



Department of Economics and Finance
Finance
Master Thesis in Empirical Finance

High Frequency Trading Illiquidity Patterns

Supervisor
Prof. Paolo Santucci De Magistris

Co-Supervisor
Prof. Stefano Grassi

Candidate
Andrea D'Amato
694321

Accademic Year
2018-2019

Index

1	Introduction	1
2	(Il-)Liquidity measures	5
2.1	Low Frequency measure	6
2.1.1	Roll covarinace measure	6
2.1.2	Relative Bid-Ask spread	7
2.1.3	Spread estimate from High and Low prices	8
2.1.4	Amihud Illiquidity measure	10
2.1.4.1	Mid-range (il-)liquidity measure	10
2.1.5	Other illiquidity measures	11
2.1.5.1	Effective Tick	11
2.1.5.2	High-low Spread Estimator	12
2.2	High frequency measures	14
2.2.1	Liquidity volatility	14
2.2.2	Estimation models	15
2.2.3	Realized Measure	16
2.2.4	High-frequency data flaws	16
2.2.5	Other high frequency sizes	16
2.2.5.1	High frequency volume	16
2.2.5.2	High frequency relative bid-ask spread realized variation	17
3	Realized Volatility theory	18
3.1	Quadratic Variation	18
3.2	Return Realized Variance	20
3.3	Market Microstructure	22
3.3.1	High frequency realized variance	24
4	Stationary and Integrated process	27
4.1	Stationary process	27
4.1.1	Unit-root tests	27
4.1.1.1	Augmented Dickey-Fuller	28
4.1.1.2	Variance Ratio	28
4.2	Cointegration	29
4.2.1	Engle and Granger test	30
4.2.2	Johansen Cointegration test	32

5	Data set and preliminary study	35
5.1	TAQ Data	35
5.2	Descriptive analysis	36
5.3	(Il-)Liquidity measure display	36
5.3.1	Low frequency descriptive measures	36
5.3.2	High frequency measures	37
5.4	Stats comparison	40
6	Empirical analysis	45
6.1	Stationarity tests	45
6.2	Cointegration test	46
6.3	Best Liquidity measure proxies	50
6.4	Stock illiquidity premium	53
7	Conclusions	55
	References	57
	Summary	66

Chapter 1

Introduction

Since the end of the Second World War, the growth rate of the economy has enhanced year by year. This growth rate has risen thanks to the globalization of the economies. Previously, transactions required hours or even days to fulfill the request, which implied that it was not possible to dis-invest in a short time frame in order to satisfy momentary needs. Over the years, investors' liquidity needs has got a primary role for any investment opportunity. At present, the need to divest a position can occur also in seconds, or fractions, and the fear to have an illiquid instrument has risen after the most recent crisis. High frequency trading (HFT) has clearly stressed the speed at which a transaction is executed, but concerning only on the speed aspect let underestimate the implications that it carries.

Liquidity and trading activity are important features of present financial markets, but little is known about their evolution over time or about their time-series determinants. This is caused by the lack of data provision of U.S. stock markets, which has supplied this data information only on some assets for roughly four decades. Their fundamental importance has been represented by the influence that they have on trading costs on specific returns (Amihud and Mendelson, 1986; Jacoby, Fowler, and Gottesman, 2000) which implies a direct link between liquidity and corporate costs of capital. More generally, exchange organization, regulation, and investment management could all be improved by improving the knowledge of factors that influence liquidity and trading activity. Understanding the scheme of unobservable features of the market such as the magnitude of returns to liquidity provision (Nagel, 2012), the impact that asymmetric information has on prices (Llorente et al., 2002), and if trades are more likely to have a permanent or transitory impact on price (Wang, 1994; Vayanos and Wang, 2012) should increase investor confidence in financial markets and thereby enhance the efficacy of corporate resource allocation.

Market microstructure is the study of the trading mechanisms applied in financial securities transactions. There are few historical cases on this field. Some instances can be found as a mere "draft" in the second half of the past century's literature, but over the years, it has acquired importance to understand financial market dynamics.

Microstructure analyses typically touch on one or more of the following aspects of trades: *i*) sources of value and reasons for trade, *ii*) economic setting mechanisms; *iii*) several prices for the same instrument. Market microstructure has grown during the past four decades since many orders are executed, for the same instrument, hundreds of times during a normal trading day. These trading volume frequency is known as High Frequency trading (HFT). These trading data are also more sensitive to be influenced by an irregular price formation or by a trading anomaly activity. These issues characterize, mostly, by an irregular time interval within trading, which reflect the easiness of liquidating an instrument within

a short-time interval.

Indeed, liquidity is one of the most important subjects in the secondary markets. Accurate definitions only exist in the contexts of particular models, but these are widely accepted and understood practical and academic discourse. Liquidity impounds, in its meaning, the economic concept of elasticity. *In a liquid market, a small shift in demand or supply does not result in a large price change.* Liquidity relates also to the cost of trading, something distinct from the price of the security being bought or sold. Liquid markets are characterized by low trading costs. It has also dynamic traits, since, accomplishing a purchase or sale over a short horizon does not cost more than spreading the trades over a longer time interval.

Liquidity is usually measured by "depth, breadth, and resiliency." In a deep market, by focusing on the current market price, there is a large incremental quantity available for sale (buy), above (below) the current market price. In a broad market, considering that there are many participants, none of whom is presumed to have a significant market power. In a resilient market, the price effects that are associated with the trading process are small and last for very narrow periods.

With the rise of global economy, the role of liquidity demand and supply has enhanced through "liquidity externality", that can be identified as a network externality. The attributes of liquidity are important since individual agents can trade at lower cost, when the number of participants increases. This force favors market consolidation, the concentration of trading activity in a single mechanism or venue. However, differences in market participants (e.g., retail versus institutional investors) and innovations by market designers militate in favor of market segmentation (fragmentation). While liquidity is seen kindly its counterpart, illiquidity, comes from adverse selection and inventory costs. As an instrument becomes more liquid, the chances to dis-invest it are higher in order to obtain a "modest" premium, while as an instrument is more illiquid, more it is sensitive to sudden changes that may impair the investor. The more an instrument is illiquid, the more an investor may move its price by entering with a huge position in order to satisfy his needs. In order to study in which instrument an investor should invest, a new high frequency field is born in order to understand the liquidity patterns of any instrument.

Size, or the market value of the stock, has seen to be related to liquidity since a larger stock issue has smaller price impact for a given order flow and a smaller bid-ask spread. Stock expected returns are negatively related to size, which is considered as a liquidity driver. The negative return-size relationship may also result from the size variable being related to a function of the reciprocal of expected return.

Despite the stress that investors put on this concept there has been a scarce quantity of studies about these argument and at the present the concept is analyzed by many academics. Roll (1984) was one of the first pioneer in this field that developed a spread measure that has been used over and over, with its own limitations. Many scholars modeled from this finding in order to overcome the measure limits. Hasbrouck (2004, 2009) proposes a Gibbs sampler Bayesian estimation of the Roll model; Lesmond, Ogden, and Trzcinka (1999) introduce an estimator based on zero returns. Following the same line of reasoning of this, Fong, Holden, and Trzcinka (2017) formulate a new estimator that simplifies the existing the previous measures. Holden (2009), jointly with Goyenko, Holden, and Trzcinka (2009), introduces the Effective Tick measure based on the concept of price clustering. High and low prices have been usually used to proxy volatility (Garman and Klass, 1980; Parkinson, 1980; Beekers, 1983). Corwin and Schultz (2012) use them to put forward an original estimation method for transaction

costs where they assume that the high (low) price are buyer- (seller-) initiated, and their ratio can be disentangled into the efficient price volatility and bid-ask spread. Abdi and Ranaldo overcome the limits that Corwin's measure had, since it implicitly relied on the same assumptions made by Roll, to introduce a Mid-range price made as a trivial average price of the high and low prices. There is, in addition, also the Amihud (2002) illiquidity measure formulated as the daily ratio of absolute stock return to its dollar volume, averaged over some period. This can be viewed as the daily price response associated with one dollar of trading volume, which roughly reflect the price impact.

However, while these measure needs "only" data with a lower frequency, usually a day, there are others that require microstructure data. Market microstructure has been analyzed for almost four decades since it has been a game changer in the trading industry. Usual measures like variance and correlation were not applicable also in this context so new measure had to be modeled. Jacod (1994) and Jacod and Protter (1998) were one of the first studies in this new field and they introduced the new concept of realized variance, which is the realized counterpart of the conceptual variance. Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001 and 2002) re-arrange the realized measure and model it from short-term price changes, which treat volatility as observable, rather than as a latent variable. This has been proved to be an useful estimate of fundamental integrated volatility by Hasbrouck (2018), Hwang and Satchell (2000), Zhang (2010). However, Hansen and Lunde (2006) pointed out that, when dealing with market microstructure, the presence of market microstructure noise in high frequency data complicates the estimation of financial volatility and makes standard estimators, such as the realized variance, unreliable. So market microstructure noise challenges the accountability of theoretical results that rely on the absence of noise. The best remedy for market microstructure noise depends on the properties of the noise. The time dependence in the noise and the correlation between noise and efficient price arise naturally in some models on market microstructure effects, including (a generalized version of) the bid-ask model by Roll (1984) (Hasbrouck, 2004) and models where agents have asymmetric information (Glosten and Milgrom, 1985 and Easley and O'Hara, 1987, 1992). Market microstructure noise has many sources, including the discreteness of the data (see Harris 1990, 1991) and properties of the trading mechanism (Black, 1976; Amihud and Mendelson, 1987 and O'Hara, 1995). Other studies were focused on one of the most used sizes that plainly represents illiquidity: the bid-ask spread. Particularly, starting from Amihud and Mendelson's (1986) liquidity study, it has been running several researches on the evolution of the volatility of this size. Studies on liquidity volatility are found mostly on foreign exchange bid-ask spreads. Glassman (1987) and Boothe (1988) study the statistical properties of the bid-ask spread, Bollerslev and Melvin (1994) evaluate the distribution of bid-ask spreads and their ability to explain foreign exchange rate volatility based on a tick-by-tick data set. Furthermore, Andersen et al. (2001) and Cuonz (2018) model a realized variance measure on the intra-day bid-ask spread which satisfies the realized variation assumptions and gives consistent results, even if stock returns are not accounted.

By considering all these measure, it is advisable to check if these do not follow a random walk process, or, worse, an explosive behavior. Many unit-root tests have been established but the ones that are the most used and accountable are the Augmented Dickey-Fuller (Dickey and Fuller, 1981) and the Variance Ratio test (Campbell and Mankiw, 1987; Cochrane, 1988; Cogley, 1990). If a process has an unit root, i.e. it follow a random walk, is referred as integrated of order one, denoted as $I(1)$. To overcome this issue, lots of academics tried to solve the issue that a non-stationary process brings with it. By applying usual regression concept and de-trending techniques on non-stationary processes does

not solve the problem, and spurious correlation shows up, Granger (1983). Lately, Engle and Granger (1987) coined the cointegration term and a two-step approach, by having two $I(1)$ processes, non-stationary, their linear combination would be an $I(0)$ process, stationary. Johansen (1991), developed an augmented version of the two-step procedure made by Engle and Granger, based on the Dickey-Fuller test for unit roots in the residuals, which allows to test more than one cointegrating relationship. The Johansen test is subject to asymptotic properties, i.e. large samples. Indeed, if the sample size is too small, results will not be reliable and Auto Regressive Distributed Lags (ARDL) should be used.

Studies have been done across the academia to study what are the the best proxies that explain the reward that an investor should obtain in exchange of the illiquidity risk that he sustains. This study investigate the relation that runs within the illiquidity premium and low frequency liquidity measures. The measure that are used in this work are classified as low, daily, frequency and high frequency measures. The daily measure cover the Amihud illiquidity measure, the relative bid-ask spread and, a "fresh-formulated" liquidity measure, the mid-range illiquidity price, which is the ratio of the daily mid-price, developed by Abdi and Ranaldo (2017), over the related daily volume. Their counterpart, high frequency, measure, instead, embrace the Realized Variance of the stock returns, the Realized liquidity variation and the intra-day log volume.

Firstly, non-stationary-based tests computed over the measures confirms that these neither follow a random walk or an explosive process. However, both the bid and ask prices, as noted in the literature, are non-stationary. Nevertheless, by their linear combination results to be stationary, which allows to state that these are cointegrated. Despite this, their mid-prices are not cointegrated, and their combination does not satisfy the stationarity requirements. Any possible combination within the stocks has been made but these seems to not satisfy either the stationary conditions. Anyway, this result should be expected due to the low relationship that resides within the firm's industries.

Then, tests on the relation that elapses with, one of the most accountable illiquidity measures, bid-ask spread and low and high frequency measure prove that the low ones seem to be good estimates while the latter give good proxies, but with less explanatory power than the daily ones. Anyhow, a "comprehensive" test with all variables exposes a slight significance to the previous ones, but stressed the role of the "new-developed liquidity measure". Lastly, the core test of this study, the relation with the daily liquidity measures and the illiquidity risk premium, gives consistent results that confirms Amihud's finding. Indeed, as a stock is more illiquid, it requires an higher premium that it should pay to its owner. This result highlights the positive relation with the illiquidity measures, Amihud and relative bid-ask spread measure, and the negative relation with the liquid one, mid-range liquidity price. The paper proceeds as follows. Chapter 2 discusses briefly the liquidity-based measures used in this work and other measures that have been used in the literature. Chapter 3 exposes the theory behind the realized measure and the microstructure effect on this. Chapter reports a short explanation of stationary tests and the role of cointegration over non-stationary processes. Chapter 5 highlights the database and descriptive statistics of the sample and the measures. Chapter 6 embodies the analysis ran over the measures and the illiquidity reward. Chapter 7 concludes the paper.

Chapter 2

(Il-)Liquidity measures

Liquidity is an subtle concept. It cannot be observed directly, but rather has different features, which, cannot be shorted in a single measure. The definition of illiquidity concerns an asset that is difficult to sell because of its expense, lack of interested buyers, or some other reason, which generates adverse selection costs and inventory costs that affect stock price impact (Amihud and Mendelson, 1980; Glosten and Milgrom, 1985). For standard-size transactions, the price impact is the bid-ask spread, whereas larger excess demand induces a greater impact on prices (Kraus and Stoll, 1972; Keim and Madhavan, 1996), which may reflect some informed trading actions (Easley and O'Hara, 1987). Since it is difficult to disentangle order flows generated by informed traders or by liquidity (noise) traders, market makers set prices that are an increasing function of the asymmetries in the order flow, transaction volume, which may indicate market manipulation, informed trading, actions. This buyer/seller imbalance has been found to be positively related with price change, price impact (Kyle, 1985).

It is known that higher the volatility is less liquid is the market. This suggest that the market can face three different, but sometimes complementary, situations: *a)* there is a low levels of trading volume, *b)* the current volume is buyer/seller-skewed and *c)* the market is characterized by intermittent transactions, usually characterized by heavy volume positions. The focus of this work is to analyze the latter aspect of these and analyze the the evolution of its variation.

In the literature, (il-)liquidity measure can be categorized into: high frequency and low (daily) frequency measures.

The first are characterized by loads of information that resides in the price process and some of the main measures are the liquidity second moment measure and the average intra-day bid-ask spread. Even if the first tries used raw liquidity measure, the liquidity data still is difficult to extrapolate in a clean and useful way, also nowadays. Many processes measure the liquidity deviation without considering the link that liquidity risk has with it. However, these approaches are often very complex to analyze and manage. Furthermore, since the magnitude of this data is huge, the price process is usually affected by microstructure noise which must be accounted during the estimate of the measures.

Instead, the latter use daily frequency data that are easily available in any market for every stock, sometimes even not listed ones. While the first give consistent and accurate estimates of market and stock liquidity, they are also characterized by not optimal, more precisely availability, data that are not given for many instruments; the latter can give also consistent measures which can be calculated easily, even if these are not accurate as the other.

This chapter is organized as follows: at first, the main low frequency measures will be exposed and the possible advantages and disadvantages of these will be listed. Lastly high frequency measures will

be introduced: the liquidity variation, which is based on the concept of realized variance for a liquid instrument, the log of the total number of intra-day trades and the intra-day bid-ask spread.

2.1 Low Frequency measure

Usually, illiquidity measures require, for their estimate, microstructure data on transactions and quotes for very long time periods at highly frequency sample, also at frequencies lower than a minute. Although these exhibits consistent and robust estimate of the implied patterns, these microstructure data are unavailable in most markets around the world at high frequencies for many years. Instead of using these, the illiquidity measures used in this study are calculated from daily data which are readily available over long periods of time for most markets. In spite of the fact that these are roughly accurate, it is readily available for the study of the time series effects of liquidity.

2.1.1 Roll covarinace measure

Roll (1984) covariance measure has been one of the first estimates that has been used to infer the effective bid-ask spread directly from a time series made by market prices. This estimator works, and can be estimated, only if two assumptions are satisfied:

1. The examined asset is traded in an Informationally Efficient Market;
2. The probability distribution of observed price changes is stationary, at least in the short run.

It illustrates a dichotomy fundamental to many microstructure models, the distinction between price components due to fundamental security value and those attributable to the market organization and trading process. If the market is informationally efficient, and trading costs are assumed to be zero, then the market price should reflect, and contain, all the relevant information about the traded security. A price change occurs only if an unanticipated information is disclosed to market participants. Although the transaction costs are not equal to zero, the market maker (dealer) should be compensated by, usually, establishing a bid-ask spread, in which it relies the true market value of the asset. If the value of any security fluctuates, within this region, randomly, then it may be seen as setting the true value in the "middle point" in the spread, computed by taking the average between the bid and the ask price.

Roll model follows a geometric Brownian motion (GBM), and the observed (log) asset price at time t evolves according to:

$$\begin{aligned} p_t &= p_t^* + I_t \frac{s}{2}, \\ p_t^* &= p_{t-1}^* + \epsilon_t, \end{aligned} \tag{2.1}$$

where p_t^* is the underlying fundamental log price with innovation, given by ϵ_t , and the trade direction indicators $\{I_t\}$ are i.i.d. and take value ± 1 with equal probability of one half. If the transaction is a purchase $I_t = 1$ (*bid*), if it is a sale $I_t = -1$ (*ask*). Only the price p_t is observed, while all the other variables are not. The innovation term ϵ_t is serially uncorrelated and uncorrelated with the trade direction indicator.

Since recorded transactions occur at the boundaries of the spread, the bid and ask prices, and not in the "middle point", it occurs that observed market price changes cannot be supposed to be independent

anymore.

If there are no new information about the asset, transactions should arrive into the market randomly at the buy or sale position by the traders. Since each position is equally likely to occur, the joint probability of next price change ($\Delta p_t \equiv p_t - p_{t-1} = \epsilon_t + \Delta I_t \frac{s}{2}$) in trades depends on the position taken, if it was at the bid or at the ask. As Roll showed, if the transaction at $t - 1$ is at the bid (ask) price, next price change cannot be negative (positive) since there is no new information. So, it is not possible to see two successive price increases (declines).

Since these transactions are based on sides of the spread on one side, or the other, Roll computed the covariance between the price change of trades between the "current" time and the following one.

$$Cov(\Delta p_t, \Delta p_{t+1}) = -\frac{s^2}{4}, \quad (2.2)$$

by rearranging the equation, in order to obtain the spread, it has been found that:

$$Spread = 2\sqrt{-Cov}, \quad (2.3)$$

To have a deeper understanding of the relation between price changes, Roll estimated the variance of the price change is equal to one half the squared spread (i.e. $Var(\Delta p) = \frac{s^2}{2}$). By knowing this information it has been found that the autocorrelation coefficient is equal to minus one half.

At first sight, the autocorrelation coefficient may seem wrong, but it must be noted that, since the covariance is divided by the variance of the unconditional price changes. From the efficient market hypothesis about new information, the variance of observed changes is likely to be influenced by the new set of information, while the covariance between forward price changes cannot be explained by new information, if markets are efficient.

Roll's measure is an intuitive and easy measure to compute. It provides estimates with intra-day data. However, even with a long time series of daily observations, the covariance of price changes is usually positive, forcing the researchers to convert implement and imaginary number into the spread estimate. Indeed, since Roll found that cross-sectional average covariances are positive for some years, for these cases, researchers usually do one of three things:

1. treat the observation as missing;
2. set the Roll spread to zero;
3. multiply the covariance by negative one and multiply the spread by negative one, in order to produce a negative spread.

This measure has helped the literature to deal with price movement patterns of trades. However, as Rinaldo and Abdi noticed, Roll's assumptions can be seen also as its limits that highlight two issues:

1. It relies on the occurrence of bid-ask bounces;
2. It relies on the assumption of equally likely serially independent trade directions

2.1.2 Relative Bid-Ask spread

In the literature, the bid-ask spread has always been an indicator of the market liquidity. Indeed, it is known that as the spread widens, the liquidity of the market narrows, due to high transaction costs

that affect the market. It is the difference that an investor would earn if he executes an immediate sale (ask or offer) and an immediate purchase (bid) on the same instrument. To study the development of market liquidity, Andersen, Bollerslev, Diebold and Labys (2001) and Cuonz (2018) noticed that the concept of realized liquidity variation shares the same empirical assumption of the realized volatility.

$$BAS_t = \log P_{ASK,t} - \log P_{BID,t}, \quad (2.4)$$

where P_{ASK} and P_{BID} are respectively the ask and bid prices. To overcome the dimension (currency) limit that may affect this index in different markets with different currencies, relative spread is denoted as the ratio of the bid-ask spread over the mid price. Working with relative bid-ask spreads has various advantages. Although the bid-ask spread itself is measured in terms of the underlying currency, the relative bid-ask spread is dimensionless. Hence, relative spreads from different markets (respectively currencies) can directly be compared with each other. The relative bid-ask (RBA) is then computed as

$$RBA_t = \frac{BAS_t}{\log(\eta_t)}. \quad (2.5)$$

where η_t is the mid price of the respective stock at time t . Relative spread is thus a measure-free estimate that, within its easiness, allows to infer about the liquidity of the analyzed stock. Nonetheless, in an high-frequency market, market market makers can slightly move in their favor the price if they are sufficiently fast, such as the spread would be slightly biased for the informed party.

2.1.3 Spread estimate from High and Low prices

As mentioned, Roll measure kicked off the finance literature in order to estimate the spread in order to understand its pattern. Corwin and Schultz (2012) established a model based on the estimation of Bid-Ask Spreads through the use of daily close high and low prices. Later, Rinaldo and Abdi (2017) modified furthermore Corwin's estimate in order to overcome Roll's covariance limits.

Daily high prices are usually buyer-initiated trades while low prices are usually seller-initiated trades. The ratio of high-to-low prices of a day reflects the fundamental volatility and the bid-ask spread of a stock. The volatility component of the high-to-low ratio increases proportionally with the length of the trading interval, while the bid-ask spread component is not affect by this scheme.

Rinaldo's measure overcomes Roll's limit since it exploits a larger information set, by using daily high and low prices while Roll used only daily close prices, and it is completely independent of any trade direction dynamic.

The model scheme has similar assumptions to those made by Roll. Efficient price follows a GBM and the observed price can be either buyer or seller initiated. The observable close log-price is equal to:

$$c_t = c_t^e + q_t \frac{s}{2}, \quad q_t = \pm 1, \quad (2.6)$$

while the respectively high and low prices are:

$$\begin{aligned} h_t &= h_t^e + \frac{s}{2}, \\ l_t &= l_t^e - \frac{s}{2}, \end{aligned} \quad (2.7)$$

the superscript e indicate the efficient price, so, c_t^e refers the efficient log-price at the closing time and the q_t indicators is similar to the ones introduced by Roll $\{I_t\}$, where it takes values of ± 1 , 1 if the trade is buyer initiated, -1 if it is seller initiated.

By defining the mid-range as the average of daily high and low log-prices:

$$\eta_t \equiv \frac{l_t + h_t}{2}, \quad (2.8)$$

As the authors exposed in their studies, the mid-range has some characteristics:

1. The mid-range of observed prices matches with the mid-range of efficient prices:

$$\eta_t = \frac{l_t^e + h_t^e}{2}, \quad (2.9)$$

and the efficient price hits η_t at least once during the day;

2. The mid-range of the current day and the following one is an unbiased estimator for the closing mid-quote:

$$E\left[\left(c_t^e - \frac{\eta_t + \eta_{t+1}}{2}\right)\right] = 0, \quad (2.10)$$

3. The squared distance of t^{th} day and the mid-point is:

$$E\left[\left(c_t^e - \frac{\eta_t + \eta_{t+1}}{2}\right)^2\right] = \frac{s^2}{4} + \sigma_e^2\left(\frac{1}{2} - \frac{k_1}{8}\right), \quad k_1 \equiv 4\ln(2), \quad (2.11)$$

which is made up by two components:

- i. squared effective spread: $\frac{s^2}{4}$

which is the bid-ask spread component, that represents the squared distance between the close price and the midquote when the market closes;

- ii. transitory variance: $\sigma_e^2\left(\frac{1}{2} - \frac{k_1}{8}\right)$

which is the efficient price variance component, that represents the squared distance of the midquote from its approximation.

4. The variance of changes in mid-ranges is a linear function of the efficient price variance:

$$E\left[\left(\eta_t + \eta_{t+1}\right)^2\right] = \sigma_e^2\left(2 - \frac{k_1}{2}\right). \quad (2.12)$$

Since the mid-ranges are independent of the spread, their difference reflects the volatility. The estimated efficient price volatility follow the "true" efficient price volatility and this measure is still unbiased than other high-low volatility estimates.

Furthermore, the squared effective spread evaluated follows

$$s^2 = 4E\left[\left(c_t - \frac{\eta_t + \eta_{t+1}}{2}\right)^2\right] - E\left[\left(\eta_t + \eta_{t+1}\right)^2\right] = 4E\left[\left((c_t - \eta_t)(c_t - \eta_{t+1})\right)\right]. \quad (2.13)$$

At first sight the squared spread looks like the Roll's. However, this measure does not depend on constrained assumptions on the serial independence of trades and a similar likelihood of the close price on the buy and sell side.

