Centre for Computational Finance and Economic Agents

Working Paper Series

WP030-08

Wing Lon Ng

Spectral Densities of Ultra-high Frequency Data

September 2008

University of Essex

CCFEA

www.ccfea.net

Spectral Densities of Ultra-high Frequency Data

Wing Lon Ng [*]
Centre for Computational Finance and Economic Agents
University of Essex
Wivenhoe Park
Colchester
Essex CO4 3SQ
United Kingdom
E-mail: wlng@essex.ac.uk

September 2008

[*]The author is very grateful to Mark Trede and to seminar participants at the University of Muenster and the University of Essex for their valuable comments and helpful discussion. He is particularly indebted to conference participants at the International Conference on High-Frequency Finance (2006), the Conference Forecasting Financial Markets (FFM 2006), the International Conference on Time Series Econometrics, Finance and Risk (TSEFAR 2006), the Annual Meeting of the DStatG (2006) and the Annual Meeting of the SCE (2007) for their extensive suggestions.

Spectral Densities of Ultra-high Frequency Data

Abstract

This paper suggests the application of advanced methods from Fourier Analysis in order to describe ultra-high frequent data in limit order books. Using Lomb's normalized periodogram and Scargle's Discrete Fourier Transforms to take account of the irregularity in spacing, the power spectra of different time series processes can be easily estimated, without immense computational effort caused by the large amount of observations. With empirical data extracted from the German XETRA system, the spectral analysis shows that the entire trading process contains various different periodic components. While duration and volume processes have a strong cyclical behavior in the low-frequency domain, seasonalities of price differences arise in the high-frequency domain. Contrarily, the time series of the spread reveals no periodicity, neither in the long term, nor in the middle or short term.

Key Words: Ultra-high frequency transaction data, limit order book, irregularly spaced data, Lomb-Scargle Fourier Transforms, spectral density.

JEL Classifications: $C22, C32, C63$.

1 Introduction

For the last decade, the literature on financial econometrics has witnessed a growing interest in the analysis of order book data provided by electronic exchange trading systems. Since this so-called "ultra-high-frequency" data is observed in real-time, yielding the highest possible sampling limit, it is characterized by one key feature, namely the irregularity of time intervals between two consecutive events. Therefore, the econometric analysis of this unevenly spaced data always focus on the modelling of durations in order to avoid any loss of important information stored in the temporal structure of the entire transaction process. Treating these special time series as point processes, Engle and Russell (1997) have introduced the *Autoregressive Conditional Duration* (*ACD*) model, which describes a dynamic duration process with a conditional expectation that is written as a linear function of past durations. Based on their seminal work, many studies have concentrated on its further development in order to describe limit order book activities more accurately (for a survey, see Fernandes and Grammig (2006) or Hautsch (2004)). According to Cox and Isham (1980), alternative approaches to deal with point processes are count models (see Grammig, Heinen, and Rengifo (2005)) or intensity models (see Hall and Hautsch (2006)).

Although these three types of models have been improved by many authors and shown a good performance in numerous previous studies, they still have their drawbacks and limits. Duration models, for example, can not consider more than one single process because of the asynchronization problem of multivariate point processes (otherwise truncation is required as shown in Engle and Lunde (2003)). In contrast, the other approaches are easier to be extended to multidimensional settings, but they either lose too much information due to the aggregation over discrete time intervals (count models) or involve complex model specifications and computational burdens by reasons of the assumption of continuity (intensity models). Furthermore, there are many additional variables and information recorded in the order book that has to be considered as well. Since point processes naturally focus on points, they sometimes lose sight of their "marks", in which economists are more interested. In fact, in many models these order book variables serve as regressors explaining the duration or the intensity of the process, or, the relationship of different covariates are analyzed by decomposition

