon feedback rules. Such an implication would have real economic effects and it is thus vital to incorporate this time period into the study.

III.b. Key Explanatory Variables

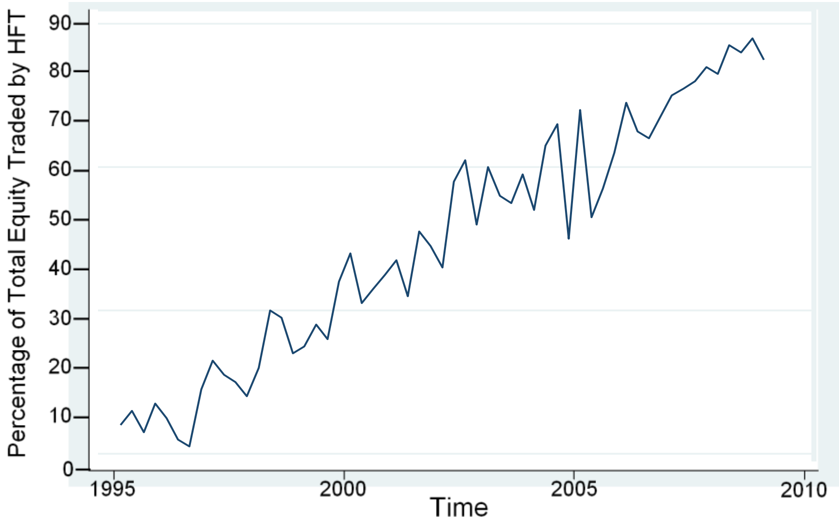
The data approximating the proportion of total equity traded via high-frequency algorithms, kindly provided by Dr. X. F. Zhang of Yale University, is, to date, the only academically valid measure of its kind. Although Zhang (2010) constructs this data across 4073[20] equities, this measure can be used as a valid proxy in this paper as by the nature of HFT and US equity markets, the vast majority of HFT activity occurs in the "investable universe", and thus within only a small number of the largest of equities (Chordia and Swaminathan, (2000)). Consequently, the majority of the 4073 equities would have a negligible proportion of overall HFT activity with the vast majority of total HFT and trade volume occurring in the S&P100: this study's sample.

[19] Note: the realised volatility measure for the purpose of the figure has been averaged across all constituent stocks, weighted by each stocks market capitalisation.
[20] See Zhang (2010) for specific calculation methodology.

Moreover, although this measure is not company specific, HFT does not differ between companies based on fundamentals as – noted by Brogaard (2010) – its millisecond time horizon strategies result in only a consideration for noise so can be approximated to equal across the investable universe spectrum.

Figure 2 illustrates several interesting aspects of the HFT data:

Figure 2: Proportion of Equity Traded via HFT over Time



Firstly, HFT has a clear trend from below 5% to 78% over the 14 years, with an average growth rate of 2.45% per quarter. This is in line with the TABB Group's estimates[21], supporting validity of the data. Secondly, the data appears trend-stationary and this is confirmed by an Im-Peseran-Shin unit root test[22]. As a direct consequence, the HFT data will have to be detrended to prevent spurious results upon estimation given simultaneous trends in volatility.

Finally, HFT is considered in the literature – see Hendershott, Jones and Menkveld (2010), Zhang (2010), Chaboud et al. (2009) and Brogaard (2010) – to be endogenous with volatility. The principle reasons for this centre on two-way causality stimulated by the aggressive momentum strategies of short-horizon traders[23]. Figure A1 and Table A6 in the Appendix visually and statistically confirm a high correlation between the HFT and volatility data used in this paper, alongside Granger Causality tests that infer substantial endogeneity between these two variables; thereby supporting previous studies. Section IV.c. explains the instrumental variable estimation technique that shall be used to account for this endogeneity in the data.

[21] Jeremy Grant (Sept. 02, 2010). "High-frequency trading: Up against a bandsaw". *Financial Times*.
[22] See Table A5 in Appendix.
[23] Morris and Shin (2003) theoretically model these short-horizon strategies that are substantially supported by empirical study.

An additional key variable for this study is overconfidence. Naturally, overconfidence is unobservable, thus presenting difficulty in its measurement. Nevertheless, as per Chuang and Lee (2006) and theoretically valid by the same argument as given in the literature review, this paper uses the fitted values from the following regression to construct the overconfidence data:

$$TradingVolume_{it} = constant + \beta \sum_{k=1}^{n} \left[returns_{i,t-k} \right] + \varepsilon_{it}$$

where there are $k = n$ lags of returns. The overconfidence data yields a mean of 3.5031 – statistically significant at the 1% level – which implies an inherent level of overconfidence in the US markets. Moreover, as expected given the related overconfidence literature, there is a positive correlation between overconfidence and implied and realised volatility of 0.0758 and 0.1335 respectively.

IV. Methodology

IV.a. Outline of Methodology

Three separate regression techniques will be used to enable an interesting comparison of HFT's implications for volatility that are intuitively appealing but progress to allow for firm-specific variation, serial correlation and potential endogeneity bias. The analysis shall begin by utilising an Aggregated Ordinary-Least-Squares Model; followed by a Fixed Effects Model that controls for both time- and firm-fixed effects; and finally, an Arellano-Bover/Blundell-Bond Dynamic Panel GMM Model[24].

IV.b. Motivation for Methodology

The Ordinary-Least-Squares estimation technique is utilised to provide an intuitive initial analysis and will take the general form:

$$\begin{aligned} Volatility_t = \alpha + \beta_1 HFT_t + \beta_2 Overconfidence_t + \beta_3[Overconfidence * HFT]_t \\ +\beta_4[HighUncertainty * HFT]_t + \beta_5[LowUncertainty * HFT]_t \ + \\ \textstyle\sum_{k=1}^{5} \delta_k [Firm\ control]_t + \sum_{j=1}^{4} \delta_j [Macroeconomic\ control]_t + \varepsilon_t \end{aligned}$$

where $[x_1 * x_2]$ for any two variables x_1 and x_2 represents an interactive variable between the two variables and t corresponds to the specific time period. A comprehensive list indicating the firm-level and macroeconomic control variables is given in Table A2 of the Appendix. Lags of variables are only included if there is a clear economic rationale behind their inclusion and to also limit serial correlation issues, however, given unobserved heterogeneity, it is likely this initial estimation method will yield somewhat biased coefficients.