Nevertheless, this model may have some estimation errors, that may result the RHS with a negative value. To overcome this issue, as already stated in the literature, three dealing approach must be accounted:

1. Set negative estimates to zero, and then calculate the spread;
2. Set negative estimates to zero and then take the average of the previous two days spreads;
3. Remove negative estimates and compute the spread for only the positive estimates and take their average.

TAQ data simulations and comparisons indicate that the first two approaches provide better outcomes, both in terms of bias correction and estimation errors.

2.1.4 Amihud Illiquidity measure

Amihud (2002) drafted an illiquidity measure that shows that stock excess return is positively affected by expected market illiquidity and negatively by unexpected illiquidity.

Stock illiquidity is defined as the average ratio of the daily absolute return over the trading volume on that day:

$$AI_t = \frac{1}{N} \sum_{t=1}^N \frac{|r_t|}{Vol_t}, \quad (2.14)$$

where r_t is the stock return over a day in a year, Vol_t is the respective daily volume and N is the number of days for which trading data are available.

This measure has been used widely in the literature thanks to its easy composition and understanding structure: higher is the ratio lower is the trading volume of the corresponding stock (i.e. lower liquidity) compared to its return.

The studies found that stock excess return, also known as "*risk premium*", includes a premium for illiquidity; since stocks are riskier and also less liquid than short-term Treasury securities because:

- i. Bid-ask spread and brokerage fees are higher on stocks, i.e. illiquidity costs have a greater influence on stocks;
- ii. Treasuries transactions' size is greater.

So, Amihud found that expected stock excess return is an increasing function of expected market illiquidity. This result has been found to be stronger on small firm stocks rather on larger firms. This effect is enhanced during period in which liquidity is stressed, such that there is a "flight to liquidity" that makes large stocks more attractive. Small stocks' sensitivity to illiquidity indicates that these stock are subject to higher illiquidity risk, that translates in higher illiquidity risk premium.

2.1.4.1 Mid-range (il-)liquidity measure

To put an heavier stress on the illiquidity spin, on the reading of the market return, it has run a re-model between the mid-range price measure, made by Ranaldo, and the Amihud one. This measure has been developed as a mix of Amihud and Ranaldo' s. The main goal of this rephrase is to discover if an instrument is relatively illiquid, by having at one's disposal only "trivial" stats of any listed stock (daily

high and low prices and volume). This reinforce has been devised since results with just the mid-range price were not crucial in the estimation of the return. As the AI, this fit is estimated as:

$$MI_t = \frac{1}{N} \sum_{t=1}^N \frac{\eta_t}{Vol_t}. \quad (2.15)$$

As already mentioned, larger is the measure more sensible the stock price is to a large position entrance in the market, and viceversa.

2.1.5 Other illiquidity measures

In this section other low-frequency illiquidity measures are briefly summarized. These are listed in order to give a complete view of many empirical tests made on this subject. However, since the main interest of this paper is to compare estimation methods based on price data when quote data are not available, tests on these would not take place in this work since this has been already treated by Abdi and Rinaldo (2018).

Roll (1984) introduced the first measure, made by price data for bid-ask spread estimation. To return a non-negative spread, the first-order autocovariance of the price changes must be negative. However, as mentioned, Roll (1984) found positive estimated autocovariances for several stocks. Harris (1990) found that the positive autocovariance values occurred when the spread is likely to be narrow. This has motivated the common practice of replacing the positive autocovariance values with zero, which allows to have a zero spread estimate.

To overcome the negative spread estimate issue, Hasbrouck (2004) introduces a Gibbs sampler estimation of the Roll model using prices from all days. Hasbrouck assumes that the public information shock ϵ in the Roll model is normally distributed with mean of zero and variance of σ_ϵ^2 . He denotes the half-spread in the Roll model as $c = \frac{S}{2}$.

Hasbrouck uses the Gibbs sampler to numerically estimate the model parameters $\{c; \sigma_\epsilon^2\}$, the latent buy/sell/no-trade indicators $Q = [Q_1; Q_2; \dots; Q_T]$; and the latent "efficient prices" $V = [V_1; V_2; \dots; V_T]$, where T is the number of days in the time interval develops a Gibbs sampler Bayesian estimator. By using annual estimates, he shows that the spreads originated from the Gibbs method have higher correlations with the high-frequency benchmark.

Fong, Holden, and Trzcinka (2017) develop an estimator which relies on the assumption that price movements that are smaller than the bid-ask spread are unobservable and manifest in the days with zero returns. They elaborated this measure to simplify the LOT measure developed by Lesmond, Ogden, and Trzcinka (1999). This has proved to give accurate liquidity estimates of the global equity market.

2.1.5.1 Effective Tick

Holden (2009) and Goyenko, Holden, and Trzcinka (2009), develop The effective tick estimator, which is based on the idea that wider spreads are associated with larger effective tick sizes. Their model assumes that, when both the tick size and the bid-ask spread are one-eighth, all possible prices are used, but when the tick size is one-eighth and the spread is one quarter, only prices ending on even-eighths are used.

Goyenko, Holden, and Trzcinka (2009) show that their assumed relation between spreads and the

effective tick size allows researchers to use price clustering to infer spreads. Lets suppose that there are four possible bid-ask spreads for a stock: $\$ \frac{1}{8}$, $\$ \frac{1}{4}$, $\$ \frac{1}{2}$, and $\$1$. The number of quotes with odd-eighth price fractions, associated only with $\$ \frac{1}{8}$ spreads, is given by N_1 . The number of quotes with odd-quarter fractions, which occur with spreads of either $\$ \frac{1}{8}$ or $\$ \frac{1}{4}$, is given by N_2 . The number of quotes with odd-half fractions, which can be due to spreads of $\$ \frac{1}{8}$, $\$ \frac{1}{4}$ or $\$ \frac{1}{2}$ is N_3 . Lastly, the number of whole-dollar quotes, which can occur with any spread width, is given by N_4 .

To calculate an effective spread, the proportion of prices observed at each price fraction is calculated as

$$F_j = \frac{N_j}{\sum_{j=1}^J N_j}, \quad \text{for } j = 1, \dots, J. \quad (2.16)$$

The unconstrained probability of the j^{th} spread, U_j , is given by

$$U_j = \begin{cases} 2F_j, & \text{for } j = 1 \\ 2F_j - F_{j-1} & \text{for } j = 2, \dots, J-1 \\ F_j - F_{j-1}, & \text{for } j = J \end{cases} \quad (2.17)$$

The effective tick model directly assumes price clustering (i.e., a higher frequency on rounder increments). However, in small samples it is possible that reverse price clustering may be realized (i.e., a lower frequency on rounder increments). Reverse price clustering unintentionally causes the unconstrained probability of one or more effective spread sizes to go above one or below zero. So, constraints are added to generate proper probabilities. Let $\hat{\gamma}_j$ be the constrained probability of the j^{th} spread. It is computed in order from the smallest to the largest as follows:

$$\hat{\gamma}_j = \begin{cases} \text{Min}[\text{Max}(U_j, 0), 1], & \text{for } j = 1 \\ \text{Min}[\text{Max}(U_j, 0), 1 - \sum_{k=1}^{j-1} \hat{\gamma}_k] & \text{for } j = 2, \dots, J \end{cases} \quad (2.18)$$

Lastly, the effective tick measure is a simple probability-weighted average of each effective spread size divided by \bar{P}_i , which is the average price at time i

$$\text{EffectiveTick} = \frac{\sum_{j=1}^J \hat{\gamma}_j s_j}{\bar{P}_i} \quad (2.19)$$

2.1.5.2 High-low Spread Estimator

Corwin and Schultz (2012) develop an estimator based on daily high and low prices. The high-low (HL) price ratio reflects both the true variance of the stock price and the bid-ask spread. While the variance component grows proportionately with the time period, the spread component does not. This allows to solve for both the spread and the variance by deriving two equations, the first a function of the high-low ratios on 2 consecutive single days and the second a function of the high-low ratio from a single 2-day period.

They assume that the true or actual value of the stock price follows a diffusion process. They assume that there is also a spread of $S\%$, which is constant over the 2-day estimation period. Because of the spread, observed prices for buys are higher than the actual values by $\frac{S}{2}$, while observed prices for sells are lower than the actual value by $\frac{S}{2}$. They hypothesize that the daily high price is a buyer-initiated

trade and is therefore grossed up by half of the spread, while the daily low price is a seller-initiated trade and is discounted by half of the spread. Hence, the observed high-low price range contains both the range of the actual prices and the bid-ask spread. Let $H_t^A(L_t^A)$ denotes the actual high (low) stock price on day t and $H_t^o(L_t^o)$ the observed high (low) stock price for day t , so the HL measure is given by

$$\left[\ln \left(\frac{H_t^o}{L_t^o} \right) \right]^2 = \left[\ln \left(\frac{H_t^A(1 + \frac{S}{2})}{L_t^A(1 - \frac{S}{2})} \right) \right]^2 \quad (2.20)$$

HL estimator captures transitory volatility at the daily level and will closely approximate the effective spread, or the cost of immediacy. It also does not require data on trading volume. It can therefore be applied in settings such as emerging markets, where the quality or availability of volume data may be suspect. Furthermore the power to predict cross-sectional differences in returns is very similar to that of the AI.

However, this measure has been found to suffer of two drawbacks: not being robust to the price movements in non-trading periods, such as weekends, holidays, and overnight price changes and it does need some ad-hoc overnight price-making adjustments., and, being based on the price range, the model is sensitive to the number of observed trades per day since it adjusts the estimate by including the gap between observable prices.

Measure	Input	Description
Roll	Close price	$s = 2\sqrt{\max\{-cov(\Delta p_{t+1}, 1Deltap_t), 0\}}$
RBA	Bid and Ask price	$\frac{BAS_t}{\log(\eta_t)}$
Gibbs	Close price	Bayesian estimation of spreads by setting a nonnegative prior density for the spreads.
AI	Closing price and volume	$\frac{1}{N} \sum_{t=1}^N \frac{ r_t }{Vol_t}$
EffTick	Close price	$\frac{\sum_{j=1}^J \hat{\gamma}_j s_j}{P}$
FHT	Close price	$FHT = 2\sigma \mathcal{N}^{-1}(\frac{1+z}{2})$, $z = \frac{ZRD}{TD+NTD}$, where ZRD is the number of zero returns days, TD is number of trading days and NTD is number of no-trade days in a given stock-month.
HL	High and Low prices	$[\ln \frac{H_t^o}{L_t^o}]^2 = [\ln \frac{H_t^A(1+S/2)}{L_t^A(1-S/2)}]^2$
MR	High and Low prices	$\eta_t = \frac{l_t + h_t}{2}$

Table 2.1: Low frequency illiquidity estimators.

All uppercase variables are expressed with their market observation while all lowercase variables are the respective log transformation.

Table 2.1 summarizes all the main measures that had significant results. It must be clarified that all

of these are merely bid-ask spread estimators, except AI. Nonetheless, the magnitude of measures the ones that are used are AI, RBA and MI

2.2 High frequency measures

As already mentioned, high frequency estimates are the ones that exhibit consistent and robust results in the study of price dynamics in the markets. Although these are usually difficult to estimate and to model, due to their scarce availability for many stocks, these are widely used to verify and to test liquidity assumptions that reside in the market.

In this section the high frequency measure that will be treated are: the realized liquidity variation, the intra-day relative spread and the log of the total number of intra-day trades. While the first measure, the realized liquidity variation, requires more attention and will be treated among this section, the last two variables have been already treated and analyzed and require less explanation, that will be highlighted at the end of the liquidity variation.

2.2.1 Liquidity volatility

In the literature the topics of volatility and liquidity have always been interconnected. It is worth to repeat that periods which have been featured with an high volatile environment, usually characterized by large bid-ask spreads, face a lack of liquidity supply which has, more or less, affected the price processes. Indeed, a liquid instrument can be analytically defined as an instrument that has a consistent and sustained trading volume along many periods. Instrument that are cataloged as liquid usually have a low risk premium, since these are assumed to be "safe". On the other hand, illiquid instruments are associated with a higher risk when an investor holds these, which translates in a higher reward that he wants to obtain in exchange of the risk that he has taken. The latter are distinguished from the first for their higher fundamental volatility. However, in a high frequency world, price changes, and also return, occur hundreds of times during a single day, which implies that a new dispersion measure should be used to analyze price schemes.

Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001 and 2002) introduced the respective realized measure calculated from short-term price changes. This is an useful estimate of fundamental volatility, fundamental integrated volatility (IV), and typically serve as inputs to longer-term forecasting models (Hasbrouck, 2018; Hwang and Satchell, 2000; Zhang, 2010). Realized Volatility (RV) can be constructed directly from trade, bid and offer prices, which are typically noisy due to the presence of microstructure components. However, by locally averaging, their effects can be smoothed (Hansen and Lunde, 2006).

While the RV estimate will be carefully exposed along chapter 3, here it will be presented the importance that the variance of liquidity has to access information about the liquidity of a stock.

The liquidity second moment can be measured by giving concern on the measures, on which, the estimate must be made. Liquidity concept has deep meaning, but the main feature that define it can be categorized in: funding liquidity and market liquidity, when the focus is all assets liquidity, when looking at a single asset it can be referred simply as asset liquidity. Although the topic of funding liquidity can influence even matters tied to the financial stability of markets, it is not interest in this work to analyze, directly, it. Therefore, unless explicitly disclosed, with the term liquidity it will be

referred as asset liquidity.

Even-though, asset liquidity is a shrank fraction of the broader one, it has still different aspects that must be accounted. One facet reads the level of liquidity as a measure for the added costs per transaction associated with trading large quantities of a financial asset. Another-one reads the liquidity as the ability to trade an asset without causing significant changes in its market price. Market and asset liquidity is can be assessed on three aspects: tightness, depth and resilience. Tightness is usually measured by the bid-ask spread, which defines the round-trip cost (cost of the instantaneous purchase and sale of the same asset), but it excluding operational costs. Depth measures the size of a transaction required to change the price of an asset. Resilience is the speed at which prices return to their old equilibrium after a shock event in the demand or supply. Certainly, depth and resilience are concept that are more difficult to measure, than tightness, because they require continuous information concerning the order book, which is the information on every single asset trade that runs in the market. Indeed, market tightness is most closely monitored by the investor community, while the others are more relevant for other features that characterize the market participant or the context that wants to be analyzed.

In the previous academia works, liquidity has been measured mostly on low-frequency samples. At present, the increasingly supply of high-frequency data, for many listed assets, stress that by relying on low-frequency liquidity data would be unwise. Some measures can be derived also at high frequencies, which offer information on the intra-day level of liquidity. It can be said that the most known ones are the bid-ask spreads and the order flow.

In this paper, different measures have been used to study high-frequency liquidity.

Bid-ask spreads are acknowledged as being one of the most accurate and direct measures of liquidity, and they have been used as the primary driver of the Amihud and Mendelson's (1986) liquidity study. Bid-ask spreads are available to most investors and for a many asset classes, and so are one of the most used measures by researchers.

Studies on the liquidity volatility can be found in the literature, especially on foreign exchange bid-ask spreads. On the one hand, Glassman (1987) and Boothe (1988) have analyzed the statistical properties of the bid-ask spread, on the other hand, Bollerslev and Melvin (1994) evaluated the distribution of bid-ask spreads and their ability to explain foreign exchange rate volatility based on a tick-by-tick data set. Furthermore, many other researchers have focused their interest on the determinants of observed patterns in bid-ask spreads.

2.2.2 Estimation models

Since the second moment cannot be directly observed, it must be extracted from the available data-set. However, the best estimate approach which concerns high-frequency data is elaborated. In addition, the methods used to estimate the return volatility are not evenly suited to quantify spreads variation. One of the most applied approaches, of obtaining daily volatility, is to estimate it through the computation of historical volatility. However, historical volatility method suffers from the lag feature afflicting all historical window-based estimation methods. For instance, Amihud (2002) formulated an illiquidity measure, which has been used widely in the years that followed the sub-prime crisis. Amihud measure has a simple construction that uses the absolute value of the daily return-to-volume ratio to capture price impact. Second, the measure has a strong positive relation to expected stock return. Indeed, in the literature, the return premium associated with the widely used Amihud (2002) illiquidity measure

is generally considered a liquidity premium that compensates for price impact or transaction cost. Other used approaches to measure and model volatility are treated in the GARCH and stochastic volatility literature. However, both GARCH and stochastic volatility measures depend on some parametric models that are too sophisticated and, sometimes, even self-limiting. The multitude of parameters that must be estimated may cause some issues during the model selection and computation. Therefore, volatility estimates based on these models can be applied under several, restrictive, assumptions. The (implied) asset volatility can be estimated, also, on options written on the underlying asset. Still, this approach is used to profit on the asset's position but has a limited link to the mere liquidity concept. Liquidity cannot be traded directly; therefore, there are no option contracts written on liquidity.

2.2.3 Realized Measure

To overcome the drawbacks of the above mentioned approaches, Andersen, Bollerslev, Diebold, and Labys (2001c) introduced the concept of realized volatility. It is characterized with a framework that enables to estimate the volatility measure at daily, and even lower, frequency by using high-frequency intra-day data. It has been proved that Realized volatility is a good measure of the unobservable, true volatility. Under some conditions, it represents a model free and accurate volatility estimates. Moreover, it does not suffer from the lag issue that affects historical volatility because a *"precise estimation of diffusion volatility does not require a long calendar span of data. Rather, volatility can be estimated well from an arbitrarily short span of data, provided that returns are sampled sufficiently frequently"* (Andersen, Bollerslev, Diebold and Labys, 2000).

Due to the arrival of thick high-frequency data, the concept of the realized measure to liquidity changes allows to obtain valuable information related to high-frequency liquidity data.

2.2.4 High-frequency data flaws

High-frequency data are not free of issues that may impair variance estimates. Different market microstructure effects and implicit noise can show up as autocorrelated changes in the respective time series. These cases can cause realized volatility-type measures to be biased and inconsistent respect the true volatility measure. Usually, this manifestation exhibits in the form of intra-day seasonality.

It was seen that the unconditional volatility, based on the returns taken at short time intervals, is greater than the volatility that refers to larger time intervals. This volatility anomaly is caused by a non-zero auto-correlation of the returns, which enhances when the sampling frequency is increased. Precisely, a positive bias is a consequence of the in negative auto-correlation of the intra-day returns, and, vice-versa.

2.2.5 Other high frequency sizes

Here the other two high frequency measure will be briefly treated in order to have a complete view of the forthcoming study.

2.2.5.1 High frequency volume

Admati and Pfleiderer (1988) provide one of the first studies that investigate volume intraday patterns as being a standard liquidity measure. Trading volume, however, could be measured with different

dimensions. The most popular measurements of trading volume are: number of exchanged shares, share monetary value and number of share transactions. Those measures are, however, closely interrelated to each other. Total volume in a particular period is the sum of the sizes of individual trades. The main measure used in this work will be the latter, number of share transactions, that is easily obtained as

$$lVol_t = \log(Vol_t) = \log\left(\sum_{i=1}^{mT} NST_i\right), \quad (2.21)$$

where NST is the number of share traded intra-daily.

Although volume measures are the most common liquidity proxy used in stock exchanges and in literature describing inter-day and intra-daily patterns, alone it fails to sufficiently reflect the market impact of price movements and variant sizes of different trades because it considers small and large size trades as having the same effect (Ranaldo, 2000). Other studies claim that information contained in trading frequency is higher than in TS or NT.²⁸ Therefore, other liquidity proxies are employed in order to capture the different liquidity dimensions (Alabed and Al-Khour, 2009).

2.2.5.2 High frequency relative bid-ask spread realized variation

To complete the analysis, the relative intra-day bid-ask spread, which is the high frequency counterpart of the one already treated in 2.1.2, is used to study illiquidity patterns

To operate with high-frequency data, a rearrangement must be made. Given m intra-day spread observations per day for T number of days, time series consists of $m \times T$ quotes for the data set. To let the example be practical, the variation will be assessed within the day. The relative bid-ask spread over $[t - \frac{1}{m}, t]$ is given by

$$\Delta BA(m)(t) = BA(t) - BA(t - \frac{1}{m}), \quad \text{for } t = \frac{1}{m}, \frac{2}{m}, \dots, T. \quad (2.22)$$

The realized measure of the relative spread gives information about the liquidity aspect of the analyzed object. This is obtained by summing the squares of the instantaneous changes in the relative bid-ask spreads, as for the RV. However, while in terms of realized volatility, changes are expressed as logarithmic or percentage returns, this measure has absolute changes in the dimensionless relative bid-ask spreads. The time- t h -periodic realized liquidity variation is dimensionless and is given by

$$lRBA_h(t; m) = \sum_{i=1}^{mh} \Delta BA_{(m)}^2(t - h + \frac{i}{m}), \quad (2.23)$$

for $t = h, 2h, \dots, T$. Regardless of the sampling frequency, realized liquidity variation is directly observable, which is not true for its theoretical counterpart, the quadratic variation. As for the RV, an appropriate sampling frequency must be chosen when dealing with high-frequency observations. Indeed for sufficiently large m , the realized liquidity variation converges in probability towards the quadratic variation of the relative bid-ask spreads.

Chapter 3

Realized Volatility theory

The notion of realized variance (realized volatility) can be tracked down by Andersen, Bollerslev, Diebold and Labys. This measure is typically discussed in a continuous time framework where logarithmic prices are characterized by a semi-martingale. More restrictive specifications have been considered by Barndorff-Nielsen and Shephard (2001). According to their assumptions, the quadratic variation (QV) of the return process can be consistently estimated as the sum of squared intra-period returns. When some studies talk about realized variance it is associated the notion of QV. Importantly, Andersen et al. show that QV is the crucial determinant of the conditional return (co-) variance thereby establishing the relevance of the realized variance measure. If the conditional mean of the return process is deterministic, or function of variables contained in the information set, the QV is equal to the conditional return variance, which can be estimated consistently as the sum of squared returns. However, this exposition impose a random intra-period evolution of the instantaneous mean. But, Andersen et al. showed that these effects are likely to be trivial in magnitude and that the QV can be considered as the main determinant of the conditional return variance.

Through the academia the link that interconnects this measure and illiquidity has got some attention. Not surprisingly, the wider this is more illiquid is an instrument during the event window (Engler and Jeleskovic, 2016, and Amiram, Cserna and Levy, 2016).

This chapter focus on the exposure of the Realized Variance and the market microstructure patterns that affect HFT. Lastly, it will be discussed the role that these have during the liquidity measurement and how these will be used for forward tests.

3.1 Quadratic Variation

Lets consider an n -dimensional price process defined on a complete probability space, $(\Omega, \mathcal{F}, \mathbb{P})$, established in continuous time over the interval $[0, T]$, where T denotes a positive integer. We consider an information filtration, i.e., an increasing family of σ -fields, $(\mathcal{F}_t)_{t \in [0, T]} \subseteq \mathcal{F}$, which satisfies the usual conditions of P-completeness and right continuity. Finally, we assume that the asset prices through time t , including the relevant state variables, are included in the information set \mathcal{F}_t .

Under the standard assumptions, that the return process does not allow any arbitrage position and has a finite instantaneous mean, the asset price process can be categorized into the class of special semi-martingales. The stochastic integration theory states that processes like these can have a unique canonical decomposition. Indeed, the log-price process can be decomposed into $p = (p(t))_{t \in [0, T]}$.

For any n -dimensional arbitrage-free vector price process with finite mean, the logarithmic vector price process, p , may be written as the sum of a finite variation and predictable mean component, $A = (A_1, \dots, A_n)$, and a local martingale, $M = (M_1, \dots, M_n)$. These can be decomposed in their turn into a continuous sample-path and jump process like:

$$\begin{aligned} p(t) &= p(0) + A(t) + M(t) = \\ &= p(0) + A^c(t) + \Delta A(t) + M^c(t) + \Delta M(t), \end{aligned} \quad (3.1)$$

where A^c and ΔA are respectively the continuous and pure jump processes of the finite-variation predictable component, while M^c and ΔM are respectively the continuous sample-path and compensated jump process; by definition it is known that $M(0) = A(0) \equiv 0$.

By denoting the (continuously compounded) return over the interval $[t-h, t]$ by $r(t, h) = p(t) - p(t-h)$. The cumulative return process from $t=0$ onward, $r = (r(t))_{t \in [0, T]}$, can be derived as $r(t) \equiv r(t, t) = p(t) - p(0) = A(t) + M(t)$. Obviously, $r(t)$ inherits all the main properties of $p(t)$ and may be decomposed into the predictable and integrable mean component, A , and the local martingale, M . The predictability of A allows for some properties to still hold in the (instantaneous) mean process (i.e. it may evolve stochastically and display jumps). Nonetheless, the continuous component of the mean return must have smooth sample paths compared to those of a non-constant continuous martingale, such as a Brownian motion, and any jump in the mean must be accompanied by a corresponding predictable jump in the compensated jump martingale, ΔM . Consequently, there are two types of jumps in the return process:

- i. Predictable jumps where $\Delta A(t) \neq 0$;
- ii. Purely unanticipated jumps where $\Delta A(t) = 0$ but $\Delta M(t) \neq 0$.

The first type of jump comes in place when new information is released according to a predetermined schedule, such as macroeconomic news releases or company earnings reports. Instead, the latter jump event usually occurs when unanticipated news hit the market. However, any uncertainty about the precise timing of the news nulls the predictability assumption and removes the jump in the mean process. If there are no perfectly anticipated news releases, the predictable, finite variation mean return, A , may still evolve stochastically, but it will have continuous sample paths.

