Exploring Irregular Time Series
Through Non-Uniform Fast Fourier Transform

Jung Heon Song
Marcos Lopez de Prado
Horst D Simon
Kesheng Wu

Lawrence Berkeley National Laboratory
One Cyclotron Road
Berkeley, CA 94720

DISCLAIMER
This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.
Exploring Irregular Time Series
Through Non-Uniform Fast Fourier Transform

Jung Heon Song, Marcos López de Prado, Horst D. Simon, and Kesheng Wu

Aug. 29, 2014

Abstract

One of the fundamental shortcomings of the popular analysis tools for time series is that they require the data to be taken at uniform time intervals. However, the real-world time series, such as those from financial markets, are mostly from irregular time intervals. It is a common practice to resample the irregular time series into a regular one, but, there are significant limitations on this practice. For example, if one is to resample the trading activities on a stock into hourly series, then the time series can only last through the trading day because there usually is no trading in the night. In this work, we directly explore the dynamics of irregular time series through a tool known as Non-Uniform Fast Fourier Transform (NUFFT). To illustrate its effectiveness, we apply NUFFT on the trading records of natural gas futures contracts for the last seven years. Results accurately capture well-known structural features in the trading records, such as weekly and daily cycles, and at the same time also reveal unknown or unexplored features, such as the presence of multiple power laws. In particular, we observe a new power law in the Fourier spectra in recent years.

1 Introduction

Many different types of data are collected automatically by computers. Such data records in time are known as time series. A large number of tools and techniques have been developed to analyze such time series [4, 16, 33]. The majority of these tools assume the data is collected on regular time intervals. However, in a broad range of use cases, such as those from financial applications, the time series contain values at irregular time points. Common financial data such as trading records of a stock or a futures contract often have highly bursty activities during the trading day and lightly traded or not traded at all in after-hours. To use the existing analysis tools, the irregular time series are restructured into regular time series [16]. This process of transforming irregular time series into regular ones is useful, but, has significant limitations. For example, if the stock trading records are binned into days, then we can not study the hourly activity patterns. If we bin the data into hours, then we have to contrive ways of dealing with after-hours and non-trading hours, which can create spurious seasonal effects. In this work, we apply a technique, named NUFFT that can work with irregular time series directly [15]. To demonstrate its usefulness, we apply NUFFT to trading records of natural gas because natural gas is widely used and its trading is known to be volatile.

NUFFT is an efficient technique for performing Fast Fourier Transform (FFT) on irregular time series [15]. We use it to study the time domain features of trading records. In literature, studies of intraday characteristics of financial data have to limit their time series to be within a day, typically only including eight hours of the active trading day, not the after-hours [12]. Such time series are too short to give sufficient support for features that occur a small number of times a day. It also separates the studies on multi-day
activity patterns from those on the intraday ones, which may miss important opportunity of connecting the characteristics of patterns from different time scales.

To demonstrate the effectiveness of NUFFT, we use it to study the impact of high-frequency trading on the natural gas futures contracts. With advances in computing technology, and legislative changes in both the United States and Europe, high frequency trading (HFT) strategies were made profitable and popular [8]. Figure 1 shows an estimate of the fraction of stock trades due to HFT firms reported by TABB Group[21]. This figure shows that between 50% and 60% of the trades are initiated by HFT firms, while others estimates put this fraction much higher [6, 21].

Natural gas futures have remained one of the most heavily traded energy contracts for many years. From 2007 to 2012, the number of futures contracts traded totaled more than 50 million [35]. Therefore, we believe natural gas futures contract to be a good representative of commonly traded financial instruments.

There are two key contributions in this work. We introduce a novel tool called NUFFT that works with irregular time series, and demonstrate its effectiveness by applying it to records of natural gas futures trading. To our knowledge, this is the first time that NUFFT has been applied to study financial time series. Because of this, we are also able to observe certain features of trading activities that have not reported before.

The rest of the paper is organized as follows. In Section 2, we briefly review related work. In Section 3, we provide a more detailed description of the data used in this study. A description of NUFFT method is given in Section 4. The output from NUFFT, known as Fourier spectrum, is extensively discussed in Sections 5 and 6. We provide a brief summary in Section 7.

2 Related Work

In this brief review of related work, we concentrate on two aspects that are most intimately relevant to our work, the Fast Fourier Transform and high-frequency trading.