The Fixed Effects estimation accounts for the aforementioned unobserved heterogeneity bias, enabling a far more robust analysis and potentially generating far greater significance regarding coefficient estimates. For completeness, this model shall take the general form:

$$\begin{aligned} Volatility_{it} = \alpha + \beta_1 HFT_{it} + \beta_2 Overconfidence_{it} + \beta_3[Overconfidence * HFT]_{it} \\ +\beta_4[HighUncertainty * HFT]_{it} + \beta_5[LowUncertainty * HFT]_{it} \ + \\ \textstyle\sum_{k=1}^{5} \delta_k [Firm\ control]_{it} + \sum_{j=1}^{4} \delta_j [Macroeconomic\ control]_{it} + \varepsilon_{it} \end{aligned}$$

where i corresponds to a specific firm listed on the S&P100. This model also includes relevant lags of the appropriate variables but Fixed Effects estimation may still incur bias given potential two-way

[24] It should be noted, given the explanatory variables, no bias in any estimation technique can by due to multicollinearity of regressors – see Table A4 in Appendix. Moreover, both Pooled OLS and Fixed Effects estimations use robust standard errors to control for heteroskedasticity in the error term – see Table A8 in the Appendix. For this study, the Dynamic Panel estimation cannot use robust standard errors given the variance-matrix is near singularity, thus uses GMM standard errors.

causality of HFT and volatility, in addition to autocorrelation via lagged dependent variables (Arellano and Bond (1991)).

This paper utilises Arellano-Bover/Blundell-Bond Dynamic Panel GMM estimation to help control for the aforementioned perceived two-way causality bias whilst accounting for dynamic effects in the dependent variable. To this end, the estimation technique incorporates lags of the endogenous variable, HFT, as instruments in addition to first-differencing past volatility levels[25]. The general form is[26]:

$$\Delta Volatility_{it} = \alpha + \beta_1 \Delta HFT_{it} + \beta_2 \Delta Overconfidence_{it} + \beta_3 \Delta[Overconfidence * HFT]_{it} + \beta_4 \Delta[HighUncertainty * HFT]_{it} + \beta_5 \Delta[LowUncertainty * HFT]_{it} + \sum_{k=1}^{5} \delta_k \Delta[Firm\ control]_{it} + \sum_{j=1}^{4} \delta_j \Delta[Macroeconomic\ control]_{it} + \Delta\varepsilon_{it}$$

where Δx_{it} represents the first-difference of variable x for firm i at time t[27]. This technique, unique to the literature in this field, provides an extremely robust and far more efficient[28] analysis to support the more naturally intuitive analysis provided by the former two methods.

IV.c. Instrumental Variable Approach

The regression analysis specified in Section IV.b. will allow for a relationship between HFT and volatility to be implied, however, despite partially controlling for endogeneity bias in the Arellano-Bover/Blundell-Bond estimation, the endogeneity between these two variables will require an exogenous shock to HFT to be used as an instrument before *causality* can be inferred.

This paper takes inspiration from Hendershott, Jones and Menkveld (2010) by utilising the NYSE automated quote dissemination system – Autoquote – that was first implemented in 2003, quarter one, as an exogenous shock to HFT. This technological development "provides quicker feedback to algorithms and results in more electronic message traffic"[29], thereby facilitating a step-change increase in HFT activity in NYSE listed stocks.

Consequently, the NYSE Autoquote in the form of a structural change dummy is used as an instrument for a positive step-change in HFT activity in affected stocks. The tests for instrument relevance and argument for exogeneity are given in Table A7 of the Appendix.

[25] This method is widely considered as an improvement upon the Hausman and Taylor (1981) instrumental variables approach.
[26] Lagged Instruments are excluded for observational simplicity.
[27] Note: the constant, α, now represents the differenced time trend.
[28] See Green (2002) for details of efficiency gains with dynamic panel estimation.
[29] Hendershott, T., Jones, C. and Menkveld, A. (2009). Does Algorithmic Trading Improve Liquidity? *Journal of Finance*, 66(1), pp. 3.

V. Empirical Results and Discussion

V.a. Outline of General Results

Initial comparisons[30] of the results across the two types of dependent variable presented in Table 1 suggest that, although the coefficients appear of greater significance concerning the realised volatility relative to the implied volatility dependent variable, the coefficients are in the same direction on aggregate. The difference in significance between the coefficients concerning the realised and implied volatility regressands is likely due to the implied measure (VXO) being constructed across companies that differ to the 100 sampled companies for this study at certain, infrequent points in the sample period, whilst the implied measure is also not firm-specific. More importantly, the fact that both sets of coefficients tend to be in the same direction substantially supports the robustness of the following results.

It can be observed, moving from column (1) to column (3), that both the firm-level and macroeconomic controls become increasingly significant in explaining volatility. This is the expected result as the three estimation techniques progressively control for unobserved heterogeneity bias, serial correlation and simultaneity bias[31]. Additionally, the structural change dummies concerning the end of the *Dot Com Bubble* and beginning of the *Credit Crisis* appear consistent and significant across all specifications, thus emphasising the importance of such historical events for equity market volatility.

Furthermore, it should be noted that the empirical results support the existing literature regarding the remaining control variables; Uncertainty, Volume and Overconfidence. Specifically, across all three estimation techniques and both dependent variable measures, the coefficients for uncertainty that are statistically significant at least at the 5% level are positive for high levels of uncertainty and negative for low levels. Moreover, in accordance with Bessembinder and Seguin (1993), the results imply a positive short-run effect of trade volume on equity volatility, with this effect diminishing in the long-run albeit remaining positive. Perhaps more ambiguous are the results for overconfidence whose coefficients can be seen to become negative once unobserved heterogeneity bias and dynamic effects are accounted for. This result opposes both seminal papers such as Chuang and Lee (2006) and also economic intuition. However, this finding is likely a product of endogeneity bias as once this is eliminated via the use of an exogenous shock to HFT – see Table 2 – the significant coefficients for overconfidence become positive, thereby following conventional logic and the consensus in the literature.

---

[30] In accordance with the focus of this paper, the analysis will pay specific attention to the direction and significance of coefficients rather than their absolute values.

[31] See Table A9 of Appendix for details of serial correlation test. It should be noted that, at a 1% significance level against which the tests are evaluated, serial correlation is only present in estimations that use *implied* volatility as the dependent variable. This implies serial correlation should not greatly bias the results as the *implied* dependent variable estimation results are – as previously mentioned – predominantly used as a robustness check and are not the sole focus for this paper's key findings.

