High-Frequency Data and scaling laws to forecast liquidity risk

Department of Economics and Finance
Master's Degree in Finance
Chair of Empirical Finance

Supervisor:
Prof. Giacomo Morelli

Co-supervisor:
Prof. Federico Carlo Eugenio Carlini

Candidate:
Francesco Maria Ventriglia
ID: 748941

Academic Year 2022/2023

Abstract

This thesis investigates new liquidity risk forecasting models based on high-frequency data, incorporating the relationship between trading volumes and price impact through the use of scaling laws methods. With the advent of high-frequency data, it has become necessary to develop new risk measurement and forecasting methodologies that take into account the dynamic nature of financial markets. The analysis was conducted with a 1-minute frequency dataset of the NASDAQ Composite Index from 05-May-2022 to 03-May-2023, considering only data when the NASDAQ market was open and a "rolling window" framework to estimate the model parameters. The empirical results show that one month in-sample data are sufficient to obtain good risk prediction performance, especially when volatility is more concentrated and market conditions are stable, demonstrating that scaling law models can be valuable tools in the field of financial risk management.

Keywords: Liquidity Risk, Market Microstructure, High-frequency data, Scaling laws, Scale invariance.

Contents

List of Figures

List of Tables

Introduction

In the field of financial analysis, liquidity risk management is a crucial element for financial institutions, investors and regulators. Liquidity, defined as the ease with which financial assets can be bought or sold without causing significant changes in market prices, is considered the central aspect of market microstructure and is fundamental to the stable functioning of financial markets. History shows that one of the most frequent causes of market crises is lack of liquidity, as in the case of the financial crisis of 2007-2008 or the Flash Crash of 6 May 2010. The latter highlighted the importance of risk management in financial markets by shifting the focus to high-frequency trading and the use of algorithms in stock markets. The continuing evolution of high-frequency trading, which has greatly revolutionized the structure and speed of financial markets, has led to the view that risk models based on low-frequency data are no longer sufficient, and to focus on finding new liquidity risk management models that take into account the massive amount of data that high-frequency trading generates in the market. These data, which capture details of financial transactions at extremely short time intervals, have opened up new perspectives in the field of risk management and market microstructure, providing a real-time view of financial markets.

This thesis investigates new liquidity risk forecasting models based on high-frequency data, incorporating the relationship between trading volumes and price impact through the use of scaling laws methods, which provide a better understanding of the dynamic nature of markets and, due to their scale invariance property, allow to evaluate and predict liquidity risk at different time scales. The combined use of high-frequency data and scaling law methods allows to not consider assumptions about the price distribution and rely on a single month's high-frequency in-sample dataset to obtain good results and forecasts, rather than using time series of a few years that may not provide a complete picture of market information. The models presented in this analysis are based on the existence of scaling laws that relate the main liquidity measures of market microstructure to the time variable, denoted as the size of the time interval Δt. The objective is to evaluate liquidity as price impact, i.e. how much a given amount of trading volume can affect the price of an asset, and for this the liquidity variables under analysis are Kyle's Lambda (λ), presented by Kyle (1985), and the ratio between the price of an asset P and the corresponding Volume Weighted Average Price $VWAP$, called Liquid Ratio. Measuring liquidity risk with scaling laws methods that relate risk to the time variable improves the flexibility and accuracy of models because it allows to evaluate the liquidity of an order for both traders with intraday and longer investment horizons. The comparison of the forecasts obtained and the actual out-of-sample liquidity values shows that the models presented exhibit a good representation of the dynamic trend for both liquidity risk variables.

In recent years, there has been a growing application of scaling laws in the risk management framework (see Muller et al. (1990), Glattfelder et al. (2011), Qi (2011), Qi et al.

(2018)), although there is a lack of an exhaustive literature in the field of liquidity risk. The majority of studies focus on liquidity risk estimation without considering scaling laws. Kyle's model, introduced by Kyle (1985), is one of the most important models in market microstructure and asset pricing theory, and is concerned to analyze the role of information asymmetries among investors by measuring liquidity as the OLS regression coefficient of price change ΔP on trading volume. Many other works are based on this model: Ekren et al. (2022) constructs an equilibrium for Kyle's continuous-time model with stochastic liquidity; Brennan and Subrahmanyam (1996) have measured the liquidity of securities by price impact using the methods of Glosten and Harris (1988) and Hasbrouck (1991), which are extensions of the Kyle's model in which a variable is added to account for the fixed part of the trading costs, represented by the bid-ask spread. Emna and Chokri (2014), on the other hand, propose in their work an enhanced liquidity adjusted intraday value at risk, named LIVaR, based on the application of high-frequency data but without considering the use of scaling laws. This empirical study contributes to the literature by being a starting point for research into new methodologies for measuring liquidity risk using scaling laws methods to better understand the dynamic nature of financial markets.

The rest of the paper is organized as follows: Section 1 presents a brief description of the key topics covered in the paper and a literature review of the main measures of liquidity risk. The main aspects of market microstructure are reported in Section 1.1, while Section 1.2 introduces the importance of liquidity risk management. Section 2 is the heart of the work because all the methodology and mathematical concepts applied in the study are explained. Section 2.1 presents the concepts of scaling law and scale invariance that are the basis of this analysis, without which it would be possible to apply the models in Section 2.2. In this section, in fact, four scaling law models (two basic M1 & M2 and two extended models M1.1 & M2.1) are presented with the objective of forecasting liquidity risk through power laws that relate liquidity measures of market microstructure to different time intervals. Section 3 examines the results obtained through the application of the empirical models, first presenting a detailed description of the high-frequency data used in Section 3.1, and then evaluating the validity and performance of the models in Section 3.2. In the latter section, the forecasts obtained from the models will be compared graphically with the actual values of liquidity risk measures, and a residuals diagnostic will be presented to understand which model exhibits better performance in forecasting risk. Finally, conclusions are discussed in Section 4.