Because the return process is a semi-martingale it has an associated quadratic variation process. For any n -dimensional arbitrage-free price process with finite mean, the quadratic variation $n \times n$ matrix process of the associated return process, $[r, r] = \{[r, r]_t\}_{t \in [0, T]}$, is well-defined. The i^{th} diagonal element is called the quadratic variation process of the i^{th} asset return while the ij^{th} off-diagonal element, $[r_i, r_j]$, is called the quadratic covariation process between asset returns i and j . The covariance measure follow the process:

$$Cov[r(t+h, t) | \mathcal{F}_t] = E([r, r]_{t+h} - [r, r]_t | \mathcal{F}_t) + \Gamma_A(t+h, h) + \Gamma_{AM}(t+h, h) + \Gamma'_{AM}(t+h, h), \quad (3.2)$$

where $\Gamma_A(t+h, t) = Cov(A(t+h) - A(t) | \mathcal{F}_t)$ and $\Gamma_{AM}(t+h, h) = E(A(t+h)[M(t+h) - M(t)]' | \mathcal{F}_t)$. If the finite variation component, A , in the canonical return decomposition is continuous, then it has the following property

$$[r_i, r_j] = [M_i, M_j] = [M_i^c, M_j^c] + \sum_{0 \leq s \leq t} \Delta M_i(s) \Delta M_j(s). \quad (3.3)$$

This statement explains that the quadratic variation of continuous finite-variation processes is zero. If this holds, then the mean component becomes irrelevant for the quadratic variation. Moreover, jump components only contribute to the quadratic covariation if there are simultaneous jumps in the price path for the i^{th} and j^{th} asset, whereas the squared jump size contributes directly to the quadratic variation. The quadratic variation process measures the realized sample-path variation of the squared return processes. Under this property, this variation is caused by the innovations to the return process. So, the quadratic covariation builds a unique and invariant ex-post realized volatility measure that is essentially model free.

Despite this property the quadratic variation has another useful property that helps it to deal with high-frequency data.

Lets consider an increasing sequence of random partitions of $[0, T]$, with $0 = \tau_{m,0} \leq \tau_{m,1} \leq \dots$, and so on, and such that $\sup_{j>1}(\tau_{m,j+1} - \tau_{m,j}) \rightarrow 0$ and $\sup_{j>1}\tau_{m,j} \rightarrow T$ for $m \rightarrow \infty$ with $p = 1$, we have that:

$$\lim_{m \rightarrow \infty} \left\{ \sum_{j>1} [r(t \wedge \tau_{m,j}) - r(t \wedge \tau_{m,j-1})][r(t \wedge \tau_{m,j}) - r(t \wedge \tau_{m,j-1})]' \right\} \rightarrow [r, r]_t, \quad (3.4)$$

where $t \wedge \tau \equiv \min(t, \tau)$, with $t \in [0, T]$.

This suggest that it is possible to approximate the quadratic variation by summing the cross-products of high-frequency returns. This measure, obtained from high-frequency data, are denoted as realized volatility.

3.2 Return Realized Variance

After having exposes the continuous time model, in this section it will be explained the discrete time one, and how to estimate efficiently the return realized volatility.

Let $S_{t,j}$, for $j = 1, \dots, N$, represent the price of a security in the j^{th} intra-day moment, fraction of day t , by assuming that observations are equally-time spaced ¹, the daily log return is defined as:

$$r_t = \ln S_{t,N} - \ln S_{t-1,N}, \quad (3.5)$$

for $t = 1, \dots, T$. By sampling with a frequency f , then $N_f = \frac{N}{f}$ intra-day returns can be obtained as:

$$r_{f,t,i} = \ln S_{t,i,f} - \ln S_{t,(i-1),f}, \quad (3.6)$$

for $i = 1, \dots, N_f$ and $S_{t,0} = S_{t-1,N}$. It is important to notice that this model relies also on a continuous time model, but, as above mentioned, here it will be treated in its discrete aspect.

The first two moments of the return are assumed to exist, and can be denoted as $\mu_t = E[r_t | \mathcal{F}_{t-1}]$ and $\sigma_t^2 = \text{var}[r_t | \mathcal{F}_{t-1}]$ where \mathcal{F}_t denotes the information set available at the beginning of the day $-t$, which includes all the previous information.

The information set contains many observations that relate to the prices, quotes, trading volume and

¹Generalization of irregularly time spaced returns.

market depth. It is also assumed that the log (excess) return of the security follows:

$$r_t = \sigma_t \epsilon_t, \quad (3.7)$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ and σ_t^2 is the return variance of the day- t . By the additive property of returns, the t^{th} day return is given by:

$$r_t = \sum_{i=1}^{N_f} r_{t,f,i}. \quad (3.8)$$

To give a clue on the working of these measure here is summarized its behavior. Let the return process can be written as a standard Ito process like:

$$dr(t) = \mu(t)dt + \sigma(t)dW(t), \quad (3.9)$$

where $\mu(t)$ is the process drift, $\sigma(t)$ is the spot volatility and $W(t)$ is a standard Brownian motion. This model can be denoted also as a stochastic volatility (SV) model. This representation follow a special semi-martingale model, which relies on useful properties. One of these states that, when $\mu(t)$ and $\sigma(t)$ are jointly independent from $W(t)$, then:

$$[r_t | \mu_t, IV_t] \sim \mathcal{N}(\mu_t, IV_t), \quad (3.10)$$

where

$$\mu_t = \int_{t-1}^t \mu(s)ds \quad \text{and} \quad IV_t = \int_{t-1}^t \sigma(s)ds$$

where IV_t denotes the integrated variance, which is the return variance.

The quadratic variation of a stochastic process over the interval $[t-1, t]$ is equal to:

$$QV_t \equiv \int_0^t \sigma_s^2 ds \equiv \text{plim}_{N \rightarrow \infty} \sum_{j=1}^N [S_{t,j} - S_{t,j-1}]^2. \quad (3.11)$$

As previously mentioned, the empirical equivalent measure of QV_t is the realized variance, which is defined as the sum of the squared intra-period returns, with frequency f ,

$$RV_t \equiv \sum_{j=1}^{N_f} r_{t,f,j}^2. \quad (3.12)$$

Since returns are serially uncorrelated at any frequency f , it follows that:

$$\text{var}[r_t | \mathcal{F}_t] = E[r_t^2 | \mathcal{F}_t] = E \left[\sum_{i=1}^{N_f} r_{t,f,i}^2 | \mathcal{F}_t \right]. \quad (3.13)$$

So the Realized Volatility is an *unbiased estimator of the conditional return variance*.

The conditional return variance over a fixed period can thus be estimated by summing the squared intra-period returns sampled at increasingly high frequency. While this result does not depend on the choice of period (i.e. one day), it does crucially rely on the property that returns are serially uncorrelated at any sampling frequency. This suggests that the use of high frequency returns can reduce the measurement error in the return variance estimates, provided that the return series is a martingale

difference sequence.

The RV_t can be related to the conditional variance only if, for instance, $\{\mu(s), \sigma(s)\}_{s=t-1}^t$ is \mathcal{F}_{t-1} -measurable, then $IV_t = \sigma_t^2$. If there is no microstructure noise, or measurement error, the realized volatility can be approximated as the second uncentered sample moment of the return process. As shown by Barndorff-Nielsen and Shepard (2002), if there are no price jumps and microstructure noise is not impounded in the process, the realized volatility is a consistent non-parametric measure of the Notional Volatility:

$$RV^{(n)} \xrightarrow{p} IV \quad \text{as } n \rightarrow \infty. \quad (3.14)$$

3.3 Market Microstructure

When market liquidity is discussed in market microstructure theory, it is often the case that more practical concepts are introduced, such as the cost of changing positions (tightness), the trade size or thickness of the order book-profile (order book refers to a panel which provides traders with bid-ask prices and volume offered per price) required for changing prices (market depth), and the required period of time to recover from price fluctuation caused by a sudden shock or to reach a new equilibrium (market resiliency).

The process by which market microstructure affects price discovery in a market and market liquidity via changes in market participants' behavior is composed of two stages: *i*) when market participants hold potential trade needs, based on their individual reasons to when they actually decide to place orders in the market, and *ii*) when the stage in which such orders are accumulated in the market and trades are executed.

Muranaga and Shimizu (1999) found that when many investors followed short-term price movements an increase in the number of trades and a decrease in the volume of accumulated order flows showed up. When market participants become more risk-averse on average, market liquidity decreases. A sudden fall in market liquidity results when market participants lose confidence in their expectations on future prices, due to their change of investors' sensitivity on market information.

Real world high-frequency data can differ quite substantially from the processes assumed in the formal setting of realized liquidity variation. Although the use of high-frequency data certainly offers insights that low-frequency data cannot reveal, by dealing with loads of data that arise "continuously", microstructure analysis can become challenging if not applied correctly to real world data. A wide range of effects, mainly belonging to the market microstructure category, may cause a gap between the observed bid and ask price process from the true bid and ask price processes, which cannot be observed directly. The observed high-frequency bid-ask spread process is likely to not follow a semi-martingale process. Therefore, realized liquidity variation can be a biased and inconsistent estimator of the integrated volatility of bid-ask spreads when the estimation procedure does not consider observation errors. It has been found that, when choosing a small sample frequency, the realized volatility estimator fails to hold its robustness' properties, and the bias of the estimate increases as the sample frequency narrows (Andersen, Bollerslev, and Meddahi, 2005; and Andersen, Bollerslev, and Meddahi, 2011).

Concerning security returns, the bid-ask bounce that runs within security prices has been found to be a significant source of market microstructure noise. In the early studies, Roll (1984) and Blume and Stambaugh (1983) described the negative first-order autocovariance in observed price changes, which is caused by the existence of bid and ask prices. However, by studying bid-ask spreads, the bid-ask

bounce itself cannot be a possible source of noise.

Bid and ask prices discreteness can cause significant disturbances. Particularly, foreign exchange rates are usually quoted out to four digits after the decimal place. This rounding process can present some rounding errors. In the stock market, even transaction happens even at small frequencies, this issue seems to be not relevant.

Another factor that can distort realized volatility is the intra-day seasonality. Intra-day seasonality is a well-documented phenomenon and appears in many other quantities of interest such as foreign exchange mid price returns, volatility, trading volume or market and limit order arrivals. Research has found a link between the asset pricing and, both, human activity and the arrival of new information, the latter is usually associated with private information. Admati and Pfleiderer (1988) argue that high volume during peculiar periods occurs when, due to asymmetric information, noise traders smooth the activities of informed traders, which rise the volatility. In contrast, Brock and Kleidon (1992) hold that trading halts and different trading strategies at the open and close of markets are accountable for patterns in liquidity, which they measure using trading volume as a proxy.

In the high frequency world, the usual mean of all actions is the same for all traders, since these still have to trade to profit from information, and other traders still try to learn what they know by analyzing market data. However, market data are not the same. Algorithmic trading means that trades are not the basic unit of market information, but the underlying orders are the characters of these information. Adverse selection is problematic because the real information that move the prices is not always focused.

Microstructure models have been vague, portraying private information as a signal of the underlying asset's true value. In the high frequency world, it is uncertain which information relates to the fundamental trading information. This problem has risen since the time dimension that affects high speed trading also affects market makers movements. Whereas the time horizon of the NYSE specialist was at one point measured in weeks (Hasbrouck and Sofianos, 1993), currently it is measured in seconds, even in milliseconds in some cases. During these narrow time frames, information might be asset-related but also order-related. Haldane (2011) states that now being informed relates, mostly, on the speed which a trade is executed (more a party is informed, faster the trader executes the negotiation). Informed trading is multidimensional in that traders can know more about the asset or about the market (or markets) or even about their own order flow and use this information to take advantage of liquidity providers. Even large traders who know nothing special about the asset's value can impact market makers' position since they know more about their own trading plans. Trade imbalances are problematic for market makers because they are always on the other side: buying if traders are selling and selling if they are buying. Trading that is heavily skewed to buys or sells is can lead market makers to withdraw from trading as their inventory or short positions reach preset parameters. Over the short time intervals of interest to market makers, even these classically uninformed traders are informed traders in the new high frequency world.

These definitions of informed trading are worrisome. Now, it is not clear what drives the adjustment of prices or, more to the point, where they are going. Analyses of market efficiency suggest that markets generally remain informationally efficient, which should allay at least some concerns for asset pricing researchers. But momentary instability, presently, can affect markets sporadically. Markets are more inter-connected, tied together by market making/statistical arbitrage that operates across, not just within, markets. These characteristics suggest that liquidity factors play an increased role in asset

pricing. What these liquidity factors capture, and how to even measure them, is problematic.

In a friction-less market, microstructure does not matter and profits come solely from longer term views on the market.

Transaction costs can be added, leading to the clearing equation with transaction costs:

$$\delta_n W = L_n \delta_n p \pm \frac{s_n}{2} |\delta_n L|, \quad (3.15)$$

where W is a trader's wealth before n , L is his net position in the traded asset and δ_n represents the difference of the successive moment of the target variable and the actual one (i.e. $\delta_n L = L_{n+1} - L_n$), s_n is the bid-ask spread at time n , p the mid-range price before the execution of the trade and the \pm is "+" if the trade is made with limit order and "-" if made with market orders. This equation, even if considers the spread impact, ignores the price impact. By ignoring such component this is the characterization of low frequency traders, who do not optimize their execution and use just long-term views to trade. By incorporating the price impact, high frequency traders can model their views by:

$$\delta_n W = L_n \delta_n p \pm \frac{s_n}{2} |\delta_n L| + \delta_n L \delta_n p \quad (3.16)$$

where this expression represents the notion of microstructure noise, since wealth can be tracked by using measurable market quantities.

3.3.1 High frequency realized variance

In the absence of noise the RV, $[r, r]_T$, consistently estimates the quadratic variation σ_t^2 . The sum converges to the integral, with a known distribution, dating back to Jacod (1994) and Jacod and Protter (1998).

Lets consider a fixed time period h (a trading day) and write the observed price process at the end of the i^{th} period as

$$\tilde{p}_{ih} = p_{ih} \xi_{ih}, \quad \text{for } i = 1, \dots, n, \quad (3.17)$$

where p_{ih} is the friction-less equilibrium price, (price that would prevail in the absence of market microstructure frictions) and $\xi_{i,h}$ denotes microstructure noise. By taking the difference with the previous value and applying logarithmic properties, the process would be

$$\ln(\tilde{p}_{ih}) - \ln(\tilde{p}_{(i-1)h}) = \ln(p_{ih}) - \ln(p_{(i-1)h}) + \vartheta_{ih} - \vartheta_{(i-1)h}, \quad (3.18)$$

where $\tilde{r}_{ih} = \ln(\tilde{p}_{ih}) - \ln(\tilde{p}_{(i-1)h})$, $r_{ih} = \ln(p_{ih}) - \ln(p_{(i-1)h})$ and $\vartheta_{ih} - \vartheta_{(i-1)h} = \ln(\xi_{ih}) - \ln(\xi_{(i-1)h})$. Each period can be divided in M sub-periods, and the observed high-frequency continuously compounded return is given by

$$\tilde{r}_{j,i} = \ln(\tilde{p}_{(i-1)h+jh\Delta}) - \ln(\tilde{p}_{(i-1)h+(j-1)h\Delta}), \quad \text{for } j = 1, \dots, M, \quad (3.19)$$

where $\Delta = \frac{1}{M}$ is the horizon over which the continuously-compounded returns are computed. Hence, $\tilde{r}_{j,i}$ is the j^{th} intra-period return over the i^{th} period. So

$$\tilde{r}_{j,i} = r_{j,i} + \varepsilon_{j,i}, \quad (3.20)$$

where $\varepsilon_{j,i} = \vartheta_{(i-1)h+jh\Delta} - \vartheta_{(i-1)h+(j-1)h\Delta}$ is the microstructure noise that affect the data. However, both $r_{j,i}$ (equilibrium return) and $\varepsilon_{j,i}$ are unobservable.

To simplify the exposure, it will be treated just the case in which $i = 1$. As Bandi and Russel (2003) outlined, under the previous setting, the realized variance of the observed return is made up of three components

$$\hat{RV} = \sum_{j=1}^M r_j^2 + \sum_{j=1}^M \varepsilon_j^2 + 2 \sum_{j=1}^M r_j \varepsilon_j. \quad (3.21)$$

If the true price process was directly observable, only the first term ($\sum_{j=1}^M r_j^2$) would lead the determination of \hat{RV} . However, microstructure noise introduces also two terms. The first added term ($\sum_{j=1}^M \varepsilon_j^2$) diverges to infinity almost surely as the number of observations increases asymptotically (or, equivalently, as the sampling frequency increases in the limit) since more and more noise is being accumulated for a fixed period of time h .

By selecting Δ as small as possible (i.e., M as large as possible) is optimal. But ignoring market microstructure noise leads to an even more dangerous situation than assuming constant volatility and $T \rightarrow \infty$. After suitable scaling, RV based on the observed log-returns is a consistent and asymptotically normal estimator, but of the quantity $2ME[\varepsilon^2]$ rather than σ^2 . Said differently, in the high frequency limit, market microstructure noise totally swamps the variance of the price signal at the level of the realized volatility. Since the expressions above are exact small-sample ones, they can, in particular, be specialized to analyze the samples at increasingly higher frequency ($\Delta \rightarrow 0$, say, sampled every minute) over a fixed time period (T fixed, say, a day).

In the absence of market microstructure noise (i.e. $\varepsilon_j = 0, \forall j$), the estimation error between the realized variance estimator and integrated variance converges weakly to a mean-zero mixed Gaussian distribution, with speed \sqrt{M} .

$$\frac{\sqrt{M}}{h}(\hat{RV} - RV) \sim MN(0, 2Q), \quad \text{as } M \rightarrow \infty,$$

where $Q = \int_0^h \sigma_s^4 ds$.

When microstructure noise plays a role, the realized variance estimator does not consistently estimate the integrated variance over any given period. Intuitively, the summing of an increasing number of contaminated return data entails infinite accumulation of noise as the frequency increases. Specifically, while the first sum term converges to the integrated variance over the period, the second sum term diverges to infinity, almost surely, while the third sum is stochastically dominated by the second.

$$\hat{RV} \xrightarrow{a.s.} \infty, \quad \text{as } M \rightarrow \infty,$$

The limiting result in (ii) is an asymptotic approximation suggesting that for large M , as is the case for high-frequency data, the researcher must be wary of microstructure contamination as the effect of the noise can be substantial. Hence, any statement about the informational content of the conventional realized variance estimator as a measurement of the integrated variance of the underlying logarithmic price process ought to be a finite sample statement.

Nevertheless, sample moments of the observed return series can be used to learn about population moments of the unobserved noise returns at high frequencies. Indeed, if the noise variance has a finite fourth moment, $E(\varepsilon^8) < \infty$, then

$$\frac{1}{M} \sum_{j=1}^M \hat{r}_j^q \xrightarrow{P} E(\vartheta^q),$$

for large M , and any fixed period h , moments of the unobserved noise process can be estimated at any frequency, by using data sampled at the highest frequency.

With $N = \frac{T}{\Delta}$, it has been found that the variance of the noise is essentially unrelated to σ^2 . It has long been known that sampling at very high frequencies, 1, 5, 10 secs, is not a good idea. The recommendation in the literature has then been to sample sparsely at some lower frequency, by using a realized volatility estimator constructed by summing squared log-returns at some lower frequency: 5, 10, 15, 30 minutes, (Andersen et al 2001; Barndorff-Nielsen and Shephard, 2002; Gençay et al., 2002). Reducing the value of n , from 23400 (1 second sampling) to 78 (5 min sampling over the same trading day), has the advantage of reducing the magnitude of the bias term $2ME[\epsilon^2]$ (Aït-Sahalia and Yu, 2009).

Chapter 4

Stationary and Integrated process

This section introduces econometric methods used in analysis of volatility and liquidity time series from the previous section. Firstly, unit root and stationarity tests will be outlined, since time series stationarity is required before testing more complex patterns. Then the concept of cointegration is exposed, since it is a useful tool to apply if time series are found to be non-stationary. Both Engle-Granger and Johansen methods are presented, because the former is suitable and easier to apply in this analysis, while the latter is used mainly as an input into further, and articulated, econometric methods.

4.1 Stationary process

Many time series econometric methods require that the studied series to be stationary, if its joint probability distribution does not change over time. Precisely, let $(y_{t1}, y_{t2}, \dots, y_{tk})$ have the same joint probability distribution function as $(y_{t1+\tau}, y_{t2+\tau}, \dots, y_{tk+\tau})$, $\forall k, \tau$ such that

$$F(y_{t1}, y_{t2}, \dots, y_{tk}) = F(y_{t1+\tau}, y_{t2+\tau}, \dots, y_{tk+\tau}). \quad (4.1)$$

This formulation implies strong assumption on the moments of the process, i.e. it requires all moments to be constant, which are difficult to satisfy, but weaker form, covariance stationarity, is usually satisfactory. To overcome this limit, the concept of covariance stationary suits as well, i.e. first and second moments do not change over time. Even though it is a weaker assumption than the previous one, it is a good connotation since it implies that stochastic process' moment should be: $E(y_t) = \mu$ and $Cov(y_t, y_{t+h}) = \gamma_h$. The first moment is assumed constant as for the second condition as well. The latter implies that variance is constant since it relies only on $h \geq 0, \forall t$.

It is said that weakly dependent stationary process is integrated of order zero $I(0)$, while nonstationary process, which can be transformed into stationary process by taking its first difference, is said to be integrated of order one $I(1)$. More generally, a stochastic process is said to be integrated of order d , $I(d)$, if by differencing the process d -times yields a stationary process.

4.1.1 Unit-root tests

Unit root and stationarity tests have to be performed at the beginning of any time series analysis, because if time series are not stationary (or at least covariance stationary), inference from econometric

analysis is generally incorrect, mainly because of inflated t-statistics (and therefore lower p-values) and high R^2 . There are many root tests but Augmented Dickey-Fuller (ADF) test is the one that is used in this work. Since it is ease computing it gives useful results for this kind of tests.

4.1.1.1 Augmented Dickey-Fuller

Original unit root Dickey-Fuller test was developed by Dickey and Fuller (1979), the augmented version of their test is used more often nowadays, since it allows to test larger and more complex time series models. The testing procedure is the same as the "classical" Dickey-Fuller test but applied to a different pattern of an autoregressive process. Let y_t be an autoregressive process of order p , $AR(p)$

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t, \quad (4.2)$$

An $AR(p)$ process has unit root if $\sum_{i=1}^p \alpha_i = 1$. A transformation of equation 4.2 can be easily made by subtracting y_{t-1} and a scaling procedure, which is used for ADF test of order p

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{p-1} \gamma \Delta y_{t-i} + \epsilon_t, \quad (4.3)$$

ADF test statistics is trivially obtained as a "classical" t-statistics of coefficient β , i.e. $\frac{\beta}{SE(\beta)}$. However, this statistic does not have neither a t-distribution nor asymptotically standard normal distribution. This is the reason why critical values have to be obtained by simulations. If null hypothesis $H_0 : \beta = 0$ is rejected in favor of the alternative $H_1 : \beta < 0$, it means that the time series does not have unit root, while if H_0 holds, is not rejected, the time series has unit root, or there is not enough evidence that the time series does not have unit root.

4.1.1.2 Variance Ratio

Testing the presence of a process with an unit root is the counterpart of testing the random walk (RW) hypothesis, which provides a mean to test the weak-form efficiency of financial markets (Fama, 1970; 1991). This helps to measure the long-run effects of shocks on the path of real output in macroeconomics (Campbell and Mankiw, 1987; Cochrane, 1988; Cogley, 1990). Indeed, a RW is a trivial example of a non-stationary process.

Given a time series y_t , the process follows a RW if the autoregressive coefficient is equal to one, $\psi = 1$

$$y_t = \mu + \psi y_{t-1} + \epsilon_t \quad (4.4)$$

where μ is an unknown drift parameter and the error terms ϵ_t are usually neither independent nor identically distributed (i.i.d.).

Many statistical tests were designed to test the RW but a class of test, based on the variance-ratio (VR) methodology (see, e.g., Campbell and Mankiw, 1987; Cochrane, 1988; Lo and MacKinlay, 1988; Poterba and Summers, 1988; Charles and Darné, 2009). The VR methodology consists of testing the RW against stationary alternatives, by exploiting the fact that the variance of random walk increments is linear in all sampling intervals, i.e., the sample variance of k -period return (or k -period differences), $y_t - y_{t-k}$, of the time series y_t , is k times the sample variance of one-period return (or the first difference), $y_t - y_{t-1}$.

The VR at lag k is then defined as the ratio between $(\frac{1}{k})^{\text{th}}$ of the k -period return (or the k^{th} difference) to the variance of the one-period return (or the first difference). A k -period VR is then computed as

$$VR_k \equiv \frac{1}{k} \frac{\text{Var}(r_{t+k})}{\text{Var}(r_{t+1})} \quad (4.5)$$

Hence, for a RW process, the variance computed at each individual lag interval k ($k = 2, 3, \dots$) should be equal to unity. If the VR is larger than 1, the price series shows a tendency to form trends, i.e. changes in one direction are more often followed by changes in the same direction. Instead, if the VR is less than 1, the price series shows some degree of mean reversion, which is the equivalent of a stationary process. Changes in one direction are more often followed by changes in the opposite direction.

4.2 Cointegration

Once variables have been classified as integrated of order 0,1, etc. is possible to set up models that lead to stationary relations among the variables, and where standard inference is possible. The necessary criteria for stationarity among non-stationary variables is called cointegration. Testing for cointegration is necessary step to check if your modelling empirically meaningful relationships. If variables have different trends processes, they cannot stay in fixed long-run relation to each other, implying that the long-run models cannot be forecasted, and there is usually no valid base for inference based on standard distributions. If any cointegration has been found, then it is necessary to continue to work with variable differences.