Table 1: Empirical Estimation Results

	Dependent Variable = Realised Volatility			Dependent Variable = Implied Volatility		
	(1)	(2)	(3)	(1)	(2)	(3)
HFT	0.4342	0.4709**	0.8236***	3.6161***	3.6729***	14.0461***
HFT*Overconfidence	-0.0010*	-0.0107**	-0.1731***	0.0067	0.0052	-0.2309***
HFT*HighUncertainty	0.0161***	0.0143**	0.0107*	0.0603**	-0.0619*	-0.0221
HFT*LowUncertainty	0.0033	0.0020	-0.0072	-0.0230	-0.0242	-0.1298**
Overconfidence	0.1575***	0.1365***	-0.0081	-0.2189**	-2.1313**	-0.3299*
Volume	0.0881***	0.0753***	0.0074***	0.0164	0.0221	0.1240***
L.Volume	-0.0861***	-0.0828***		-0.0159	-0.0087	
HighUncertainty	0.0697*	0.1707***	0.3949***	-0.0569	-0.0985	3.4014
LowUncertainty	0.0477*	0.0150	0.0369	-0.0824	-0.0890	-1.0924***
Earnings Surprise	0.7031	0.6094	0.3474***	-0.0007	-0.0005	0.3403**
Book-to-Market Ratio	0.3036*	0.6520**	0.8533***	0.0861	0.1381	-5.6055***
Firm Size	0.0307	0.1133***	-0.0269	0.0730	0.3690***	-4.0686***
Leverage	0.0630*	0.1444***	0.2889***	-0.0238	-0.0828*	0.6238***
Inverted Price	3.6802**	3.3029	2.9184***	4.3204	8.4086***	64.4288***
Money Supply (M1)	0.0262***	0.0369***	0.0351***	0.4880***	0.4825***	0.2591***
Real GDP (QoQ)	-0.2019***	-0.1934***	-0.3551***	-2.3153***	-2.2692***	-5.7287***
FedFunds Rate	-0.1245**	-0.1284***	0.1319***	-8.4640***	-8.5001***	0.5094***
L. FedFunds Rate	0.6370***	0.6327***		13.8180***	13.8121***	
L2. FedFunds Rate	-0.4501***	-0.4148***		0.0098	0.0441	
L3. FedFunds Rate	-0.3653***	-0.3288***		-11.4787***	-11.4870***	
L4. FedFunds Rate	0.3694***	0.3372***		4.8325***	4.8233***	
FX Index	0.0393***	0.0416***	0.0140***	0.2030***	0.2005***	0.2408***
L. FX Index	-0.0376	-0.0381***		0.0226	0.0215*	
L2. FX Index	-0.0228***	-0.0195***		-0.3066***	-0.3062***	
L3. FX Index	0.0050	0.0014		-0.0669*	-0.0685***	
L4. FX Index	0.0097*	0.0101**		-0.0593*	-0.0641**	
L.Dep. Variable	0.2535***	0.1878***	0.0896***	0.1680***	0.1682***	0.3016***
L2.Dep. Variable	0.0794	0.0469		0.0599***	0.0574***	
L3.Dep. Variable	0.0673	0.0471*		0.5498***	0.5556***	
L4.Dep. Variable	0.0943	0.0509		-0.3312***	-0.3343***	
S-Change Dummy1	-0.2206***	-0.3211***		-12.7652***	-12.9776***	
S-Change Dummy2	0.0489	0.1176***		6.8859***	6.7645***	
Constant	-0.1200	-0.3337	-1.6957***	43.2063***	44.0640***	1.1966
Firm/Time Fixed Effects	NO	YES	YES	NO	YES	YES
R^2	0.6091	0.5648	0.3886	0.8862	0.8855	0.5996
#Observations	4446	4446	4533	4545	4545	4629

- Columns (1), (2) and (3) give the coefficients yielded by the pooled OLS, Fixed Effects and Dynamic Panel estimations respectively.
- ***,**,* correspond to the coefficient being significant at the 1%, 5% and 10% significance levels respectively.
- Any variable, X_1*X_2 is an interactive variable between variables X_1 and X_2.
- Several variables omitted from column (3) due to estimation method – see Arellano and Bond (1991).

Table 2: Instrumental Variable Results

	Dependent Variable = Realised Volatility		Dependent Variable = Implied Volatility	
	(1)	**(2)**	**(1)**	**(2)**
HFT	0.3078	0.7540**	0.2006	0.4346
HFT*Overconfidence	-0.0994	-0.1171	-0.0196	-0.0340
HFT*HighUncertainty	0.3041***	0.3492***	-0.0750	0.1251
HFT*LowUncertainty	-0.0098	-0.0603	-0.6576**	-0.7347***
Overconfidence	0.1930***	0.1718***	-0.1544	-0.1603
Volume	0.0894***	0.0768***	0.0151	0.0179
L.Volume	-0.0879***	-0.0795***	-0.0179	-0.0112
HighUncertainty	-0.0702*	0.0173	-0.0751	-0.0996
LowUncertainty	0.0419	0.0320	0.2021	0.2384
Earnings Surprise	0.6363	0.5006	0.0157	0.0245***
Book-to-Market Ratio	0.3364**	0.6857**	0.1080	0.2240
Firm Size	-0.0240	0.0968**	0.0533	0.1631
Leverage	0.0604**	0.1294**	-0.0250	-0.0767**
Inverted Price	4.3689**	4.5521**	4.3204	7.0714***
Money Supply (M1)	0.0246***	0.0395***	0.1572***	0.1608***
Real GDP (QoQ)	-0.1902***	-0.1790***	-2.0171***	-1.9909***
FedFunds Rate	-0.1381***	-0.1448***	-7.4940***	-7.5082***
L. FedFunds Rate	0.6327***	0.6268***	13.5182***	13.5026***
L2. FedFunds Rate	-0.3368***	0.3451***	-2.8739***	-2.8344***
L3. FedFunds Rate	-0.4784***	-0.3973***	-9.1797***	-9.1797***
L4. FedFunds Rate	0.4346***	0.3382***	4.8618***	4.8407***
FX Index	0.0389***	0.0411***	0.3538***	0.3526***
L. FX Index	-0.0373***	-0.0381***	-0.2657***	-0.2663***
L2. FX Index	-0.02503***	-0.0218***	-0.0746	0.0734***
L3. FX Index	0.0036	0.0009	-0.2171***	-0.2176***
L4. FX Index	0.0119**	0.0130***	-0.0510	-0.0524***
L.Dep. Variable	0.2210***	0.1548**	0.3483***	0.3467***
L2.Dep. Variable	0.0807	0.0234	-0.1035***	-0.1051***
L3.Dep. Variable	0.1827**	0.1036	0.6456***	0.6469***
L4.Dep. Variable	0.0646	0.0285	-0.2815***	-0.2819***
S-Change Dummy1	-0.3044***	-0.8053***	-8.8923***	-9.2011***
S-Change Dummy2	0.1298**	0.1762***	7.8295***	7.7695***
Constant	-0.0688	-0.1614	28.0056***	28.5337***
Firm/Time Fixed Effects	NO	YES	NO	YES
R^2	0.6166	0.5621	0.8690	0.8688
#Observations	4277	4277	4372	4372

- Columns (1) and (2) give the coefficients yielded by the pooled OLS and Fixed Effects estimation techniques respectively.
- ***,**,* correspond to the coefficient being significant at the 1%, 5% and 10% significance levels respectively.
- Any variable, X_1*X_2 is an interactive variable between variables X_1 and X_2.

