Centre for Computational Finance and Economic Agents

Working Paper Series

WP070-13

Ata Türkoğlu

Normally Distributed High-Frequency Returns: A Subordination Approach

2013

University of Essex

CCFEA
Centre for Computational Finance and Economic Agents

Normally Distributed High-Frequency Returns: A Subordination Approach

Ata Türkoğlu

Centre for Computational Finance and Economic Agents
University of Essex, Wivenhoe Park, Colchester, Essex, CO43SQ
aturko@essex.ac.uk

Abstract

In this paper, market microstructure models are considered to assess the influence of the key components that derive asset prices. Synchronization techniques and sampling methods are reviewed. Alternatives to normally distributed asset returns, specifically subordinated Brownian motion is considered. The influential components of the price processes are then combined under tick time to recover normality of asset returns via subordination, a process I denote as "natural time". Normally distributed returns are obtained with the natural time approach which is also found to dominate GARCH results.

Keywords: *Market Microstructure, Time Synchronization, Normal Distribution, Subordination, Natural Time, GARCH*

I. Introduction

Lying at the heart of numerous financial studies, the assumption of normally distributed financial returns has been increasingly challenged. The assumptions of the efficient market hypothesis have been undermined by the microstructure manifestations observed in equidistantly time-spaced financial series. Several reasons emerge as responsible for the inability of the random walk model to account for empirically observed market dynamics. The lack of arbitrage, liquidity constraints, trading frictions and transaction costs, dependence of successive observations and non-stationarity are some of the key elements that contribute to the non-normality of empirical series (Mandelbrot (1963); Fama (1965); Engle (1982); Bollerslev 1986)).

The normal distribution assumption is central to many financial theories. However, empirical results, especially for high frequency data, often provide evidence against the normal distribution assumption. In this study, I will provide an alternative explanation to the empirical divergence of financial returns from the normal distribution. The first key observation one needs to make when evaluating the distribution of asset returns is that most statistical analysis in this area is conducted using a physical time approach. However, the superimposition of a time grid on the actual time of trades distorts the actual timing of trades. A second factor that is often overlooked is how the environment in which the prices are formed evolves over time.

The "natural time" approach that is detailed in Section III, addresses these two key observations and aims to test the validity of normal distribution under a high frequency setting. Under the natural time approach, instead of sampling in physical time, transaction time is used to record each trade as it materializes. By moving to the frequency (tick) domain, the need to force each trade into a time slot is removed. Furthermore, methods of diurnalization, which remove deterministic intraday patterns seen in physical time studies, are no longer necessary.

Natural time approach also addresses the trade "environment". By focusing on the factors through which prices are formed, the seemingly erratic behavior of volatility is accounted for. Variables derived from the limit order book are used to a form a gauge for volatility which is used to subordinate raw returns resulting in normally distributed return series. The ultimate goal of this research is to find the best approximation for the "natural time" that results in normally distributed subordinated returns. The choice of sampling frequency and variables used in the subordinator function are the key to the success of this method.

Previous calendar time based subordination studies have found volume and number of trades to contain volatility related information such that normality could be recovered under certain periods (Clark (1973); Ané and Geman (2000); Velasco-Fuentes and Ng (2010)). Corresponding variables under the frequency domain, namely volume and duration, are used here as well. However, by using only these variables, the literature has neglected important information contained in the order book which can be used to explain the price formation process. For this reason, in addition to volume and duration, order book variables such as the imbalance in the standing order book and the difference in the number of bid and offer initiated trades are used to augment the models mentioned above. Asymmetric versions of the same subordination procedure are also tested.

The natural time approach is applied to 3 highly liquid LSE listed stocks for 4 quarters each. In more than half the cases normality of returns is achieved. Since subordination essentially accounts for the volatility in the data, the natural time approach is tested against the standard GARCH(1,1) model. Natural time is found to dominate GARCH results with respect to normality, with the exception of a single case. These findings suggest that volatility can be modeled efficiently under tick time sampling so much so that subordination results in normally distributed returns.

The results in this study support the normally distribution assumption that is central to finance. However, they also point to the changes one needs to make in the standard model such as the sampling methodology. In addition to providing evidence for the normal distribution assumption, this research contributes to the literature by focusing on order book variables which contain relevant information that may be used to forecast volatility. The variables found to be influential here can be employed by market players to adjust their leverage or by financial regulators to assess the health of market. Either use will contribute to the efficiency of financial markets.

Section 2 reviews key market microstructure models to identify the instrumental elements of the price process. Sampling methodologies and synchronization techniques are discussed along with previous subordination studies. It is in this section, that it will become apparent to the reader why the move to tick time sampling is necessary. Moreover, the rationale behind variables used to subordinate the raw returns becomes clear. In Section 3, the natural time model is introduced and preliminary analysis of the empirical data is presented. Section 4 is dedicated to the model results and Section 5 concludes.

II. Literature Review

The first subsection of this three-part literature review focuses on how information is conveyed in financial markets. Several market microstructure models that explain trading patterns are reviewed and variables that effect price variance are identified. The natural time approach used in this research combines the variables presented in this subsection while accounting for variance related information. The second subsection reviews synchronization methods under physical time and identifies the inherent problems of working in the time domain. Alternate sampling methodologies are reviewed and the benefits of using tick time sampling, which forms an integral part of the natural time approach, are discussed. Finally, the last subsection reviews previous subordination based studies aimed at recovering normality. Natural time combines the variables identified in the first subsection under a stochastic subordination setting to recover normality of returns under tick time sampling.

a. Trading Patterns in Market Microstructure

Numerous market microstructure effects that cause return anomalies have been documented in detail in the extensive microstructure literature (Aït-Sahalia et al. (2010); Bandi & Russell (2008); Dacorogna et al. (2001)). Much focus has been given to bid-ask spread with two main strands of models. Inventory-based models argue market makers adjust their quotes to mirror their inventory positions, while information-based models focus on the costs associated with adverse selection.

Glosten and Milgrom (1985) mapped the bid-ask spread as the market maker's tool against traders with insider information. Easley and O'Hara made important extensions to this model introducing the possibility of no information, varying trade sizes and choice of not trading (Easley & O'Hara (1987); Easley & O'Hara (1992)). Information-based models underscore important market dynamics where trade size, duration between consecutive trades and lack of trades reveal information about price dynamics.

These models also gave rise to the "stealth trading hypothesis" where market participants with insider information try to avoid information leakage while submitting orders (Barclay & Warner (1993); Chakravarty (2001)). Insiders are forced to find a balance between the risk of effecting prices adversely with block trades - impact risk - and price risk, due to order slicing. The stealth trading hypothesis shows that volume of trades affect market impact.

Kyle (1985) identifies three major components to market liquidity, namely tightness, depth and resiliency. Given this setup "Kyle's λ" becomes a measure of market sensitivity to transaction size where order book imbalances can be used to infer impact of order size (Aldridge (2010)). Ng (2008) tested the absorption limits of financial markets within a nonlinear framework and report that markets are incapable of absorbing large block trades introducing additional "time costs of liquidity". These findings suggest market participants actively try to balance liquidity and information costs, and necessitate the need to use some form of liquidity measure in order to account for high frequency dynamics.

Additional microstructure effects have surfaced with greater availability of high frequency data. The effect of scheduled macroeconomic announcements on diurnal financial returns and volatility is one such area. Scheduled announcement studies show the drying of liquidity and a sudden spike right before and after new releases. Savaşer (2011) finds evidence of price contingent stop-loss and take profit orders surrounding scheduled announcements. Her findings highlight the role of order book imbalances in accounting for news effects.

Announcement reactions, scheduled or unscheduled, bring another dimension to the price process. The intensity of trades following sudden changes of sentiment reflects an inevitable herding behavior following important news. The number of transactions spikes and important adjustments to asset prices are realized during these short time intervals. This inherent correlation between number of transactions and return variance has been previously tested in a physical time setting (Ané and Geman (2000)).

b. Synchronization & Sampling

High frequency time series, unevenly spaced in physical time, requires synchronization for statistical inference. Two major synchronization methods emerge in literature for homogenizing high frequency series of a single asset in calendar time.