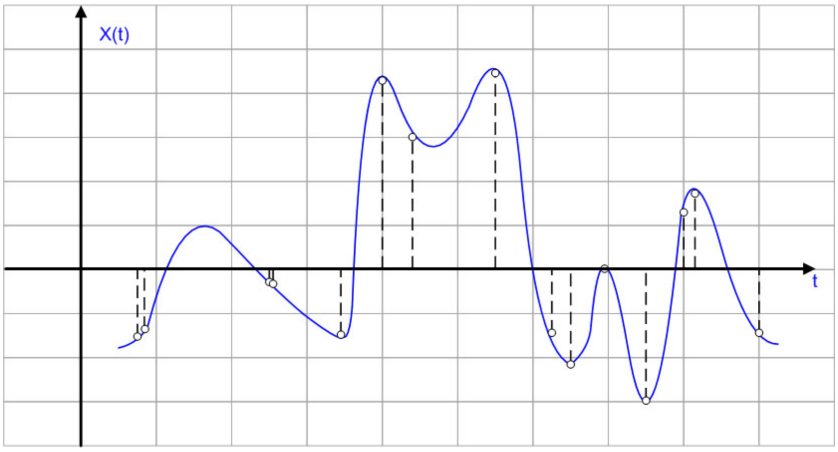

Figure 1: Fitting the FFT on non-equispaced order book data

methods (see Engle (2000) or Rydberg and Shephard (2003)). However, the key variables are never treated as single self-contained processes.

In order to solve these problems, this paper suggests using advanced modelling techniques from the Fourier analysis, minimizing the computation time of the large amount of transaction data. Fast Fourier algorithms afford a rapid calculation of the Fourier values of a given time series that represent the same process in the frequency domain without any loss of information. Applying modified Fourier transforms which are especially designed for non-equidistant spaced data, this framework allows to reveal the stochastics of every single process without loss of any information. Hence, the main goal of this study is to recover the underlying stochastic behaviour of the entire transaction process observed in an electronic trading system. Compared to models of the autoregressive conditional framework, where the estimation procedure always require recursive evaluations of the likelihood, the adopted methods in this research make an enormous reduction of the computational effort possible though analyzing large sets of order book data.

Finally, there is one more important reason that motivates this application and is referred to the periodicity of financial time series. It is well-known that intraday data have a consistent diurnal pattern of trading activities over the course of a trading day due to institutional characteristics of organized financial markets, such as opening and closing hours or intraday auctions. Indeed, most researchers only take the regular daily seasonality of durations into account, assuming that all other variables (price limit, order volume, etc.) are not influenced by the diurnality or other seasonal patterns. Since this study concentrates on the investigation in the frequency domain, we can easily detect characteristic cyclical patterns of all interested variables by estimating their power spectra which captures the periodicity of stochastic processes as is generally known.

The paper is structured as follows: In Section 2, the modified Fourier transforms for irregularly spaced data and the estimation of spectral density are described. Section 3 represents the data and the empirical results with regard to the economic implications. Section 4 concludes.

2 Spectral Analysis of Order Book Data

The general aim of spectral analysis is the decomposition of a time series into its periodic constituents and seasonalities in order to reveal the cyclical behavior of economic processes (see Priestley (1981), a survey can be found in Granger and Engle (1983) or Iacobucci (2003)).

Consider the finite time series x_t with length T and N observations. In case of periodic sampling, the temporal distance between two realizations is always constant

$$t_j - t_{j-1} = \Delta t = \frac{1}{T} \qquad \forall j \in \mathbb{N} \quad ,$$

and the time series can be regarded as a sum of trigonometric polynomials

$$\begin{aligned} x_t &= \sum_{k=-N/2}^{N/2-1} a_k \cos(2\pi k t) + b_k \sin(2\pi k t) && (1) \\ &= \sum_{k=-N/2}^{N/2-1} c_k e^{i2\pi kt/N} \qquad , && (2) \end{aligned}$$

where the Fourier coefficients

$$c_k = \frac{1}{N}\sum_{t=1}^{N} x_t e^{-i2\pi kt/N} \tag{3}$$

can be easily computed by the well-known Fast Fourier Transforms (FFT) that needs only $\mathcal{O}(N \log N)$ arithmetical operations (see, for example, Bloomfield (1976) and Warner (1998)). Since high frequency transaction data is observed in real-time and, thus, arrives in irregular time intervals, the implementation of the common FFT can not be done for the unevenly sampled data. Because Δt is no more constant but stochastic, the computation of (3) will require $\mathcal{O}\left(N^2\right)$ operations. The problem here is to find an algorithm that is able to compute (3) faster.