V.b. Focused Discussion of Key Findings

The empirical results of particular interest, presented in Table 1 and Table 2, provide a robust set of inferences concerning the relationship between HFT and equity market volatility, thus facilitating an informed discussion when analysed alongside the related empirical and theoretical literature.

Firstly, all three estimation methods produce positive coefficients for the HFT variable across both measures of volatility, with the lowest significance being at the 5% level once unobserved heterogeneity is accounted for[32]. This robustly implies a strong, positive relationship between HFT and equity volatility. Moreover, the Instrumental Variable results given in Table 2 support this finding, suggesting an increase in the proportion of equity traded via HFT, *causes* the level of equity market volatility to rise. Consequently, this paper endorses hypothesis 1. This finding parallels the results of Zhang (2010), Jovanovic and Menkveld (2010) and Kirilenko et al. (2010), whilst disputing Brogaard (2010) and Hasbrouk and Saar (2010). Despite using a unique dataset and methodology; the adoption of the NYSE Autoquote system as an exogenous shock, in addition to the specific HFT data used by this study, perhaps offers an econometric explanation for why this paper's analysis supports the findings of one side of the literature as opposed to the other. Importantly, the result is uniformly supported by theoretical study, most notably Biais and Woolley (2011), and complimented by recent regulatory reports that suggest "stuffing", "smoking" and "spoofing"[33] strategies as definitive explanations for this pernicious effect of HFT on volatility.

Secondly, this paper's methodology yields both positive and negative coefficients for the interactive variable between HFT and overconfidence. Critically however, the significance of these coefficients varies, with only those that are negative being significant below the 10% significance level. Additionally, the estimation results in Table 2 reveal how it is only these negative coefficients that persist once endogeneity is instrumented for. This key result implies a rise in HFT reduces the positive influence of overconfidence on volatility. The author proposes this finding, in support of hypothesis 2, is a result of the perfect rationality, inherent in the algorithmic trading systems necessary to operate at such microsecond frequencies, replacing the force of cognitive error in the markets. Put explicitly, and somewhat insinuated by Jovanovic and Menkveld (2010), the rational, short-horizon algorithms may replace the impact of human irrationality through arbitrage of their price irregularities.

Finally, this investigation's unique examination of the interaction between HFT and volatility at extreme levels of uncertainty provides an interesting set of results. The reported coefficient

[32] Refers to columns (2) and (3). For reasons noted on page 13, GMM rather than robust standard errors are used to test the results of the Dyamic Panel estimation presented in column (3). Consequently, although the high significance of the results and the one-step, as opposed to the more sensitive two-step estimation, being used would imply that eliminating any heteroskedastic downward bias would still render the results of considerable significance, the fact that some degree of downward bias is likely should be taken into account whilst analysing these results.

[33] These are different types of market manipulation strategy uniquely employed by High-Frequency Traders. For a detailed explanation see Biais and Woolley (2011), pp. 8-9.

concerning HFT interaction with high levels of uncertainty is positive but only significant over both volatility measures before controlling for dynamic panel effects[34], thus to draw conclusions regarding this relationship from Table 1 alone is potentially misleading. Nevertheless, once the simultaneity bias induced by HFT is controlled for using the Instrumental Variable approach, the aforementioned coefficient retains its direction, whilst also becoming significant at the 1% level[35]. Furthermore, the coefficients reported in Table 1 for the interaction of HFT and low levels of uncertainty are negative for the most robust specification – Dynamic Panel estimation – across both measures of volatility and are highly significant for the implied measure. These results represent the first empirical findings specifically isolating this implication of HFT and imply an additional positive (negative) effect of HFT on equity market volatility at particularly high (low) levels of market uncertainty. The author suggests this is a direct consequence of the ability of High-Frequency Traders – unlike traditional market-makers – to withdraw from market-making activity, and thus withdraw liquidity, as desired to avoid adverse selection. This key result and potential explanation parallel the related works of Easley, Prado and O'Hara (2011) and Kirilenko et al. (2011) that suggest if order flow becomes increasingly toxic[36], HFT market-makers will leave the market, setting the stage for "episodic illiquidity".

[34] See column (3) in Table 1.
[35] See Table 2.
[36] An increase in trade toxicity means an increase in expected loss from trading with better informed counterparties.

VI. Limitations and Potential Extensions

The limitations concerning this study are essentially derived from one source: the data.

HFT is unobservable, thus estimations of the actual level of HFT have had to be approximated. Naturally, the methods for constructing these estimates differ across papers and therefore, it is likely – noted by Zhang (2010) – that some amount of measurement error is in operation. Despite this study using a justifiably robust measure of HFT, the comparison of the results achieved in this study with those of previous studies should thus be treated with this caveat in mind.

Moreover, despite justifying the rationale behind the use of quarterly frequency data for the aims of this study, it should be noted that this frequency makes the investigation unlikely to detect, and thus evaluate, potentially large microsecond price fluctuations. Firstly, this may result in a negative bias in the realised volatility data, subsequently reducing validity of the results. Secondly, this limits the relevance of the paper's findings for some market participants as, although such temporary, microsecond volatility is of relatively little importance to the *real* economy, it presents a major concern for investors who may be at risk of hitting stop-losses.

Lastly, the inference made by this paper of HFT stimulating volatility corresponds to the net-effect of HFT. This limits the value of these findings in shaping future policy as certain HFT functions, such as liquidity provision, are widely regarded as beneficial to efficient price realisation. Accordingly, any policy aimed at reducing volatility – improving efficient resource allocation – by preventing all HFT activity, as may be incentivised by the raw empirics of this paper, could in-part be counter-productive.

Given the surprisingly limited literature concerning the topic of HFT, the potential for insightful extension to this study is abundant.

Perhaps most notably, the use of ultra-high frequency tick-by-tick data (UHFTBTD) in answering the hypotheses of this study would extend the findings to be relevant at the microsecond level, whilst also improving the measurement accuracy of variables of interest. For example, regarding the latter point, if volatility is observed at a daily frequency, as done for construction of the realised volatility measure, volatility is a "hidden variable" (Fabozzi et al. (2006)). UHFTBTD provides sufficient data points that such a variable can be estimated almost as an average of instantaneous volatility, thereby arguably making the variable "observable".