The "previous tick" method (e.g., Wasserfallen and Zimmerman (1985)) which takes into account the most recent price at or before a given time, is perhaps the most frequently used method of fitting inhomogeneous data into an evenly spaced time grid. One major shortcoming of this method however is spurious jumps observed in case of extended periods of missing data (Dacarogna, Gençay, Müller, Olsen and Pictet (2001)).

The second method, "linear interpolation" forms the homogenous time series by interpolating between the nearest tick data observed just before and after the grid time. While the difference between the two methods might be negligible, linear interpolation violates causality.

Sampling in physical time further requires data to be adjusted for deterministic market patterns. Diurnalization, which accounts for the deterministic intraday patterns in the data, utilizes splines, Fourier transforms or kernel based estimators. Fourier transforms employed in Andersen et al. (2003) are very smooth processes whereas spline methods employed by Engle and Russell (1998) are much more flexible. However, the choice of nodes may present problems. Kernel based estimators as in Ng (2008) face similar setbacks as to the choice of the kernel bandwidth. Empirically, diurnalization may produce satisfactory results but its exact effects on the data is little explored (Martens et. al (2002); Allen et al. (2009)).

Despite the fact that the literature predominantly focuses on sampling in physical time, this is not the only option. Generally referred to as tick time, transaction time accounts for trades as they materialize. Contrary to calendar time methods which require synchronization, in tick time each transaction falls nicely on the tick grid. Hence, sampling in tick time inherently eliminates the need for synchronization. Furthermore, by sampling at a fixed number of ticks, one completely avoids the processes of diurnalization as the clock moves faster (slower) when market activity is high (low). Given its advantages in adjusting for market seasonality, tick time will be used in this research while evaluating the distribution of stock prices in the high frequency domain (Oomen (2006); Dacorogna et. al (1993)).

c. Time Deformation

The empirical divergence of asset returns from normality, excess skewness and fat tails, has long spurred interest in alternate distributions such as the exponential and t-distribution. Merton (1976) proposed the addition of jumps to the original continuous stochastic diffusion process in Black and Scholes (1973) to account for fat tailed asset returns. Tauchen and Pitts (1983) explored the possibility of normal mixture distribution, while Mandelbrot (1963) examined the stable distribution. Mandelbrot posited that although asset returns were approximately independent they were characterized by unbounded second moments and advocated the use of stable Paretian distribution. However, substantial evidence against unbounded first and second moments undermines the applicability of stable processes to financial return series (Perry (1983); Cont (2001)).

Clark (1973) was the first to apply the subordination process to asset prices to recover normality of returns. He conjectured that financial return series, which are semimartingales, could be defined as subordinated Brownian motions. Clark tested the applicability of trade volume as a subordinator for cotton futures and found evidence in favor of the Gaussian distributed asset returns within an i.i.d. subordinator increments setting using cumulative trade volume. Karpoff (1987) also documented the connection between large trades and large price swings and conjectured that it might be linked to both factor's shared link to the underlying information process.

Ané and Geman (2000) generalized the subordination framework by relaxing Clark's i.i.d. assumption in a finite variance jump setting. They have found transaction frequency to be a better subordinator compared to volume for S&P future contracts. Geman (2002) has also shown that the directing process can also be interpreted as the "mixing factor" within a normal mixture distribution setting, an often used distribution to account for excess skewness and kurtosis in stock returns. Murphy and Izzeldin (2006) however questioned the reliability of moment estimation methods in Ané and Geman (2000) and presented counter evidence on recovery of normality using re-centered number of trades or volume.

Similar to Ané and Geman (2000), Huth and Abergel (2012) used number of transactions to subordinate the returns for multiple assets. In a multivariate framework, Huth and Abergel (2012) chose to sample each time a trade occurs in any one of the assets creating a "common stochastic clock". Then by subordinating with an event time N, which represents the total number of trades in all assets under consideration, they obtained results that support normally distributed returns for 4 asset pairs. However, the large number of trades they have used to obtain normality, which in one case reaches almost 6,000, and the fact that the joint stochastic clock used only produces reliable results if the asset pairs have similar trading patterns suggest that their findings may be mostly attributed to aggregation.

Velasco-Fuentes and Ng (2010) further investigated the use of volume and number of trades as stochastic time changers. In a study using FTSE-100 futures tick data they have tested cumulative volume, total number of trades and their linear and quadratic combinations to recover normality. They have also explored the possibility of asymmetric market response to the sign of returns in order to reduce skewness. Using first and second order functions of volume and number of trades Velasco-Fuentes and Ng recover normality in two of the four sub-periods.

III. Methodology

In this research, I take an atypical approach to stochastic subordination. Unlike its predecessors in literature, this research will be conducted under tick time. By moving to transaction time, the applicability of the normal distribution assumption is tested for the first time under the frequency domain. When sampling in tick time, daily deterministic patterns present under physical time disappears naturally and the need to use diurnalization methods is eliminated. Moreover, additional errors introduced while conforming to a calendar time grid is no longer present under tick time.

Furthermore, by using volume-weighted transaction prices and their returns, I also avoid using quotes which react asymmetrically during unidirectional market swings. Hence, log-returns calculated from volume-weighted prices are not subject to microstructure contaminations such as non-synchronous updating of quotes. Instead the effects of the bid-ask bounce are reduced substantially, and the price on which market players agree is determined by the volume of transactions.

Four important components affecting price evolution emerged from previous sections, namely volume, duration, market liquidity and order imbalance. The use of these components under tick time in a unified subordination framework will be a key contribution of this work. I will account for various market dynamics by extending the arsenal of possible factors that are most closely related with information arrival and intrinsic time.

Volume, as per its impact to push prices in a given direction is the first of these factors. However, as shown in Gillemot, Farmer and Lillo (2006), volume and number of trades cannot totally account for the volatility observed in the stock markets. This may be caused by the imperfect correlation these variables have with the latent process which drives volume, number of trades and volatility. Hence, as per the findings of information-based models, duration between trades is also added to the subordination framework to account for the speed with which market participants act in physical time.

The use of duration augments the model in two respects. Given the stealth trading reasoning presented in the previous sections, and the information-based MMS models, the duration between trades not only helps capture the speed with which the market moves in real-time but also reveals the private information content. By including duration between trades I allow physical time related information to be included in the frequency domain.

In addition to volume and duration, proxies for the liquidity component of the market are included in my model, namely the net order book imbalance, the difference between standing bid and ask orders, net traded volume imbalance, the volume difference between bid and ask initiated trades, and net initiator imbalance, the difference between the number of aggressors on buy and sell sides. The addition of liquidity variables sets the scene in which the trades occur and adjusts for the impact of block or frequent trades given market depth or resiliency. However, it is highly unlikely for liquidity conditions to affect prices like volume of trades, where one consistently drives prices while the other acts as a determinant (multiplier) of price impact for a given trade. For this reason, the effects of order book imbalance on the price process is likely to be non-linear. We will test this assumption during the subordination process.

By including these possibly omitted variables in the subordinator, I aim to regain normality of asset returns during all states of the world. The use of an asymmetric response function similar to the one in Velasco-Fuentes and Ng (2010) is also examined. Thus, in addition to linear combinations of the three factors identified, the importance of nonlinear models will also be tested, given the inability of linear models in explaining asset price fluctuations.

The stochastic time change that will be applied to the raw return series can be described as follows. Define the log-price series of an asset sampled under calendar time as:

$$X_{cal}(c_i) = \{X(c_1), X(c_1), X(c_1), \dots, X(c_{n-1}), X(c_n)\}. \tag{1}$$

Similarly define a stochastic (parent) process:

$$W_{cal}(q_i) = \{W(q_1), W(q_2), W(q_3), \cdots, W(q_{m-1}), W(q_m)\}, \tag{2}$$

where q denotes market's intrinsic time. The stochastic parent process, W_{cal}, is Brownian Motion in our case.