1 The relationship between Liquidity Risk Management and Market Microstructure

In the following section the relationship between liquidity risk management and market microstructure will be explored in details. The continuous evolution of financial markets has greatly increased the relevance of these two concepts, making them fundamental pillars of the analysis and management of financial transactions. Understanding their fundamental relationship has become essential for market participants who want to manage risk effectively and make informed decisions in their trading activities. Market microstructure focuses on the study of the internal structure and functioning of the market, seeking to understand the mechanisms that lead investors to make their trading choices, while liquidity risk concerns the ability to execute trades efficiently and without significantly impacting financial asset prices. In this analysis, the main components of market microstructure will play a crucial role in the management and forecast of liquidity risk.

Section 1.1 will focus on the definition of market microstructure and its main aspects such as bid-ask spread, market depth, price impact, and the role of information. Section 1.2 will introduce the importance of liquidity in risk management and then examine the main measures of liquidity risk in section 1.2.1.

1.1 The field of Market Microstructure

Market microstructure is a branch of finance that deals with the study of the structure and functioning of financial markets, focusing on the trading dynamics and price formation of financial instruments. This field of research focuses on the most detailed level of market analysis, considering factors such as the interaction between traders, trading mechanisms, liquidity, and the impact of information on market dynamics.

One of the main goals of market microstructure is to understand how traders act within financial markets and how their decisions affect the prices of financial instruments. Market microstructure studies the trading strategies adopted by traders, such as the use of buy (bid) and sell (ask) orders, automated trading algorithms, and limit or market orders. These strategies can have a significant impact on liquidity and price formation. O'Hara (1995), in fact, defines market microstructure as *"[...] the study of the process and outcomes of exchanging assets under explicit trading rules. While much of economics abstracts from the mechanics of trading, microstructure literature analyzes how specific trading mechanisms affect the price formation process."*[1]

In recent decades, this field of research has gained increasing importance due to the bigger complexity and speed of financial markets. The most important aspect that research wants to focus on is the price formation. Market microstructure studies explore how bid

[1] O'Hara (1995)

and ask prices are determined, considering factors such as supply and demand, trading orders, and information available to market participants: the goal is to understand how prices reflect information and how they adapt to changes in market conditions.[2] In order to answer these questions, an in-depth study of spreads, market depth, and price impact is needed.

The spread, also known as the bid-ask spread, is the amount by which the ask price exceeds the bid price for an asset in the market. The ask price reflects a trader's willingness to sell an asset while the bid price reflects a trader's willingness to buy it. Mathematically, it represents the difference between the lowest demand price (best ask) and the highest bid price (best bid). A wide bid-ask spread can reduce the profitability of transactions because traders must overcome a larger price difference to generate tradings. It is a crucial indicator because it can be seen as a measure of supply and demand for a specific asset: when the two prices diverge, price action reflects a shift in supply and demand. Traders wishing to buy or sell a financial asset must consider the bid-ask spread as part of the transaction costs associated with order execution.

Market depth refers to the ability of a market to absorb relatively large market orders without significantly affecting the price of the security. It indicates the amount of orders present at various price levels and is based on the presence of buy (bid) and sell (ask) orders. Greater market depth indicates a significant presence of bid and ask orders at different price levels, suggesting that traders can execute large transactions without causing a large price change. Market depth data may be used to analyze the bid-ask spread for a security, which can assist traders predict where the price of a certain asset may be headed. A narrower bid-ask spread indicates greater market depth and can be considered a positive signal, as it implies greater liquidity and a smaller price difference between buying and selling a security. To access market depth information, traders can consult the order book, which is a computerized record of pending buy and sell orders at different price levels, updated in real time to reflect current market activity. While in the past this information may have been chargeable, many brokers and trading platforms now offer free views of market depth. This allows traders to see the full list of pending buys and sell orders, along with their size, allowing for greater transparency and information in market assessment.

Price impact is a more complex concept that refers to the effect that a buying or selling transaction has on the price of a financial asset. It indicates the relationship between trading volume and the change in market price. When a large buy order is executed in the market, there is a chance that the price of the financial asset will increase due to the additional demand generated by that order (*"upward impact"*). On the other hand, when a large sell order is executed, the price may decrease due to oversupply (*"downward impact"*). The magnitude of the price impact depends on several factors, including the size of the order relative to the average trading volume, market liquidity, price sensitivity to trades, and

[2] Russell and Engle (2010)

information asymmetries present, in a directly proportional way: the larger the size of the order relative to market volume, the greater the price impact. Price impact is an important concept for market participants because it can affect transaction costs and profitability of trades. Therefore, managing price impact is a crucial aspect of order execution strategy for traders, especially in the case of large orders, as its mismanagement can reduce expected returns or increase trading costs.