Standard econometric analysis cannot be used whenever two or more time series are integrated. Moreover, any linear combination of series which are integrated of the same order is generally integrated and has the same order of integration as individual series (Engle and Granger 1987). Namely, if $x_t \sim I(1)$ and $y_t \sim I(1)$, usually also $x_t + \beta y_t \sim I(1)$. However, it is possible that both processes are driven by the same deterministic or stochastic trend, integration and cointegration are concepts that allow to some variable to be stationary. Integrated variables, identified by unit root and stationarity tests, can be differenced in order to obtain stationarity. Cointegrated variables, identified by cointegration tests, can be combined to form new, stationary variables. In other words they evolve in a similar way. In such a case, linear combination $x_t + \beta y_t$ is integrated of a lower order than individual time series, i.e. the combination is $I(0)$, and so x_t and y_t are said to be cointegrated.

In the presence of cointegration, simple differencing is a model miss-specification, since long-term information appears in the levels. Since, the cointegrated VAR model provides intermediate options, between differences and levels, by mixing them together with the cointegrating relations. So, all terms of the cointegrated VAR model are stationary, unit roots problems are erased in the process.

Cointegration modeling is often adopted in the economic theory. Some variables that are described with a cointegrated vector autoregressive (VAR) model are:

- i. Money stock, interest rates, income, and prices ;
- ii. Spot and forward currency exchange rates and interest rates;
- iii. Interest rates and inflation.

These macroeconomic-models must consider the possibility of structural changes in the underlying data-generating process during the sample period.

However, Financial data, as mentioned, is available at high intra-day frequency .The Law of One Price suggests cointegration, otherwise some arbitrage opportunities may arise, among the following groups of variables:

- i. Bid and ask prices;
- ii. Prices of assets with identical cash flows;
- iii. Prices of assets and dividends;
- iv. Spot, future, and forward prices.

Financial markets are characterized by "price multiplicity". In particular, different investors can provide different valuations and attach different prices to the same asset. Also, different market venues can be available for the same asset. Following Hasbrouck (2002), we can guess a statistical model for the joint behavior of two prices for the same asset linked together by a no-arbitrage or equilibrium relationship. Basically, the two prices incorporate a single long-term component that takes the form of a cointegrating relation. Cointegration involves restrictions stronger than those implied by correlation. Two stock prices can be positively correlated but not cointegrated. If stock A is cointegrated with stock B, there exists an arbitrage relationship that ties together the two stocks. In addition, the ask and bid quotes for stock A are also cointegrated. The reason is that the difference between the quotes can often be characterized as a stationary variable, meaning that it cannot explode in an unbounded way. The price of a stock on two different exchanges can be different at any point in time, but it is natural to assume that this difference reverts to its mean over time. There are several cointegration tests that can be used. The Johansen test is the most fundamental and used. Engle and Granger (1987) developed a first cointegration test, based on common stochastic trends. This test is also know as the two-step procedure test.

4.2.1 Engle and Granger test

To start the test, it takes to consider the first step by considering a cointegrating regression,

$$x_{1,t} = \beta_1 + \sum_{i=2}^p \beta_i x_{i,t} + u_t, \quad (4.6)$$

where p is the number of variables in the equation. In this regression we assume that all variables are $I(1)$ and might cointegrate to form a stationary relationship, and thus a stationary residual term $\hat{u}_t = x_{1,t} - \beta_1 - \sum_{i=2}^p \beta_i x_{i,t}$. This equation represents the assumed economically meaningful (or understandable) steady state or equilibrium relationship among the variables. If the variables are cointegrated, they will share a common trend and form a stationary relationship in the long run. Furthermore, under cointegration the estimated parameters can be seen as the correct estimates of the long-run steady state parameters, and the residual (lagged once) can be used as an error correction term in an error correction model.

The second step is to test for a unit root in the residual process of the cointegrating regression above.

To test for unit roots, the ADF test is employed to run

$$\Delta \hat{u}_t = \alpha + \pi \hat{u}_{t-1} + \sum_{i=1}^k \gamma_i \hat{u}_{t-i} + v_t, \quad (4.7)$$

where the constant term α can be often left out to improve the efficiency of the estimate. Under the null hypothesis of no cointegration, the estimated residual is $I(1)$ because $x_{1,t}$ is $I(1)$, and all parameters are zero in the long run. The empirical t-distribution is not identical to the Dickey-Fuller one, although the tests are similar. The reason is that the unit root test is applied to a derived variable, the estimated residual from a regression. Thus, new critical values must be tabulated through simulation. The null hypothesis is still to check that the process has no cointegration, while the alternative hypothesis is that the equation is a cointegrating equation, meaning that the integrated variable $x_{1,t}$ cointegrates at least with one of the variables on the right hand side. A significant π value would imply co-integration. If the dependent variable is integrated with $d > 0$, and at least one variable is also integrated of the same order, cointegration leads to a stationary $I(0)$ residual. But, the test does not tell us if $x_{1,t}$ is cointegrating with all, some or only one of the variables on the right hand side. Lack of cointegration means that the residual has the same stochastic trend as the dependent variable. The integrated properties of the dependent variable will if there is no cointegration pass through the equation to the residual. The test statistics for $H_0 : \pi = 0$ (no co-integration) against $H_1 : \pi < 0$ (co-integration), changes with the number of variables in the co-integrating equation, and in a limited sample also with the number of lags in the augmentation ($k > 0$).

Asymptotically, the test is independent of which variable occurs on the left hand side of the cointegrating regression. By choosing one variable on the left hand side the cointegrating vector are said to be normalized around that variable, implicitly it is assumed that the normalization corresponds to some long-run economic meaningful relationship. But, this is not always correct in limited samples, there are evidence that normalization matters (Ng and Perron 1995). If the variables in the cointegrating vectors have large differences in variances, some might be near integrated (having a large negative $MA(1)$ component), such factors might affect the outcome of the cointegration test. If testing in different ways give different conclusions it is necessary to use more complex tests on cointegration.

There are three main problems with the two-step procedure. First, since the tests involves an ADF test in the second step, all problems of ADF tests are valid here as well, especially choosing the number of lags in the augmentation. Second, the test is based on the assumption of one cointegrating vector, captured by the cointegrating regression. Thus, care must be taking when applying the test to models with more that two variables. If two variables cointegrate adding a third integrated variable to the model will not change the outcome of the test. If the third variable does not belong in the cointegrating vector, OLS estimation will simply put its parameter to zero, leaving the error process unchanged. Logical chains of bi-variate testing is often necessary (or sufficient) to get around this problem. Third, the test assumes a common factor in the dynamics of the system. To see why this is so, rewrite the simplest two-variable version of the test as,

$$\Delta u_t = \pi u_{t-1} + v_t = [x_{1,t-1} - \beta_2 x_{2,t-1}] + v_t. \quad (4.8)$$

If this restriction does not hold, the test should perform badly.

The advantage of the procedure is that it is easy, and relatively cost-less to apply compared with other

approaches. Especially for two variables it can work quite well, but it should be reminded that the common factor restriction is a severe restriction since all short-run dynamics are forced to the residual process.

4.2.2 Johansen Cointegration test

A more advanced test for cointegration is Johansen's test. This is a test which has many desirable statistical properties, but its weakness is that it relies on asymptotic properties, which let it to be sensitive to specification errors in limited samples. Lets consider a vector autoregressive (VAR) representation of order p of the variables. It is a p -dimensional process, integrated of order d , $\{y\}_t \sim I(d)$, with the VAR representation

$$y_t = \mu + A_1 y_{t-1} + \dots + A_p y_{t-p} + \epsilon_t, \quad (4.9)$$

where y_t is an $n \times 1$ $I(1)$ vector and ϵ_t is an $n \times 1$ vector of innovations. By using the difference operator $\Delta = 1 - L$, where L is the lag operator, the VAR in levels can be transformed to a vector error correction model (VECM) However, the use of the lag operator let to "lose" one lag, leading to a $k - 1$ lags in the VECM

$$\Delta y_t = \mu + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \epsilon_t, \quad (4.10)$$

where $\Pi = \sum_{i=1}^p A_i - I$ and $\Gamma_i = -\sum_{j=i+1}^{p-1} A_j$ are coefficients matrices.

The number of cointegrating vectors, are identical to the number of stationary relationships in Π . If there is no cointegration, all rows in Π must be filled with zeros. If there are stationary combinations, or stationary variables, in some parameters in the matrix will be non-zero. By a trivial mathematical property, the rank of Π determines the number of independent rows in the matrix, which are also the number of cointegrating vectors. The rank of Π is given by the number of significant eigenvalues found in its estimate ($\hat{\Pi}$). Each significant eigenvalue represents a stationary relation. Under the null hypothesis of $y_t \sim I(d)$, with $d > 1$, the test statistic for determining the significance of the eigenvalues is non-standard, and must be simulated.

If Π has reduced rank there are co-integrating relations among the y 's. $Rank(\Pi) = 0$, implies that all y 's are non-stationary. There is no combination of variables that leads to stationarity. The conclusion is that modeling should be done in first differences instead, since there are no stable linear combination of the variables in levels.

If Π has full rank, then all variables must be stationary.

If Π has reduced rank r , with $0 < r < p$, the cointegrating vectors are given as, $\Pi = \alpha \beta'$ where β_i represents the i^{th} co-integration vector, and α_j represents the effect of each co-integrating vector on the $\Delta y_{p,t}$ variables in the model. Once the rank of Π is determined and imposed on the model, the model will consist of stationary variables or expression, and estimated parameters follows standard distributions. Since Π has reduced rank r , then then matrices α and β have both rank r such that the coefficient matrix can be disentangled into $\Pi = \alpha \beta'$ and $\beta' y_t$ represents a stationary process. The number of cointegrating relationships are denoted in r , the adjustment parameters are denoted in α elements and each column of β is a cointegrating vector.

However, the reduced rank test of Π determines only the number of co-integrating vectors (r) and the number of common trends ($p - r$). The econometrician must identify through normalization and tests of the α and β matrices the cointegrating vector(s) so they represent meaningful economic relationships.

The test statistics varies depending on the inclusion of constants and trends in the model. The standard model can be said to include an unrestricted constant but no specific trend term. For a p -dimensional vector of variables (y_t) , since it has been proved that Π can be decomposed into $\alpha\beta'$, the estimated model results,

$$\Delta y_t = \mu + \alpha\beta'y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \Phi d_t + \epsilon_t, \quad (4.11)$$

where μ is an unrestricted constant term and $\epsilon_t \sim \mathcal{N}_p(0, \Sigma)$. The number of cointegrating relationships are denoted by r , the adjustment parameters are denoted in α elements and each column of β is a cointegrating vector. It can be shown that for a given r , the maximum likelihood estimator of β defines the combination of y_{t-1} that yields the r largest canonical correlations of Δy_t with y_{t-1} after correcting for lagged differences and deterministic variables when present.

Johansen proposes two different likelihood ratio tests to test the significance of these canonical correlations and also the reduced rank of the Π matrix: the trace test and maximum eigenvalue test.

$$J_{TRACE} = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (4.12)$$

$$J_{EIGEN} = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (4.13)$$

where T is the sample size and $\hat{\lambda}_i$ is the i^{th} largest canonical correlation. The trace test tests, under the null hypothesis, that the coefficient matrix has rank less or equal than r , which resembles r cointegrating vectors, against the alternative hypothesis that the matrix has full rank n , which resembles n cointegrating vectors.

The maximum eigenvalue test tests the null hypothesis of r cointegrating vectors against the alternative hypothesis of $r+1$ cointegrating vectors. Usually, neither the trace and the eigenvalue tests follow a chi square distribution. Although both the maximum eigenvalue and trace test statistics are useful to test strict unit-root assumption, they lack on testing on systems characterized by near-unit-root processes. Johansen's test is used when all variables are $I(1)$, however as he states there is little need to pre-test the variables in the system to establish their order of integration. If a single variable is $I(0)$, instead of $I(1)$, this will reveal itself through a cointegrating vector whose space is spanned by the only stationary variable in the model.

Lets consider a model in which $y_t = (y_{1,t}, y_{2,t})'$ where $y_{1,t}$ is an $I(1)$ and $y_{2,t}$ is an $I(0)$ process. It is expected that there should be a cointegrating vector as $\beta = (0, 1)'$. If the coefficient matrix Π has full rank n , all the n variables are stationary.

The assumption that states that any variable that is not $I(1)$, or a pure unit-root process, is a stationary $I(0)$ process allows to avoid preliminary tests on the classification of the variables as $I(1)$ or $I(0)$ processes. Nevertheless, this does not make the method robust to near-integrated variables, since these do not fall into none of the above mentioned classifications. However, the above specification tests of the cointegrating vector suggest a way of making inference more robust in the potential presence of near-unit-root variables. For instance, considering the bivariate case described above, explicitly testing whether $\beta = (01)'$ will help to rule out spurious relationships that are not rejected by the initial maximum eigenvalue or trace test. Although we argue that such specification tests should be performed in almost every kind of application, they are likely to be extra useful in cases where the variables are likely to have near-unit-roots and the initial test of cointegration rank is biased.

As mentioned in the literature, the results should prove that the quotes are cointegrated by rejecting the null hypothesis of no cointegration. Nevertheless, rejecting the null hypothesis does not necessarily mean that the two variables are cointegrated. Indeed, rejecting the hypothesis of $r = 0$ would not imply that the null hypothesis, no cointegration, is rejected if:

- i. $r = 1$ is also rejected, which implies that matrix Π has full rank, because the variables are stationary, which means that it cannot be decomposed ($\Pi = \alpha\beta'$);
- ii. $r = 1$ cannot be rejected, the restrictions *i.* $\beta' = (1, 0)$ and *ii.* $\beta' = (0, 1)$ must be, also, accounted in the possible results. If one of these holds, it would mean that between x_{1t} and x_{2t} there is no cointegration. If *i)* holds, the result exploits that x_{1t} is stationary and that it does not have a long-run relationship with x_{2t} , if *ii)* cannot be rejected the case would be symmetric.

Chapter 5

Data set and preliminary study

Intra-day data are collected from the NYSE Trade and Quote (database) for the Walmart (WMT), The Coca-Cola company (KO), JPMorgan (JPM) and Caterpillar (CAT) stocks. As event window it has been chosen to collect data from January 29, 2001 to June 29, 2018. As time window it has been chosen to analyzed the period that starts right after the burst of the Dot-com bubble in order to observe their evolution and, respective, firm net-implementation during the years until the end of the first half 2018. These stocks have been selected due to their trading volume during the day, which is large enough in order to test the idea behind this paper.

5.1 TAQ Data

In order to start the estimation procedure, data has been cleaned, since these had to satisfy several condition in order to be categorized in TAQ database.

When trades and quotes are matched together, obviously, some data errors may arise and so they should be filtered out. To remove such types of errors, a set of filters is applied in order to clean the data, some of these are:

- Delete observations which are indicated to be incorrect, delayed or corrected.
- Delete entries outside the regular trading hours.
- Delete entries with a negative, or zero, quote or transaction price.
- Delete all entries with negative spreads.
- Delete entries whenever the price is outside the interval $[Bid - 2Spread ; Ask + 2Spread]$.
- Delete all entries with the spread being greater or equal than 15 times the median spread of that day.
- Delete all entries with the price being greater or equal than 5 times the median mid-quote of that day.
- Delete all entries with the mid-quote being greater or equal than 10 times the mean absolute deviation from the local median mid-quote.
- Delete all entries with the price being greater or equal than 10 times the mean absolute deviation from the local median mid-quote.

5.2 Descriptive analysis

In this section, the main descriptive statistics of the data-set are exposed in order to have a light touch on the data that may be encountered during the work.

	Mean	Median	SE	Skew.
WMT	0.0104%	0.0243%	1.33%	0.0461
KO	0.0098%	0.0300%	1.18%	-0.0441
JPM	0.0145%	0.0000%	2.44%	0.2320
CAT	0.0420%	0.0445%	1.97%	-0.1068

Table 5.1: Stocks return descriptive statistics.

Table 5.1 reports the descriptive estimate of the return of each stock, by trading it with a frequency of one second. The results are not statistically significant, but show a link in their return process.

5.3 (Il-)Liquidity measure display

In this part, the main liquidity measures, both low and high frequency, will be refreshed and a brief descriptive analysis will be taken upon these, separately.

Low Frequency	High Frequency
Relative Bid-Ask Spread (RBA)	Realized Variance (RV)
Amihud Illiquidity (AI)	High-Frequency RBA (HRBA)
Mid-range (Il-)Liquidity (MI)	log-Volume (IVol)

Table 5.2: (Il-)Liquidity measures.

5.3.1 Low frequency descriptive measures

Lets consider firstly the low frequency liquidity measures. These are usually easier to compute rather than the high frequency counterpart. Nevertheless, these can give consistent results that are similar to the latter, even if it has to be be loose the accuracy of the estimates, in some instances.

RBA	WMT	KO	JPM	CAT
Mean	0.0157	0.0159	0.0144	0.0245
Standard Deviation	0.0255	0.0235	0.0313	0.0289
Skewness	4.1406	3.4492	6.4136	2.8973
Kurtosis	25.4123	18.1577	77.0857	13.1297
Min	0.0022	0.0026	0.0021	0.0022
Max	0.3200	0.2396	0.7048	0.2499

Table 5.3: Descriptive Relative Bid-Ask measure.

As Table 5.3, 5.4 and 5.5 exhibit, all stocks present low values for the low frequency measures, which do not seem to be significant, except MI. However, the main focus of this stats is to understand the boundaries of each measure in order to have a further comprehension of this work's topic.

	AI	WMT	KO	JPM	CAT
Mean		$5.74e^{-4}$	$8.87e^{-4}$	$8.82e^{-4}$	$8.90e^{-4}$
Standard Deviation		$5.83e^{-4}$	$5.13e^{-4}$	0.0011	$8.77e^{-4}$
Skewness		2.9666	3.3083	3.7827	2.4407
Kurtosis		19.4411	24.9432	26.3069	13.5972
Min		0	0	0	0
Max		0.0066	0.0072	0.0123	0.0102

Table 5.4: Descriptive Amihud illiquidity measure.

	MI	WMT	KO	JPM	CAT
Mean		0.2534	0.2073	0.2297	0.2657
Standard Deviation		0.0174	0.0176	0.0239	0.0340
Skewness		0.3288	0.2259	0.3030	-0.3424
Kurtosis		2.1306	1.8467	3.3057	2.3631
Min		0.2060	0.1682	0.1480	0.1803
Max		0.3030	0.2567	0.3037	0.3581

Table 5.5: Descriptive Mid-range (II-)Liquidity measure.

These values may imply that, on average, the firms that had a lower spread and AI were the ones that had the most liquid behavior in the sample. Instead of merely, "observing" the values it is useful to verify that these measure that reports liquidity status do not follow a Gaussian distribution, while only MI gets closer to these.

5.3.2 High frequency measures

Considering illiquidity as the lack of prospect to close the relative position in a short period, it is pertinent to have an overlook on the evolution of the trading volume that elapses during the day, and over the whole period. Lets start by examining the log-volume and the intra-day log-volume of a sample trading day.

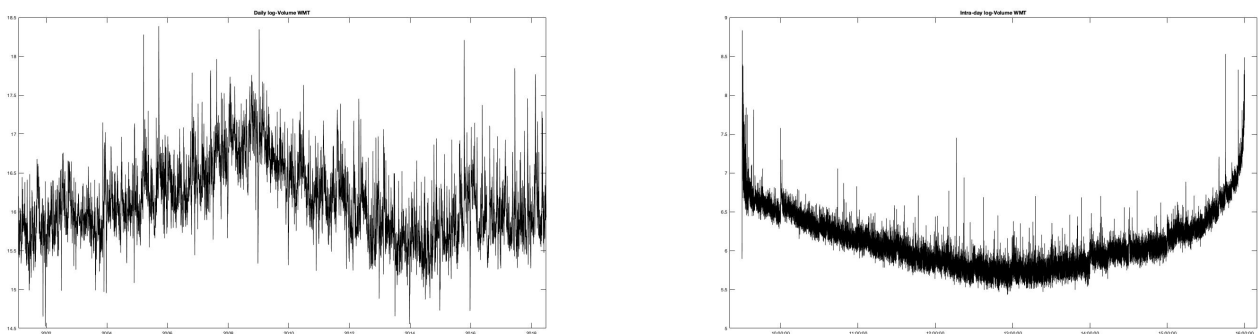
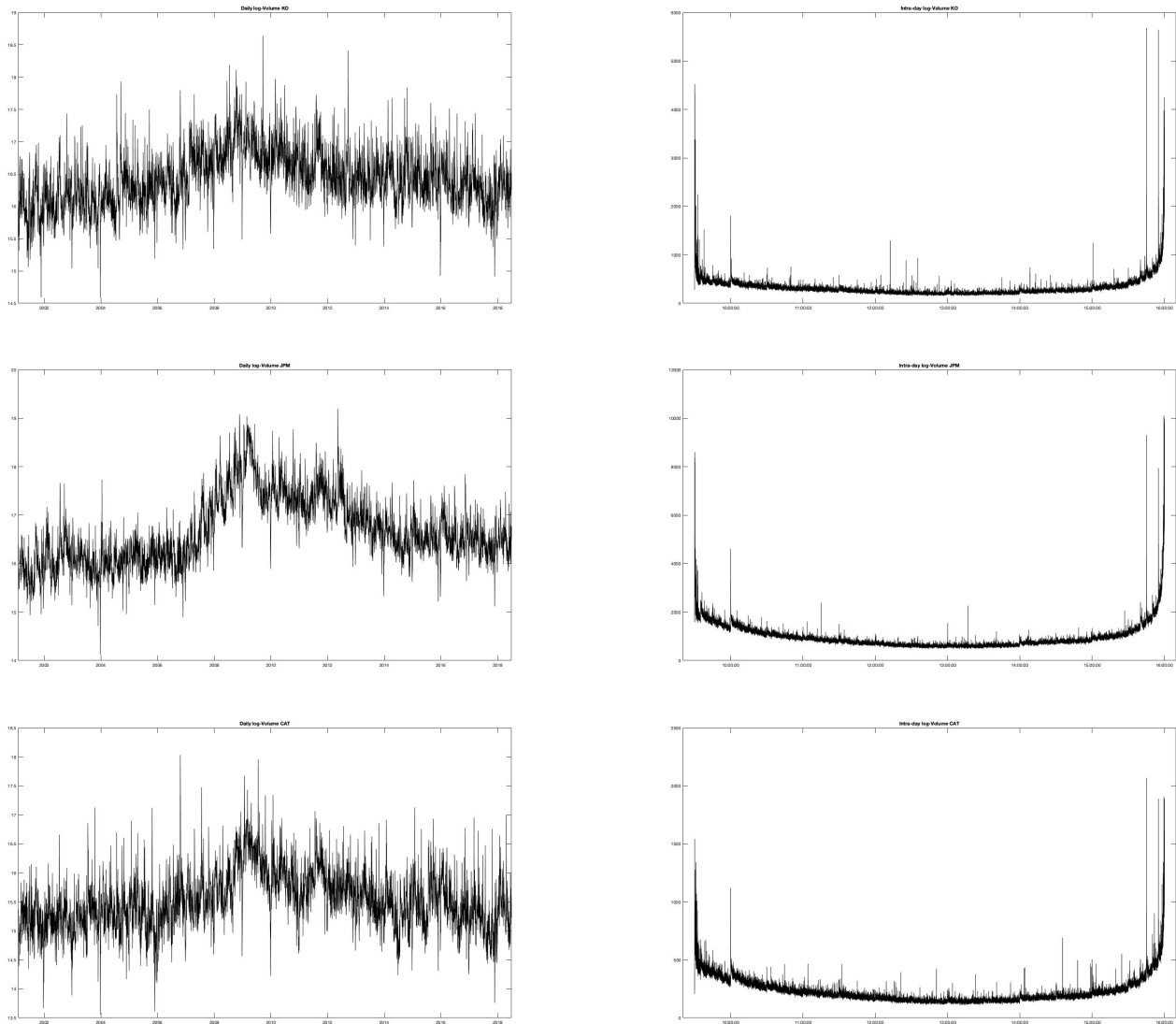


Figure 5.1: Daily and Intra-day log-Volume.



Continued Figure 5.1: Daily and Intra-day log-Volume.

Figure 5.1 shows how the log-volume of the four stocks has evolved during the event window and how it is distributed during the intra-day transactions. The daily trading volume has enhanced, steadily, during the years until the 2010, from which, it started to decay sharply in order to recover, and hold the previous values, along the following year. While KO trading volume was pretty stable, the other three had a peak around during the year 2009. This evidence may be caused by the US real estate problem that lead to the 2007 financial crisis that spread in all the world. The rise in the real estates allowed the three companies to ride the event in order to satisfy their client needs, for what concerns the principal family (WMT), credit (JPM) and real estate (CAT) needs. Concerning the intra-day volume, it is clear that all instruments behave like any other since they display peaks at the open and closure of any trading day.

Once described the trading volume, return is analyzed to read its model and its high frequency variations. Firstly, Figure 5.2 shows the evolution of stocks' RV along the period. By giving a first look, it can be seen that illiquidity periods arise when several peaks show up, particularly during the period 2008-2010 and during 2003. Since it is made up by two components behavior, it is useful to have a brief overlook at their individual patterns, such as its comprehension should be as clear as possible. It is also briefly exposed the correlation that runs within these firm assets, to arrange for any possible

portfolio-build idea between these objects, and how the portfolio composition may change according to a prospect of a long-term investment or a shorter one.