If a strictly increasing stochastic process:

$$s(c_i) = \{s(c_1), s(c_2), s(c_3), , \dots, s(c_{n-1}), s(c_n)\}, \tag{3}$$

exists such that :

$$q = s(c), \tag{4}$$

then, the price process can be summarized as:

$$X_{cal}(c) = W_{cal}\big(s(c)\big). \tag{5}$$

In Equation (5), the price series $X_{cal}(c)$, is said to be subordinated to the parent process $W_{cal}\big(s(c)\big)$ and the subordinator $s(c)$ is a càdlàg process that measures market's intrinsic time which flows at variable rates (Velasco-Fuentes and Ng (2010)).

Alternatively, the return series, $r_{cal}(c)$ can be expressed as:

$$r_{cal}(c) = \Delta W_{cal}\big(s(c)\big), \tag{6}$$

where $\Delta W_{cal}\big(s(c_i)\big) = W_{cal}\big(s(c_i)\big) - W_{cal}\big(s(c_{i-1})\big)$.

Sampling under tick time, where t represents transaction time, asset returns, $r_{tick}(t)$ can then be expressed as:

$$r_{tick}(t) = \Delta W_{tick}\big(s(t)\big) \tag{7}$$

where $\Delta W_{tick}\big(s(t_i)\big) = W_{tick}\big(s(t_i)\big) - W_{tick}\big(s(t_{i-1})\big)$.

Then, given that subordinated parent process $\Delta W_{tick}\big(s(t)\big)$ in Equation (7), is a Brownian Motion, normally distributed returns should be obtained by using the following transformation:

$$R_t = \frac{r_{tick}(t)}{\sqrt{s(\cdot)}} \sim N(\mu_{sub}, \sigma_{sub}), \tag{8}$$

where R_t and $r_{tick}(t)$ represents time deformed and tick returns respectively, and $s(\cdot)$ is the subordination function.

Let the unique subordinator $s_N(\cdot)$, with which the return series achieve perfect "normality" under tick-time sampling, be called "natural time". Then the goal of this study is to find the best approximation for natural time via the choice of sampling frequency and subordinator s_N, using various linear and nonlinear combinations of volume, duration and order book imbalance parameters.

The linear subordinator utilized for in this study can be summarized as:

$$s(\cdot) = \rho\ subordinator. \tag{9}$$

Maximum likelihood estimation (MLE) methodology is used to estimate the coefficient α where a (subordination-adjusted) normal distribution is specified as the resulting distribution. The subordination-adjusted log likelihood function that is employed in MLE estimations takes into account the fact that this subordinated return series follow a normal distribution with unknown but finite mean and variance. Additionally, the use of 100-tick sampling frequency dampens autocorrelations within the tick data.

Given this structure, the joint probability distribution function for the subordinated tick time series can be expressed as:

$$f\big(R_{t_1}, R_{t_2}, \dots, R_{t_n} \big| \cdot, \mu_{tick}, \sigma_{tick}\big) \tag{10}$$

where R_t is the time deformed returns described in Equation (8).

Equation (10) can also be expressed as:

$$f\big(R_{t_1}, R_{t_2}, \dots, R_{t_n} \big| \cdot, \mu_{tick}, \sigma_{tick}\big) = \frac{1}{\sigma_{tick}{}^{n}\left(\sqrt{2\pi}\right)^{n}}\ exp\left\{-\frac{1}{2}\sum_{tick=1}^{\infty}\frac{\left(\frac{r_{tick}(t)}{\sqrt{s(\cdot)}}-\mu_{tick}\right)^2}{\sigma_{tick}^2}\right\}. \tag{11}$$

Then the log-likelihood function is:

$$\ln LF\,(\cdot, \mu_{tick}, \sigma_{tick}) = -\frac{n}{2}\ln\sigma_{tick}^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{tick=1}^{\infty}\frac{\left(\frac{r_{tick}(t)}{\sqrt{s(\cdot)}}-\mu_{tick}\right)^2}{\sigma_{tick}^2}. \tag{12}$$

To evaluate the ability of the linear subordinators to transform the tick returns into a normally distributed series, time deformed return series are tested with Kolmogorov-Smirnov (KS) and Jarque-Bera (JB) tests. The use of KS test is justified by the favorable aspects of the dataset used in this study. KS test, which requires a large dataset, is capable of providing useful estimations due to the large number of data points present. Furthermore, unlike the Jarque-Bera statistics, which focuses on skewness and kurtosis, KS test is known to be sensitive to the location and shape parameters. By using both KS and JB statistics, I account for all of the first four moments while testing for normality.

To better assess the value of proposed subordinators, the MLE procedure is augmented to include multiple subordinators as well as additional structural changes to the subordinator function itself. An asymmetric subordination function is added into the MLE procedure to account for possible differences in the behavior of the subordinator to the sign of returns.

The returns and their corresponding subordinators are classified according to the sign of returns. The positive and negative return series are then used to estimate the coefficients for the subordinators. The corresponding results are combined with the two original return series, classified according to the sign of returns, to produce the subordinated return distribution.

Additionally, given the existing literature on the autoregressive nature of variance, the past values of squared returns were used to augment the subordinator. AR(1) terms are used to test this hypothesis. Finally, the asymmetric and autoregressive extensions to the original linear model are combined to produce the fourth structural model for the subordinator.

The respective formulas[1] for each of the four subordinator functions, namely linear, autoregressive, asymmetric and autoregressive asymmetric, are presented in Equations (13)-(16).

The linear subordinator is of the form:

$$s(\cdot) = \alpha\, volume + \beta\, duration + \gamma\, Init\ Imb^2 + \delta\, Vol\ Imb^2. \tag{13}$$

The autoregressive subordinator function includes past values of the squared returns:

$$s(\cdot) = \vartheta\, {r_{tick}}^2(t-1) + \alpha\, volume + \beta\, duration + \gamma\, Init\ Imb^2 + \delta\, Vol\ Imb^2. \tag{14}$$

The asymmetric subordinator is of the form[2]:

$$s(\cdot) = \begin{cases} \alpha^+\, volume^+ + \beta^+\, duration^+ + \gamma^+\,(Init\ Imb^+)^2 + \delta^+\,(Vol\ Imb^+)^2, & r \ge 0 \\ \alpha^-\, volume^- + \beta^-\, duration^- + \gamma^-\,(Init\ Imb^-)^2 + \delta^-\,(Vol\ Imb^-)^2, & r < 0 \end{cases}. \tag{15}$$

Finally, the autoregressive asymmetric subordinator function can be expressed as:

$$s(\cdot) = \begin{cases} \vartheta^+\, {r_{tick}}^+)^2(t-1) + \alpha^+\, volume^+ + \beta^+\, duration^+ + \gamma^+\,(Init\ Imb^+)^2 + \delta^+\,(Vol\ Imb^+)^2, & r \ge 0 \\ \vartheta^-\, {r_{tick}}^-)^2(t-1) + \alpha^-\, volume^- + \beta^-\, duration^- + \gamma^-\,(Init\ Imb^-)^2 + \delta^-\,(Vol\ Imb^-)^2, & r < 0 \end{cases}. \tag{16}$$

[1] The formulas presented in Equations (13)-(16) represent the final functional forms used in estimations.

[2] The + and − signs indicate the respective series for positive and negative returns.

Subordination essentially aims to account for the heteroscedasticity in returns, by utilizing volatility related information. Thus, in many respects, this study could be classified as a volatility based approach. The use of past square returns then naturally brings to mind the GARCH model (Bollerslev (1986)). Hence, to make an accurate comparison, a GARCH(1,1) model is separately estimated. Returns are then subordinated using these estimated GARCH parameters to construct a benchmark model.