Interestingly, this issue has been addressed in many different areas of science, such as astrophysics, geology, computer tomography and other fields of applied mathematics (for an overview, see Ware (1998)). The general purpose is to transform the data on the non-equispaced grid into the frequency domain in order to get a unbiased estimation of the power spectrum. Recent tools especially designed for these problems are, for example, the advanced $NFFT$ algorithms proposed by Kunis, Potts, and Steidl (2006) (see also Steidl (1998) and Fenn and Potts (2005)). More simply, one can directly calculate the "normalized" periodogram

$$P(\omega) \equiv \left(\frac{C_1 \cos(\omega\tau) + S_1 \sin(\omega\tau)}{N + C_2 \cos(2\omega\tau) + S_2 \sin(2\omega\tau)} + \frac{S_1 \cos(\omega\tau) - C_1 \sin(\omega\tau)}{N - C_2 \cos(2\omega\tau) - S_2 \sin(2\omega\tau)}\right)$$

with

$$\tau = \frac{1}{2\omega} \arctan\left(\frac{\sum_{j=1}^{N} \sin(2\omega t_j)}{\sum_{j=1}^{N} \cos(2\omega t_j)}\right) \tag{4}$$

$$S_1 = \sum_{j=1}^{N} (x_j - \bar{x}) \sin(\omega t_j) \tag{5}$$

$$C_1 = \sum_{j=1}^{N} (x_j - \bar{x}) \cos(\omega t_j) \tag{6}$$

$$S_2 = \sum_{j=1}^{N} \sin(2\omega t_j) \tag{7}$$

$$C_2 = \sum_{j=1}^{N} \cos(2\omega t_j) \tag{8}$$

as suggested by Lomb (1976) (see also Press, Teukolsky, Vetterling, and Flannery (1992), p. 581, a generalization for the non-sinusoidal case can be found in Bretthorst (2001)). The advantage of this estimator is that it considers the inequidistance of observations by multiplying the Fourier-frequencies ω with τ. Based on this method, Scargle (1982) proposed an improved version with modifications, resulting in the altered discrete Fourier transforms

$$f(\omega_k) = f_0(\omega_k) \sum_{j=1}^{M} a(\omega_k^*) x_j \cos(\omega_k^*) + ib(\omega_k^*) x_j \sin(\omega_k^*) \tag{9}$$

(see also equation (1)) with

$$\omega_k = 2\pi k/N \qquad k = 1, 2, ..., K \tag{10}$$

$$\omega_k^* = \omega_k (t_j - \tau(\omega_k)) \tag{11}$$

$$\tau(\omega_k) = \frac{1}{2\omega_k} \arctan\left(\frac{\sum_{j=1}^{N} \sin(2\omega_k t_j)}{\sum_{j=1}^{N} \cos(2\omega_k t_j)}\right) \tag{12}$$

$$f_0(\omega_k) = \frac{1}{\sqrt{2}} \exp(i\omega_k \tau) \tag{13}$$

$$a(\omega_k^*) = \left(\sqrt{\sum_{j=1}^{N} \cos^2(\omega_k^*)}\right)^{-1} \tag{14}$$

$$b(\omega_k^*) = \left(\sqrt{\sum_{j=1}^{N} \sin^2(\omega_k^*)}\right)^{-1} \tag{15}$$

(see also Scargle (1989)). Due to the (re-)shifting of all N sampling times t_j with τ, the time invariance of $f(\omega_k)$ is ensured. The shift in the time domain that causes the irregular grid can be easily considered by the phase shift in the frequency domain induced by f_0 (see also Schulz and Stattegger (1997)).

In financial transaction data, a plethora of additional information $\mathbf{X} = (X_1, ..., X_m)'$ can be observed at the arrival times t (for example, price, volume, quotes, depth, etc.). These variables provide the transparency in a market and, thus, have important economic value. In this study, we concentrate on the variables stored in the limit order book without conditioning them on the time varying duration, intensity or the filtration of the process as required in existing autoregressive point process models in the literature. Hence, the stochastic processes of price differences, transaction volume and spread will be "directly" investigated. After computing their Scargle-DFTs ($SDFT$), their power spectra are estimated with the common Schuster-