It is well-known that the financial markets exhibits complex dynamics in different time scales [25, 28, 30]. One of the best tools for studying such dynamics is the Fourier transform, which decomposes a function in time into a summation of a number of simple oscillations. Each of this simple oscillation can be described by a frequency and its amplitude. This collection of frequencies and their amplitudes are collectively known as the Fourier spectrum of the original function.
Given a function over time $f(t)$, the Fourier transform is

$$g(k) = \frac{1}{2\pi} \int f(t)e^{i2\pi kt}dt.$$  

The function $g(k)$ is the Fourier transform of $f(t)$, which is also known as the Fourier coefficient of $f(t)$. In computing applications, a discrete form of the above transformation is used.

In many applications, the function $f(t)$ is sampled uniformly in a time period. On such a regular time serie, the summations used to compute the Fourier coefficients share a significant amount of common expressions, therefore the amount of computation can be significantly reduced. This faster way of computing the Fourier coefficients is known as Fast Fourier Transform (FFT) [34, 32].

The usefulness of FFT is recognized by many researchers [2, 5, 7, 29]. Since trades in real markets happen at unpredictable time points, there is no easy way to use the conventional FFT to analyze financial time series. The technique we consider in this work, NUFFT, was specifically designed to work with irregular time series.

To demonstrate the usefulness of NUFFT, we use it to study a series of trading records on natural gas futures. A key object is to find the evidence or the impact of HFT. Frequently, researchers study the microstructure of the markets to understand the impact of HFT [17, 26]. A number of high-impact market events, such as the Flash Crash of May 6, 2010, are considered to be associated with HFT [9, 22, 21]. In particular, the volume synchronized probability of informed trading (VPIN) was shown to give a strong signal more than an hour ahead of the Flash Crash of 2010 [9]. In a number of later tests, strong VPIN signals were shown to indeed proceed high volatility events in some of the largest market microstructure studies [31, 35].

Like many other analysis techniques for time series data, VPIN was based on an assumption of uniformity in the input time series. To deal with irregular “real” time, the authors of VPIN developed the idea of “volume time” [10], and transform the input data into uniform bins in “volume time.” Operationally, this transformation places trades into bins with equal volumes. The algorithm then works with the bins instead of the original time series. Each of these bins is known as a “volume bar.” Each of this “volume bar” is considered as a unit of “volume time.” This treatment has the effect of stretching out a time period of heavy trading into more “volume bars” and allow more in-depth analysis. The time periods with little or no trading activity are broken into a smaller number of “volume bars” so one skips over these time periods quickly. The success of VPIN proves the concept of “volume time” is a useful theoretic construct.

There are a number of sophisticated methods to deal with irregular time series by redefining time, for example, autoregressive conditional duration [11] and stochastic conditional duration model [1]. In addition to regularizing the time, there are also alternative analysis tools, such as approximate entropy [27] and fractal theory [18]. In comparison to these techniques, NUFFT is much simpler and its output is much easier to understand.

3 Natural Gas Futures Contracts

We choose this instrument for our study because natural gas is widely used in homes and in businesses, and its prices are notoriously volatile. Natural gas is used for cooking, heating, generating electricity and so on. Some of these uses, such as cooking, are relatively stable throughout the year, while others, such as heating and generating electricity, vary according to the weather and other conditions. The anticipated fluctuation in demand, supply and transportation all can affect the prices of futures contracts. In addition, the releases information including weather forecasts by the National Oceanic and Atmospheric Administration (NOAA)
and Weekly Natural Gas Storage Reports by US Energy Information Administration are also known to affect trading of natural gas futures contracts.

Natural gas futures contracts are traded at Chicago Mercantile Exchange (CME) under the code NG \(^1\). The electronic markets Globex and ClearPort operate six days a week Sunday – Friday 6:00PM – 5:15PM (Eastern US Time) with a 45-minute break each day beginning at 5:15PM. Open Cry trading operates Monday – Friday 9:00AM – 2:30PM (Eastern US Time). The trading records do not distinguish which venue has conducted the trades. This trading schedule means that during the normal business hours between Monday and Friday, the trading volume is considerably higher than during the after-hours. Since the 5PM hour has only 15 minutes of trading time, it is likely to be the least active time during a work day. We refer to the daily and weekly operating schedule as the structural pattern of the market in later discussions.

The trading records used for our work include all trades from the beginning of 2007 to the end of 2013. The basic information on each trade includes price, volume and trading time.

For every month of the year, there is a different NG futures contract. As one contract expires, a new one with a longer maturity date replaces it. In this process, the price of the a contract would change abruptly. The price jumps associated with this periodic change of maturity date can be problematic for many data analysis methods and require some form of special handling. With NUFFT, these jumps are treated as the natural part of the pricing function. No additional processing is needed. In the Fourier spectra, we expect to see these jumps translated to a strong amplitude associated with the frequency of once per month.