The GARCH(1,1) model used in estimations can be summarized as follows:

Let error term ϵ_{tick} represent the mean-adjusted returns which can be decomposed into a time-varying standard deviation σ_{tick} and a stochastic component $Z_{tick} \sim N(0,1)$.

$$\epsilon_{tick}(t) = \sigma_{tick}(t)\, Z_{tick}(t). \tag{17}$$

Then the conditional variance under a GARCH(1,1) specification can be expressed as:

$$\sigma_{tick}{}^2 = \varphi_0 + \varphi_1 \epsilon_{tick}{}^2(t-1) + \omega_1 \sigma_{tick}{}^2(t-1), \tag{18}$$

where $\varphi_0 > 0$, $\varphi_1 \geq 0$, $\omega_1 \geq 0$ and $\varphi_1 + \omega_1 < 1$.

Then the log-likelihood function for GARCH(1,1) estimation becomes:

$$\ln LF\,(\mu_{tick}, \sigma_{tick}) = \sum_{i=1}^{n} -\frac{1}{2}\ln 2\pi - \frac{1}{2}\sigma_{tick}{}^2(i) - \frac{1}{2}\frac{\epsilon_{tick}{}^2(i)}{\sigma_{tick}{}^2(i)}. \tag{19}$$

IV. Data & Analysis

The high frequency dataset utilized in this study uses Level 2 SETS data provided by the London Stock Exchange (LSE) where stocks are traded in a continuous-time double auction system. The LSE sorts and matches orders first by their price competitiveness and then by their time of submission. The Level 2 dataset includes the whole order book depth at any given point in time as well as the actual trade times and prices for realized trades. The order book data includes "public" orders that appear on the order book and excludes order types such as non-persistent or Iceberg orders. Hence, the bulk of the information contained in the order book stems from limit and market orders.

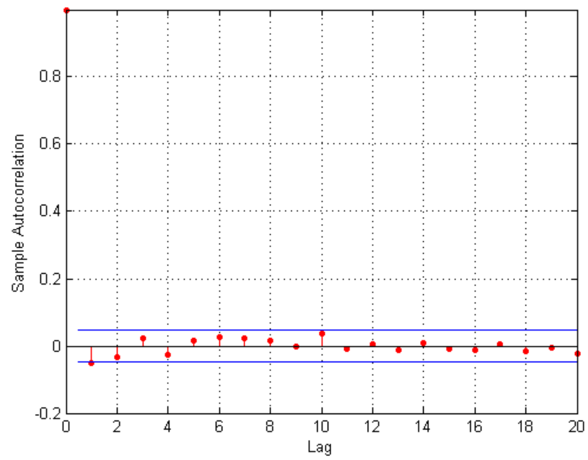
The period under study spans from July 2007 to June 2008. Taking into account the large market swings during this time, the whole dataset is split into four 3-month periods where the first period (P1) spans from July 2007 to September 2007. Similarly, P2 covers October 2007 – December 2007, P3 January 2008 – March 2008 and P4 April 2008 – June 2008. Three highly liquid stocks, HSBC, SAB Miller and Royal Dutch Shell, are selected for the purpose of this study. Each stock is analyzed on a period by period basis so as to not include irrelevant past data in high frequency analysis.

The first obstacle one needs to address when working with financial series is autocorrelation. This phenomenon becomes even worse as the sampling frequency is increased. The fourth period for HSBC stock was chosen for exemplification purposes and Ljung-Box test was applied to several sampling tick sizes using a lag size of 20. Autocorrelation was present up to a sampling frequency of 100 ticks. Autocorrelation and partial autocorrelation functions for HSBC P4 with a sampling frequency of 100-ticks were also mapped via a correlogram and ACF and PACF decay rate did not converge albeit being small. Similar results were obtained for other stocks. ACF functions and Ljung-Box (LB) test statistics for HSBC are presented in Table 1 and Figure 1 respectively.

Table 1: Ljung-Box Test (Lag=20)

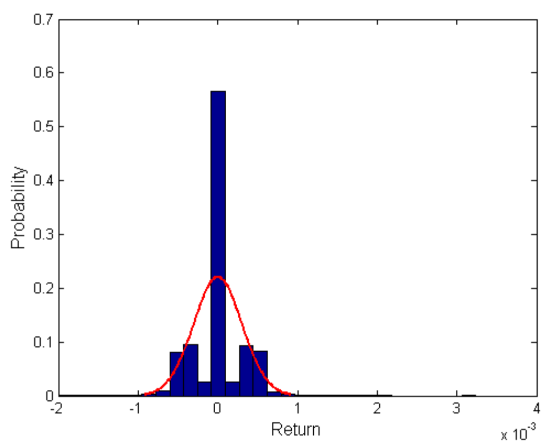
Sampling Tick	LB Test	p-value
5	226.8479	0
10	105.6981	< 0.0001
20	56.3034	< 0.0001
50	33.5694	0.0292
100	16.1198	0.7092
200	17.0837	0.6475

Figure 1: Autocorrelation Function for Returns of HSBC Stock Prices in Period 4 Sampled at 100 Ticks



Due to the nature of ultra-high frequency data, additional measures to deal with price discreteness were necessary. Figure 2 shows the return histogram fitted on a normal distribution curve for tick returns.

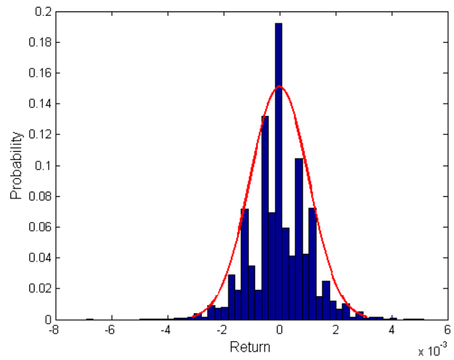
Figure 2: Histogram for Tick Returns for HSBC Stock Prices in Period 4



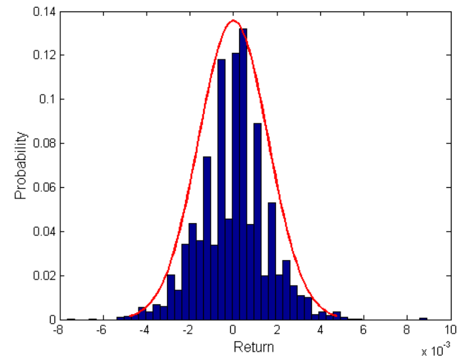
As is apparent from Figure 2, raw returns at the single-tick sampling frequency, is dominated by price discreteness. Hence, several sampling frequencies were tested to ascertain the exact effects of sparse sampling on the distribution of returns, results can be found in Table 2. The graphs in Figure 3 illustrate the relationship between decreasing sampling frequency and return distribution.

Figure 3: Distribution vs. Sampling Frequency: HSBC Stock Returns in Period 4

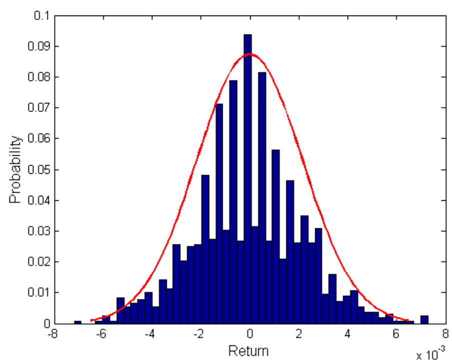
Sampling Frequency: 20-Ticks



Sampling Frequency: 50-Ticks



Sampling Frequency: 100-Ticks



Sampling Frequency: 200-Ticks

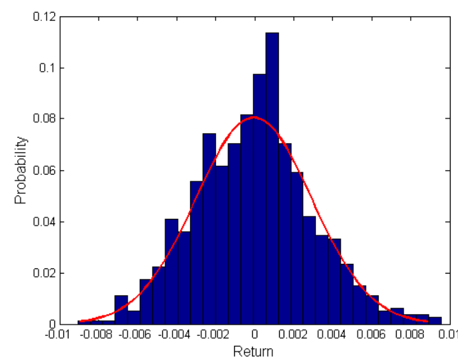


Figure 3 shows that at sparser sampling frequencies the return distribution approaches normality. However, neither simple aggregation of returns to produce normality is new to the literature, nor would it be feasible in a subordination study testing the limits of the sampling frequency under which subordination still produces normality. Thus, to assess the exact effects of sampling frequency on price discreteness and the distribution of returns, the first moments are computed. Table 2 contains the results.