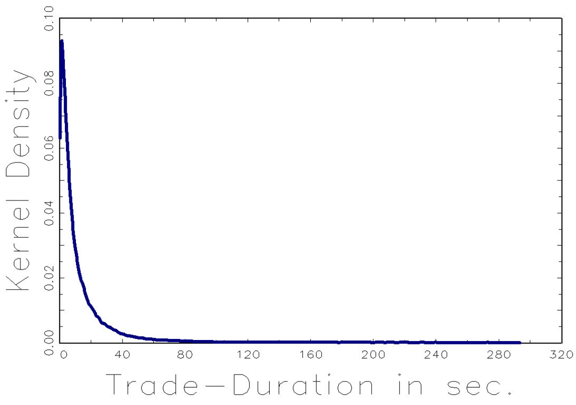

Figure 2: Kernel density of trade durations

Periodogram

$$P(\omega) = \left| \hat{f}(\omega) \right|^2$$

and their spectral densities can be obtained by standardizing $P(\omega)$ with σ_X^2, i.e. the variance of the variable.

For reasons of better comparison with existing ACD models, the duration variable will be analyzed as well. The distribution of financial durations is generally skewed to the right and unimodal (see Figure 2). In case of asymmetric information as often assumed in the common market microstructure theory, the appearance of insiders will lead to duration clusters as illustrated in O'Hara (1997). Hence, the spectral density of the duration process should have a declining slope in the $[0; 0.5]$-interval, which indicates that the "energy" (i.e. the variance) of the process comes from the lower frequencies, i.e. the long periods.

3 Empirical Results

The transaction data of the Deutsche Telekom stock is extracted from the open order book of the German XETRA system, which is an order-driven market. The sample includes the whole history of $N = 71902$ transactions from 31^{st} July until 1^{st} September 2000, observed for 25 trading days over 5 weeks. The continuous trading phase starts after the opening auction at 9 a.m. and ends before the closing auction at 8 p.m. Further, it is interrupted by (at least) two intraday auctions at 1 p.m. and 5 p.m., each lasting at most

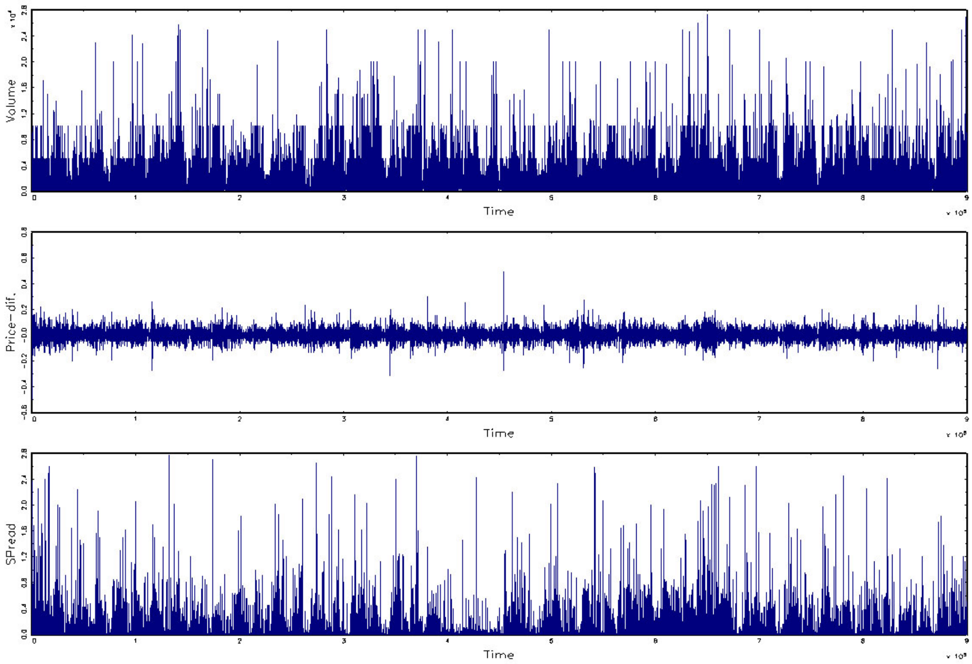

Figure 3: The original (irregularly spaced) time series

120 seconds. These and other delays involving deterministic waiting times, such as overnight durations, are cut off. The electronic trading is based on an automatic matching algorithm, generally following a strict price-time priority of orders. The ultra-high frequency order book data does not only show the price, the volume and the time stamp of the transaction (with an accuracy as of one hundredth second), but also the initial buy or sell order.