Market microstructure also considers the role of information in the price formation process and trading strategies. Information plays a key role in market microstructure and influences traders' trading decisions, distinguishing itself into public and private information. Public information is generally accessible to all market participants and includes financial news, quarterly company reports, economic data, and other information that can influence the prices of financial assets relatively quickly and widely. On the other hand, private or insider information is that to which only some market participants have access. This information can be obtained through internal research, data analysis, or possession of confidential information and may include, for example, acquisition plans, information on future earnings of companies, or information on the activities of key investors. Information asymmetries occur when some participants have insider information compared to others. Market microstructure is concerned with studying how this information, both public and private, is incorporated into the prices of financial asstes, exploring how information asymmetries affect market dynamics, liquidity, and asset prices. For example, when information is released to the public, traders who quickly interpret it and act on it can take advantage of the price before they fully adjust to the new information.

Trading strategies are often based on analysis of available information. Traders, in fact, may try to exploit market inefficiencies based on information that they believe to be under- or over-estimated by the current price of financial instruments. Market microstructure recognizes the fundamental role of information in the price formation process and traders' trading strategies, making, therefore, information management a crucial component for market participants seeking to gain a competitive advantage and make informed trading decisions.

1.2 The importance of Liquidity Risk Management

In the recent years, the relationship between risk management e market microstructure has become increasingly important, gaining more attention especially in times of financial crises. Many recent studies[3] show that one of the most frequent causes of many market crises is precisely lack of liquidity. Indeed, the financial crisis of 2007-2008 can be seen as a global financial crisis marked by a liquidity crisis that affected both financial institutions and states. During the first phase of the crisis, many financial institutions faced severe difficulties specifically because of the lack of liquidity in the market, despite having adequate levels of capital. These market crises reminded us that liquidity plays a very important role in financial markets: studying the impact of liquidity variations due to rapid changes in market conditions has become essential for risk management nowadays, both because it is critical to ensuring the proper functioning of financial markets by enabling traders to execute their transactions quickly and efficiently, manage risk, and reduce transaction costs, but also because with the advent of High Frequency Data, traders now are forced to construct new strategies to beat a market in which volatility and trading intensity have increased significantly.

The advent of High Frequency Trading (HFT) over the 1990s profoundly revolutionized the structure and functioning of financial markets. The intensive use of computer algorithms and the ability to process large amounts of data in real time radically changed the way trading was done. In a high frequency world, the speed with which information must be incorporated into the market has increased dramatically, making many learning models used in the past obsolete and increasing the complexity of the HFT strategies to be followed. O'Hara (2015), in her work, gives us a clear view of how some aspects of this change have affected the market microstructure, making it an increasingly prominent player in a world that goes faster and faster.

The assumption that an asset can be traded at a certain price for a certain quantity for a fixed period of time can no longer be considered realistic, and that is why risk management, in the past two decades, is focusing strongly on considering liquidity as one of the key risk factors to consider in an investment, trying to incorporate and predict liquidity risk through the tools it has available.

However, a paradox in the market microstructure research is the lack of an objective definition of liquidity. Black (1971) explains the concept of a liquid market by saying: *"... a liquid market is a continuous market, in the sense that almost any amount of stock can be bought or sold immediately, and an efficient market, in the sense that small amounts of stock can always be bought and sold very near the current market price, and in the sense that large amounts can be bought or sold over long periods of time at prices that, on average, are very near the current market price."*. Kyle (1985), on the other hand, describes the concept of market liquidity by introducing three different aspects:

[3] Matz and Neu (2007)

- *"tightness"* refers to the cost of turning over a position in a short period of time. Obviously the higher market tightness the lower will be the liquidity, since the transaction costs to close the position will be more.

- *"depth"* refers to the ability of the market to absorb quantities without having a large effect on price. It is nothing more than the size of an order flow required to move the price by a given amount.

- *"resiliency"* refers to the speed with which prices tend to converge towards the underlying liquidation value of the asset, generally it measures the rate at which prices recover from a random shock.

Subsequently, Harris (2002) introduces another important aspect of the market liquidity that is the *"immediacy"*, which refers *"... to the ability to trade large size quickly at a low cost when you want to trade"*. It could be considered as a measure of the time opportunity cost, that increases for all those investors who are unable to find a suitable counterpart to conclude the transaction.[4] Therefore, we can conclude that liquidity refers to the ease with which financial assets can be bought or sold without causing significant changes in market prices and is considered the central aspect of the market microstructure.

As a matter of fact, liquidity risk refers to the possibility that a financial asset cannot be traded easily or at a desired price without causing a significant reduction in its value. In other words, liquidity risk indicates the difficulty of converting a financial asset into cash or executing large orders without significantly affecting the price of the asset. The greater the illiquidity of a security, the more investors demand an illiquidity risk premium in addition to the security's yield.[5]

Bangia et al. (1999) argue that the liquidity risk is an important component in capturing the overall risk. They divide the overall market risk, defined as uncertainty in market value of an asset, into two parts: uncertainty about future asset returns, which is described by the "pure" form of market risk and concerns uncertainty about prices and returns due to market movements, and uncertainty about liquidity, described by the additional liquidity risk component and concerns the uncertainty of liquidation costs. These costs arise from the fact that in a real friction market, traders, when liquidating a position quickly, realize a liquidation price that is lower than the mid-price.