Furthermore, there would be particular value added in expanding this paper's focus into a European market context. This would be a valid extension as the existing literature has tended to focus upon the US environment, with only Menkveld (2011) and Jovanovic and Menkveld (2010) investigating any influence of HFT upon European equities.

VII. Conclusion

Empirical study, as discussed at length in the literature review, has not yet reached a consensus regarding the net implication of HFT for equity price volatility. This paper, highly motivated by the growing influence of HFT in US markets and the potential consequences for the *real* economy[37], has used a unique dataset and methodology to investigate this implication of HFT, focussing for the first time on HFT's additional effects on volatility via interaction with cognitive error and extreme market uncertainty.

This study finds empirical evidence that, in terms of a share of total trading volume, a rise in HFT increases equity market volatility, and subject to this, increasing HFT has an additional positive (negative) impact on equity volatility if market uncertainty is particularly high (low). Furthermore, this paper provides evidence that an increase in HFT has a reductive influence on the positive impact of the cognitive error of overconfidence on market volatility.

In light of the above findings, complimented by the existing literature, it is understandable that regulatory bodies are becoming increasingly wary of the potential requirement to limit the price impact and systematic risk influences generated by HFT. Notably, the Markets in Financial Instruments Directive (MIFID II) has endorsed capital requirements for HFT firms. This should help increase the potential costs of HFT activity for banks, hedge funds and "pure-play" HFT firms, thus reducing the moral hazard problem associated with limited liability.

Potential policy responses to the suggested impacts of HFT include introducing focused taxation, for example, on collocation[38]. Biais and Woolley (2011) state that "such taxes would be akin to Pigovian taxes, leading High-Frequency Traders to internalise the adverse selection costs they impose on slow traders". Further to this, such tax revenues could be utilised by regulatory bodies to improve monitoring of this notoriously opaque activity, thereby diminishing HFT's volatility impact by reducing the market manipulation strategies referred to in Section V. It may also be beneficial for market efficiency at times of high uncertainty to introduce minimum latency limits. This could help incapacitate the aggressive feedback characteristic of HFT, therefore diminish the additional positive impact of HFT within high uncertainty environments that was implied by this study.

As the High-Frequency Trader continues to evolve, and the significance of its implications become increasingly clear, regulators will need to respond or the arms-race against latency will go unchecked: a risky outcome at best.

[37] Details are specified in Section I.

[38] Collocation is the location of the high-frequency trading computers – black boxes – into the actual exchanges themselves in order to further reduce latency.

VIII. References

Andersen, T and Bondarenko, O. (2007). Construction and Interpretation of Model-Free Implied Volatility. *In*: Nelken, I., ed. *Volatility as an Asset Class*. London: Risk Books, p. 141-181.

Arellano, M., and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, **58**(2), pp. 277-297.

Beltratti, A., and Morana, C. (2006). Breaks and Persistency: macroeconomic causes of stock market volatility, *Journal of Econometrics*, **131**(2), pp. 151-177.

Bencivenga, V., Smith, B. and Starr R. (1995). Transactions Costs, Technological Choice, and Endogenous Growth. Journal of Economic Theory, **57**(1) , pp. 53-177.

Benos, A. (1998). Aggressiveness and Survival of Overconfident Traders. Journal of Financial Markets, 1, pp. 353-383.

Bessembinder, H., and Seguin, P. (1993). Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets. *Journal of Financial and Quantitative Analysis*, **28**(1), pp. 21-39.

Biais, B. and Woolley, P. (2011). High Frequency Trading. *Working Paper*, Toulouse School of Economics.

Brogaard, J. (2010). High Frequency Trading and its Impact on Market Quality. Working paper, Northwestern University.

Chaboud, A., Chiquoine, B,. Hjalmarsson, E., and Vega, C. (2009). Rise of the Machines: Algorithmic Trading in the Foreign Exchange Markets. *Working paper*, Board of Governors of the Federal Reserve System.

Chordia, T., and Swaminathan, B. (2000). Trading Volume and Cross-Autocorrelations in Stock Returns. *Journal of Finance*, **55**(2), pp. 913-936.

Chuang, W. and Lee, B. (2006). An Empirical Evaluation of the Overconfidence Hypothesis. *Journal of Banking and Finance*, **30**(9), pp. 2489-2515.

Cvitanic, J. and Kirilenko, A. (2010). High Frequency Traders and Asset Prices. *Working paper*, California Institute of Technology.

De Long, J., Shleifer, A., Summers, L. and Waldman, R. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, **98**(4), pp. 703–738.

Drukker, D. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal*, **3**(2), pp. 168-177.

Devereux, M,. Smith, B. and Gregor, W. (1994). International Risk Sharing and Economic Growth. *International Economic Review*, **35**(4), pp. 535-50.

Easley, D., Prado, M., and O'Hara, M. (2011). The Microstructure of the "Flash Crash": Flow Toxicity, Liquidity Crashes, and the Probability of Informed Trading. *The Journal of Portfolio Management*, **37**(2), pp. 118-128.

Fabozzi, F., Focardi, S., Focardi, S. and Kolm, P. (2006). Financial modeling of the equity market: From CAPM to Cointegration. 1st ed. Hoboken, NJ: John Wiley & Sons. pp. 529-552.

Froot, K., Perold, A. and Stein, J. (1992). Shareholder Trading Practices and Corporate Investment Horizons. *Journal of Applied Corporate Finance*, **5**(2), pp. 42-58.

Froot, K., Scharfstein, D. and Stein, J. (1992). Herd on the street: Informational inefficiencies in a market with short-term speculation. *Journal of Finance*, **47**(4), pp. 1461-1484.

Greene, W. (2003). Econometric Analysis. 5th ed. Upper Saddle River: Prentice Hall, pp. 285-304.

Hasbrouck, J. and Saar, G. (2009). Technology and Liquidity Provision: The Blurring of Traditional Definitions. *Journal of Financial Markets*, **12**(2), 143-172.

Hausman, J., and William, T. (1981). Panel Data and Unobservable Individual Effects. *Econometrica*, **49**(6), pp. 1377-1398.

Hendershott, T., Jones, C. and Menkveld, A. (2009). Does Algorithmic Trading Improve Liquidity? *Journal of Finance*, **66**(1), pp. 1-33.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *Journal of Finance,* **56**(4), pp. 1533-1597.

Jovanovic, B., and Menkveld, A. (2010). Middlemen in Limit-order Markets. *Working paper*, New York University.

Karoui, A. (2006). The correlation between foreign exchange rate volatility and stock exchange returns volatility: an emerging markets overview, *working paper*: HEC Montreal.

Kaufman, H. (2000). On Money and Markets: a Wall Street Memoir. 1st ed. New York: McGraw-Hill.