Following the daily and weekly trading schedule outlined above, we also expect the spectra of trading activities to contain strong amplitudes for frequencies of once per day, once per week, and their multiples. However, the existing analyses \([12]\) do not give us information about the relative strengths of these expected frequencies.

### 4 NUFFT

Given \(N\) data points in a time series, the straightforward version of Discrete Fourier Transform requires \(O(N^2)\) operations to complete. FFT algorithms require \(O(N \log(N))\) operations instead of \(O(N^2)\) operations. For large values of \(N\), FFT can be much faster than the straightforward algorithm. NUFFT can be regarded as a variant of Fast Fourier Transform developed for irregular time series \([15]\). It also requires \(O(N \log(N))\) operations to complete. Since the authors of the NUFFT algorithm have made an effective software implementation available to the public \(^2\), we will not describe the algorithm, but instead provide some practical details on how to use this software.

In this work, we use the 1-dimensional type 1 FFT from NUFFT software library \(^3\). Our time series contains only real numbers, therefore the Fourier spectrum returned is Hermitian. In other word, for a given frequency \(k\), \(g(-k) = g(k)^H\). For our analysis, there is no need to distinguish frequency \(k\) from frequency \(-k\). To study \(K\) frequencies, we need to tell NUFFT to compute \(2K\) of them.

For our seven years of trading records, we split them into seven separate data sets to be passed to NUFFT separately. By splitting the trading records by year, we lose the opportunity to observe the lower frequencies that span multiple years. However, since we expect the year-over-year trading pattern to be non-periodic, this splitting of data will not lose important information. Furthermore, this splitting of data, will give us an opportunity to compare the differences among the trading patterns from different years.

---


\(^2\)The NUFFT software is available at http://www.cims.nyu.edu/cmcl/nufft/nufft.html.

\(^3\)The name of 1D FFT function in NUFFT is dirft1d1.
Given our splitting of data, the lowest frequency other than 0 is once per year (also recorded as 1/year). All other frequencies computed are multiples of 1/year. The frequencies returned by NUFFT can be expressed as 0, ±1, ±2, ±3, . . . . Conceptually, the pair of ±k frequencies represent waves of same frequency but traveling in opposite directions. In this work, we treat this pair of traveling waves as a single one of frequency k with amplitude \( \| g(k) \| \), which denote the 2-norm of \( g(k) \). For each group of one-year-worth of trading records, we compute 2 million different frequencies. These 2 million different frequencies form 1 million pairs. All Fourier spectra reported next each contains 1 million amplitudes and frequencies, excluding the frequency of 0 which has amplitude equal to the average price of the year.

5 Prices Of Trades

To demonstrate the usefulness of NUFFT, we first apply it on prices over time. Figures 2 – 8 show the Fourier spectra produced by NUFFT on prices from 2007 to 2013. In each plot, the small blue dots indicate the amplitude, i.e., the norm of Fourier coefficient \( g(k) \) for frequency \( k \). The horizontal axes of these figures are the frequency \( k \). The frequency of 12 is 12 times a year, which is once per month. The other frequencies marked on the horizontal axes are: once per week (52), once per day (365, or 366 for leap year), once per hour (8,760 or 8,784), and once per minute (525,600 or 527,040). Note that both horizontal axes and the vertical axes are in log-scale. We have chosen to use the log-log plot because it is effective for displaying values spanning many orders of magnitudes and for the power-law relationship. To reduce clutter, we have cut off the vertical axis so as not to show frequencies with amplitudes less than 0.01. This threshold is small enough that all the important features are captured in the plots, but large enough so that a good fraction of the blue dots are excluded.
Figure 3: Fourier spectrum of NG prices in 2008.

Figure 4: Fourier spectrum of NG prices in 2009.
Figure 5: Fourier spectrum of NG prices in 2010.

Figure 6: Fourier spectrum of NG prices in 2011.
Figure 7: Fourier spectrum of NG prices in 2012.

Figure 8: Fourier spectrum of NG prices in 2013.
5.1 Dominant frequencies

A number of points are marked with plus signs ’+’ to indicate that they are local peaks. A local peak is a frequency \( k \) whose amplitude is larger than \( \delta \) neighboring frequencies immediately below or above it, i.e.,

\[
g(k) > g(k + j),
\]

where \( j = \pm 1, \ldots, \pm \delta \),

and \( \delta = \max(2, k/10) \).