Liquidity risk is, in turn, divided into two types: exogenous and endogenous liquidity. Exogenous liquidity risk is common to all market participants and cannot be affected by the actions of any one trader. It is influenced and reflects the market characteristics such as market depth and bid-ask spread. Endogenous liquidity risk, on the other hand, is related to the position taken by the trader. It is different for each market participants and it is mainly driven by the size of the position in a directly proportional way. The effect of

[4] Emna and Chokri (2014)
[5] Qi (2011)

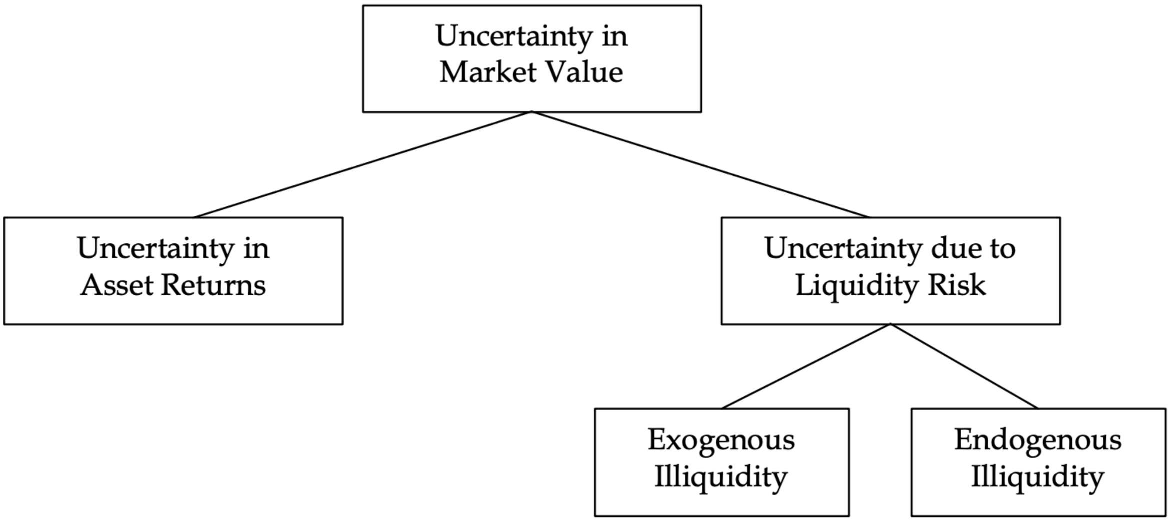


Figure 1: *Bangia et al. (1999): Taxonomy of Market Risk*

endogenous liquidity is most evident when the volume traded by the investor exceeds the quote depth (the volume of shares available at the quoted price): the endogenous liquidity risk becomes higher as the position size increases. It is usually appropriate to consider both components to best investigate liquidity risk. Figure 1 illustrates schematically the taxonomy of Market Risk just described.[6]

1.2.1 Measures of Liquidity Risk

Liquidity is a broad and multidimensional concept that encompasses various aspects important to financial traders and investors that do not allow it to be captured in a single measure. There are several measures used to calculate liquidity risk, most of them based on microstructure data.

The bid-ask spread is the most common indicator for measuring market liquidity risk. It is inversely correlated with the liquidity: a wider bid-ask spread indicates a greater distance to match price intentions of sellers and buyers, suggesting less liquidity in the market. This implies that traders may have to pay a higher premium, as they have to overcome a larger price gap between the bid and ask price, or accept a larger discount to execute transactions making it more difficult to buy or sell a financial asset at the desired price. It is a natural measure of liquidity because it represents the cost of immediate execution, that is the difference between the price at which a trader can buy and the price at which he can sell the asset. Because of its simplicity, intuitiveness, and comprehensibility, it is one of the most widely used measures for extracting information about market conditions. However, it does not take into account a number of crucial aspects of liquidity such as: trading volume and market depth, as the bid-ask spread only measures the difference between the ask price

[6] See Bangia et al. (1999) for more details.

and the bid price at the time of measurement but does not take into account the amount of securities available at those prices.

For this reason, another widely used measure for calculating liquidity is trading volume. It represents the number of securities or financial contracts traded in a given time period and is closely related to liquidity, as it is a key indicator of market activity. High volume can indicate greater liquidity, as there are more participants willing to buy and sell, thus reducing the risk of significant price impact during trades. In contrast to the bid-ask spread, trading volume is sensitive to supply and demand in the market and takes into account the number of trades made in a given time period, allowing for a measure of liquidity that reflects market activity over a specific period. In fact, while the bid-ask spread focuses on the immediate cost of trading a security or financial asset and provides a direct measure of the difference between buying and selling prices, trading volume provides a broader view of overall market activity over a specific period. Both measures are important in assessing the liquidity of a security or market, but each has a different approach and provides complementary information.

Since liquidity is a very broad and varied concept, trading volume should also be used in conjunction with other measures of liquidity, such as price impact, to obtain a comprehensive assessment of liquidity and make informed decisions in the context of trading and investing. It is a crucial measure for assessing liquidity because it indicates how much the price of an asset may vary in response to a given amount of trading volume. In academic research, many researchers have tried to investigate methods and models to calculate this aspect of liquidity risk. One of the pioneers who tried to study this aspect is Albert S. Kyle.