Kirilenko, A., Kyle, A., Samadi, M. and Tuzun, T. (2011). "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market. *Working paper*, CFTC and the University of Maryland.

Kyle, A. (1985). Continuous Auctions and Insider Trading. *Econometrica*, **5**, pp. 1315-1335.

Levine, R. (1991). Stock Markets, Growth, and Tax Policy. *Journal of Finance*, **46**(4), pp. 1445-65.

Lo, A. and Wang, J. (2000). Trading Volume: Definition, Data Analysis, and Implications of Portfolio Theory. *Review of Financial Studies*, **13**, pp. 257-300.

Morris, S., and Shin, H. (2004). Liquidity Black Holes. *Review of Finance*, **8**(1), pp. 1-18.

Obstfeld, M. (1994). Risk-Taking, Global Diversification, and Growth. *American Economic Review*, **84**(5), pp. 1310-29.

SEC (2010). Concept Release on Equity Market Structure. Release No. 34-61358; File No. S7-02-10.

Tversky, A. and Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, **185**, pp. 1124–1131.

Vayanos, D. (1999). Strategic Trading and Welfare in a Dynamic Market, *Review of Economic Studies*, **66**(2), pp. 219-254.

Vives, X. (1995). Short-term Investment and the Informational Efficiency of the Market. *Review of Financial Studies*, **8**(1). pp. 125-160.

Wang, F.A., (2001). Overconfidence, Investor Sentiment, and Evolution. *Journal of Financial Intermediation*, **10**(2), pp. 138-170.

Zhang, X. (2011). High-Frequency Trading, Stock Volatility, and Price Discovery. *Working paper*, Yale University.

Zhang, X. (2006). Information Uncertainty and Stock Returns. Journal of Finance, **61**(1), pp. 105-137.

IX. Appendix

Table A1: **List of S&P100 Firms in Sample.** This table provides a comprehensive list of the sampled companies and their corresponding "ticker" symbols.

AA	Alcoa Inc.	EXC	Exelon	NOV	National Oilwell Varco Inc
AAPL	Apple Inc.	F	Ford Motor	NSC	Norfolk Southern Corp.
ABT	Abbot Laboratories	FCX	Freeport-McMoran	NWSA	News Corp.
AEP	American Electric Power	FDX	FedEx	NYX	NYSE Euronext
ALL	Allstate Corp	GD	General Dynamics	ORCL	Oracle Corporation
AMGN	Amgen Inc.	GE	General Electric Co.	OXY	Occidental Petroleum Corp.
AMZN	Amazon.com	GILD	Gilead Sciences	PEP	Pepsico Inc.
AVP	Avon Products Inc.	GOOG	Google Inc.	PFE	Pfizer Inc.
AXP	American Express Inc.	GS	Goldman Sachs	PG	Procter & Gamble Co.
BA	Boeing Co.	HAL	Halliburton	PM	Phillip Morris International
BAC	Bank of America Corp	HD	Home Depot	QCOM	Qualcomm Inc.
BAX	Baxter International Inc.	HNZ	H. J. Heinz Company	RF	Regions Financial
BHI	Baker Hughes	HON	Honeywell	RTN	Raytheon Co.
BK	Bank of New York	HPQ	Hewlett Packard Co.	S	Sprint Nextel
BMY	Bristol-Myers Squibb	IBM	Int. Business Machines	SLB	Schlumberger
BRK.B	Berkshire Hathaway	INTC	Intel Corporation	SLE	Sara Lee Corporation
CAT	Caterpillar Inc.	JNJ	Johnson & Johnson Inc.	SO	Southern Company
C	Citigroup Inc.	JPM	J.P. Morgan Chase & Co.	T	AT&T Inc.
CL	Colgate-Palmolive Inc.	KFT	Kraft Foods Inc.	TGT	Target Corp
CMCSA	Comcast Corporation	KO	The Coca-Cola Company	TWX	Time Warner Inc.
COF	Capital One Financial Corp	LMT	Lockheed-Martin	TXN	Texas Instruments
COP	ConocoPhillips	LOW	Lowe's	UNH	United Health Group
COST	Costco	MA	Mastercard	UPS	United parcel Service Inc.
CPB	Campbell Soup Company	MCD	McDonalds Corp.	USB	US Bancorp
CSCO	Cisco Systems	MDT	Medtronic Inc.	UTX	United Technologies Corp.
CVS	CVS Caremark	MET	Metlife Inc.	VZ	Verizon Communications Inc
CVX	Chevron	MMM	3M Company.	WAG	Walgreens
DD	DuPont	MO	Altria Group	WFC	Wells Fargo
DELL	Dell	MON	Monsanto	WMB	Williams Companies
DIS	The Walt Disney Company	MRK	Merck & Co.	WMT	Wal-Mart
DOW	Dow Chemical	MS	Morgan Stanley	WY	Weyerhauser Co.
DVN	Devon Energy	MSFT	Microsoft	XOM	Exxon Mobil Corp.
EMC	EMC Corporation	NKE	Nike	XRX	Xerox Corp.
ETR	Entergy				

Figure A1: **Correlation between Quarter-on-Quarter Changes in HFT and Realised Volatility.** This figure is included to illustrate the relationship between HFT and volatility, and given lack of causal inference, to provide some level of support for potential endogeneity of the two variables.

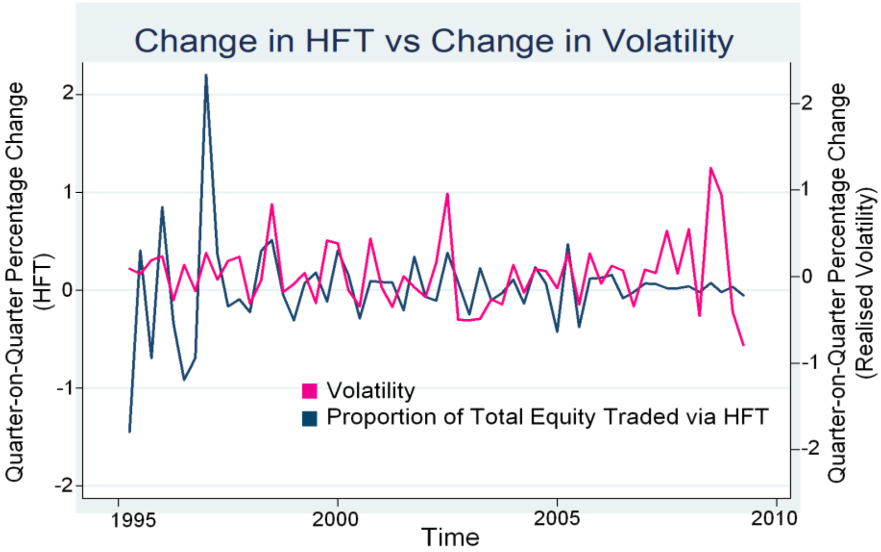


Table A2: **Definitions of Variables.** Several of the definitions are adopted from Zhang (2010). The names of variables that are included in brackets the variable description correspond to their names in the Compustat database such that the reader can reconstruct the relevant variables with ease if so desired.