Descriptive statistics of the transaction volume, price difference and spread are listed in Table 1. The original (irregularly spaced) times series of these three variables are presented in Figure 3. Their power spectra estimation, smoothed with a Bartlett window, are shown in Figure 5, 6 and 7, a comparison of the spectral densities can be found in Figure 8. The time unit is one second. Since t_N ends with 910386.17 (seconds), the overall observation period has a length of $T = 252.8850$ hours. For purposes of better orientation and interpretation, Table 1 also reports the descriptive statistics of the single trade durations (the non-parametrically estimated kernel density is shown in Figure 2).

For a given time series, the power spectrum gives a plot of the contri-

Variable	Volume	Price-dif.	Spread	Duration
Mean	1522.8664	0.0007	0.0248	12.6614
Median	1000.0000	0.0000	0.0000	6.1600
Variance	3759.83e3	0.0008	0.0125	417.1340
Std. dev.	1939.0288	0.0287	0.1120	20.4238
Dispersion	1.2732	39.1836	4.5030	1.6130
Skewness	3.4666	0.2908	9.7503	4.9969
Kurtosis	24.7195	9.9741	142.7619	40.7198

Table 1: Descriptive statistics

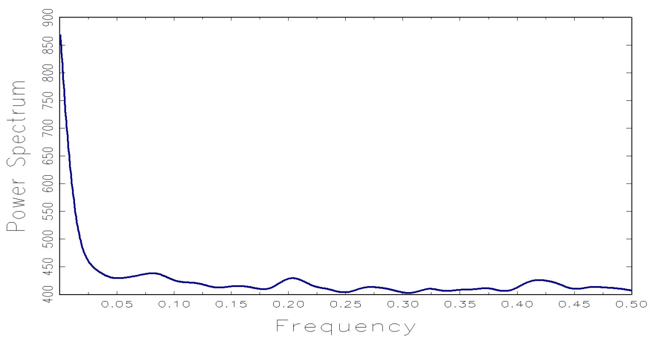

Figure 4: Power spectrum of duration

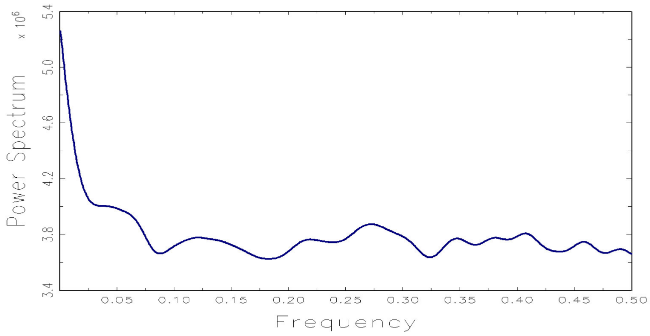

Figure 5: Power spectrum of traded quantity

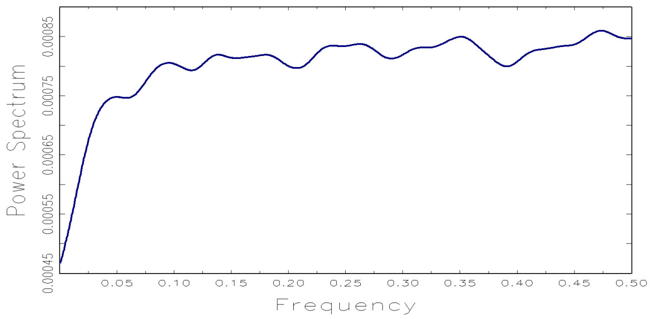

Figure 6: Power spectrum of price differences

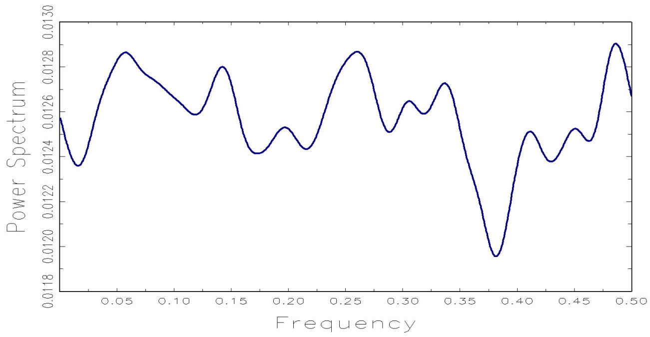

Figure 7: Power spectrum of spread

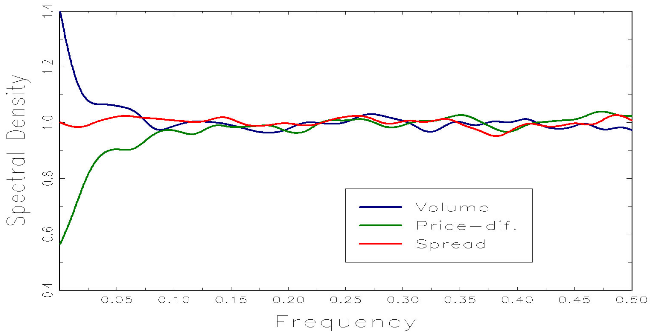

Figure 8: Comparison of the spectral densities

bution of a “signal’s energy”, i.e. the variance per unit time, falling within given frequency bins. The power spectrum of the duration process is shown in Figure 4 and has a decreasing slope as expected. This result implies high disturbance of the process in the “long term” (and no disturbance in the “short term”) and is in line with most ACD estimation results in the literature on intraday trading. Representing duration clusters in the frequency domain, it confirms the hypothesis of asymmetric information as stated in the market microstructure theory (see O’Hara (1997)).

For the process of the transaction volume, Figure 5 also exhibits a declining slope and signalizes that the essential contribution to the variance of the process comes from the low frequencies, that means the long run development of the trading quantity. In economic terms, this indicates that financial markets have a limited absorption speed. Each time passing this limit, it induces a recurring volume shock causing temporary market illiquidity. Contrarily, the periodogram of the price differences reveals a demanding proportion of spectral energy for the high and middle frequencies, but not for the low ones, as discernible in Figure 6. This effect is reported in the literature on realized volatility and is called as market microstructure noise, which appears when sampling data at the highest frequencies (see Aït-Sahlia and Mykland (2003) and Andersen, Bollerslev, and Diebold (2005)). Hence, the ascending slope shows that there is no evidence of long-term seasonalities of returns in the data.