Kyle's model, introduced by Kyle (1985), is one of the most important models of market microstructure and asset pricing theory. It was developed to analyze the role of information asymmetries among investors in determining asset prices and market liquidity. The model seeks to explain how interactions between informed traders (insiders), who knows private information, and a market maker can influence asset prices and market liquidity. The key intuition that can be deduced from Kyle's model concerns precisely the concept of price impact. Kyle demonstrated how the transactions of insiders affect asset prices because they act on their private information while trying to maximize their expected profit. This generates a positive relationship between transaction volume and price change (precisely the price impact, as described in Section 1.1) because the market maker, unable to distinguish the volume generated by the insider trader from that of the noise trader, sets the price as an increasing function of the total quantity traded without taking into account the imbalance in order flow resulting from information asymmetries.

The equilibrium proposed by Kyle's model can be synthetically represented by the regression of price change ΔP_t on trading volume as:

$$\Delta P_t = \mu + \lambda v_t$$

where μ is a constant, v_t is the total quantity traded (transaction volume) and the constant λ is the so-called Kyle's Lambda which is the sensitivity of price to total demand and is affected by the standard deviations of the fundamental price and noisy trades.[7] In summary, Kyle's model mathematically expresses how market liquidity is directly proportional to the volume of total transactions. It has provided a solid theoretical basis for understanding the role of asymmetric information and trading in determining asset prices and market liquidity, and its influence in the academic literature is considerable.

Brennan and Subrahmanyam (1996) measured stock liquidity by price impact using the methods of Glosten and Harris (1988) and Hasbrouck (1991), both based on Kyle's model. The main difference with Kyle's model is that in this case the price impact is measured by taking into account both the variable trading cost, measured by Kyle's λ, and the fixed cost, represented by the bid-ask spread. Trivially, a variable is added to the equilibrium equation of the model to account for the fixed part of the trading costs.

All of these measures of liquidity require intra-daily transaction data for their calculation that might be unavailable in many stock markets. To overcome this problem, it is possible to consider rough measures of price impact that do not need microstructure data: the most important is the illiquidity measure, called *ILLIQ*, described by Amihud (2002). He defined it as *"the daily ratio of absolute stock return to its dollar volume, averaged over some period and can be interpreted as the daily price response associated with one dollar of trading volume"*.[8] It is calculated using daily transaction data and volumes that are available in most markets over long periods of time and it has been shown by the author himself that this measure is positively and strongly related to measures of liquidity based on microstructure data, thus indicating that both of them provide an indicative estimate of price impact.

Another measure widely used by high-frequency traders to measure liquidity risk is the Volume Weighted Average Price ($VWAP$). $VWAP$ is a widely used indicator in market microstructure to measure the volume-weighted average trading price of a financial asset during a given time period. Since it is a technical analysis indicator, it is used for intra-day trading and is therefore calculated for the trading day by resetting itself at the beginning of each new trading session. It is an indicator used mainly by investors who prefer to adopt a passive investment style[9], and its main function is to represent a trading benchmark[10], as it

[7]Ekren et al. (2022)

[8]The formula is defined by the following equation:

$$ILLIQ_{iy} = \frac{1}{D_{iy}} \sum_{t=1}^{D_{iy}} \frac{|R_{iyd}|}{VOLD_{iyd}}$$

where D_iy is the number of days for which data are available for stock i in year y, R_iyd is the return on stock i on day d of year y and $VOLD_iyd$ is the respective daily volume in dollars. See details in Amihud (2002).

[9]Berkowitz et al. (1988)

[10]Madhavan (2002)

is able to show the relationship between the price of the asset and its total trading volume, allowing traders to execute orders that are in line with market volume at a fair price. Executing orders at this price allows investors to reduce transaction costs by minimizing market impact costs, which are all such costs that arise from market liquidity.

Since it is a weighted average price, its formula will have the same structure as the moving average:

$$VWAP = \frac{\sum(P_i \cdot V_i)}{\sum P_i}$$

where $VWAP$ is the Volume Weighted Average Price, P_i is the trading price at time i and V_i is the trading volume at time i.

In summary, the Volume Weighted Average Price is a versatile indicator that provides key information about the liquidity of a market or financial asset. Its sensitivity to trading volume makes it a valuable tool for market participants and analysts in assessing market conditions and developing effective trading strategies.

2 Methodology

This section explains the methodology, the type of data used and the mathematical concepts behind the empirical models applied. In the era of high frequency and advanced technologies, market microstructure has evolved to understand the impact of high-frequency data on the dynamics of financial markets. High-frequency data provide detailed transaction information, such as prices, volumes, and execution times that provide a better understanding of trading patterns, traders' strategies, and the evolution of liquidity over time.

In this empirical study, ultra-high-frequency data are used to estimate the parameters of empirical scaling laws that provide predictions of liquidity risk. The scaling law method does not consider assumptions about distributions of returns by managing to capture all that information, given the abundance of available data, that would be lost if low-frequency, and time-equidistant data were used. Section 2.1 will introduce the concept of scaling laws, focusing on its property of scale invariance, which is the underlying assumption of this study.

Empirical models based on scaling laws have already been applied to risk management and volatility modeling.[11] Section 2.2 presents the new empirical models, introduced in this study, used to estimate liquidity risk. The methodology is based on the paper presented by Qi et al. (2018) and aims to incorporate the relationship between trading volume and price impact to assess the liquidity risk of an asset. The main objective of this work is precisely to find a scaling law that relates liquidity measures with different time intervals in order to predict their values at any future time horizon.