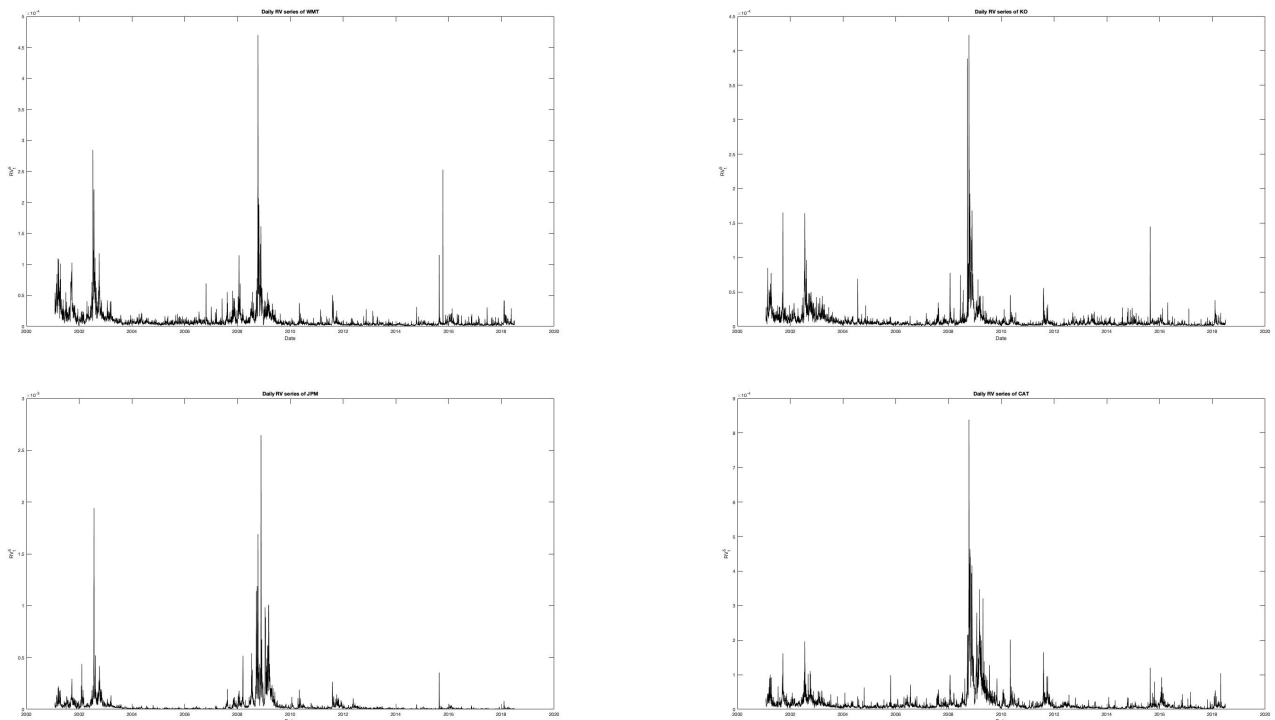


Figure 5.2: Stocks daily RV with 5 minutes frequency.

Figure 5.2 displays the RV of the instruments along the event window. It is not a surprise to see peaks around years like 2002 (Dot-com bubble) and 2008 (subprime crisis and Lehman Brothers bankruptcy). While during the whole period the RV stayed pretty flat and stable, in 2016 WMT and KO had a lift due to industry related problems. KO faced a production issue in several nations all over the world, which reduced the production temporarily. WMT’s one was, and is, related to the poor customer service and employees relation caused by its politic of having always the lowest possible price of their commodity needs.

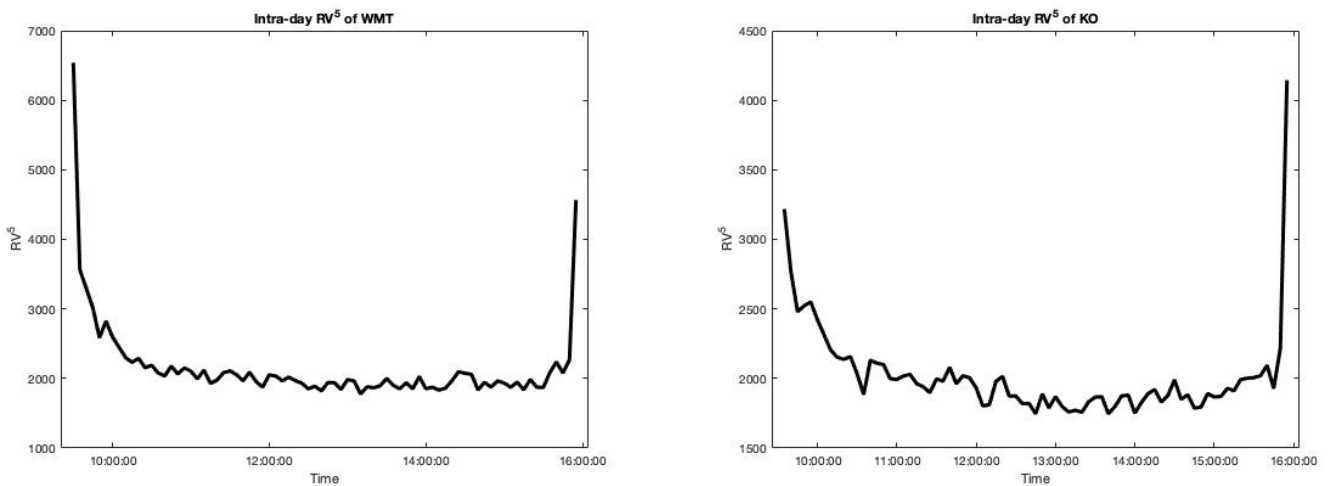
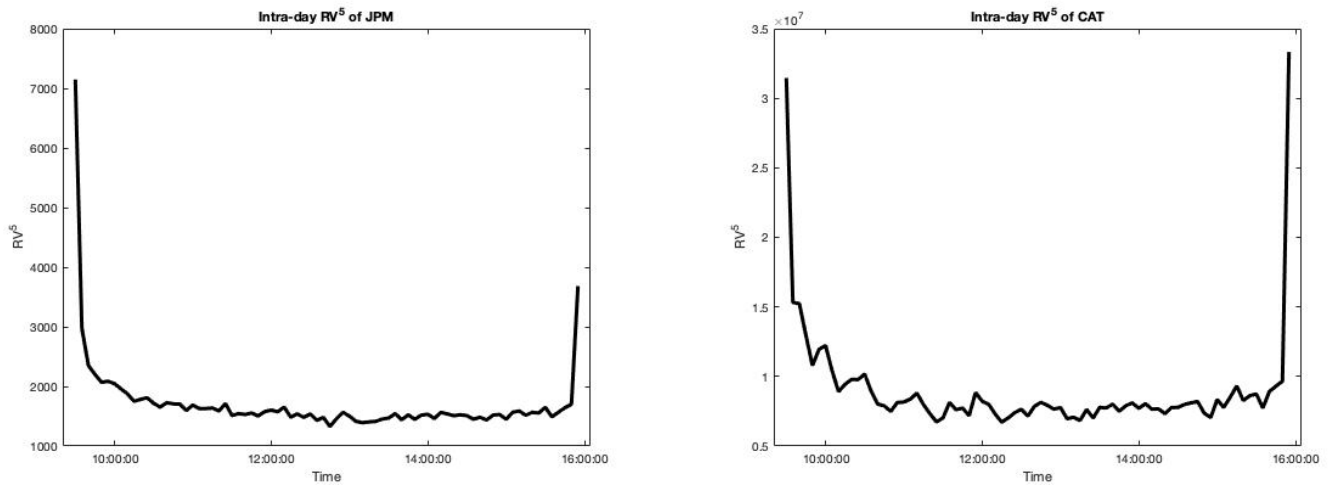


Figure 5.3: Stocks Intra-day RV series with 5 minutes frequency.



Continued Figure 5.3: Stocks Intra-day RV series with 5 minutes frequency.

Figure 5.3 exhibits the intra-day RV of the four shares, by considering a sample frequency of 5 minutes ($\Delta = \frac{1}{300}$) within each trade. The market opening, and closing, effect has already been analyzed in the literature. As a common listed security, price variance enhances during the first trading minutes after the trading day opens. After this time frame, as the time passes, the variation narrows during the day and stays pretty stable for the day. However, as the day gets near its closure, price changes occur in the last minutes, since every trader wants to close all his positions.

5.4 Stats comparison

After computing the return realized measures, it is useful to check if these are characterized by a (cor-)relation, which characterizes measure's sensibility against the others that are going to be used. To understand the behavior of the two measures it has been used the concept of the Realized Volatility to check the variance property:

$$\text{Var}(A - B) = \text{Var}(A) + \text{Var}(B) - 2\text{Cov}(A, B). \quad (5.1)$$

For the sake of knowledge, it has run a relation estimate of the RV of the illiquidity measures, the AI and Bid-Ask spread (BAS) measures. Through the concept of RV, it has been estimated the realized covariance within these factors to check if these can be considered exchangeable for the following study.

As supposed, the results show that the variables are characterized by a a perfectly positive correlation among the two measures.

Since the two measures are correlated, to spare space and tests, the AI selection, to be the comparative variable, is intentional since this may further stress the paper topic, illiquidity. To give a deep knowledge about the argument, it has been put in place a comparison between the (usual) descriptive correlation measure between the two measures of the stock and the "realized correlation" ($R\rho$).

As seen, the correlation measures are very different due to frequency at which the stocks are traded during the day. Andersen, Bollerslev, Diebold and Labys (1999) found that the realized variance and covariance are highly skewed, while the standard deviation (volatility) is almost symmetric, but its log transformation behaves almost as a Gaussian. Realized correlation is usually positive, sometimes even

ρ_{xy}	WMT	KO	JPM	CAT
WMT	1	-32.11%	69.56%	62.18%
KO	-32.11%	1	-24.24%	2.96%
JPM	69.56%	-24.24%	1	68.31%
CAT	62.18%	2.96%	68.31%	1

Table 5.6: Stocks Correlation.

$R\rho_{xy}$	WMT	KO	JPM	CAT
WMT	1	15.54%	15.04%	9.23%
KO	15.54%	1	19.04%	9.63%
JPM	15.04%	19.04%	1	9.48%
CAT	9.23%	9.63%	9.48%	1

Table 5.7: Stocks Realized Correlation.

with a strong weight, and it displays substantial variation during periods. It is, also, highly correlated with the realized volatility. Particularly, return correlations tend to rise on high-volatility days.

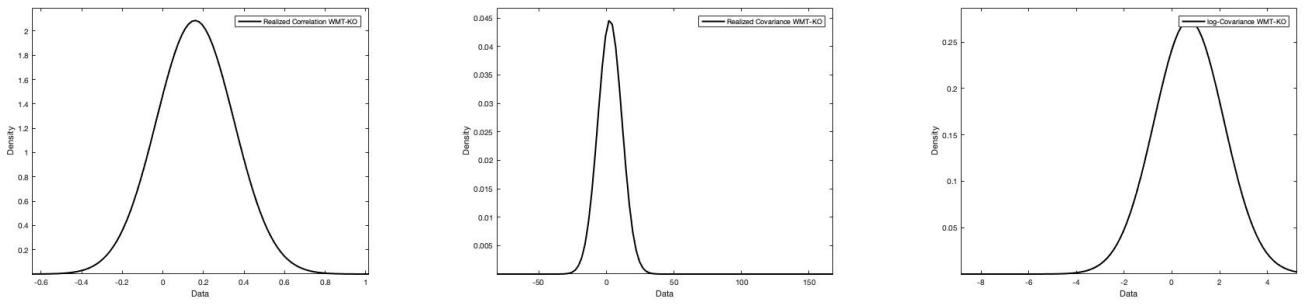


Figure 5.4: WMT-KO Realized Correlation, Realized Covariance and Realized log-Covariance.

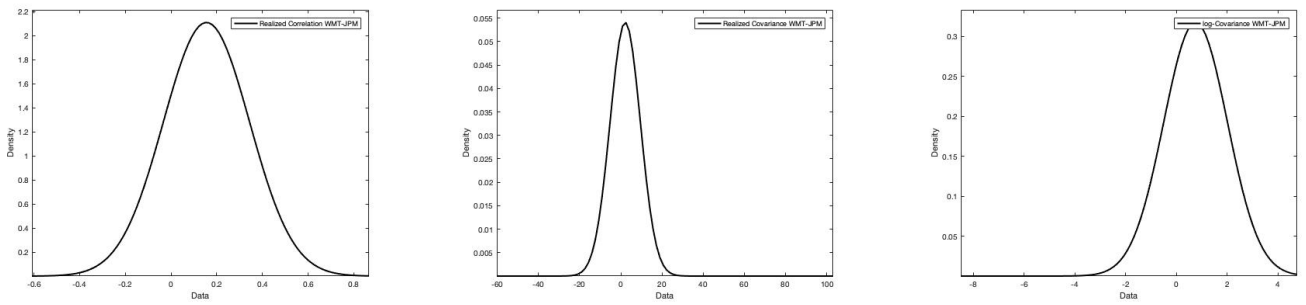


Figure 5.5: WMT-JPM Realized Correlation, Realized Covariance and Realized log-Covariance.

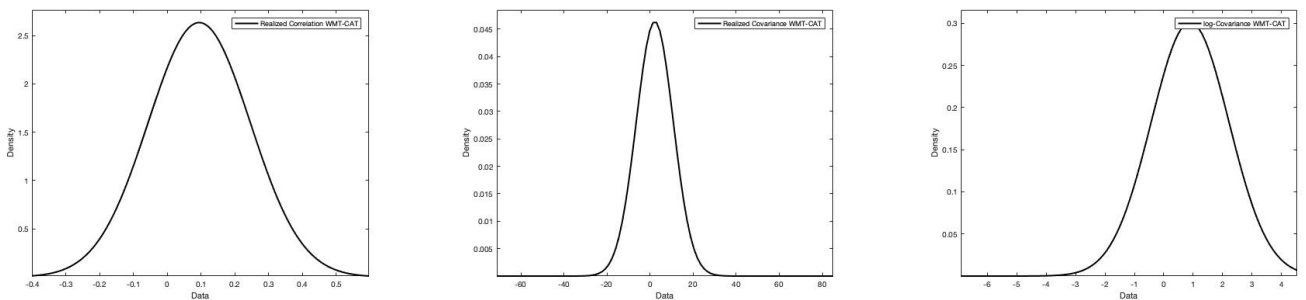


Figure 5.6: WMT-CAT Realized Correlation, Realized Covariance and Realized log-Covariance.

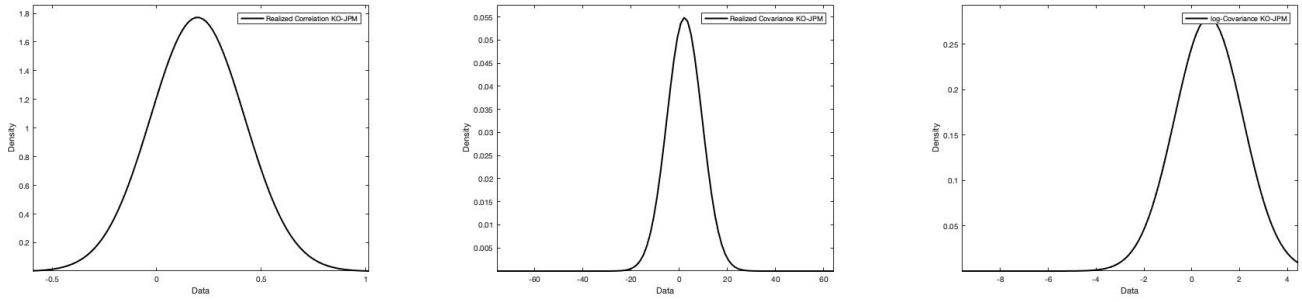


Figure 5.7: KO-JPM Realized Correlation, Realized Covariance and Realized log-Covariance.

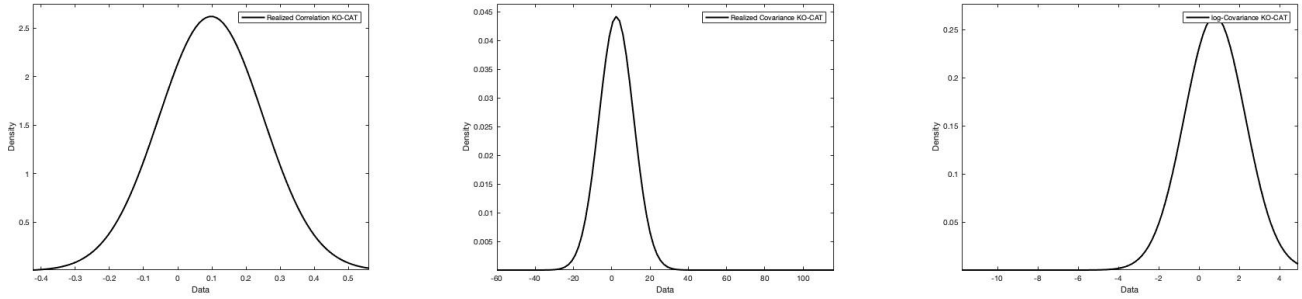


Figure 5.8: KO-CAT Realized Correlation, Realized Covariance and Realized log-Covariance.

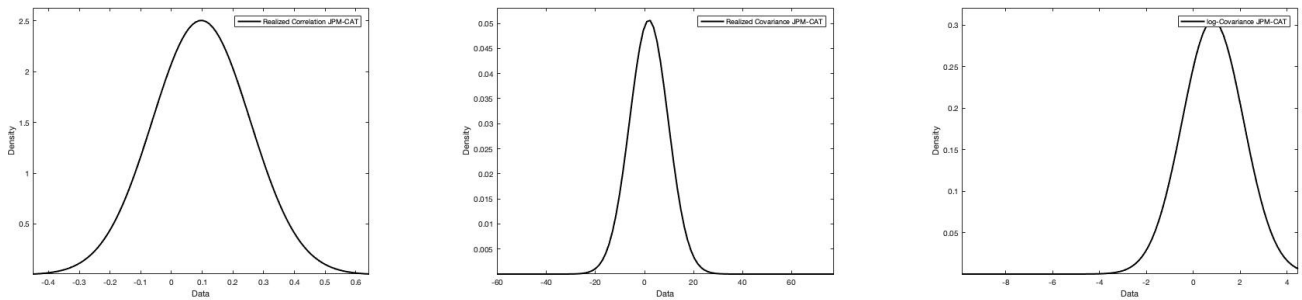


Figure 5.9: JPM-CAT Realized Correlation, Realized Covariance and Realized log-Covariance.

As the graphs shows from Figure 5.4 to Figure 5.9 the Realized measures behave exactly like the Andersen et al. said. By considering the idea of building a portfolio made with every stock, the result will change consistently by considering which type of time-frame investment someone wants to enter. The share are, more or less, correlated with all significant values as for the long-term investment as for the shorter one.

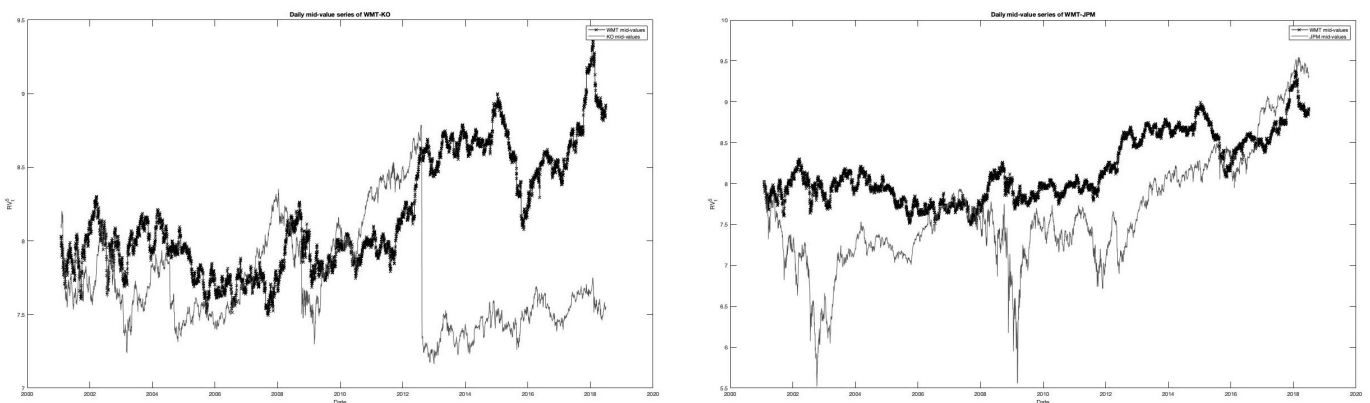
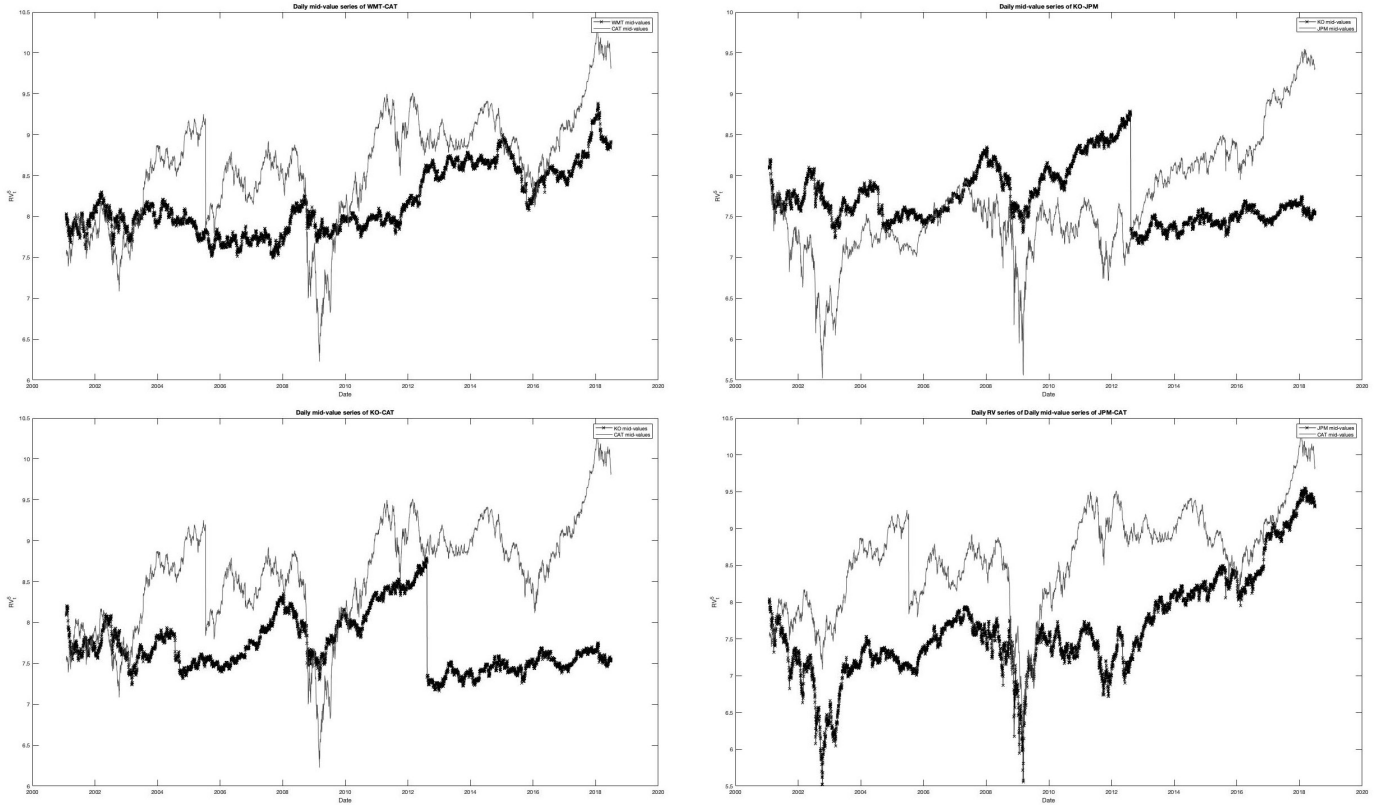


Figure 5.10: Daily stock mid-values.



Continued Figure 5.10: Daily stock mid-values.

Figure 5.10 shows the combination of pairs of mid-values between two of the different shares. Their scheme seems to follow, in some instances, a pattern that relates the movement of one respect to the other.

To further encourage the learning on the two measures' behavior, it has been used the above stated variance decomposition. To ascertain that these fits are consistent, and capture any possible dependence, it is used the variation decomposition between two objects, by checking if the equality holds. Let Var_{A-B} be the LHS of the stated equation and Var_{AB} the RHS of the same formula (obtained by summing the two variances and subtracting their covariance). The measures used to test this are, for the discrete statistic, the mid-range daily prices, while for the realized one it has been used the mid-price variation with 5 minutes frequency.

σ_{WMT-Y}^2	Var_{A-B}	Var_{AB}	RV_{WMT-Y}	Var_{A-B}	Var_{AB}
KO	36.04%	36.04%	KO	5.39%	5.39%
JPM	25.64%	25.64%	JPM	14.20%	14.20%
CAT	26.62%	26.62%	CAT	8.48%	8.48%

Table 5.8: Mid-range (WMT–Stock) Price Variance and Realized Variance.

σ_{KO-Y}^2	Var_{A-B}	Var_{AB}	RV_{KO-Y}	Var_{A-B}	Var_{AB}
WMT	36.04%	36.04%	WMT	5.39%	5.39%
JPM	71.75%	71.75%	JPM	14.55%	14.55%
CAT	53.61%	53.61%	CAT	8.38%	8.38%

Table 5.9: Mid-range (KO–Stock) Price Variance and Realized Variance.

σ_{JPM-Y}^2	Var_{A-B}	Var_{AB}	RV_{JPM-Y}	Var_{A-B}	Var_{AB}
WMT	25.64%	25.64%	WMT	14.20%	14.20%
KO	71.75%	71.75%	KO	14.55%	14.55%
CAT	29.12%	29.12%	CAT	14.59%	14.59%

Table 5.10: Mid-range (JPM–Stock) Price Variance and Realized Variance.