Concerning the three dimensions of the classical liquidity concept, these findings imply that (high-frequent) price shocks appear more often than (low-frequent) volume and durations shocks in the market. Hence, in case of liquidity trading, a market participant should (a) avoid high-volume-orders, which the market need to much time to absorb, and (b) prefer aggressive order limits to decrease the inside spread. However, the power spectrum of the spread shows several peaks at different frequencies, which indicates that there are cyclical patterns with (relative) high energy in short, middle and long periods respectively. Indeed, after rescaling all periodograms by the variance of the investigated variable, Figure 8 shows that time series of the spread seems to have a uniformly distributed spectral density similar to White Noise processes, implying a constant proportion of the variance in all frequencies.

4 Conclusion

This paper applies advanced methods from Fourier Analysis in order to describe the stochastic processes of non-equispaced order book data in the frequency domain. In contrast to existing autoregressive conditional models (*ACD*, *ACI*, etc.) in the literature, this approach has the advantage that it can (a) directly investigate all economic variables without conditioning on the filtration of the underlying point process and (b) save a lot of computation time that is usually needed for modeling the large data sets. In order to take account of the irregularity in spacing, the power spectra of different time series processes are estimated by means of Scargle's Discrete Fourier Transforms ($SDFT$) without any loss of information. It allows a reduction of computational efforts and has no complex model specifications or obligatory deseasonalisation. Using empirical limit order book data from the German XETRA system, the spectral analysis shows that various parts of the whole transaction process display different periodic patterns, revealed by the energy of the process in the respective frequency domain. While duration and volume processes have cycles of "long periods" which indicates duration clusters and absorption limits in the intraday trading, seasonalities of price differences are relatively short, characterizing high-frequent oscillations at the microstructure level of financial markets. However, the time series of the spread seems to have no periodic patterns at all.