2.1 The concept of scaling law

A scaling law is a common concept in several scientific disciplines that describes a mathematical relationship between two or more variables that change proportionally with respect to a specific scale of magnitude. In particular, a scaling law expresses how the properties of a system change with respect to the size or scale of the system. Its most important property is precisely scale invariance, i.e. the ability to express this relationship as a power law that remains unchanged with respect to the scale magnitude. The concept of scale invariance is the basis of this work, without it all the analyses performed would be worthless. It refers to the property that the distribution of the quantities in a specific phenomenon remains the same at a different scales of measurement. A phenomenon observed at different scales of magnitude will assume a distribution with a similar general shape for each magnitude, even though the specific values may vary.

The concept of scale invariance has applications in many scientific and technological fields. In geometry it is referred to when talking about fractals. They are geometric objects that repeat their shape in the same way on different scales, so when you enlarge or reduce

[11] Muller et al. (1990), Glattfelder et al. (2011), Qi et al. (2018)

it you get a figure similar to the original complete object. In the field of economics, the Pareto distribution is a common example that clearly demonstrates scale invariance. This distribution is often used to model the distribution of wealth of different subjects and suggests that the frequency of people with a given amount of wealth follows a power law, regardless of the level of wealth. In the field of finance, power laws and scale invariance are often used to describe the distribution of financial returns and price movements in financial markets. Hence the importance of these aspects for this work. The ability to identify risk and return models that are based on scale invariance allows financial analysts to assess and predict risk for any time interval, improving the flexibility and accuracy of the models.

In a power law, one variable is proportional to the power of another variable, often expressed through an equation of the type:

$$y = Cx^{-\alpha}$$

where y is the dependent variable, x is the independent variable, C is a constant coefficient, α is a power exponent that determines the slope of the curve.
Usually, the general form of the power law distribution is:

$$p(x) \propto L(x)x^{-\alpha}$$

where $L(x)$ is a function that slowly varies due to the scale invariance of $p(x)$.[12] When $L(x)$ is a constant, the distribution becomes:

$$p(x) = Cx^{-\alpha}$$

and so given a relation of this type, scaling the argument x by a constant factor k causes only a proportionate scaling of the function itself. That is,

$$p(kx) = C(kx)^{-\alpha} = k^{-\alpha}p(x) \propto p(x).$$

A distinguishing feature of power laws is that when the data are represented on a logarithmic scale, the relationship between the variables appears as a straight line.

Figure 2 is a log-log plot that shows an example of the power law regression line for the NASDAQ Composite Index using 1-minute data from May 2022 to My 2023. The red dots represent the observed value of the variables related by the power law, while the blue line represents the fitted value of the regression line that explains the relationship between the variables. The graph shows, on a logarithmic scale, the good fit of the variables to a straight line, indicating that the power law is a good representation of the data. The scale invariance property should assert the law for any chosen set of thresholds: in the example,

[12] Clauset et al. (2009)

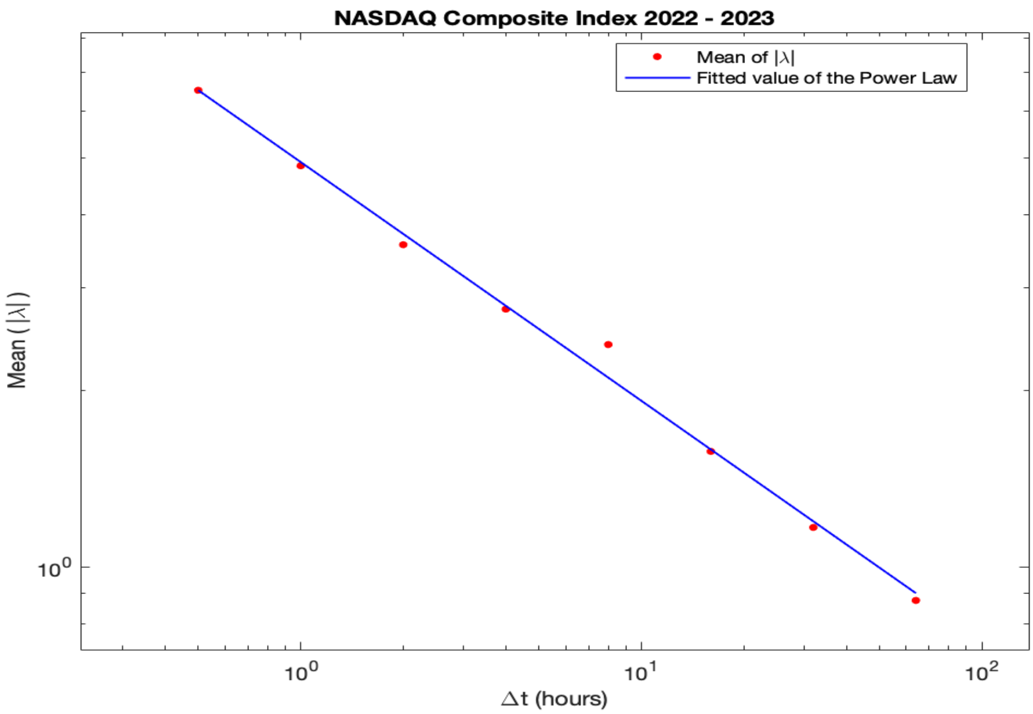


Figure 2: *An example of the estimated scaling law regression line* (1.3) *for the NASDAQ Composite Index using 1-minute data from May 2022 to May 2023. The x-axis represent the* Δt *and the y-axis the average* $|\lambda|$ *for the chosen time interval.*

the threshold time intervals $\Delta t = \{0.5, 1, 2, 4, 8, 16, 32, 64\}$ (hours) are applied to obtain the average $|\lambda|$ for the single sampling window to demonstrate their power relationship.[13]

Looking at the graph, it is evident how the basic hypothesis for the validity of the scale invariance is acceptable in this empirical work.