σ_{CAT-Y}^2	Var_{A-B}	Var_{AB}	RV_{CAT-Y}	Var_{A-B}	Var_{AB}
WMT	26.62%	26.62%	WMT	8.48%	8.48%
KO	53.61%	53.61%	KO	8.38%	8.38%
JPM	29.12%	29.12%	JPM	14.59%	14.59%

Table 5.11: Mid-range (CAT–Stock) Price Variance and Realized Variance.

As the tables from Table 5.8 to Table 5.11 show, the descriptive variance and RV formula holds for all the stocks possible pairwise combinations. The RV exposes that with a 5 minutes frequency, i.e. dividing the day in 78 sub-periods, the realized measure is consistent and accurate and should, hence, be accounted to operate in short time frame of hours, or even seconds, of trading. These outcomes may suggest that these shares co-move in an observable, accountable and measurable way, by considering the "usual" conditional variance concept.

Concerning the ex-post measure, it is clear at first sight (as observed in the literature) that the price discrepancy shrinks as the infra-day trading frequency narrows.

Therefore, because both the conditional descriptive measure and the realized one verify the variance equality, anyone, who concerns to open a position with a combination of these, should not face sudden, unexplained, price variations by accounting these measures.

Chapter 6

Empirical analysis

In this chapter, the event studies and measure analysis are summarized. In order to give a more comprehensive understanding of the study this chapters is divided as follows:

- i. Unit-root test on liquidity measures;
- ii. Cointegration of the four stocks and their combinations;
- iii. Relevance of the liquidity measure;
- iv. Measures estimate power of the on the market illiquidity risk premium.

6.1 Stationarity tests

Before entering the main analysis of this work. Unit-root tests are computed on each measure to see if there is a need of a transformation of these measures.

	WMT	KO	JPM	CAT
Low frequency				
AI	0.5886	0.5906	0.6107	0.5764
RBA	0.7991	0.8104	0.7887	0.8428
MI	0.9998	0.9998	0.9998	0.9998
High frequency				
RV	0.9998	0.9998	0.9998	0.9998
IRBA	0.9227	0.9300	0.9152	0.9328
IVol	0.9227	0.9300	0.7652	0.8189

Table 6.1: ADF test coefficients.

At first sight Table 6.1 gives significant results for all the measures, but for the RV and MI ones exhibits a near-unit-root result. However, since this results may lead us to doubt about the reliability of ADF test, VR confirms the rejection of the unit root by giving very low values, which report that all process are stationary. Therefore, both Table 6.1 and 6.2 show that both ADF and VR test gives significant results for all measure, that allows to refuse the hypothesis of the presence of unit-root in the time series process of each size.

	WMT	KO	JPM	CAT
Low frequency				
AI	0.5078	0.4887	0.5105	0.4671
RBA	0.5304	0.5286	0.5615	0.5521
MI	0.6912	0.6781	0.7476	0.6876
High frequency				
RV	0.5218	0.4993	0.6943	0.5534
HFRBA	0.5444	0.5382	0.5325	0.5803
IVol	0.6758	0.6552	0.7014	0.6708

Table 6.2: VR test.

Since all the "old" measures have been proved in the literature to be stationary, it is worth to have a further study of the behavior of the new-developed measure, MI.

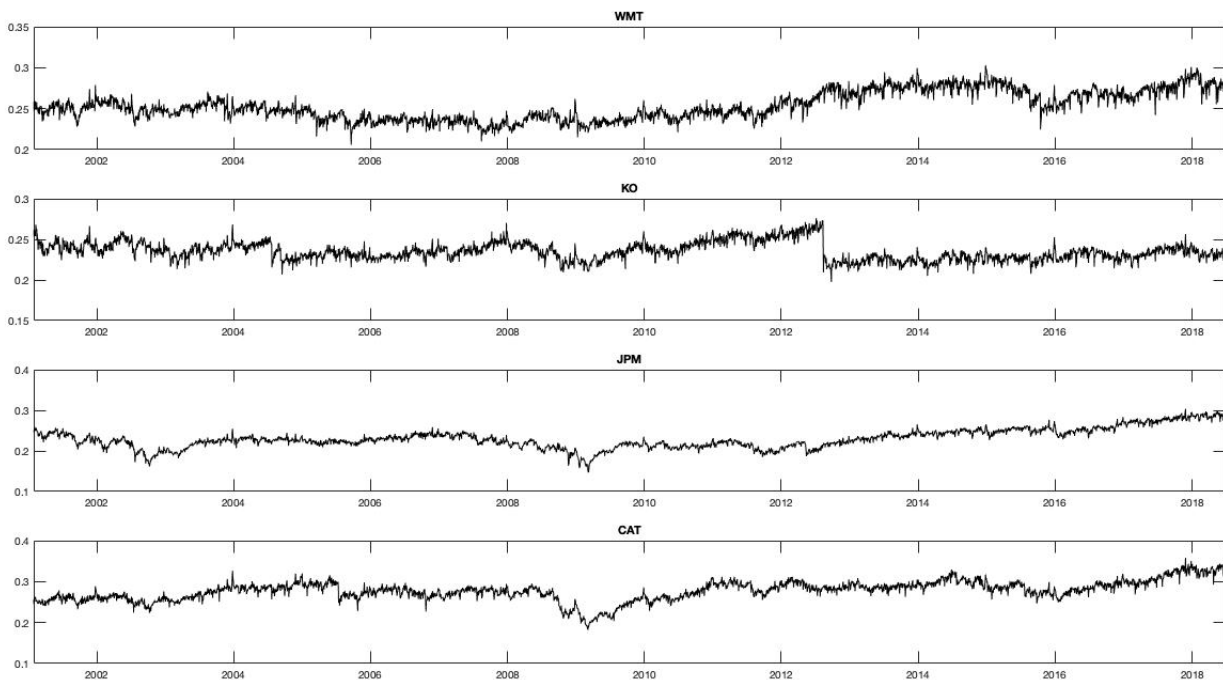


Figure 6.1: Stocks Mid-range liquidity series.

Figure 6.1 shows the evolution of MI during the event period. Even through rough illiquid sub-periods, the index stayed pretty stable for all the instrument within the boundary $[0.2; 0.35]$, which may result in a difference stationary process.

6.2 Cointegration test

To start the study on the target stocks, it is useful to know if the bid and ask prices are cointegrated. As Frederick (1995) and Bollerslev (1997) discovered, often, bid and ask prices are cointegrated. To analyze their behavior, and thanks to the previous tests on the data-set, it is used the mid-range bid and ask price of each day to analyze their movement pattern.

Usually bid and ask prices are cointegrated, so it is beneficial to study their evolution and variation over the event window.

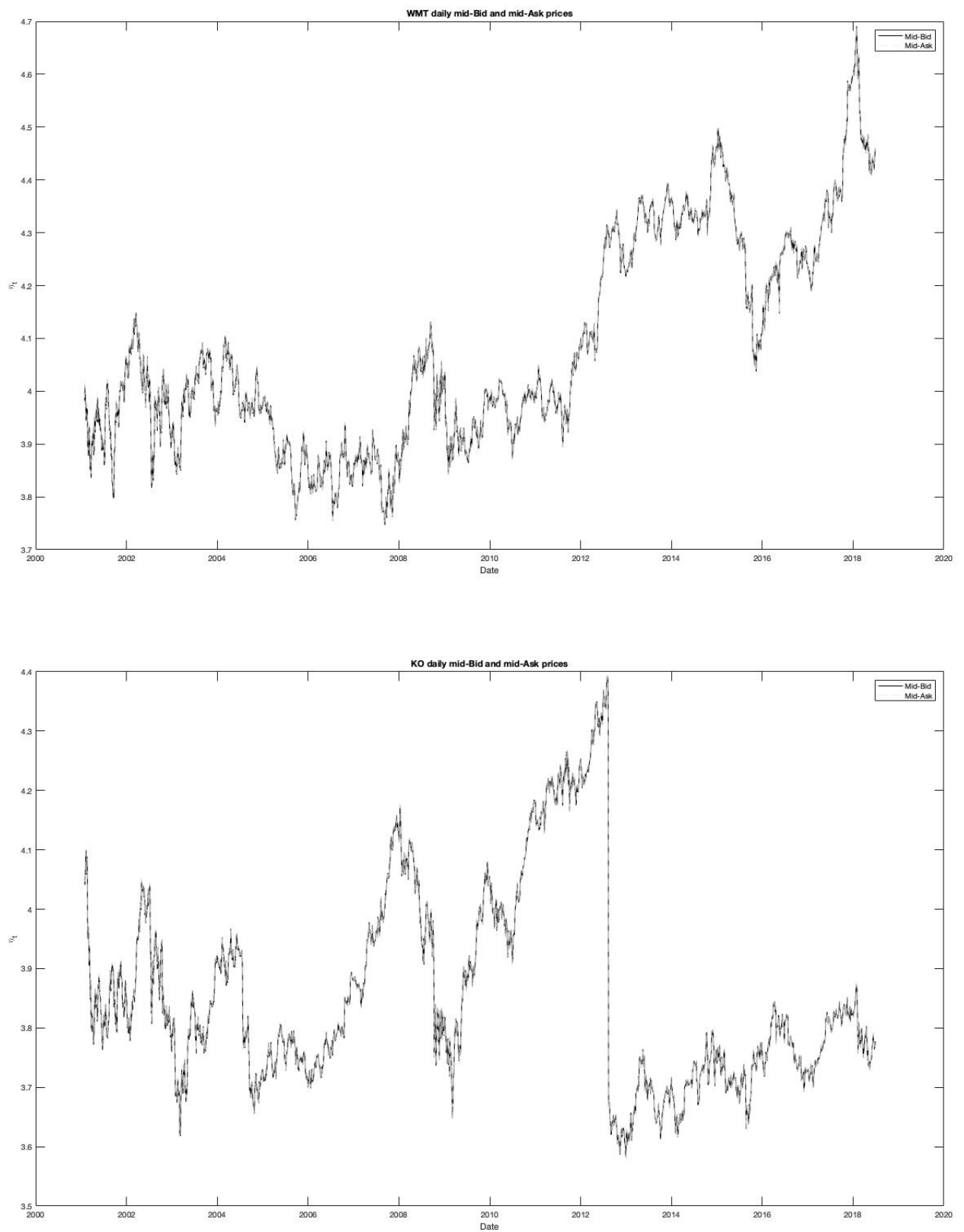
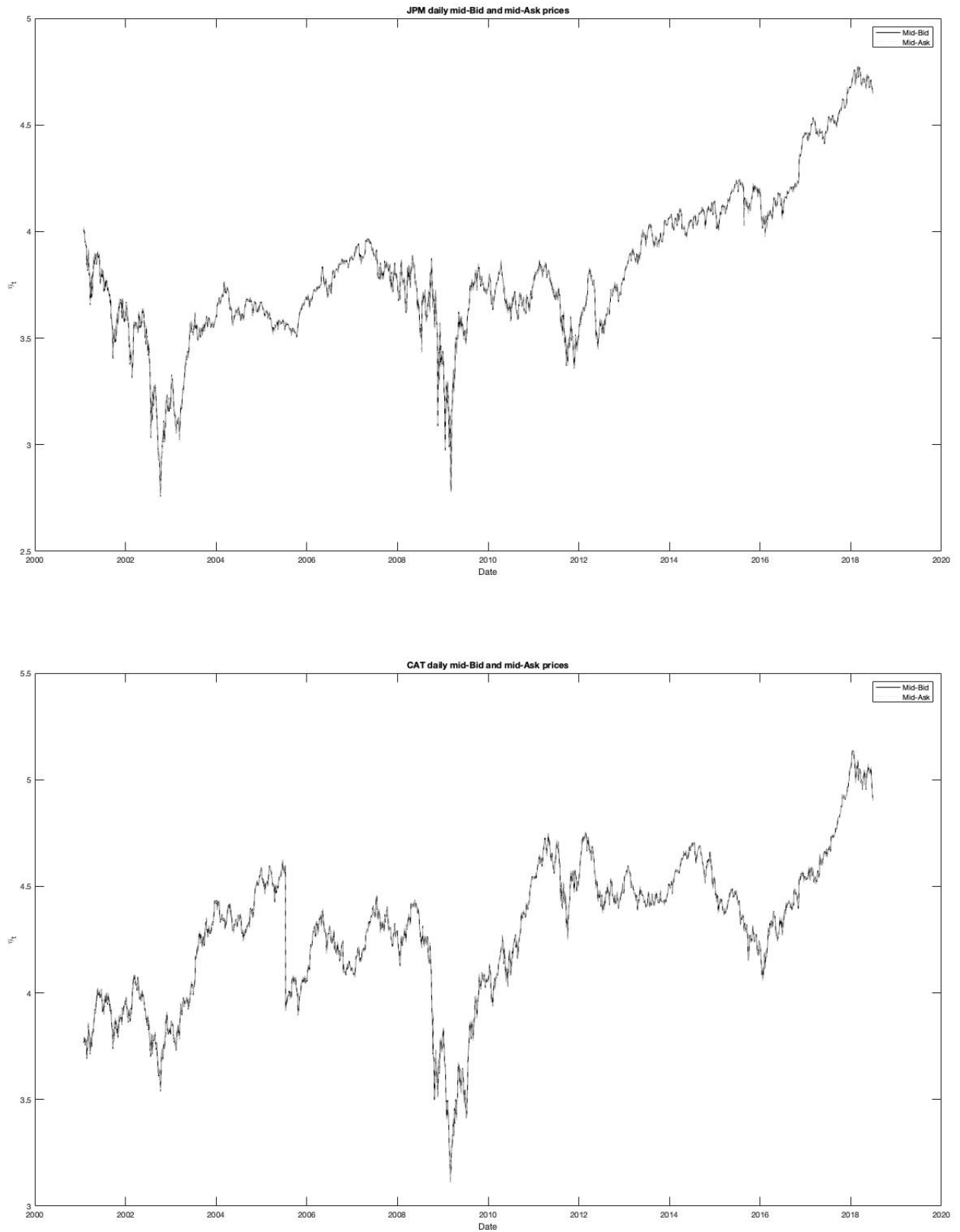


Figure 6.2: Daily stock bids and asks.



Continued Figure 6.2: Daily stock bids and asks.

Figure 6.2 shows the respective bid and ask mid-prices of firms share. At first sight, as already mentioned, these are likely to follow a similar price process. Considering this, the previously stated variance decomposition has been re-arranged, as before, to the concept of RV to observe if there is an observable dependence within the quotes.

Because of the stress that is intentional to give to the decomposition of the coefficient matrix Π , it has run a Johansen test that include also intercepts, linear trends (in the cointegrated series) and deterministic quadratic trends. The decomposition of the cointegrated process would be:

$$\Delta x_t = \alpha\beta'(x_{t-1} + \gamma + \delta t) + \lambda + \Psi d_t + \epsilon_t. \quad (6.1)$$

Nevertheless, here the focus is given to the coefficients of the decomposed matrices, $\alpha\beta'$, because they are the argument that allows to verify if a process stationary or not. The test used for all the Johansen's test is the trace test.

Rank	<i>BA_{WMT}</i>		<i>BA_{KO}</i>	
	Stat	p-value	Stat	p-value
$r = 0$	601.2412 ^a	0.0010	480.3424 ^a	0.0010
$r = 1$	9.0393 ^a	0.0034	8.8578 ^a	0.0037
Stationary process test				
ADF	-6,4634	0.0010	-22.7678	0.0010
$ 1 + \alpha'\beta $	0.7473		0.7963	

Table 6.3: Johansen cointegration test of stocks B–A Spread.

Rank	<i>BA_{JPM}</i>		<i>BA_{CAT}</i>	
	Stat	p-value	Stat	p-value
$r = 0$	412.4965 ^a	0.0010	898.0167 ^a	0.0010
$r = 1$	11.8115 ^a	0.0010	6.8392 ^a	0.0092
Stationary process test				
ADF	-19.7918	0.0010	-23.2897	0.0010
$ 1 + \alpha'\beta $	0.8257		0.6320	

Continued Table 6.3: Johansen cointegration test of stocks B–A Spread.

All variables are taken with logs. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

Table 6.3 reports Johansen cointegrating test for all the stocks. The results exploits a strong evidence to reject the null hypothesis (cointegrating rank equal to zero) in favor of the alternative. Although the null hypothesis is rejected, the coefficient vectors of the decomposed matrix have been analyzed to verify that the alternative hypothesis holds. To verify the stationarity of the process it has been used a test based on the verification of the unit root, Augmented Dickey-Fuller (ADF), and the assessment that the product of the two matrices lies inside the unit disk. The outcomes validate the finding of alternative hypothesis. So it can be stated that bid and ask prices of these stocks are cointegrated.

Rank	WMT-KO		WMT-JPM		WMT-CAT	
	Stat	p-value	Stat	p-value	Stat	p-value
$r = 0$	27.7538 ^a	0.0025	20.7757 ^b	0.0230	15.1087	0.1365
$r = 1$	6.0434 ^b	0.0141	8.0651 ^a	0.0047	6.1529 ^b	0.0134
Stationary process tests						
$ 1 + \alpha'\beta $	0.9931		0.9956		0.9963	

Table 6.4: Johansen cointegration test of stocks couples.

Rank	KO-JPM		KO-CAT		JPM-CAT	
	Stat	p-value	Stat	p-value	Stat	p-value
$r = 0$	26.0687 ^a	0.0041	14.8052	0.1483	20.5753 ^b	0.0245
$r = 1$	5.6910 ^b	0.0172	6.4742 ^b	0.0111	6.2011 ^b	0.0130
Stationary process tests						
$ 1 + \alpha'\beta $	0.9932		0.9963		0.9952	

Continued Table 6.4: Johansen cointegration test of stocks couples.

All variables are taken with logs. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

Table 6.4 reports Johansen Trace statistics, on stocks mid-range bid-ask price, to check if there is a cointegration relation between the instruments and if they respect the stationarity condition. These statistics provide some evidence for rejecting the null hypothesis of no cointegration for the couples' average price. In any case, the stationary process audit disclose a near unit-root output. As Duffee and Stanton (2007) outlined, a process with a very near-bound root would let shock last ,on average, one year. This result is not an issue by considering the purpose, but it is important to mention that these are heavily asymmetric when data are persistent.

6.3 Best Liquidity measure proxies

In this section liquidity proxies, both at low and high frequency, will be tested to see their relation with the most common measure of illiquidity, bid-ask spread.

Measures are categorized, firstly, in low and high frequency as follows

$$LF = [AI, RBA, MI, D_{MON}, D_{FRI}],$$

$$HF = [RV, lRBA, lVol, D_{MON}, D_{FRI}], \quad (6.2)$$

where other two dummy variables are added in the measures, D_{MON} and D_{FRI} . These take value of one, respectively, if the trading day is a Monday or a Friday, to study also any possible week seasonality.

Thereafter, it will be run a regression to state which frequency is the one that best estimates the illiquidity size. The estimates would be obtained by running an OLS regression

$$BAS = \alpha_{LF} + LF\beta_{LF}, \quad (6.3)$$

$$BAS = \alpha_{HF} + HF\beta_{HF}. \quad (6.4)$$

Table 6.6 reports the results of the individual regressions run for both the sample frequency measures.

	WMT	KO	JPM	CAT
Low frequency				
AI	13.4344 ^a (4.45)	5.4209 ^c (2.76)	2.7611 ^c (1.68)	-2.7682 ^b (-2.01)
RBA	2.7841 ^a (17.81)	2.1671 ^a (21.94)	2.7461 ^a (12.01)	2.0728 ^a (26.86)
MI	0.0692 ^b (2.31)	-0.2839 ^a (-7.62)	0.1495 ^a (3.73)	-0.3224 ^a (-7.86)
D_{MON}	0.0001 (0.02)	0.0020 (0.97)	0.0013 (0.35)	0.0020 (0.75)
D_{FRI}	-0.0004 (-0.16)	-0.0002 (-0.11)	0.0009 (0.35)	0.0007 (0.27)
Const.	-0.058 (-0.68)	0.0728 ^a (8.22)	-0.0232 ^b (-2.22)	0.1289 ^a (10.43)
R^2	52.52%	52.42%	55.48%	49.79%
High frequency				
RV	0.0943 ^a (3.92)	17.8142 ^a (4.11)	5.0102 ^a (4.90)	5.4451 ^a (6.46)
IRBA	0.0123 ^a (22.46)	0.0072 ^a (13.51)	0.0090 ^a (10.80)	0.0234 ^a (35.54)
IVol	-0.0073 ^a (-8.05)	-0.0119 ^a (-13.38)	-0.0087 ^a (-7.13)	-0.0019 ^b (-2.47)
D_{MON}	-0.0005 (-0.68)	0.0005 (0.68)	-0.0002 (-0.16)	0.0002 (0.29)
D_{FRI}	-0.0006 (-0.85)	-0.0002 (-0.31)	-0.0002 (-0.15)	0.0004 (0.46)
Const.	0.0943 ^a (6.31)	0.1904 ^a (12.48)	0.1332 ^a (6.09)	-0.0315 ^b (-2.49)
R^2	42.07%	39.48%	22.88%	47.03%

Table 6.5: Individual frequency test regressions.

All variables are taken with logs except dummy variables. The t-statistics are in parentheses and the error variance are robust to heteroskedasticity and autocorrelation in the residuals. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

Table 6.5 exposes the single regression results, ran to investigate the common bid-ask spread measure, used as the main illiquidity proxy.

While high frequency measures report expected values (sign), low frequency liquidity measure report a mixture of results for this test. While AI and RBA are found to hold their previous assumptions that categorizes these as illiquid measures (since these have a positive weight with the spread), MI presents a mix of positive and negative weights. Since this measure may be affected also by other variables, its factor loading should be carefully interpreted. For both the regressions, dummy variables are not significant at any confidence level. This might suggest that these stock are not affected by weekly seasonality.

To further investigate the relation that runs within these frequency measures, an additional regression is run to analyze all the liquidity measures on the usual spread measure. The vector of all measure is then made up by

$$LM = [AI, RBA, MI, RV, IRBA, IVol, D_{MON}, D_{FRI}], \quad (6.5)$$

The results of the OLS

$$BAS = \alpha_{LM} + LM\beta_{LM} \quad (6.6)$$

are reported in Table 6.6.

	WMT	KO	JPM	CAT
AI	6.4750 ^c (1.94)	5.7242 ^a (2.59)	2.5842 (1.45)	1.0601 (0.74)
RBA	2.0200 ^a (10.77)	1.6976 ^c (14.34)	2.5117 ^c (10.12)	1.5286 ^b (15.43)
MI	-0.4220 ^a (-5.49)	-0.0904 (-1.37)	0.1628 ^a (2.14)	-0.0302 ^a (-0.57)
RV	29.9224 ^b (2.39)	14.7735 ^b (2.03)	3.4065 ^b (2.12)	2.5758 (1.11)
IRBA	0.0214 ^a (8.99)	0.0090 ^a (6.74)	0.0106 ^a (3.70)	0.0337 ^a (11.47)
IVol	-0.0287 ^a (-7.20)	-0.0199 ^a (-6.90)	-0.0079 ^c (-1.76)	-0.0063 ^b (-2.08)
D_{MON}	-0.0008 (-0.26)	0.0033 (1.14)	0.0010 (0.27)	0.0016 (0.64)
D_{FRI}	-0.0012 (-0.58)	-0.0001 (-0.05)	0.0009 (0.33)	0.0005 (0.21)
Const.	0.5300 ^a (6.53)	0.3418 ^a (5.76)	0.0783 (0.83)	0.0374 ^b (0.58)
R^2	57.50%	56.23%	56.93%	53.86%

Table 6.6: All frequency measures test regression.

All variables are taken with logs except dummy variables. The t-statistics are in parentheses and the error variance are robust to heteroskedasticity and autocorrelation in the residuals. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

As Table 6.6 displays factor loading for all measures did not change its sign and confirms that MI can be interpreted as a mere liquidity measure, in contrast of the above mentioned AI and RBA. However, running this regression does not improve consistently the explanatory power of the regression, R^2 . Such result may imply that by using low frequency estimates is a good way to analyze the market behavior, with an affordable loss.

6.4 Stock illiquidity premium

Lastly, after all the tests on the variables, the main idea behind this work can be tested. To overcome the usual variables, that usually are also difficult to manage, low frequency measures are used to explain the illiquidity risk premium (Amihud, 2002).

The test is made by a re-arrangement of the CAPM equation, already treated by analysts like Fama, French (1992) and Amihud himself, in order to price the illiquidity risk of these HFT shares. Hence, more illiquid stocks should have higher illiquidity risk premium, which can be measured as:

$$(r_{M,t} - r_{f,t}) = \gamma_0 + \gamma_1 AI_{t-1} + \gamma_2 RBA_{t-1} + \gamma_3 MI_{t-1}, \quad (6.7)$$

where r_M and r_f are, respectively, the stock market return and the risk free rate. Since, the firms' stocks are all from US companies, r_M is the daily return taken from the S&P500 and as risk free the three-months Treasury bill rates. It should be recalled that the variables coefficients should be positive for the first two measure, since illiquidity measure should reflect the higher risk, and negative for the liquid one. The goal is to run a OLS regression to quantify the weights that the market return gives to the variables. Considering that the error may be correlated to one, or all, of these variables, the White test allows to verify that the heteroscedasticity problem would not impair or bias the estimators. Through the test, it has been found that the variance of the estimators behaves like a homoscedastic estimator.

	WMT	KO	JPM	CAT
AI	1.9732 ^a (5.08)	3.9137 ^a (7.30)	1.6028 ^a (6.36)	1.0899 ^a (5.27)
RBA	0.0291 ^a (4.67)	0.0052 (0.72)	0.0252 ^a (5.56)	-0.0313 ^a (-4.75)
MI	-0.1244 ^a (-15.38)	-0.1051 ^a (-12.36)	-0.1059 ^a (-16.24)	-0.0908 ^a (-15.83)
Const.	0.0378 ^a (17.58)	0.0276 ^a (15.02)	0.0304 ^a (18.84)	0.0317 ^a (18.53)
R^2	9.60%	11.19%	17.20%	11.72%

Table 6.7: Daily Regressions on the illiquidity risk premium.