The five strongest peaks found for each year have the same frequencies. They correspond to frequencies of once a day (365, or 366 for leap years), twice a day (730, or 732 for leap years), once a week (52), six times a week (313, or 314 for leap years), and eight times a week (417, or 418 for leap years). Since their frequencies and relative strengths are the same throughout all seven years, we believe these peaks correspond to structural patterns of the trading operations mentioned in Section 3. More specifically, the once and twice per day frequencies reflects the daily cycles of starting and ending of a work day, and daily closing of electronic trading venues. The multiple of weekly frequencies correspond to the weekly operating pattern of opening late Sunday afternoon and closing Friday afternoon.

As mentioned before, we expected strong frequencies for daily and weekly cycles, however, from the structure of trading operations, we cannot deduce whether the daily cycles is stronger than the weekly cycles or not. The data shows that the daily cycles are consistently stronger than the weekly cycles. In fact, both daily cycle and twice daily cycles are stronger than the weekly cycles. We believe this observation could be explained by the daily rush of activities immediate following the start of a workday and just before the closing of a workday. While there are weekly activities of traders taking up positions during Monday and closing them by Friday, the data shows that the weekly cycle is considerably weaker than the daily cycles.

5.2 Power law

As frequencies increase, their amplitudes generally decrease in Figure 2 – 8. In each figure, we draw a thin green line passing through the local peaks. This green line is a least-squares regression of the local peaks with frequencies between once per day and once per minute. From these figures, we see that green lines fit local peaks quite well. The presence of this straight lines in the log-log plots indicates a power-law relationship between the amplitudes of local peaks and their frequencies.

Figure 9 shows the exponents of the power law represented by green lines in Figures 2 – 8. Among these seven data points, we can easily identify two groups based on the values of the exponents. The three most recent years have exponents that are noticeably larger than those of the previous four years. This trend indicates that the relatively strengths of the higher frequency activities have increased in the recent years, and there appears to be a monotonically increasing trend for the exponents of the power law.

Similar power-law distribution has been observed by many researchers [25, 3, 14, 13, 12]. A number of different explanations have been proposed in the literature [24, 19, 20, 23]. Generally, an activity pattern with a higher frequency is more likely to be created by a HFT operation. Following assumption, we postulate that the HFT activities have become more important in the recent years and the relative importance of HFT activities is increasing as well.

5.3 Once per minute cycle

There is a strong and consistent peak around the frequency of once per minute for all seven years. Table 1 shows the actual frequency of these peaks and their relative strengths compared to the amplitudes of the
Figure 9: The exponent of the observed power law for top peaks in various years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency</th>
<th>Rel Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>525600</td>
<td>6.7</td>
</tr>
<tr>
<td>2008</td>
<td>527040</td>
<td>5.1</td>
</tr>
<tr>
<td>2009</td>
<td>525600</td>
<td>13.7</td>
</tr>
<tr>
<td>2010</td>
<td>525600</td>
<td>20.3</td>
</tr>
<tr>
<td>2011</td>
<td>525600</td>
<td>15.6</td>
</tr>
<tr>
<td>2012</td>
<td>527040</td>
<td>15.7</td>
</tr>
<tr>
<td>2013</td>
<td>525600</td>
<td>15.4</td>
</tr>
</tbody>
</table>

Table 1: The frequencies and relative strength of the local peak around once per minute in the Fourier spectrum of trading prices.

nearby frequencies. Note that the nearness is defined by Equations 1 and 2. Furthermore, around the frequency of once per minute, the majority of the amplitudes are below 0.01 and not shown in the Figures 2 – 8.

From Table 1, we see that the frequencies of the local peaks around once per minute are exactly once per minute. The mechanism that is producing this peak must be very precise in its timing, and NUFFT must be accurate in picking out this cyclic pattern. We believe that the precise frequencies observed is the result of an HFT strategy based on clock time. The likely candidates are two automated trading algorithms known as TWAP (time-weighted average price) and VWAP (volume-weighted average price).

From Table 1, we see that the relative strengths of these local peaks are quite high. Furthermore, the relative strengths are much higher in the recent year than in 2007 and 2008. This again suggests that the HFT activities are increasing in the recent years.
5.4 New emerging power law

In Figures 2 – 8, we have drawn a yellow line to indicate the center of the bulk of the Fourier components. In Figures 6 – 8, we also draw a set of red lines to indicate the appearance of the second group in the Fourier spectra for year 2011 – 2013. This second group suggests that there might be another mechanism in the time dynamics. We suspect this new power law in recent years has something to do with HFT. Further studies are needed to understand the exact nature of this apparent new power law.