[13]Following the procedure adopted by Qi et al. (2018).

2.2 Empirical models used to estimate Liquidity Risk

The empirical scaling law method applied in this study is based on a multi-time scale analysis that considers the most popular microstructure liquidity measures, already used in the risk management framework. The choice of this type of measure was mainly forced by the use of high-frequency data. To demonstrate the presence of scale invariance in financial markets, there is a need for a large amount of data that provides an overall picture of the financial situation at each moment of the trading day. For this reason, the liquidity measures, subject to the analysis, must refer to the market microstructure in order to obtain results consistent with the objective of the work.

The first model (M1) is based on Kyle's Lambda (λ) and wants to explore the possible relationship between the average absolute value of λ in a certain interval and the size of the interval itself. In particular, it wants to demonstrate the existence of a scaling law between the two variables, to obtain future predictions of λ values based on the averages calculated in pre-specified time intervals.

Let denote the price and trading volume of an asset in a certain time interval Δt as $P_{\Delta t}$ and $V_{\Delta t}$[14] and calculate λ as described by Kyle (1985):

$$\Delta P_{\Delta t} = \mu_{\Delta t} + \lambda_{\Delta t} V_{\Delta t} \tag{1.1}$$

To remark the dependence with the time interval, let denote the obtained λ values with the symbol Δt. Then, to study the relationship between the liquidity variable and the time variable, we apply the following scaling law in its general form:

$$E(|\lambda_{\Delta t}|) = c(\Delta t)^{\beta} \tag{1.2}$$

where c is a constant, β is the scaling exponent and $E(.)$ is defined as the averaging operator.[15] Thus, in this analysis, the independent variable is the time interval Δt, which can be pre-specified, and the dependent variable is the average value of λ obtained, thanks to the high-frequency data, in the corresponding time interval.

To obtain the parameters of the scaling law, it is sufficient to calculate a simple Ordinary Least Squares (OLS) regression on the logarithms of both terms of Equation (1.2).[16] Taking the logarithms, in fact, we obtain a relation of the type:

$$\log(E(|\lambda_{\Delta t}|)) = \log(c) + \beta \cdot \log(\Delta t) \tag{1.3}$$

where the scaling parameters are the slope β and the intercept *log(c)* of the linear relationship

[14] Consider that $P_{\Delta t} = \{P(\tau); \tau \in [t - \Delta t; t]\}$ and the same for $V_{\Delta t}$

[15] $E(x) = \frac{1}{n}\sum_{j=1}^{n} x_j$

[16] Remember that, on a logarithmic scale, the relationship between the two variables becomes linear, grealty simplifying the calculation of the scaling parameters.

that arises from λ and the time interval Δt. Now, if the scaling law exists, given its scale invariance property, the M1 model should satisfy:

$$E(k \times |\lambda_{\Delta t}|) \propto E(|\lambda_{\Delta t}|) \tag{1.4}$$

and should hold for any chosen time interval.[17]

Once the parameters of the scaling law have been calculated, to obtain the future value of the price impact (λ) it will be sufficient to substitute the time interval of interest into the scaling law (1.2). For example, if there is interest in knowing the future daily estimates of λ it will be enough to substitute the time interval equal to the working day (in minutes) into the power law.

Note that in this scaling law only the magnitude of the price impact is analyzed and not its direction: the goal is only knowing the magnitude of the impact that volume has on the price change without determining whether this impact is positive or negative. Focus on studying the absolute value of λ allows to get a generic liquidity measure that provide information about the cost of liquidity and the volatility of the price with respect to the volume. The analysis of the magnitude of the price impact may help traders to detect market anomalies, without considering the specific direction of the price movement. Moreover, to compare liquidity risk among different securities, the magnitude of the price impact may be a standardized and neutral measure, allowing traders to assess which security might be more suitable for possible trading strategies.

The second model (M2) is based on the $VWAP$ and more precisely on the ratio between the closing price P and the corresponding $VWAP$. This ratio is a very important indicator for intraday traders because it allows them to implement trading strategies that take into account both the trade's benchmark ($VWAP$) and the price of the asset itself. Examining the Liquid Ratio ($P/VWAP$) consent to measure the efficiency of order execution but also to study the trend of the stock price. In fact, comparing the current price to the $VWAP$ can help traders to assess the strength and direction of a trend, but also to identify possible break out points by using the $VWAP$ as a reference to identify key levels of support and resistance, useful to implement intraday strategies which take advantage of trading opportunities that may occur.

The Figure 3 show an example of how price and the corresponding $VWAP$ interact with each other. The chart depicts a 1-minute time frame during a period of five working days (the red dots represent the end of each day) in which the 1-minute price of the NASDAQ Composite Index (the blue line) is compared with the corresponding 1-minute $VWAP$ (the orange line) calculated using the formula introduced in Section 1.2.1. Looking at the graph, intraday traders may obtain multiple information through the $P/VWAP$ analysis. In the first day it is possible to notice that the price is always below the $VWAP$: this

[17] With k that is a simple drift constant component.