HFT	High-Frequency Trading measured as the percentage of total quarterly trading volume conducted by High-Frequency Traders in the US. Constructed and provided by Zhang (2010).
Overconfidence	The quarterly equity trading volume due to lagged returns, as per Chuang and Lee (2006).
HighUncertainty	The upper quartile of the standard deviation in analysts' forecasts of a firm's EPS.
LowUncertainty	The lower quartile of the standard deviation in analysts' forecasts of a firm's EPS.
Earnings Surprise	Earnings per share (IBQ/(CSHOQ*AJEXQ)) in quarter q minus earnings per share in quarter q-4, deflated by stock price (PRCCQ/AJEXQ) in quarter q.
Book-to-Market Ratio	The ratio of the book value of equity (CEQQ) to its market value (CSHOQ*PRCCQ) at the beginning of quarter q.
Firm Size	The logarithm of the market value of equity (CSHOQ*PRCCQ) at the beginning of quarter q.
Leverage	The sum of the firm's short term and long term debt (DLTTQ+DLCQ) scaled by the market value of equity (CSHOQ*PRCCQ).
Inverted Price	Stock price from the CRSP monthly file, unadjusted for stock splits and dividends.
Money Supply	The quarterly sum of currency, traveller's checks, demand deposits, and OCDs, each seasonally adjusted separately. This is commonly referred to as M1.
Real GDP (QoQ)	The real quarter-on-quarter change in US Gross Domestic Product, seasonally adjusted at annual rate.
FedFunds Rate	The daily Federal Funds Rate aggregated over the quarter.
FX Index	A trade-weighted average of the foreign exchange value of the U.S. dollar against a subset of the broad index currencies that circulate widely outside the country of issue.
S-Change Dummy 1	Structural-change dummy variable accounting for the end of the dotcom bubble.
S-Change Dummy 2	Structural-change dummy variable account for beginning of the Credit Crisis.
Volume	The quarterly equity trading volume that cannot be explained by overconfidence.
Implied Volatility	Standard & Poor's VXO index, constructed from option contracts on the S&P100 index.
Realised Volatility	Market-Capitalisation weighted, quarterly average of daily price volatility of each firm's equity.

Table A3: **Descriptive Statistics.** This table presents descriptive statistics for the explanatory variables used in this paper's empirical analysis. Variable names correspond to those given in Table A2. The sample consists of 5630 firm-quarter observations between 1995q1 and 2009q2.

Variables	**Mean**	**Std. Dev.**	**Min**	**Max**
HFT	29.3918	16.3691	1.0000	57
Overconfidence	3.5031	0.4862	0.1759	6.2990
HighUncertainty	0.3306	0.4705	0.0000	1.0000
LowUncertainty	0.2753	0.4467	0.0000	1.0000
Earnings Suprise	-0.1714	1.9703	-5.6382	143.7334
Book-to-Market Ratio	0.3479	0.3042	-3.1546	4.9964
Firm Size	0.5782	0.6873	0.0034	6.0441
Leverage	0.5737	1.7657	0.0000	49.3372
Inverted Price	0.0273	0.0247	0.0000*	0.5464
Money Supply	2.2154	4.2355	-4.9000	16.6000
Real GDP (QoQ)	0.5864	0.7529	-2.3000	2.0000
FedFunds Rate	3.8375	1.9180	0.1800	6.5200
FX Index	90.5906	10.3523	70.8588	111.5926
Volume	0.0000**	7.8200	-5.6382	143.7334
Implied Volatility	22.3879	10.0360	10.8868	63.1768
Realised Volatility	0.5978	1.1028	0.0000	36.1113
Observations	**5630**			

Notes: * approximated to zero given 4 decimal places. ** Both negative and positive observations yield a mean close enough to zero to be approximated to zero given 4 decimal places.

Table A4: **Bivariate Correlation Matrix across Explanatory Variables.** This table presents correlation coefficients between all the explanatory variables used in this paper's empirical analysis. Variable names correspond to those given in Table A2, however, the variable names across the top of the figure have had to be intuitively shortened for illustrative simplicity. The consensus among the econometric literature is that any correlation between to explanatory variables below 0.8 is considered sufficiently low to ensure multicollinearity does not cause bias in the regressional estimates. This table thus describes the robustness of this paper's empirical specification with regard to potential multicollinearity bias.

	HFT	Over	Volume	LowU	HighU	EarnSup	BMR	FSize	Lev	InPrice	MS	FFRate	FX	RealGDP	SD1	SD2
HFT	1.0000															
Overconfidence	0.0351	1.0000														
Volume	0.0302	0.0032	1.0000													
LowUncertainty	0.0260	-0.0168	0.1178	1.0000												
HighUncertainty	-0.0140	0.0352	-0.0690	-0.4133	1.0000											
Earnings Surprise	0.0168	-0.0090	0.0341	0.0479	-0.1159	1.0000										
Book-to-Market	0.0413	0.0746	0.1398	-0.2514	0.2726	-0.0642	1.0000									
Firm Size	0.0098	-0.0082	0.3238	0.1736	-0.1073	-0.0773	-0.1723	1.0000								
Leverage	0.0160	0.0584	0.2716	-0.1322	0.2486	0.0164	0.4066	-0.0493	1.0000							
Inverted Price	0.0637	0.0616	0.3513	0.0392	-0.0313	0.0946	0.4190	-0.1581	0.4006	1.0000						
Money Supply	0.1367	0.0450	0.1845	-0.0020	0.0145	-0.0257	0.1973	0.0222	0.1095	0.2409	1.0000					
FedFunds Rate	-0.2001	-0.0406	-0.1567	-0.0085	0.0043	0.0231	-0.2110	0.0086	-0.0979	-0.2341	-0.8331*	1.0000				
FX Index	0.0786	-0.0471	-0.1386	0.1162	-0.0321	0.0237	-0.1302	-0.0580	-0.0535	-0.0189	-0.0258	0.1706	1.0000			
Real GDP (QoQ)	-0.0794	-0.0897	-0.2079	0.1149	-0.0555	0.0117	-0.2027	-0.0316	-0.1242	-0.2133	-0.4620	0.4271	0.2503	1.0000		
Structural D 1	-0.0178	0.0009	0.1758	-0.0426	-0.0217	-0.0372	0.1520	0.1015	0.0541	0.0867	0.3456	-0.4436	-0.8204†	-0.3214	1.0000	
Structural D 2	0.0951	0.0856	0.2160	-0.1493	0.0749	-0.0134	0.1781	0.0638	0.1045	0.1168	0.3035	-0.2650	-0.6313	-0.5549	0.4758	1.0000

Notes: * Although this is marginally over 0.8, the third and most robust estimation technique used –Arellano-Bover/Blundell-Bond Dynamic Panel Estimation – takes the differences between the explanatory variables. Consequently, the correlation between the two variables of interest in this case falls to -0.2866. † The correlation is very marginally over 0.8 for no particularly intuitive reason. This correlation cannot cause multicollinearity bias in the Arellano-Bover/Blundell-Bond estimation method used by this paper as the structural change dummies are omitted from this estimation technique due to the specific nature of the estimation technique not allowing for the binary values of a dummy variable.