Table 2: Sampling Frequency vs. Moments: HSBC Stock Returns in Period 4

Sampling Frequency	Mean	Variance	Skewness	Kurtosis
1-Tick	-2.41 e-7	9.34 e-8	0.0154	4.4958
5-Ticks	-1.30 e-6	3.09 e-7	0.0103	4.1357
10-Ticks	-2.67 e-6	5.77 e-7	0.0636	4.1606
20-Ticks	-3.71 e-6	1.08 e-6	0.0345	4.2115
30-Ticks	-6.97 e-6	1.55 e-6	0.0672	4.1877
40-Ticks	-6.12 e-6	2.07 e-6	0.0327	4.1762
50-Ticks	-4.70 e-6	2.52 e-6	0.0488	4.0486
60-Ticks	-9.25 e-6	2.93 e-6	0.0989	4.0019
70-Ticks	-6.56 e-6	3.45 e-6	-0.0017	3.6552
80-Ticks	-9.57 e-7	3.83 e-6	0.0067	3.3744
90-Ticks	-1.12 e-5	4.37 e-6	0.0056	6.4033
100-Ticks	-1.37 e-5	4.67 e-6	-0.0020	3.2000
200-Ticks	-1.86 e-5	5.96 e-6	0.0108	3.2685
300-Ticks	-3.24 e-5	6.94 e-6	-0.0046	3.2953
400-Ticks	4.17 e-6	7.93 e-6	0.0663	3.0219
500-Ticks	-4.66 e-5	8.85 e-6	0.1828	3.1304

Table 2 suggests the use of 100 ticks as the sampling frequency is appropriate, as sampling at lower frequencies causes further negative skewness and abnormally low kurtosis values for a high frequency return series. The results presented in Table 2 were reproduced for all stocks and periods but they are included here to conserve space. However, the effects of sampling frequency do not vary much from stock to stock. Thus, a sampling frequency of 100-ticks is used for all stocks and periods unless mentioned otherwise. In cases, where different sampling frequencies have been used, the moments of the resulting raw distribution was utilized to determine the new sampling frequency.

Upon selection of the sampling frequency, the influential variables discussed in the previous sections can now be tested for validity. Trade volume, cumulated across the selected number ticks, and its log transformation are used to find the impact of trade size on price formation. Duration between each sampling point is also used to assess the urgency with which orders have been filled. In order to assess how the liquidity state of the market influences price movements, the imbalance in the order book is computed in various different ways. The *Imbalance* term cumulates the volume difference between bid and ask sides for the whole depth of the order book and averages this number for across the selected sampling frequency. Similarly, *Level* 1 *Imbalance* and *Level* 3 *Imbalance* applies the same procedure to the first 1 and 3 levels from the top of the order book, respectively.

The number of transactions has been previously used by Ané and Geman (2000) to subordinate the price processes. This measure provides partial information on the number of entities involved, but does not make any distinction between the direction of trades. Thus, a more transparent measure is needed, which can be obtained by looking at the difference in the number of unique trades in a given interval. At each tick, which may include multiple buy and sell orders, the number of initiators for each side is found and the difference is recorded[3]. This number is then cumulated for the span of sampling frequency and divided by the number ticks to form *Initiator Imbalance* variable. The same process is repeated for *Volume Imbalance* taking into account the volume of trades. A negative number means excess sell side orders where as a positive number denotes buys side for these two variables. Finally, as per the non-negativity constraint dictated by Equation (8), log transformations of squared *Initiator Imbalance* and *Volume Imbalance* are added into the list of possible variables.

[3] On occasions where a single aggressor matches with multiple orders on the order book, the transaction is classified as a single trade with volume equal to the sum of all corresponding orders.

Table 3: Regression Analysis[4]: Mean Adjusted Squared Returns for HSBC in P4 Sampled at 100 Ticks

Subordinator	Regression Statistics		
	constant	coefficients	R^2
Volume	1.8288 e-6	2.5488 e-12 (0)	0.0211
Duration (sec)	6.0325 e-6	-1.2757 e-9 (0)	0.0099
Imbalance	4.6929 e-6	-7.3210 e-14 (0.8409)	0
Level 1 Imbalance	4.6660 e-6	6.5167 e-12 (0.5088)	0.0003
Level 3 Imbalance	4.6626 e-6	3.6090 e-12 (0.2688)	0.0007
Initiator Imbalance	4.6872 e-6	-2.8378 e-7 (0.3850)	0.0004
Volume Imbalance	4.6357 e-6	7.1408 e-11 (0.2607)	0.0008
Log-Volume	-3.7174 e-5	3.0175 e-6 (0)	0.0221
Log-InitImb²	4.0356 e-6	3.0759 e-8 (0)	0.0121
Log-VolImb²	-1.8930 e-6	4.6257 e-7 (0)	0.0251

Tables 3 shows a peculiar outcome. None of the standing order book variables that describe market liquidity conditions, namely *Imbalance*, *Level* 1 *Imbalance* and *Level* 3 *Imbalance* are found to be significant in explaining squared returns. This is an unexpected finding, which suggests that variables related to the active trading environment already contain the necessary liquidity information. For this reason, all standing order book variables are dropped from further study. Additionally, *Initiator Imbalance* and *Volume Imbalance*are also removed from further analysis, as per the non-negativity constraint[5]. Although the remaining five subordinators are significant in normalizing the return series at the 5% significance level, confirming the findings of Clark (1973) and Ané and Geman (2000), volume is also dropped from further subordination runs as similar results can be produced by the log-volume.

[4] The values in parentheses in Table 3 and all of the tables that follow show respective p-values for each variable.

[5] Logarithms of squared initiator and volume imbalance are referred to as initiator imbalance and volume imbalance from this point on.

The coefficients, p-values, log-likelihood function value as well as KS and JB test statistics for the subordinated returns using the produce described in Equations (13)-(16) are presented in Table 4:

Table 4: Multiple Subordination[6] Results for HSBC Returns in P4 Sampled at 100 Ticks

Subordinator	*Linear*	*Autoregressive*	*Asymmetric*		*Autoregressive Asymmetric*	
μ	1.2869 e-14 (1)	-9.7291 e-11 (0.1538)	-3.5890 e-10 (0.0332)		1.4613 e-11 (0.0591)	
σ	2.6896 e-10 (1)	6.3386 e-10 (0)	7.9000 e-9 (0)		3.1514 e-10 (0)	
r_{tick-1}^2	-	7.3386 e+4 (0)	-	-	3.7485 e+5 (0)	2.5673 e+5 (0)
Volume	1.9926 e+6 (0)	1.7613 e+5 (0)	1.7673 e+3 (0)	1.7660 e+3 (0)	9.2414 e+5 (0)	7.0897 e+5 (0)
Duration	2.4668 e+6 (0)	1.8243 e+5 (0)	1.7373 e+3 (0)	1.7701 e+3 (0)	9.6022 e+5 (0)	7.4304 e+5 (0)
Log-InitImb^2	7.5385 e+4 (0)	1.2095 e+3 (0)	2.4497 e+1 (0)	2.5722 e+1 (0)	6.6751 e+3 (0)	4.7491 e+3 (0)
Log-VolImb^2	1.9315 e+6 (0)	1.7633 e+5 (0)	1.4189 e+3 (0)	1.4644 e+3 (0)	9.0276 e+5 (0)	7.3118 e+5 (0)
Log-likelihood	-11,803	-9,904	-6,034		-11,203	
KS Test	0.0437 (0.0031)	0.0442 (0.0026)	0.0474 (9.9740 e-4)		0.0443 (0.0026)	
JB Test	18 (0.0010)	15 (0.0018)	52 (0.0010)		16 (0.0014)	

The multiple subordination results presented in Table 4 points to a striking conclusion. Neither asymmetric or autoregressive asymmetric models produce significantly different results from the remaining models.[7] Contrary to the asymmetric approach, the autoregressive model is found to augment the linear model, further supporting the use of past squared returns. Moreover, the significance of imbalance terms in addition to volume and duration parameters seems to solidify the notion that order book information is important in subordination, hence variance estimation.