References

AÏT-SAHLIA, Y., AND P. MYKLAND (2003): "The Effects of Random and Discrete Sampling when Estimating Continuous-Time Diffusions," *Econometrica*, 71(2), 483–549.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2005): "Some Like it Smooth, and Some Like it Rough: Untangling Continous and Jump Components in Measuring, Modeling and Forecasting Asset Return Volatility," Working paper, Duke University.

BLOOMFIELD, P. (1976): *Fourier Analysis of Time Series: An Introduction*, Wiley series in probability and mathematical statistics. John Wiley & Sons, New York et al.

BRETTHORST, G. L. (2001): "Generalizing The Lomb-Scargle Periodogram - The Nonsinusoidal Case," in *Maximum Entropy and Bayesian Methods in Science and Engineering*, ed. by A. Mohammad-Djafari, pp. 246–251. AIP conference proceedings.

COX, D. R., AND V. ISHAM (1980): *Point Processes.* Chapman and Hall, London.

ENGLE, R. F. (2000): "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68(1), 1–22.

ENGLE, R. F., AND A. LUNDE (2003): "Trades and Quotes: A Bivariate Process," *Journal of Financial Econometrics*, 1(2), 159–188.

ENGLE, R. F., AND J. R. RUSSELL (1997): "Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model," *Journal of Empirical Finance*, 4(2-3), 187–212.

FENN, M., AND D. POTTS (2005): "Fast summation based on fast trigonometric transforms at nonequispaced nodes," *Numerical Linear Algebra with Applications*, 12, 161–169.

FERNANDES, M., AND J. GRAMMIG (2006): "A Family of Autoregressive Conditional Durations Models," *Journal of Econometrics*, 130(1), 1–23.

GRAMMIG, J., A. HEINEN, AND E. RENGIFO (2005): "Trading activity and liquidity supply in a pure limit order book market," Working paper, University of Tübingen and CORE.

GRANGER, C. W. J., AND R. ENGLE (1983): "Applications of Spectral Analysis in Econometrics," in *Handbook of Statistics, Vol.3*, ed. by D. R. Brillinger, and P. R. Krishnaiah, pp. 93–109. Elsevier Science Publishers.

HALL, A., AND N. HAUTSCH (2006): "Order Aggressiveness and Order Book Dynamics," *Empirical Economics*, 30(4), 973–1005.

HAUTSCH, N. (2004): *Modelling Irregularly Spaced Financial Data*. Springer, Heidelberg.

IACOBUCCI, A. (2003): "Spectral Analysis for Economic Time Series," Working paper, IDEFI.

KUNIS, S., D. POTTS, AND G. STEIDL (2006): "Fast Gauss transforms with complex parameters using NFFTs," *Journal of Numerical Mathematics*, 14, 295–303.

LOMB, N. R. (1976): "Least-Squares Frequency Analysis of Unequally Spaced Data," *Astrophysics and Space Science*, 39(2), 447–462.

O'HARA, M. (1997): *Market Microstructure Theory*. Blackwell, Oxford.

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERLING, AND B. P. FLANNERY (1992): *Numerical Recipes in C*. Cambridge University Press, New York, 2 edn.

PRIESTLEY, M. B. (1981): *Spectral Analysis and Time Series: Univariate Series*, vol. 1 of *Probability and Mathematical Statistics*. Academic Press, London et al.

RYDBERG, T. H., AND N. SHEPHARD (2003): "Dynamics of trade-by-trade price movements: decomposition and models," *Journal of Financial Econometrics*, 1(1), 2–25.

SCARGLE, J. D. (1982): "Studies in astronomical time series analysis II - Statistical aspects of spectral analysis of unevenly spaced data," *Astrophysical Journal*, 263(2), 835–853.

——— (1989): "Studies in astronomical time series analysis III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data," *Astrophysical Journal*, 343(2), 874–887.

Schulz, M., and K. Stattegger (1997): "Spectrum: Spectral Analysis of Unevenly Spaced Paleoclimatic Time Series," *Computers & Geosciences*, 23(9), 929–945.

Steidl, G. (1998): "A note on fast Fourier transforms for nonequispaced grids," *Advances in Computational Mathematics*, 9(3-4), 337–352.

Ware, A. F. (1998): "Fast approximate fourier transforms for irregularly spaced data," *SIAM Rev.*, 40(4), 838–856.

Warner, R. M. (1998): *Spectral Analysis of Time-Series Data*. Guilford Press, New York.