Table A5: **Results for Im-Peseran-Shin Unit Root Test.** This table is included to support this paper's case HFT being trend stationary. The test results show that, given the null hypothesis that HFT is non-stationary, HFT is only trend-stationary.

	Z-T-Zilde-Bar Statistic	P-Value
Null: Unit Root (Assumes Common Unit Root Processes)		
HFT (No Trend)	1.6989	0.9553
HFT (Trend)	-47.2514	0.0000

Table A6: **Correlation between HFT and Volatility, and results of Granger Causality Tests.** This table is included to support this paper's case for endogeneity between the HFT and Volatility variables. This shows that the hypotheses that Volatility Granger-causes HFT and HFT Granger-causes Volatility cannot be rejected even at the 1% level. This compliments the high correlation between the two variables to strongly infer a case of two-way causality.

Volatility Measure	Correlation with HFT (Quarter-on-Quarter changes)	F-Statistic from Granger Test of Volatility on HFT (P-Value)	F-Statistic from Granger Test of HFT on Volatility (P-Value)
Realised Volatility	0.2099	5.42 (0.0002)	3.38 (0.0091)
Implied Volatility	0.1433	154.21 (0.0000)	135.52 (0.0000)

Table A7: **Correlation Test for Instrument Relevance.** This table is included to provide statistical support for the validity of the introduction of the NYSE Autoquote system as an instrument to proxy for High-Frequency Trading. The Table reveals the high correlation between the HFT and the instrument, thus implying the instrument is *relevant*.

	HFT	Instrument
HFT	1.0000	
Instrument	0.7251	1.0000

Note: There is no test for instrument exogeneity when the number of instruments – as is the case in this paper – is less than the number of endogenous variables. There is, however, no clear rationale as to why the adoption of the Autoquote System by the NYSE that facilitates a step-change increase in HFT post 2003q1, being used as a structural change dummy variable to instrument for HFT should be, in any way, related to the unobservable factors that explain stock market volatility. The appropriateness of this instrument is further supported by its use in several previous papers – see Section II.

Table A8: **Tests for Heteroskedasticity of Residuals.** This table contains two alternative tests for heteroskedasticity in both the Pooled OLS and Fixed Effects specifications, across both dependent variable measures for standard HFT estimation and Instrumented HFT estimation. The Pooled OLS and Fixed Effects specifications rely on the assumption of homoskedastic error variance. Consequently, as the results of the following tests show this does not hold, this paper uses robust standard errors for its empirical analysis, thereby diminishing efficiency of significance test but ensure they are at least correct.

Tests across *Standard* HFT Estimations

Breusch-Pagan/Cook-Weisberg Test

	Chi^2 Statistic	P-Value
Null: Error Variance is Homoskedastic		
OLS (Realised Volatility)	25850.68	0.0000
OLS (Implied Volatility)	972.88	0.0000

Modified Wald Test

	Chi^2 Statistic	P-Value
Null: Error Variance is Homoskedastic		
Fixed Effects (Realised Volatility)	12163.77	0.0000
Fixed Effects (Implied Volatility)	502.12	0.0000

Tests across *Instrumented* HFT Estimations

Breusch-Pagan/Cook-Weisberg Test

	Chi^2 Statistic	P-Value
Null: Error Variance is Homoskedastic		
OLS (Realised Volatility)	21329.44	0.0000
OLS (Implied Volatility)	84.02	0.0000

Modified Wald Test

	Chi^2 Statistic	P-Value
Null: Error Variance is Homoskedastic		
Fixed Effects (Realised Volatility)	13586.54	0.0000
Fixed Effects (Implied Volatility)	323.75	0.0000

Note: The *Modified Wald* test tests for groupwise heteroskedasticity. There is no test for heteroskedasticity in the Arellano-Bover/Blundell-Bond model as it has no relevance for the investigation given the data utilised in the study not allowing for robust standard errors in such a model should heteroskedasticity in the error term be present. See footnotes on page 13 for details.

Table A9: **Tests for Serial Correlation in the Residuals.** This table contains two alternative tests for serial correlation: the Wooldridge test for autocorrelation in panel data and the Arellano-Bond test for autocorrelation. The Wooldridge test is applicable across both Pooled OLS and Fixed Effects estimation methods, and consequently the results – assuming a 1% significance level when evaluating the F-statistics - suggest the presence of serial correlation only in specifications concerning *implied volatility*. Alternatively, the Arellano-Bond test is only applicable to Dynamic Panel twostep estimation procedures, given GMM standard errors. The results of the latter tests imply that this paper's use of Arellano-Bover/Blundell-Bond Dynamic Panel estimation eliminates the presence of serial correlation in the specification concerning *realised volatility*, however, not in the specification concerning *implied volatility*.

Wooldridge Test

Tests across *Standard* HFT Estimations

	F Statistic	P-Value
Null: No First-Order Autocorrelation		
Dep. Var. = Realised Volatility	5.577	0.0202
Dep. Var. = Implied Volatility	2359.474	0.0000

Tests across *Instrumented* HFT Estimations

	F Statistic	P-Value
Null: No First-Order Autocorrelation		
Dep. Var. = Realised Volatility	4.838	0.0302
Dep. Var. = Implied Volatility	3295.335	0.0000

Arellano-Bond Test

Test across *Realised Volatility* Estimation

	F Statistic	P-Value
Null: No Autocorrelation		
First Order	-4.295	0.0000
Second Order	0.682	0.4651

Test across *Implied Volatility* Estimation

	F Statistic	P-Value
Null: No Autocorrelation		
First Order	-9.482	0.0000
Second Order	9.036	0.0000

Note: In the Arellano-Bond test for autocorrelation, only the test of second-order autocorrelation is of importance given the nature of the estimation technique under which the error term is first-differenced. Drukker (2003) provides simulation results showing that the Wooldridge test has good size and power properties in samples that are similar in size to that observed in this paper.