6 Volumes of Trades

Next we examine the Fourier spectra of trading volumes. These spectra share many common features as those of prices. Therefore, the discussions will be much briefer in this section. We also show some additional measurements on the trading volumes to further explore the impact of HFT.

6.1 Fourier spectra of trading volumes

In Figures 10 and 11 we show the Fourier spectra of trading volumes for 2007 and 2013. We have decided to skip the figures for the intervening years because they closely resemble either the figure for 2007 or the one for 2013. Overall, we see that the dominant peaks are still the same as in Figures 2 – 8, their frequencies are 1/day, 2/day, 1/week, 6/week, and 8/week. Not only the top five peaks have the same frequencies, they are also in the same order as in the spectra for prices. This fact agrees with our hypothesis that these peaks are the results of structural pattern of the market mentioned in Section 3.
The straight lines we saw going through the peaks of the spectra of prices are also present in the spectra of trading volumes. This indicates that the time-domain dynamics for the volumes and the prices are closely related.

In addition to the similarities, we also note a few differences. The first difference is that the heights of peaks are more scattered in Figures 10 and 11. They appear to be further away from the regression line than in Figure 2 – 8. In fact, in Figure 10, the peaks are spread so far apart that we have chosen not to show the regression line on it. Another notable feature is that the peak corresponding to once per minute frequency is not present in Figure 10. In fact this peak is also missing in the spectrum for 2008. However, from 2009 onward, this peak is present in the Fourier spectra for trading volumes. Furthermore, the relative strengths of these peaks, shown in Table 2, are just as strong as in the spectra of prices shown in Table 1.
6.2 Direct evidences of algorithmic trading

Since we have seen the peak with the frequency of once per minute in almost all the spectra of both prices and volumes, we next examine this particular pattern in more detail. Figures 12 and 13 show the distribution of number of trades within a minute during different hours of a day. We see that a much large fraction of trades are conducted during the first second of a minute than any other time. This feature again suggests that automated trading are being triggered based time. More specifically, Figures 12 and 13 indicate that these automated algorithms examine their positions during the first second of a minute.
Figure 14: The number of trades with a given size (number of contract).

Figure 15: The total number of contracts traded in a year and the average number of contracts per trade.
6.3 Global statistics

Figures 14 and 15 show some high-level summary information. Here are a few observations. There are many more trades with 100 or more contracts per trade in 2007 than in later years as shown in Figure 14. At the same time, the number of trades with one contract only has increased from about 2.3 million in 2007 to more than 10 million in 2013. From 2007 to 2013, the total number of contracts traded per year has increased from just below 9.5 million to almost 19.4 million, about doubled as shown in Figure 15. On the other hand, the average number of contracts per trade has been decreasing from 2007 to 2013, though the change is much slower after 2010.

The above observations generally support the hypothesis that HFT is more heavily used in the recent years. However, this does not mean the trading has no human involvements. If the machine is making all the decisions, there will be no preference for “round lot” trades, those trades with 50, 100, or 200 contracts per trade. As shown in Figure 14, even though there are not a large number of these “round lot” trades, there are clearly more of these “round lot” trades than similar sizes. For example, the number of trades with 100 contracts is 91 in 2013, while there only 5 trades with 99 contracts and 1 trade with 101 contracts. This preference for “round lot” trades is clearly an indication of the presence of GUI traders.

7 Summary

In this work, we introduce a novel technique named NUFFT for analyzing irregular time series, and demonstrate its effectiveness by applying it to analyze trades of natural gas futures. NUFFT does not require users to resample or bin their data, and can capture characteristics missed by these resampling or binning approaches.

From the Fourier spectra of both trading prices and volumes, we observe strong daily and weekly cycles. We believe that these cycles primarily reflect the daily and weekly operating patterns of the futures contracts trading market because of their precise frequencies and their consistency over the years. We also observe clear indications of power laws in these spectra. We believe this to be the first direct observation of such power law spanning frequency range of once per day to once per minute. We are aware of different mechanisms proposed to explain this power law behavior, further research is needed to better understand the mechanism behind the power laws.

We have seen a number of different signs of increasing high-frequency trading in the natural gas futures trading. Though this trend is unsurprising, however, NUFFT allows us to extract high frequency components easily from the trading records without any additional processing of the data.

The power law trend we observed are for frequencies between once per day and once per minute. Such an observation is impossible with analysis techniques designed for regular time series, because trades in financial markets occur at unpredictable time points. Therefore, NUFFT provides a unique capability for analyzing finance market data.

References