The findings presented in Table 4 may however be subject to the ubiquitous local extrema problem as the findings are produced on a single-run. To address this possible shortcoming, the gradient-based optimization algorithm is augmented with 10^5 different starting points to cover a vast search space.[8] The results for each stock and period using this procedure (Global) are reported in Table 5.

[6] The subordination results reported here and henceforth multiplies tick returns with 1e+6 and divides log-initimb^2 term by 100 as not to compromise floating point calculations in Matlab. Likewise, results for the duration term are reported for duration measured in minutes.

[7] Asymmetric approaches are omitted from further reporting as findings extend to other stocks.

[8] All further results reported use 10^5 starting points.

Table 5: Multiple Subordination Results using Global Procedure

HSBC

Normality		P1 (100Ticks)		P2 (100Ticks)		P3 (100Ticks)		P4 (100Ticks)	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0748 (1.2026 e-7)	0.0663 (4.2692 e-6)	0.0474 (8.2207 e-4)	0.0481 (6.5517 e-4)	0.0503 (2.4969 e-5)	0.0477 (7.7809 e-5)	0.0456 (0.0018)	0.0448 (0.0022)
JB Test		77 (0.0010)	139 (0.0010)	642 (0.0010)	238 (0.0010)	554 (0.0010)	1,248 (0.0010)	2 (0.2947)	0 (0.5000)
GARCH	KS Test	0.0529 (4.9487 e-4)		0.0539 (8.3033 e-5)		0.0395 (0.0019)		0.0415 (0.0059)	
	JB Test	6 (0.0430)		59 (0.0010)		57 (0.0010)		2 (0.4278)	

SAB Miller

Normality		P1 (100 Ticks)		P2 (90 Ticks)		P3 (100 Ticks)		P4 (70 Ticks)	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0453 (0.1318)	0.0482 (0.0929)	0.0545 (0.0087)	0.0438 (0.0597)	0.0397 (0.0699)	0.0393 (0.0745)	0.0506 (0.0035)	0.0510 (0.0032)
JB Test		85 (0.0010)	26 (0.0010)	340 (0.0010)	234 (0.0010)	129 (0.0001)	383 (0.0010)	25 (0.0010)	170 (0.0010)
GARCH	KS Test	0.0561 (0.0312)		0.0625 (0.0016)		0.0324 (0.2132)		0.0501 (0.0040)	
	JB Test	7 (0.0292)		572 (0.0010)		30 (0.0010)		43 (0.0010)	

Royal Dutch Shell

Normality		P1 (70 Ticks)		P2 (100 Ticks)		P3 (100 Ticks)		P4 (100 Ticks)	
		Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
KS Test		0.0459 (0.0049)	0.0462 (0.0045)	0.0417 (0.0513)	0.0017 (0.0510)	0.0383 (0.0331)	0.0307 (0.1424)	0.0379 (0.0758)	0.0379 (0.0756)
JB Test		2,282 (0.0010)	392 (0.0010)	26 (0.0010)	60 (0.0010)	525 (0.0010)	700 (0.0010)	22 (0.0001)	43 (0.0010)
GARCH	KS Test	0.0468 (0.0039)		0.0443 (0.0320)		0.0350 (0.0652)		0.0433 (0.0280)	
	JB Test	16 (0.0015)		32 (0.0010)		164 (0.0010)		17 (0.0013)	

Table 5 shows that findings regarding asymmetric subordination of HSBC in P4 using single-run method can be extended to all periods and stocks. One possible reason for the failure of asymmetric models could be an inherent stability of information flow through the selected subordinators. As such, asymmetric models could not produce superior results by treating returns of opposite signs differently.

Furthermore, as is apparent from Table 2, the choice of sampling frequency, which constitutes an important part of the natural time approach, has a dominant effect on the distribution of raw returns. While sparse sampling mitigates price discreteness, it eventually reduces the relevance of past order book data. For this reason, a sampling frequency of 100-ticks was used for all stocks except for second and fourth periods of SAB Miller. In these periods, normally distributed returns were obtained without the need for subordination at 100-ticks. Hence, higher sampling frequencies were chosen to produce comparable raw distributions in terms of their first four moments.

Closer examination of results reveals an interchangeability between log-volume and volume imbalance terms.[9] Either one of two subordinators, when used in conjunction with others, is significant but they fail to be significant together. While volume imbalance is significant for SAB Miller and Shell in the second period, the reverse holds for HSBC. However, a combination of volume and initiator imbalance seems to be the better choice in general since volume imbalance term is ruled out in 3 out of 4 periods for SAB Miller and Shell. This interchangeability can be caused by the structural changes in the way variance related information is conveyed in the market. It might be the case that in some periods, a combination of volume and initiator imbalance captures variance related information while in others volume imbalance proves to be a better gauge. Furthermore, convergence of volume and volume imbalance terms, which would convey similar information when orders are one-sided, can also render the volume imbalance term redundant. One or both of these factors may be at work in a given period as they are by no means mutually exclusive.

The autoregressive subordination model, which uses past squared returns, was also found to perform generally better than the linear model for all stocks. Although similar results could be obtained using the linear model in several of the periods where normally distributed subordinated returns were produced with the autoregressive model, this was not possible for the second period of SAB Miller and third period of Shell.

In comparison with autoregressive subordination, GARCH(1,1) model does an marginally better job in periods where subordination fails to produce normally distributed returns. However, in cases where normally distributed returns were obtained via subordination, GARCH not only produced worse results but also failed to achieve normality with the exception of a single instance, which could very well be due to a local minima problem in the subordination procedure. While subordination resulted in normally distributed returns in 3 periods for SAB Miller and Shell and 1 period for HSBC, GARCH(1,1) based subordination could only produce normal returns in the third period for SAB Miller. [10]

[9] Further details of the subordination results can be found in Appendix A1-A3.

[10] The resulting distributions from multiple subordination and GARCH(1,1) were assumed to be normally distributed, if they have passed either one of the KS or JB tests.

Conclusion

The work presented in this research paper, focuses on the application of stochastic subordination to high-frequency returns sampled under transaction time. Previous subordination based studies have all been performed using calendar time (Clark (1973); Ané and Geman (2000); Velasco-Fuentes and Ng (2010)). To the best of the my knowledge, subordination has not been applied to financial returns under tick time sampling before. Furthermore, only a subset of the variables used in this research were employed in the above mentioned studies. Order book variables which contain information on both market liquidity and the initiator of trades, have been added into the subordination procedure, which is another novel contribution of this paper to the literature. This subordination procedure, which operates under tick time and uses order book variables to transform the return series into a normally distributed one, is referred to as "natural time" in this paper.

Previous studies have found volume and number of trades to carry relevant information to price formation under physical time (Clark (1973); Ané and Geman (2000)). Their counterparts in transaction time, volume and duration, are also found to be significant in stochastic subordination. The results show that order book variables and past squared returns also carry important variance-related information. The addition of these variables into the subordinator augments the model such that subordinated returns are normally distributed in most cases.

The consistent superiority of natural time approach to the benchmark GARCH based model across all stocks and periods has profound implications. The success of the natural time approach not only supports the normal distribution assumption but also indicates that transaction time might be the right sampling methodology when using high frequency data. Furthermore, as the ability to successfully normalize returns via subordination essentially hinges on accounting for heteroscedasticity, the variables used to subordinate returns can also be used to forecast volatility. Research, combining GARCH(1,1) with exogenous order book variables used here, was also conducted and GARCH terms were consistently found insignificant in all periods for all stocks.