All variables are taken with logs. The t-statistics are in parentheses and the error variance are robust to heteroskedasticity and autocorrelation in the residuals. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

Table 6.7 reports the explanatory weight of each measure on the illiquidity risk premium.

The AI measure is the one that has the major weight among the others with very high values, which are also highly significant at any possible level. This result renew the theory of Amihud that even in

HFT, stocks can be differently priced due to their illiquid behavior. Obviously, lower is the value more liquid is the instrument. This measure reflect that firm shares were often considered as liquid instruments along the time window. However, even highly traded securities like these cannot be considered always liquid since some exogenous events, that may affect the market or the industry in which a firm operates, can change their usual schemes.

Since RBA is a dimensionless measure, it represents the mere illiquid aspect that characterizes each firm instrument. All the coefficients are positive, except for CAT. The positive relation is expected since greater is the bid-ask spread more volatile (illiquid) is the referred stock, however the negative outcome presented by CAT may imply that this firm price is inversely related to the usual bid-ask spread–volatility relation. This feature may be due to the industry in which it operates, or by the fact that this has been, and is, the cheapest stock firm.

Liquidity measures are commonly negatively related to the premium that an investor would like to earn if he holds an illiquid instrument. MI re-validates this link, with significant outcomes, and has also a slight weight on the determination of the illiquidity risk premium. This link is characterized by an inverse which implies that more illiquid, volatile, is the object lower, even in negative terms, lower is the load factor on this measure.

Chapter 7

Conclusions

This work presents new tests on HFT asset return and liquidity measure. It is known in the literature, that illiquidity explains differences in expected returns across stocks, which holds also in this case. As Amihud (2002), this work provides that stock illiquidity affects the stock excess return. The "compensation" for the liquidity risk that these stocks generate, also known as "illiquidity risk premium", accounts in his computation also the lower liquidity respect to Treasuries. The premiums tend to vary over time as a liquidity function.

The illiquidity measures employed in this study are categorized in a two-fold classes: high (HF) and low frequency (LF) measures. While HF are more consistent and accurate, compared to the latter, LF are easier to compute and gives similar results with a minimal loss in the accuracy process. HF are made up by: *Realized Variance*, intra-day variation of stock return, *Realized Liquidity Variation* (i.e. realized variance of the high frequency relative spread) and \log *intra-day high frequency volume*. LF comprehend: the *Amihud Illiquidity (AI)* measure, ratio of a stock absolute daily return to its daily dollar volume, averaged over the corresponding period, which can be interpreted as the daily stock price reaction to a dollar of trading volume, the *relative bid-ask spread (RBA)* measure, ratio of the common bid-ask spread over stock mid-range price, which points out that wider is the measure more illiquid is the stock, and a "new-developed" liquidity measure based on the Mid-range price, the *Mid-range liquidity (MI)* measure, which is the ratio of the daily mid-range price and the correspondent daily trading volume.

Before computing the relevant liquidity tests on these, some stationarity test have been used over the bid and ask prices, which follow a cointegrated process such as, their combination, seems to be stationary. All the measures holds the hypothesis of a stationary process, for each, and the combination of price processes are not likely to be cointegrated.

Tests on the accountability of the two-class frequency measures have been run. LF outperform HF in explaining the illiquidity measure, bid-ask spread, which indicate that both AI and RBA have a positive impact on the spread while MI has a negative one. However, by adding also HF in the regression does not consistently improve the explanation power, but further stress the negative impact that MI has on the illiquidity measure.

Furthermore, since it has been verified that LF gives consistent and, mostly, accountable measure of illiquidity, it has been analyzed the impact of the above stated measures on the illiquidity risk premium and how these help to explain the "illiquidity reward". It is not new that illiquidity measures result with a positive relation while the liquid one has a negative link. The plainness of these measure should not be topic of claims against their effectiveness. Greater is AI, more the stock will be subject to market

manipulation, by entering the market with a large position with the referred instrument. This manner is due to the illiquidity behavior of the stock itself, i.e. larger is the ratio, lower is the respective trading volume which implies a larger illiquidity attitude. While AI reflected the illiquidity topic directly, MI implied it in a reverse sense. The latter reports that a liquid instrument should be characterized by a larger (i.e. smaller in absolute value) factor weight.

The illiquidity premium takes into account also the possible easiness of liquidating the position in a short amount of time. This is stressed by the effects that stock excess return differs across stocks by their liquidity or size due to their illiquidity behavior over time. This illiquidity feature has been found with a stronger relevance, particularly, on small stocks.

Hence, the results suggest that the stock excess return, usually referred to as "risk premium" (in this instance "illiquidity risk premium"), is partially a compensation for stock illiquidity measure with low frequency measure. This explains why some stock are characterized by high equity premiums. The results mean that stock excess returns reflect not only the higher risk but also the lower liquidity of stock compared to Treasury securities.

References

- [1] Abdi, F., Ranaldo, A. (2017). A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices. *The Review of Financial Studies*, 30: 4437-4480.
- [2] Acharya, V., Pedersen, L. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77(2): 375-410.
- [3] Admati, A.R. Pfleiderer, P. (1988). A Theory of Intraday Trading Patterns. *Review of Financial Studies*, 1: 3-40.
- [4] Aït-Sahalia, Y., Yu, J. (2009). High Frequency Market Microstructure Noise Estimates and Liquidity Measures. *The Annals of Applied Statistics*, 3(1): 422-457.
- [5] Alabed, M., Al-Khouri, R. (2009). The pattern of intraday liquidity in emerging markets: The case of the Amman Stock Exchange. *Journal of Derivatives & Hedge Funds*. 14: 265-284.
- [6] Amiram, D., Cserna, B., Levy, A. (2016). Volatility, Liquidity, and Liquidity Risk. Columbia Business School, University of Frankfurt, Ben-Gurion University.
- [7] Amihud, Y., Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17: 223-249
- [8] Amihud, Y., Mendelson, H., Murgia, M. (1990) Stock market microstructure and return volatility: Evidence from Italy. *Journal of Banking and Finance*, 14: 423-440.
- [9] Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5: 31-56.
- [10] Amihud, Y., Mendelson, H., Pedersen, L. (2005) Market microstructure and asset pricing. *Stern School, New York University*.
- [11] Andersen, T. G., Bollerslev, T., Diebold, F. X., Ebens H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61: 43-76.
- [12] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2001). (Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation. *RISK*, 11.
- [13] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2000b). Exchange Rate Returns Standardized by Realized Volatility are (nearly) Gaussian. *Multinational Finance Journal*, 4: 159-179.
- [14] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association*, 96: 42-55.

- [15] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2002). Modeling and Forecasting Realized Volatility. *Econometrica*, 71: 529-626.
- [16] Andersen, T. G., Bollerslev, T., Meddahi, N. (2005). Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities. *Econometrica*, 73: 279-296.
- [17] Andersen, T. G., Bollerslev, T., Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 160: 220-234.
- [18] Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22: 327-351.
- [19] Anderson, T.W., Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20: 46-63.
- [20] Anderson, T.W. (2001). Reduced rank regression in cointegrated models. *Journal of Econometrics*, 106: 203-216.
- [21] Bandi, F.M., Perron, B. (2006). Long memory and the relation between realized and implied volatility. *Journal of Financial Econometrics*, 4: 636-670.
- [22] Bandi, F.M., Phillips, P.C.B. (2007). A simple approach to the parametric estimation of potentially-nonstationary diffusion processes. *Journal of Econometrics*, 137: 354-395.
- [23] Bandi, F.M., Russell, J.R. (2008). Microstructure Noise, Realized Variance, and Optimal Sampling. *The Review of Economic Studies*, 75(2): 339-369
- [24] Bandi, F.M., Russell, J.R. (2006a). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79: 655-692.
- [25] Bandi, F.M., Russell, J.R. (2006b). Comment on Hansen and Lunde. *Journal of Business and Economic Statistics*, 24: 167-173.
- [26] Bandi, F.M., Russell, J.R. (2006c). Volatility. In J.R. Birge and V. Linetski (Eds.) *Handbook of Financial Engineering*. Elsevier North-Holland. Forthcoming.
- [27] Bandi, F.M., Russell, J.R. (2006c). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Working paper.
- [28] Bandi, F.M., Russell, J.R., Yang, C. (2007). Realized volatility forecasting in the presence of time-varying noise. Working paper.
- [29] Barndorff-Nielsen, O.E., Shephard N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64: 253-280.
- [30] Barndorff-Nielsen, O.E., Shephard N. (2003). How accurate is the asymptotic approximation to the distribution of realized variance?. In Donald W.F. Andrews, James L. Powel, Paul L. Ruud, and James H. Stock (Eds.) *Identification and Inference for Econometric Models (festschrift for Thomas J. Rothenberg)*. Cambridge University Press. Forthcoming.

- [31] Barndorff-Nielsen, O.E., Shephard N. (2004). Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72: 885-925.
- [32] Barndorff-Nielsen, O.E., Shephard N. (2006). Variation, jumps, market frictions and high-frequency data in financial econometrics. In R. Blundell, P. Torsten and W. K. Newey 39 (Eds.) *Advances in Economics and Econometrics. Theory and Applications*. Ninth World Congress. Econometric Society Monographs, Cambridge University Press.
- [33] Beekers, S. (1983). Variance of security price returns based on high, low, and closing prices. *Journal of Business*, 56: 97-112.
- [34] Berkman, H., Eleswarapu, V.R. (1998). Short-term traders and liquidity: a test using Bombay stock exchange data. *Journal of Financial Economics* 47: 339-355.
- [35] Bessembinder, H. (1994). Bid-ask spreads in the interbank exchange markets. *Journal of Financial Economics*, 35: 317-348.
- [36] Bissoondoyal-Bheenick, E., Brooks, R., Treepongkaruna, S., Wee, M. (2016). Realized Volatility of the Spread: An Analysis in the Foreign Exchange Market, *Sabri Boubaker, Bonnie Buchanan, and Duc Khuong Nguyen, eds., Risk Management in Emerging Markets (Emerald Group Publishing)*, 3-35.
- [37] Black, S. W. (1991). Transactions costs and vehicle currencies. *Journal of International Money and Finance*, 10: 512-526.
- [38] Bollerslev, T., Domowitz, I. (1993). Trading Patterns and Prices in the Interbank Foreign Exchange Market. *Journal of Finance*, 48: 1421-1443.
- [39] Bollerslev, T., Melvin, M. (1994). Bid-ask spreads and volatility in the foreign exchange market. *Journal of International Economics*, 36: 355-372.
- [40] Bollerslev, T., Domowitz, I., Wang, J. (1997). Order flow and the bid-ask spread: An empirical probability model of screen-based trading. *Journal of Economic Dynamics and Control* 21(8-9): 1471-1491
- [41] Boothe, P. (1988). Exchange rate risk and the bid-ask spread: A seven country comparison. *Economic Inquiry*. 26(3): 485-492.
- [42] Bossaerts, P., Hillion, P. (1991). Market Microstructure Effects of Government Intervention in the Foreign Exchange Market. *Review of Financial Studies*, 4: 513-541.
- [43] Breen, W.J., Hodrick, L.S., Korajczyk, R.A. (2002). Predicting equity liquidity. *Management Science*, 48: 470-483.
- [44] Breedon, F., Ranaldo, A. (2013). Intraday patterns in FX returns and order flow. *Journal of Money, Credit and Banking*, 45: 953-965.
- [45] Calamia, A. (1999). Market Microstructure: Theory and Empirics. LEM Working Paper, Pisa, 19

- [46] Campbell, J.Y., Lo, A.W., MacKinlay A.C. (1997). The Econometrics of Financial Markets. *Princeton University Press*.
- [47] Campbell, J.Y., Mankiw, N.G. (1987). Permanent and transitory components in macroeconomic fluctuations. *American Economic Review*, 77: 111-117.
- [48] Carmona, R., Webster, K. (2017). The microstructure of high frequency markets.
- [49] Charles, A., Darné, O. (2009). Variance ratio tests of random walk: An overview. *Journal of Economic Surveys*, Wiley, 23(3): 503-527.
- [50] Chordia, T., Roll, R., Subrahmanyam, A. (2001). Market liquidity and trading activity. *Journal of Finance*, 56: 501-530.
- [51] Cochrane, J. H. (1988). How big is the random walk in GNP?. *Journal of Political Economy*, 96: 893-920.
- [52] Cogley, T. (1990). International evidence on the size of the random walk in output. *Journal of Political Economy*, 98: 501-518.
- [53] Corsi, F. (2005). Measuring and Modeling Realized Volatility: from Tick-by-tick to Long Memory, Ph.D. thesis, University of Lugano.
- [54] Corwin, S. A., Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance*, 67: 719-759.
- [55] Cuonz, J. (2018). Essays on the Dynamics of Market and Liquidity Risk. Dissertation of the University of St. Gallen
- [56] Danyliv, O., Bland, B., Nicholass, D. (2014). Convenient liquidity measure for Financial markets. *Papers*, 1412.5072
- [57] Dickey, D. A., W. A. Fuller. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74: 427-431.
- [58] Dickey, D. A., W. A. Fuller. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica*, 49: 1057-1072.
- [59] Domowitz, I., and El-Gamal, M. A. (1999). Financial market liquidity and the distribution of prices. *Social Sciences Research Network*.
- [60] Duffee R., Stanton G. R. (2008). Evidence on Simulation Inference for Near Unit-Root Processes with Implications for Term Structure Estimation. *Journal of Financial Econometrics*, 6: 108-142.
- [61] Enders, W. (1995). Applied Econometric Time Series. *Hoboken, NJ: John Wiley & Sons, Inc.*
- [62] Engle, R. F., Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55: 251-276.
- [63] Engle, R.F., Kozicki, S. (1993). Testing for common factors (with comments). *Journal of Business Economics & Statistics*, 11: 278-369.

- [64] Engler, M., Jeleskovic, V. (2016). Intraday volatility, trading volume and trading intensity in the interbank market e-MID. *Philipps-Universität Marburg*.
- [65] Evans, M. D. D., Lyons, R. K. (2002). Order flow and exchange rate dynamics. *Journal of Political Economy*, 110: 170-180.
- [66] Evans, M. D. D., Rime, D. (2016). Order flow information and spot rate dynamics. *Journal of International Money and Finance*, 69: 45-68.
- [67] Fama, E. F. (1968). Risk, Return and Equilibrium: Some Clarifying Comments. *Journal of Finance*, 23(1): 29-40.
- [68] Fama E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25: 383-417.
- [69] Fama, E. F. (1990). Stock returns, expected returns, and real activity. *Journal of Finance*, 45: 1089-1108.
- [70] Fama E. F. (1991). Efficient capital markets: II. *Journal of Finance*, 46: 1575- 1617.
- [71] Fama, E.F., French, K.R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25: 23-49.
- [72] Fama, E.F., French, K.R. (1992). The cross section of expected stock returns. *Journal of Finance*, 47: 427-465.
- [73] Fama, E.F., MacBeth, J.D. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, 81: 607-636.
- [74] Fong, K., Holden, C. W., Trzcinka., C. A. (2017). What are the best liquidity proxies for global research?. *Review of Finance*, 21(4): 1355-1401.
- [75] Frederick, H. deB. Harris, McInish, T., Shoesmith, G., Wood, R. (1995). Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets. *The Journal of Financial and Quantitative Analysis*, 30(4): 563-579.
- [76] French, K.R., Schwert, G.W., Stambaugh, R.F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19: 3-29.
- [77] Foucault, T., Pagano, M., Roell, A. (2013). Market Liquidity: Theory, Evidence, and Policy. *Oxford University Press*.
- [78] Gargano, A., Riddiough, S.J., Sarno, L. (2017). The value of volume in foreign exchange. Working paper.
- [79] Gargano, A., Riddiough, S.J., Sarno, L. (2018). Volume and Excess Returns in Foreign Exchange. Working paper.
- [80] Garman, M. B. (1976). Market microstructure. *Journal of Financial Economics*, 3: 257-275.

- [81] Garman, M. B., Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53: 67-78.
- [82] Gençay, R., Ballochi, G., Dacorogna, M. Olsen, R. and Pictet, O. (2002). Real-Time Trading Models and the Statistical Properties of Foreign Exchange Rates. *International Economic Review*, 43(2): 463-492.
- [83] Glassman, D. (1987). Exchange rate risk and transactions costs: Evidence from bid-ask spreads. *Journal of International Money and Finance*, 6: 479-490.
- [84] Goyenko, R. Y., Holden, C. W., Trzcinka, C. A. (2009). Do liquidity measures measure liquidity?. *Journal of Financial Economics*, 92: 153-81.
- [85] Granger, C. W. J. (1983). Cointegrated variables and error correction models. *UCSD Discussion paper*, 83-13a.
- [86] Guloglu, Z. C., Ekinçi, C. (2016) A comparison of bid-ask spread proxies: evidence from Borsa Istanbul futures. *Journal of Economics, Finance and Accounting*, 3(3): 244-254.
- [87] Haldane, A. (2011). The race to zero. *Bank of England speeches, given to the International Economic Association 16th World Congress, July 8*.
- [88] Hamilton, J.D. (1994) Time Series Analysis. *Princeton University Press*.
- [89] Hansen, P. R., Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24: 127-161.
- [90] Hansen, B. E. (2018). Johansen's Reduced Rank Estimator Is GMM. *Department of Economics, University of Wisconsin, Madison*
- [91] Hasbrouck, J. (1999). The Dynamics of Discrete Bid and Ask Quotes. *The Journal of Finance*, 54(6): 2109-2142.
- [92] Hasbrouck, J. (2004). Liquidity in the futures pits: Inferring market dynamics from incomplete data. *Journal of Financial and Quantitative Analysis*, 39: 305-26.
- [93] Hasbrouck, J. (2007). Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading. *Oxford University Press*, 1
- [94] Hasbrouck, J. (2009). Trading costs and returns for US equities: the evidence from daily data. *Journal of Finance*, 64: 1445-77.
- [95] Hasbrouck, J. (2018). High-Frequency Quoting: Short-Term Volatility in Bids and Offers. *Journal of Financial and Quantitative Analysis*, 53(2): 613-641.
- [96] Hasbrouck, J., Sofianos, G. (1993). The trades of market makers: an empirical analysis of NYSE specialists. *Journal of Finance* 48/5: 1565-1593.
- [97] Hendry, D. F., Juselius, K. (2001). Explaining Cointegration Analysis: Part II. *The Energy Journal*, 22(1): 75-120.

- [98] Hjalmarsson, E., Osterholm, P. (2007). Testing for Co-integration Using the Johansen Methodology when Variables are Near-Integrated. *IMF Working Paper*, 07(141): 22-27.
- [99] Holden, C.W., Jacobsen, S., Subrahmanyam, A. (2014). The Empirical Analysis of Liquidity. *Foundations and Trends*, 8(4): 1-102.
- [100] Hwang, S., Satchell, S. (2000). Market Risk and the Concept of Fundamental Volatility: Measuring Volatility Across Asset and Derivative Markets and Testing for the Impact of Derivatives Markets on Financial Markets. *Journal of Banking & Finance*, 24(5): 759-785.
- [101] Jacoby, G., Fowler, D. J., Gottesman, A. A. (2000) The capital asset pricing model and the liquidity effect: A theoretical approach. *Journal of Financial Markets*, 3: 69-81.
- [102] Jacod, J. (1994). Limit of random measures associated with the increments of a Brownian semi-martingale. Working paper.
- [103] Jacod, J. and P. Protter (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability*, 26: 267-307.
- [104] Johansen, S. (1996). Likelihood-based inference in cointegrated vector autoregressive models. *Oxford University Press, Oxford*.
- [105] Johansen, S. (2002a). A small sample correction of the test for cointegrating rank in the vector autoregressive model. *Econometrica*, 70: 1929-1961.
- [106] Johansen, S., 2002b. A small sample correction for tests of hypotheses on the cointegrating vectors. *Journal of Econometrics*, 111: 195-221.
- [107] Johansen, S. (2004a). The interpretation of cointegrating coefficients in the cointegrated vector autoregressive model. *Forthcoming Oxford Bulletin of Economics and Statistics*.
- [108] Johansen, S. (2009). Cointegration: Overview and Development. *Financial Time Series, Springer*.
- [109] Johansen, S. (2014). Times Series: Cointegration. *CREATES Research Paper*, 2014-2038.
- [110] Karnaukh, N., Ranaldo, A., Söderlind, P. (2015). Understanding FX liquidity. *Review of Financial Studies*, 28(11): 3073-3108.
- [111] Khan, W.A., Baker, H.K. (1993). Unlisted trading privileges, liquidity and stock returns. *Journal of Financial Research* 16: 221-236.
- [112] Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica* 53: 1315-1335.
- [113] Lesmond, D. A., Ogden, J. P., Trzcinka, C. A. (1999). A new estimate of transaction costs. *Review of Financial Studies*, 12: 1113-41
- [114] Llorente, G., Michaely, R., Saar, G., Jiang, W. (2002). Dynamic volume-return relation of individual stocks. *Review of Financial Studies*, 15: 1005-1047.
- [115] Lou, X. (2017). Price Impact or Trading Volume: Why is the Amihud (2002) Illiquidity Measure Priced?. *The Review of Financial Studies*, 30(12): 4481-4520.

- [116] Mancini, L., Ranaldo, A., Wrampelmeyer, J. (2013). Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *Journal of Financial Markets*, 68: 1805-41.
- [117] Melvin, M., Tan, K.-H. (1996). Foreign exchange market Bid-Ask spreads and the market price of social unrest. *Oxford Economic Papers*, 48: 329-341.
- [118] Mullins Jr., D.W. (1982). Does the capital asset pricing model work?. *Harvard Business Review*: 1-2: 105-113.
- [119] Muranaga, J., Shimizu, T. (1999). Market Microstructure and Market Liquidity. *Bank for International Settlements, Market Liquidity: Research Findings and Selected Policy Implications*, 11: 1-28.
- [120] Nagel, S. (2012). Evaporating liquidity. *Review of Financial Studies*, 25: 2005-2039.
- [121] Newey, W. K., West. K. D. (1987). A Simple Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55: 703-708.
- [122] Newey, W. K., K. D. West. (1994). Automatic Lag Selection in Covariance Matrix Estimation." *The Review of Economic Studies*. *Econometrica*, 61: 631-653.
- [123] O'Hara, M. (1997). Market Microstructure Theory. *Blackwell, London*.
- [124] O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2): 257-270.
- [125] Oomen, R. A. C. (2001). Using high-frequency data stock market index data calculate, model and forecast realized return variance. *European University Institute, Economics Discussion Paper*, No. 2001/6.
- [126] Oomen, R. A. C. (2002). Modeling Realized Variance when Returns are Serially Correlated. Manuscript, Warwick Business School.
- [127] Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53: 61-65.
- [128] Ranaldo, A. (2000). Intraday Trading Activity on Financial Markets: The Swiss Evidence. PhD thesis, University of Fribourg.
- [129] Ranaldo, A., Santucci de Magistris, P. (2019). Trading Volume, Illiquidity and Commonalities in FX Markets.
- [130] Roch, A., Soner, H.M. (2013). Resilient Price Impact of Trading and the Cost of Illiquidity. *International Journal of Theoretical and Applied Finance*, 16: 1-27.
- [131] Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39: 1127-1139.
- [132] Said, S. E., Dickey D. A. (1984). Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. *Biometrika*, 71(3): 599-607.

- [133] Schwartz, R.A., Francioni, R. (2004). *Equity Markets in Action: The Fundamentals of Liquidity, Market Structure & Trading*. John Wiley & Sons.
- [134] Sims, C., Stock, J., Watson, M. (1990). Inference in Linear Time Series Models with Some Unit Roots. *Econometrica*, 58: 113-144.
- [135] Sjö, B. (2010). Testing for Unit Roots and Cointegration 2 1 Unit Root Tests: Determining.
- [136] Stock, J. H. (1987). Asymptotic properties of least squares estimates of cointegration vectors. *Econometrica*, 55: 1035-1056.
- [137] Treynor, Jack L. (1962). Toward a Theory of Market Value of Risky Assets.
- [138] Vayanos, D., Wang, J. (2012). Liquidity and asset prices under asymmetric information and imperfect competition. *Review of Financial Studies*, 25: 1339-1365.
- [139] Wang, J. (1994). A model of competitive stock trading volume. *Journal of Political Economy*, 102: 127-168.
- [140] White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48: 817-838.
- [141] Zhang, F. (2010). High-Frequency Trading, Stock Volatility, and Price Discovery. *SSRN Electronic Journal*.

Summary

Liquidity and trading activity are important features of present financial markets, but little is known about their evolution over time or about their time-series determinants. This is caused by the lack of data provision of U.S. stock markets, which has supplied this data information only on some assets for roughly four decades. Their fundamental importance has been represented by the influence that they have on trading costs on specific returns (Amihud and Mendelson, 1986; Jacoby, Fowler, and Gottesman, 2000) which implies a direct link between liquidity and corporate costs of capital. More generally, exchange organization, regulation, and investment management could all be improved by improving the knowledge of factors that influence liquidity and trading activity. Understanding the scheme of unobservable features of the market such as the magnitude of returns to liquidity provision (Nagel, 2012), the impact that asymmetric information has on prices (Llorente et al., 2002), and if trades are more likely to have a permanent or transitory impact on price (Wang, 1994; Vayanos and Wang, 2012) should increase investor confidence in financial markets and thereby enhance the efficacy of corporate resource allocation.

At present, there are many measures that are good proxies that help to understand stocks liquidity behavior. However, these are based on microstructure data which consist in high frequency trading (HFT) data. HFT data are characterized by short time horizons, in which position are opened and closed, even in fractions of second. These estimate are indeed, usually, very accurate to understand the contemporaneous trading activity. Nevertheless, these are characterized by complex computational process, assumptions on prices, in some instances, must be accounted and, mostly, they require huge amount of data for each stock that is not available for many, even for listed ones.