All in all, this research gives the reader a set of variables that are effective in volatility forecasting. Market players that have access to the type order book data used in this research may be able to foretell imminent excess volatility episodes and adjust their positions and leverage accordingly. Additionally, financial authorities which oversee stock markets could use the information contained within the order book to prevent a disorderly collapse of the system. Either use of this information will contribute to the efficiency of financial markets.

Appendix A1

Multiple Subordination Results for HSBC using Global Procedure

Table A1.1: Linear Subordination Results for HSBC

Linear	**P1** (100Ticks)	**P2** (100Ticks)	**P3** (100Ticks)	**P4** (100Ticks)
μ	2.7296 e-15 (0.0512)	-2.0346 e-15 (0.1408)	1.0424 e-13 (0)	2.7508 e-15 (0.9998)
σ	5.3767 e-14 (0.9998)	5.7397 e-14 (0.9997)	2.6790 e-13 (0.9986)	4.5997 e-10 (0)
Volume	2.6997 e+13 (0)	1.3925 e+13 (0.0032)	6.9045 e+13 (0)	1.0000 e+12 (0)
Duration	5.0572 e+12 (0.0086)	7.1242 e+13 (0)	-2.2287 e+13 (0)	-1.0000 e+12 (0)
Log-InitImb²	7.3548 e+11 (0)	4.8892 e+11 (0)	5.5295 e+11 (0.0003)	1.0000 e+10 (0)
Log-VolImb²	-	-	6.0521 e+13 (0)	1.0000 e+13 (0)
Log-likelihood	-22,627	-26,360	-31,413	-11,400

Table A1.2: Autoregressive Subordination Results for HSBC

Autoregressive	**P1** (100Ticks)	**P2** (100Ticks)	**P3** (100Ticks)	**P4** (100Ticks)
μ	8.7329 e-15 (0)	-3.0180 e-15 (0.0008)	5.1429 e-17 (0.9601)	1.2593 e-16 (0.9024)
σ	2.6825 e-14 (0.9999)	3.7524 e-14 (0.9998)	4.8381 e-14 (0.9997)	4.2150 e-14 (0.9998)
r_{tick-1}^2	8.5851 e+13 (0)	3.5343 e+13 (0.0001)	8.5057 e+13 (0)	6.1408 e+13 (0)
Volume	9.5569 e+13 (0)	-	6.3423 e+13 (0.0235)	8.8872 e+13 (0)
Duration	1.0000 e+14 (0)	9.2838 e+13 (0)	5.2372 e+13 (0.0001)	-2.7778 e+13 (0)
Log-InitImb²	6.8892 e+11 (0)	9.8567 e+11 (0)	6.6997 e+11 (0)	1.4235 e+11 (0)
Log-VolImb²	-	7.8435 e+13 (0)	9.4366 e+13 (0.0006)	9.1453 e+13 (0)
Log-likelihood	-23,508	-27,069	-34,319	-26,224

Table A1.3: Ljung-Box Test Statistics for Subordinated HSBC Stock Returns

LB Test	**P1** (100Ticks)		**P2** (100Ticks)		**P3** (100Ticks)		**P4** (100Ticks)	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
r_t	23.9789 (0.2433)	17.0894 (0.6472)	38.5263 (0.0076)	34.2855 (0.0243)	36.2941 (0.0142)	38.1251 (0.0085)	16.9765 (0.6545)	16.0319 (0.7146)
r_t^2	175.8021 (0)	350.0387 (0)	216.8248 (0)	186.4615 (0)	271.0396 (0)	132.8735 (0)	30.2203 (0.0664)	21.9554 (0.3429)

Appendix A2

Multiple Subordination Results for SAB Miller using Global Procedure

Table A2.1: Linear Subordination Results for SAB Miller

Linear	**P1** (100 Ticks)	**P2** (90 Ticks)	**P3** (100 Ticks)	**P4** (70 Ticks)
μ	8.9473 e-13 (0.2358)	-8.6463 e-14 (0)	6.6771 e-14 (0)	2.9076 e-14 (0)
σ	1.9255 e-11 (0.9776)	3.2818 e-13 (0.9989)	1.4066 e-13 (0.9995)	1.8632 e-13 (0.9993)
Volume	9.9999 e+8 (0)	6.5458 e+13 (0)	1.0000 e+14 (0)	-
Duration	4.6693 e+8 (0)	-2.1460 e+12 (0)	1.1837 e+13 (0.0001)	5.7186 e+12 (0.0002)
Log-InitImb²	1.0000 e+7 (0)	1.0000 e+12 (0)	2.7082 e+11 (0)	-3.9134 e+9 (0)
Log-VolImb²	-	-	-1.2608 e+13 (0)	8.2618 e+13 (0)
Log-likelihood	-6,176	-12,649	-15,163	-17,707

Table A2.2: Autoregressive Subordination Results for SAB Miller

Autoregressive	**P1** (100 Ticks)	**P2** (90 Ticks)	**P3** (100 Ticks)	**P4** (70 Ticks)
μ	1.5232 e-15 (0.4161)	1.5490 e-15 (0.4065)	5.9718 e-15 (0.0147)	-5.5842 e-15 (0.0005)
σ	4.7791 e-14 (0.9999)	5.6189 e-14 (0.9998)	7.9497 e-14 (0.9997)	5.5972 e-14 (0.9998)
r^2_{tick-1}	6.7977 e+13 (0)	1.8335 e+13 (0.0029)	8.1994 e+13 (0)	1.0000 e+14 (0)
Volume	-	9.9925 e+13 (0.0001)	9.9877 e+13 (0)	-
Duration	9.1697 e+13 (0)	6.8442 e+13 (0)	4.6871e+13 (0)	1.0000 e+14 (0)
Log-InitImb²	9.9065 e+11 (0)	1.0000 e+12 (0)	2.8841 e+11 (0)	-3.4439 e+3 (0)
Log-VolImb²	-	-	-	1.0000 e+12 (0)
Log-likelihood	-10,026	-13,865	-15,747	-18,791

Table A2.3: Ljung-Box Test Statistics for Subordinated SAB Miller Stock Returns

LB Test	**P1** (100 Ticks)		**P2** (90 Ticks)		**P3** (100 Ticks)		**P4** (70 Ticks)	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
r_t	19.3210 (0.5011)	21.0630 (0.3934)	18.7713 (0.5367)	23.9882 (0.2429)	25.1293 (0.1965)	21.7390 (0.3548)	26.2966 (0.1562)	27.1843 (0.1302)
r_t^2	108.5900 (3.5527 e-14)	85.2687 (4.9200 e-10)	20.6774 (0.4163)	33.2963 (0.0313)	179.0044 (0)	183.8922 (0)	133.6843 (0)	47.6079 (4.8272 e-4)

Appendix A3

Subordination Results for Royal Dutch Shell using Global Procedure

Table A3.1: Linear Subordination Results for Royal Dutch Shell

Linear	**P1** (70 Ticks)	**P2** (100 Ticks)	**P3** (100 Ticks)	**P4** (100 Ticks)
μ	4.4576 e-16 (0.9316)	-1.0791 e-16 (0.9349)	-3.7920 e-5 (0.0625)	8.5731 e-14 (0)
σ	1.9568 e-13 (0.9992)	4.2722 e-14 (0.9998)	7.5872 e-14 (0.9997)	1.9654 e-13 (0.9993)
Volume	2.6920 e+13 (0)	1.0000 e+14 (0)	8.3493 e+13 (0)	6.5100 e+13 (0)
Duration	1.0000 e+14 (0)	1.0000 e+14 (0)	2.3232 e+13 (0)	-1.0927 e+12 (0)
Log-InitImb^2	7.1329 e+11 (0.0001)	2.9118 e+11 (0)	9.8443 e+11 (0)	2.6871 e+11 (0)
Log-VolImb^2	-4.0737 e+13 (0)	-	-2.3712 e+13 (0)	-
Log-likelihood	-20,550	-16,249	-20,801	-16,219