Conversely, low frequency measures, measure based on a larger horizon that the previous, focus on analyzing and establish measures based with, usually, a daily frequency. These indeed are easier to compute and require less information to be accounted in the analysis, which is also often readily available for many instruments, listed or not. However, this lack of information gathering within observations may lead to inaccuracy in their explanatory power for the examined topic.

Even so, among the liquidity literature many measures have been developed that give consistent estimate, compared to high frequency sizes. The aim of this work is to rely on low frequency measures, which can be computed with daily data that can be easily obtained for almost all stocks, in the explication of the illiquidity risk premium (premium that an illiquid stock should compensate to their holder in exchange for the risk that the stockholder bears by holding the referred instrument). Illiquid stocks are characterized by larger illiquid premiums, the whom are incorporated in the usual concept of market risk premium. Hence, riskier (more illiquid) stocks return higher premiums.

This work hypothesis is tested by running a regression model

$$(r_{M,t} - r_{f,t}) = \gamma_0 + \gamma_1 AI_{t-1} + \gamma_2 RBA_{t-1} + \gamma_3 MI_{t-1},$$

where r_M and r_f are, respectively, the stock market return and the risk free rate, AI is the Amihud illiquidity measure, RBA is the relative bid-ask spread and MI is a "new-developed" measure that accounts stock liquidity. Since, the firms' stocks are all from US companies, r_M is the daily return taken from the S&P500 and as risk free the three-months Treasury bill rates.

Before dealing with the core hypothesis of this work, tests on measure stationarity have been made and also on the explanatory power of both frequency measure is accounted, for the purpose of "declaring" if the daily frequency can be a reliable estimators of illiquidity.

Liquidity measures are computed over four HFT stocks from different industries. These are: Walmart (WMT), The Coca-Cola company (KO), JPMorgan (JPM) and Caterpillar (CAT). The event window that has been chosen to collect data widens from January 29, 2001 to June 29, 2018. As time window it has been chosen to analyzed the period that starts right after the burst of the Dot-com bubble in order to observe their evolution and, respective, firm net-implementation during the years until the end of the first half 2018.

Firstly, the measures are briefly exposed. The presentation starts by dealing with Roll (1984) measure, which was one of the first pioneer in this field that developed a spread measure that has been used over and over, with its own limitations. Many scholars modeled from this finding in order to overcome the measure limits. Abdi and Ranaldo "craft" a measure, which implicitly relies on the same assumptions made by Roll, denoted as Mid-range price made as a trivial average price of the high and low prices. There is, in addition, also the Amihud (2002) illiquidity measure formulated as the daily ratio of absolute stock return to its dollar volume, averaged over some period. This can be viewed as the daily price response associated with one dollar of trading volume, which roughly reflect the price impact. This measure has acquired some attention in the literature and the final test made in this work is based on the same idea made by Amihud. Other measures have been established in the academia by Hasbrouck (2004, 2009), who proposes a Gibbs sampler Bayesian estimation of the Roll model; Lesmond, Ogden, and Trzcinka (1999) that introduce an estimator based on zero returns. Following the same line of reasoning of this, Fong, Holden, and Trzcinka (2017) formulate a new estimator that simplifies the existing the previous measures. Holden (2009), jointly with Goyenko, Holden, and Trzcinka (2009), introduces the Effective Tick measure based on the concept of price clustering. High and low prices have been usually used to proxy volatility (Garman and Klass, 1980; Parkinson, 1980; Beckers, 1983). Corwin and Schultz (2012) use them to put forward an original estimation method for transaction costs where they assume that the high (low) price are buyer- (seller-) initiated, and their ratio can be disentangled into the efficient price volatility and bid-ask spread. However, these have always been treated in the literature and some flaws can be accounted to these.

Secondly, high frequency measure are exposed with the related literature. These require the gathering of microstructure data. Market microstructure has been analyzed for almost four decades since it has been a game changer in the trading industry. Jacod (1994) and Jacod and Protter (1998) were one of the first studies in this new field and they introduced the new concept of realized variance, which is the realized counterpart of the conceptual variance. Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001 and 2002) re-arrange the realized measure and model it from short-term price changes, which treat volatility as observable, rather than as a latent variable. This has been proved to be an useful estimate of fundamental integrated volatility by Hasbrouck (2018), Hwang and Satchell (2000), Zhang (2010). However, Hansen and Lunde (2006) pointed out that, when dealing with market microstructure, the presence of market microstructure noise in high

frequency data complicates the estimation of financial volatility and makes standard estimators, such as the realized variance, unreliable. So market microstructure noise challenges the accountability of theoretical results that rely on the absence of noise. The best remedy for market microstructure noise depends on the properties of the noise. The time dependence in the noise and the correlation between noise and efficient price arise naturally in some models on market microstructure effects, including (a generalized version of) the bid-ask model by Roll (1984) (Hasbrouck, 2004) and models where agents have asymmetric information (Glosten and Milgrom, 1985 and Easley and O'Hara, 1987, 1992). Market microstructure noise has many sources, including the discreteness of the data (see Harris 1990, 1991) and properties of the trading mechanism (Black, 1976; Amihud and Mendelson, 1987 and O'Hara, 1995). Other studies were focused on one of the most used sizes that plainly represents illiquidity: the bid-ask spread. Particularly, starting from Amihud and Mendelson's (1986) liquidity study, it has been running several researches on the evolution of the volatility of this size. Studies on liquidity volatility are found mostly on foreign exchange bid-ask spreads. Glassman (1987) and Boothe (1988) study the statistical properties of the bid-ask spread, Bollerslev and Melvin (1994) evaluate the distribution of bid-ask spreads and their ability to explain foreign exchange rate volatility based on a tick-by-tick data set. Furthermore, Andersen et al. (2001) and Cuonz (2018) model a realized variance measure on the intra-day bid-ask spread which satisfies the realized variation assumptions and gives consistent results, even if stock returns are not accounted.

Thereafter, a presentation of the theory behind realized variance, and its link with market microstructure, and stationarity tests has been made, for the purpose of giving the main key concepts that allow to understand the pattern of these sizes.

On one hand, low frequency measures are computed as follows:

$$RBA_t = \frac{BAS_t}{\eta_t}. \quad (7.1)$$

where BAS_t is the bid-ask spread and η_t is the log mid price of the respective stock at time t .

$$AI_t = \frac{1}{N} \sum_{t=1}^N \frac{|r_t|}{Vol_t}, \quad (7.2)$$

where r_t is the stock return over a day in a year, Vol_t is the respective daily volume and N is the number of days for which trading data are available.

$$MI_t = \frac{1}{N} \sum_{t=1}^N \frac{\eta_t}{Vol_t}, \quad (7.3)$$

where Vol_t is the daily log volume in the t^{th} day.

On the other hand, high frequency measures are computed as:

$$RV_t \equiv \sum_{j=1}^{N_f} r_{t,f,j}^2, \quad (7.4)$$

where $r_{t,f,j}$ is the j^{th} intra-day return over the N_f time horizons.

$$lVol_t = \log(Vol_t) = \log\left(\sum_{i=1}^{mT} NST_i\right), \quad (7.5)$$

where NST is the number of share traded intra-daily.

$$lRBA_h(t; m) = \sum_{i=1}^{mh} \Delta BA_{(m)}^2\left(t - h + \frac{i}{m}\right), \quad (7.6)$$

Firstly, non-stationary-based tests computed over the measures confirms that these neither follow a random walk or an explosive process.

	WMT	KO	JPM	CAT
Low frequency				
AI	0.5886	0.5906	0.6107	0.5764
RBA	0.7991	0.8104	0.7887	0.8428
MI	0.9998	0.9998	0.9998	0.9998
High frequency				
RV	0.9998	0.9998	0.9998	0.9998
IRBA	0.9227	0.9300	0.9152	0.9328
IVol	0.9227	0.9300	0.7652	0.8189

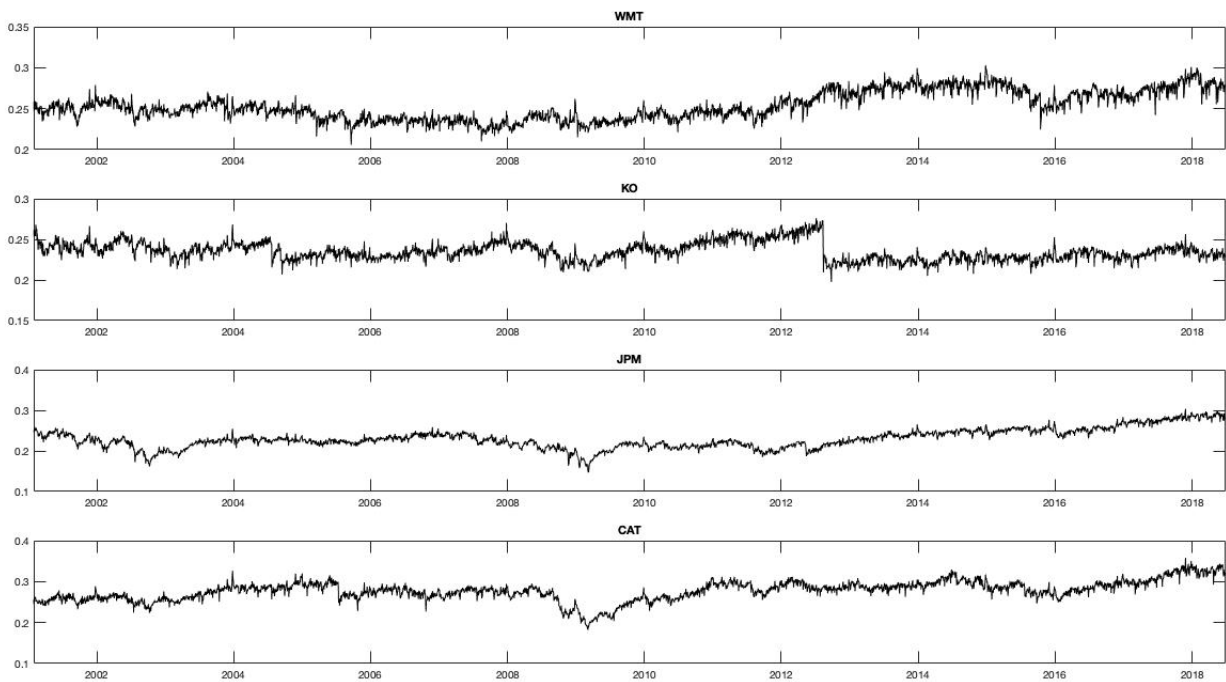
ADF test coefficients.

	WMT	KO	JPM	CAT
Low frequency				
AI	0.5078	0.4887	0.5105	0.4671
RBA	0.5304	0.5286	0.5615	0.5521
MI	0.6912	0.6781	0.7476	0.6876
High frequency				
RV	0.5218	0.4993	0.6943	0.5534
HFRBA	0.5444	0.5382	0.5325	0.5803
IVol	0.6758	0.6552	0.7014	0.6708

VR test.

Outputs give significant results for all the measures, but for the RV and MI ones exhibits a near-unit-root result. However, since this results may lead us to doubt about the reliability of ADF test, VR confirms the rejection of the unit root by giving very low values, which report that all process are stationary. Therefore, both Table 6.1 and 6.2 show that both ADF and VR test gives

significant results for all measure, that allows to refuse the hypothesis of the presence of unit-root in the time series process of each size.

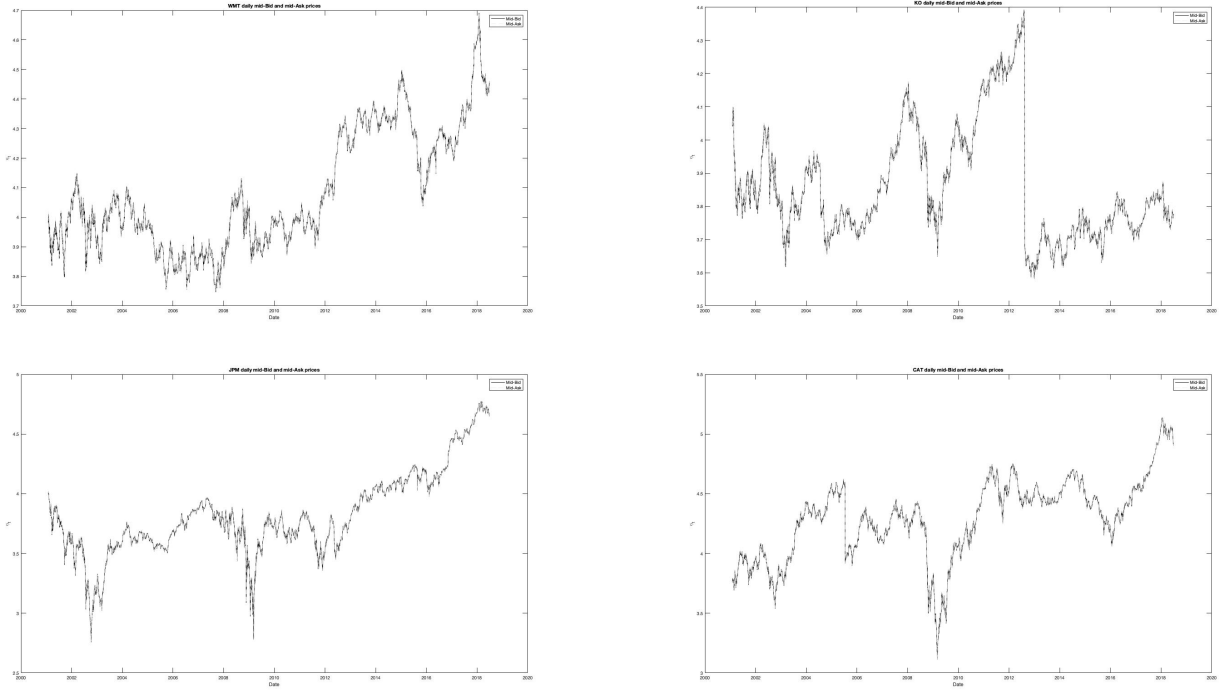


Stocks Mid-range liquidity series.

Time series path of MI during the event period exposes that, even through rough illiquid sub-periods, the index stayed pretty stable for all the instrument within the boundary $[0.2; 0.35]$, which may result in a difference stationary process.

As Frederick (1995) and Bollerslev (1997) discovered, often, bid and ask prices are cointegrated. To analyze their behavior, and thanks to the previous tests on the data-set, it is used the mid-range bid and ask price of each day to analyze their movement pattern.

Usually bid and ask prices are cointegrated, so it is beneficial to study their evolution and variation over the event window.



Daily stock bids and asks.

Bid and ask prices, of firms shares, at first sight are likely to follow a similar price process and it seems to be non-stationary. Johansen test is then applied to verify this assessment, and to find a possible combination of these which is stationary.

Rank	BA_{WMT}		BA_{KO}	
	Stat	p-value	Stat	p-value
$r = 0$	601.2412 ^a	0.0010	480.3424 ^a	0.0010
$r = 1$	9.0393 ^a	0.0034	8.8578 ^a	0.0037
Stationary process test				
ADF	-6,4634	0.0010	-22.7678	0.0010
$ 1 + \alpha'\beta $	0.7473		0.7963	

Johansen cointegration test of stocks B–A Spread.

Rank	BA_{JPM}		BA_{CAT}	
	Stat	p-value	Stat	p-value
$r = 0$	412.4965 ^a	0.0010	898.0167 ^a	0.0010
$r = 1$	11.8115 ^a	0.0010	6.8392 ^a	0.0092
Stationary process test				
ADF	-19.7918	0.0010	-23.2897	0.0010
$ 1 + \alpha'\beta $	0.8257		0.6320	

Johansen cointegration test of stocks B–A Spread.

All variables are taken with logs. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

Johansen cointegrating test for all the stocks exploits a strong evidence to reject the null hypothesis (cointegrating rank equal to zero) in favor of the alternative. Although the null hypothesis is rejected, the coefficient vectors of the decomposed matrix have been analyzed to verify that the alternative hypothesis holds. To verify the stationarity of the process it has been used a test based on the verification of the unit root, Augmented Dickey-Fuller (ADF), and the assessment that the product of the two matrices lies inside the unit disk. The outcomes validate the finding of alternative hypothesis. So it can be stated that bid and ask prices of these stocks are cointegrated.

However, both the bid and ask prices, as noted in the literature, are non-stationary. Nevertheless, by their linear combination results to be stationary, which allows to state that these are cointegrated. Despite this, their mid-price is not cointegrated, and their combination does not satisfy the stationarity requirements. Any possible combination within the stocks has been made but these seems to not satisfy either the stationary conditions. Anyway, this result should be expected due to the low relationship that resides within the firm's industries.

Then, tests on the relation that elapses with, one of the most accountable illiquidity measures, bid-ask spread and low and high frequency measure prove that the low ones seem to be good estimates while the latter give good proxies, but with less explanatory power than the daily ones. Anyhow, a "comprehensive" test with all variables exposes a slight significance to the previous ones, but stressed the role of the "new-developed liquidity measure". Measures are categorized, firstly, in low and high frequency as follows

$$LF = [AI, RBA, MI, D_{MON}, D_{FRI}],$$

$$HF = [RV, lRBA, lVol, D_{MON}, D_{FRI}],$$

where other two dummy variables are added in the measures, D_{MON} and D_{FRI} . These take value of one, respectively, if the trading day is a Monday or a Friday, to study also any possible week seasonality. Thereafter, it will be run a regression to state which frequency is the one that best estimates the illiquidity size. The estimates would be obtained by running an OLS regression

$$BAS = \alpha_{LF} + LF\beta_{LF},$$

$$BAS = \alpha_{HF} + HF\beta_{HF}.$$

	WMT	KO	JPM	CAT
Low frequency				
AI	13.4344 ^a (4.45)	5.4209 ^c (2.76)	2.7611 ^c (1.68)	-2.7682 ^b (-2.01)
RBA	2.7841 ^a (17.81)	2.1671 ^a (21.94)	2.7461 ^a (12.01)	2.0728 ^a (26.86)
MI	0.0692 ^b (2.31)	-0.2839 ^a (-7.62)	0.1495 ^a (3.73)	-0.3224 ^a (-7.86)
D_{MON}	0.0001 (0.02)	0.0020 (0.97)	0.0013 (0.35)	0.0020 (0.75)
D_{FRI}	-0.0004 (-0.16)	-0.0002 (-0.11)	0.0009 (0.35)	0.0007 (0.27)
Const.	-0.058 (-0.68)	0.0728 ^a (8.22)	-0.0232 ^b (-2.22)	0.1289 ^a (10.43)
R^2	52.52%	52.42%	55.48%	49.79%
High frequency				
RV	0.0943 ^a (3.92)	17.8142 ^a (4.11)	5.0102 ^a (4.90)	5.4451 ^a (6.46)
IRBA	0.0123 ^a (22.46)	0.0072 ^a (13.51)	0.0090 ^a (10.80)	0.0234 ^a (35.54)
IVol	-0.0073 ^a (-8.05)	-0.0119 ^a (-13.38)	-0.0087 ^a (-7.13)	-0.0019 ^b (-2.47)
D_{MON}	-0.0005 (-0.68)	0.0005 (0.68)	-0.0002 (-0.16)	0.0002 (0.29)
D_{FRI}	-0.0006 (-0.85)	-0.0002 (-0.31)	-0.0002 (-0.15)	0.0004 (0.46)
Const.	0.0943 ^a (6.31)	0.1904 ^a (12.48)	0.1332 ^a (6.09)	-0.0315 ^b (-2.49)
R^2	42.07%	39.48%	22.88%	47.03%

Individual frequency test regressions.

All variables are taken with logs except dummy variables. The t-statistics are in parentheses and the error variance are robust to heteroskedasticity and autocorrelation in the residuals. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

While high frequency measures report expected values (sign), low frequency liquidity measure report a mixture of results for this test. While AI and RBA are found to hold their previous assumptions that categorizes these as illiquid measures (since these have a positive weight with the spread), MI presents a mix of positive and negative weights. Since this measure may be affected also by other variables, its

factor loading should be carefully interpreted. For both the regressions, dummy variables are not significant at any confidence level. This might suggest that these stock are not affected by weekly seasonality.

To further investigate the relation that runs within these frequency measures, an additional regression is run to analyze all the liquidity measures on the usual spread measure. The vector of all measure is then made up by

$$LM = [AI, RBA, MI, RV, IRBA, IVol, D_{MON}, D_{FRI}], \quad (7.7)$$

The results of the OLS

$$BAS = \alpha_{LM} + LM\beta_{LM} \quad (7.8)$$

are reported in Table 6.7.

	WMT	KO	JPM	CAT
AI	6.4750 ^c (1.94)	5.7242 ^a (2.59)	2.5842 (1.45)	1.0601 (0.74)
RBA	2.0200 ^a (10.77)	1.6976 ^c (14.34)	2.5117 ^c (10.12)	1.5286 ^b (15.43)
MI	-0.4220 ^a (-5.49)	-0.0904 (-1.37)	0.1628 ^a (2.14)	-0.0302 ^a (-0.57)
RV	29.9224 ^b (2.39)	14.7735 ^b (2.03)	3.4065 ^b (2.12)	2.5758 (1.11)
IRBA	0.0214 ^a (8.99)	0.0090 ^a (6.74)	0.0106 ^a (3.70)	0.0337 ^a (11.47)
IVol	-0.0287 ^a (-7.20)	-0.0199 ^a (-6.90)	-0.0079 ^c (-1.76)	-0.0063 ^b (-2.08)
D_{MON}	-0.0008 (-0.26)	0.0033 (1.14)	0.0010 (0.27)	0.0016 (0.64)
D_{FRI}	-0.0012 (-0.58)	-0.0001 (-0.05)	0.0009 (0.33)	0.0005 (0.21)
Const.	0.5300 ^a (6.53)	0.3418 ^a (5.76)	0.0783 (0.83)	0.0374 ^b (0.58)
R^2	57.50%	56.23%	56.93%	53.86%

All frequency measures test regression.

All variables are taken with logs except dummy variables. The t-statistics are in parentheses and the error variance are robust to heteroskedasticity and autocorrelation in the residuals. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

Factor weights for all measures did not change its sign and confirms that MI can be interpreted as a mere liquidity measure, in contrast of the above mentioned AI and RBA. However, running this regression does not improve consistently the explanatory power of the regression, R^2 . Such result may imply that

by using low frequency estimates is a good way to analyze the market behavior, with an affordable loss. Lastly, the core test of this study, the relation with the daily liquidity measures and the illiquidity risk premium, gives consistent results that confirms Amihud's finding.

The test is made by a re-arrangement of the CAPM equation, already treated by analysts like Fama, French (1992) and Amihud himself, in order to price the illiquidity risk of these HFT shares. Hence, more illiquid stocks should have higher illiquidity risk premium, which can be measured as:

$$(r_{M,t} - r_{f,t}) = \gamma_0 + \gamma_1 AI_{t-1} + \gamma_2 RBA_{t-1} + \gamma_3 MI_{t-1},$$

It should be recalled that the variables coefficients should be positive for the first two measure, since illiquidity measure should reflect the higher risk, and negative for the liquid one. The goal is to run a OLS regression to quantify the weights that the market return gives to the variables. Considering that the error may be correlated to one, or all, of these variables, the White test allows to verify that the heteroscedasticity problem would not impair or bias the estimators. Through the test, it has been found that the variance of the estimators behaves like a homoscedastic estimator.

	WMT	KO	JPM	CAT
AI	1.9732 ^a (5.08)	3.9137 ^a (7.30)	1.6028 ^a (6.36)	1.0899 ^a (5.27)
RBA	0.0291 ^a (4.67)	0.0052 (0.72)	0.0252 ^a (5.56)	-0.0313 ^a (-4.75)
MI	-0.1244 ^a (-15.38)	-0.1051 ^a (-12.36)	-0.1059 ^a (-16.24)	-0.0908 ^a (-15.83)
Const.	0.0378 ^a (17.58)	0.0276 ^a (15.02)	0.0304 ^a (18.84)	0.0317 ^a (18.53)
R^2	9.60%	11.19%	17.20%	11.72%

Daily Regressions on the illiquidity risk premium.

All variables are taken with logs. The t-statistics are in parentheses and the error variance are robust to heteroskedasticity and autocorrelation in the residuals. Superscripts a, b and c reflect, severally, significance at 1%, 5% and 10% significance level.

AI measure is the one that has the major weight among the others with very high values, which are also highly significant at any possible level. This result renew the theory of Amihud that even in HFT, stocks can be differently priced due to their illiquid behavior. Obviously, lower is the value more liquid is the instrument. This measure reflect that firm shares were often considered as liquid instruments along the time window. However, even highly traded securities like these cannot be considered always liquid since some exogenous events, that may affect the market or the industry in which a firm operates, can change their usual schemes.

Since RBA is a dimensionless measure, it represents the mere illiquid aspect that characterizes each firm instrument. All the coefficients are positive, except for CAT. The positive relation is expected since greater is the bid-ask spread more volatile (illiquid) is the referred stock, however the negative outcome presented by CAT may imply that this firm price is inversely related to the usual bid-ask

spread–volatility relation. This feature may be due to the industry in which it operates, or by the fact that this has been, and is, the cheapest stock firm.

Liquidity measures are commonly negatively related to the premium that an investor would like to earn if he holds an illiquid instrument. MI re-validates this link, with significant outcomes, and has also a slight weight on the determination of the illiquidity risk premium. This link is characterized by an inverse which implies that more illiquid, volatile, is the object lower, even in negative terms, lower is the load factor on this measure.

Hence, the results suggest that the stock excess return, usually referred to as "risk premium" (in this instance "illiquidity risk premium"), is partially a compensation for stock illiquidity measure with low frequency measure. This explains why some stock are characterized by high equity premiums. The results mean that stock excess returns reflect not only the higher risk but also the lower liquidity of stock compared to Treasury securities.