Table A3.2: Autoregressive Subordination Results for Royal Dutch Shell

Autoregressive	**P1** (70 Ticks)	**P2** (100 Ticks)	**P3** (100 Ticks)	**P4** (100 Ticks)
μ	-6.1471 e-15 (0)	6.1171 e-14 (0)	-7.3507 e-16 (0.6480)	-4.5436 e-15 (0.0003)
σ	3.5711 e-14 (0.9998)	7.58291 e-14 (0.9997)	6.0001 e-14 (0.9997)	4.1984 e-14 (0.9998)
r^2_{tick-1}	1.0000 e+14 (0)	6.5455 e+13 (0.0072)	1.2339 e+13 (0.0002)	5.2155 e+13 (0)
Volume	7.0068 e+13 (0)	1.0000 e+14 (0)	4.8654 e+13 (0.0162)	9.9192 e+13 (0)
Duration	9.2564 e+13 (0)	7.9835 e+13 (0)	4.2889 e+13 (0)	7.1199 e+13 (0)
Log-InitImb^2	8.2570 e+11 (0)	-2.2156 e+9 (0)	7.1681 e+11 (0)	7.9620 e+11 (0)
Log-VolImb^2	-	-	6.5802 e+13 (0.0033)	-
Log-likelihood	-22,168	-15,635	-21,063	-17,512

Table A3.3: Ljung-Box Test Statistics for Subordinated Royal Dutch Shell Stock Returns

LB Test	**P1** (70 Ticks)		**P2** (100 Ticks)		**P3** (100 Ticks)		**P4** (100 Ticks)	
	Linear	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*	*Linear*	*Autoregressive*
r_t	75.3423 (2.3794 e-8)	56.0454 (2.8615 e-5)	17.0768 (0.6480)	17.2084 (0.6394)	21.9221 (0.3448)	20.7530 (0.4118)	13.2405 (0.8668)	12.9193 (0.8808)
r_t^2	475.5357 (0)	260.4708 (0)	40.2501 (0.0046)	33.2846 (0.0314)	239.8774 (0)	145.2595 (0)	49.2733 (2.8118 e-4)	31.9224 (0.0441)

References

Aït-Sahalia, Yacine, Fan, Jianqing, & Xiu, Dacheng. 2010. High-Frequency Covariance Estimates With Noisy and Asynchronous Financial Data. Journal of the American Statistical Association, **105 (492)**, 1504–1517.

Aldridge, Irene. 2010. High-Frequency Trading : A Practical Guide to Algorithmic Strategies and Trading Systems. New Jersey: Wiley.

Allen, David, Lazarov, Zdravetz, McAleer, Michael, & Peiris, Shelton. 2009. Comparison of Alternative ACD Models via Density and Interval Forecasts: Evidence from the Australian Stock Market. Mathematics and Computers in Simulation, **79 (8),** 2535-2555.

Andersen, Torben G., Bollerslev, Tim, Diebold, Francis X., & Vega, Clara. 2003. Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange. The American Economic Review, **93 (1)**, 38–62.

Ané, Thierry, & Geman, Hélyette. 2000. Order Flow, Transaction Clock, and Normality of Asset Returns. Journal of Finance, **55 (5)**, 2259–2284.

Bandi, Federico M., & Russell, Jeffrey R. 2008. Microstructure Noise, Realized Variance, and Optimal Sampling. Review of Economic Studies, **75 (2)**, 339–369.

Barclay, Michael J., & Warner, Jerold B. 1993. Stealth Trading and Volatility: Which Trades Move Prices? Journal of Financial Economics, **34 (3)**, 281–305.

Black, Fischer, & Scholes, Myron. 1973. The Pricing of Options and Corporate Liabilities. Journal of Political Economy, **81 (3)**, 637–654.

Bollerslev, Tim. 1986. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics, **31 (3)**, 307–327.

Chakravarty, Sugato. 2001. Stealth-Trading: Which Traders' Trades Move Stock Prices? Journal of Financial Economics, **61 (2)**, 289–307.

Clark, Peter K. 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica, **41 (1)**, 135–155.

Cont, Rama. 2001. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance, **1 (1)**, 223–236.

Dacorogna, Michel M., Müller, Ulrich A., Nagler, Robert J., Olsen, Richard B., & Pictet, Olivier V. 1993. A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market. Journal of International Money and Finance, **12 (4)**, 413–438.

Dacorogna, Michel M., Gençay, Ramazan, Müller, Ulrich A., Olsen, Richard B., & Pictet, Olivier V. 2001. An Introduction to High-Frequency Finance. San Diego: Academic Press.

Easley, David, & O'Hara, Maureen. 1987. Price, Trade Size, and Information in Securities Markets. Journal of Financial Economics, **19 (1)**, 69–90.

Easley, David, & O'Hara, Maureen. 1992. Time and the Process of Security Price Adjustment. Journal of Finance, **47 (2)**, 577–605.

Engle, Robert F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the

Variance of United Kingdom Inflation. Econometrica, **50 (4)**, 987–1007.

Engle, Robert F., & Russell, Jeffrey R. 1998. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica, **66 (5)**, 1127–1162.

Fama, Eugene F. 1965. The Behavior of Stock Market Prices. Journal of Business, **38 (1)**, 34–105.

Geman, Hélyette. 2002. Pure jump Lévy Processes for Asset Price Modelling. Journal of Banking & Finance, **26 (7)**, 1297–1316.

Gillemot, László, Farmer, J. Doyne, & Lillo, Fabrizio. 2006. There's More to Volatility Than Volume. Quantitative Finance, **6 (5)**, 371–384.

Glosten, Lawrence R., & Milgrom, Paul R. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. Journal of Financial Economics, **14 (1)**, 71–100.

Huth, Nicolas & Abergel, Frederic. 2012. The Times Change: Multivariate Subordination. Empirical Facts. Quantitative Finance, **12(1)**, 1-9.

Karpoff, Jonathan M. 1987. The Relation Between Price Changes and Trading Volume: A Survey. Journal of Financial and Quantitative Analysis, **22 (1)**, 109–126.

Kyle, Albert S. 1985. Continuous Auctions and Insider Trading. Econometrica, **53 (6)**, 1315–1335.

Mandelbrot, Benoît B. 1963. The Variation of Certain Speculative Prices. Journal of Business, **36 (4)**, 394–419.

Martens, Martin, Chang, Yuan-Chen, & Taylor, Stephen J. 2002. A Comparison of Seasonal Adjustment Methods When Forecasting Intraday Volatility. The Journal of Financial Research, **25 (2)**, 283-299.

Murphy, Anthony, & Izzeldin, Marwan. 2006. Order Flow, Transaction Clock, and Normality of Asset Returns: Ané and Geman (2000) Revisited. Oxford University / Lancaster University.

Ng,Wing Lon. 2008. Analysing Liquidity and Absorption Limits of Electronic Markets with Volume Durations. Quantitative Finance, **8 (4)**, 353–361.

Oomen, Roel C. A. 2006. Properties of Realized Variance under Alternative Sampling Schemes. Journal of Business & Economic Statistics, **24 (2)**, 219–237.

Perry, Philip R. 1983. More Evidence on the Nature of the Distribution of Security Returns. The Journal of Financial and Quantitative Analysis, **18 (2)**, 211–221.

Savaşer, Tanseli. 2011. Exchange-Rate Response to Macro News: Through the Lens of Market Microstructure. Journal of International Financial Markets, Institutions & Money, **21 (1)**, 107–126.

Tauchen, George E., & Pitts, Mark. 1983. The Price Variability-Volume Relationship on Speculative Markets. Econometrica, **51 (2)**, 485–505.

Velasco-Fuentes, Rafael, & Ng, Wing Lon. 2010. Nonlinearities in Stochastic Clocks: Trades and Volume as Subordinators of Electronic Markets. Quantitative Finance, **11(6)**, 863–881.

Wasserfallen, Walter, & Zimmermann, Heinz. 1985. The Behavior of Intra-Daily Exchange Rates. Journal of Banking and Finance, **9 (1)**, 55–72.