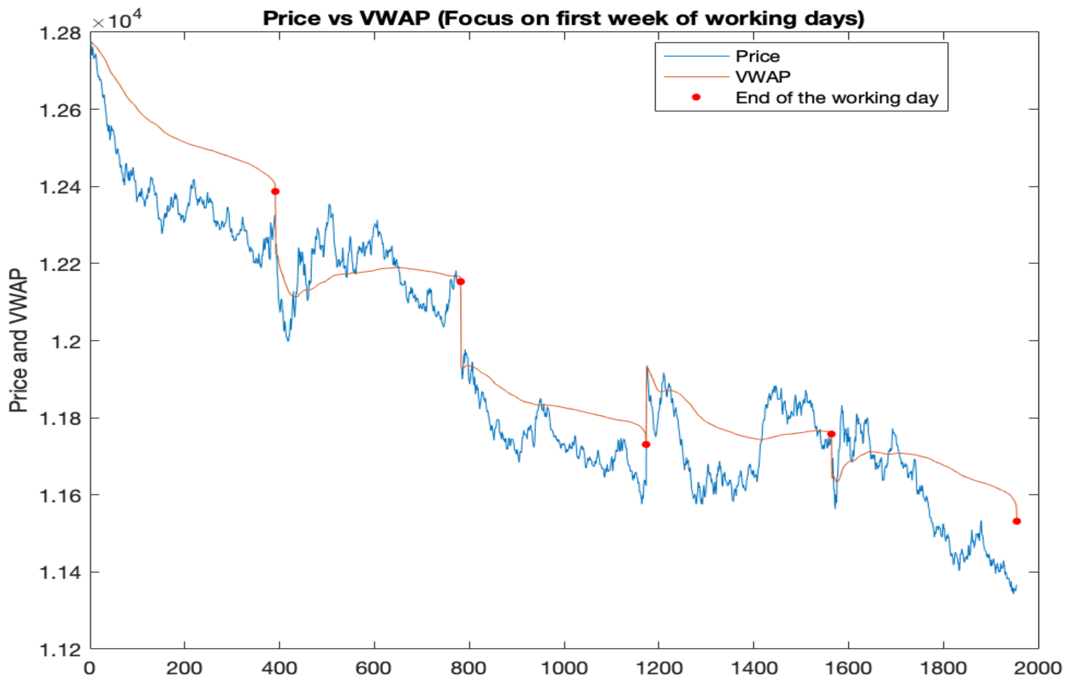


Figure 3: *The comparison between the Price and the VWAP of the NASDAQ Composite Index for five working days using 1-minute frequency data.*

situation may reflect a bearish trend in the asset price, mainly due to a lack of liquidity. Remember that $VWAP$ is calculated by weighting price for the trading volume, which means that stocks traded with higher volumes will have a greater impact on the $VWAP$. If the price remains consistently below the $VWAP$, it could indicate that most of the traded securities were trading at lower prices, suggesting possible limited liquidity. In the opposite case, where price fluctuates around the $VWAP$ line, a typical strategy of intraday traders is to use $VWAP$ in conjunction with mean reversion strategies. If price deviates significantly from the $VWAP$, as in the second and in the fourth day of the example, they might take actions to exploit a possible mean reversion in which price returns toward the $VWAP$. In fact, many trading algorithms often seek executions close to the $VWAP$ to minimize market impact and reduce liquidity risk. In conclusion, studying the deviation of the price from the $VWAP$ through the Liquid Ratio $P/VWAP$ may help traders to asses the liquidity of a securities by using it as a rough measure of the cost of liquidity.

The M2 model takes its inspiration from the original Maximal Price Change scaling law of Glattfelder et al. (2011), which is also found in Qi et al. (2018). In this case, in order to capture price movements relative to $VWAP$ within the time interval and study the volatility and dynamic nature of the asset over time, we use the concept of maximum change. The aim of this scaling law is to study the trend of the Liquid Ratio and in particular the maximum difference of it, to capture the volatility of the liquidity cost and to assess its variability over time.

First, the daily 1-minute $VWAP$ is calculated using the high-frequency data and the general $VWAP$ formula:

$$VWAP_t = \frac{\sum_{i=1}^{t} P_i \cdot V_i}{\sum_{i=1}^{t} V_i} \tag{2.1}$$

where $t = 1, 2, \ldots, n$ represents the minute of the trading day, $VWAP_t$ represents the Volume Weighted Average Price at time t, P_i represents the price time series from $i = 1$ to t, V_i represents the volume trading time series from $i = 1$ to t. Note that since $VWAP$ is a measure used in technical analysis by intraday traders, the calculation of this indicator is reset at the beginning of each trading session and recalculated for the next day.

Next, let LR denote the 1-minute Liquid Ratio between the price P and the corresponding $VWAP$ as:

$$LR_t = \frac{P_t}{VWAP_t}.$$

Now it is possible to define the Maximal Liquid Ratio Change (MLRC) as the difference between the highest and the lowest value of the Liquid Ratio within a certain time interval Δt:

$$\Delta LR_{\Delta t}^{MAX} = \max\{LR(\tau); \tau \in [t - \Delta t; t]\} - \min\{LR(\tau); \tau \in [t - \Delta t; t]\} \tag{2.2}$$

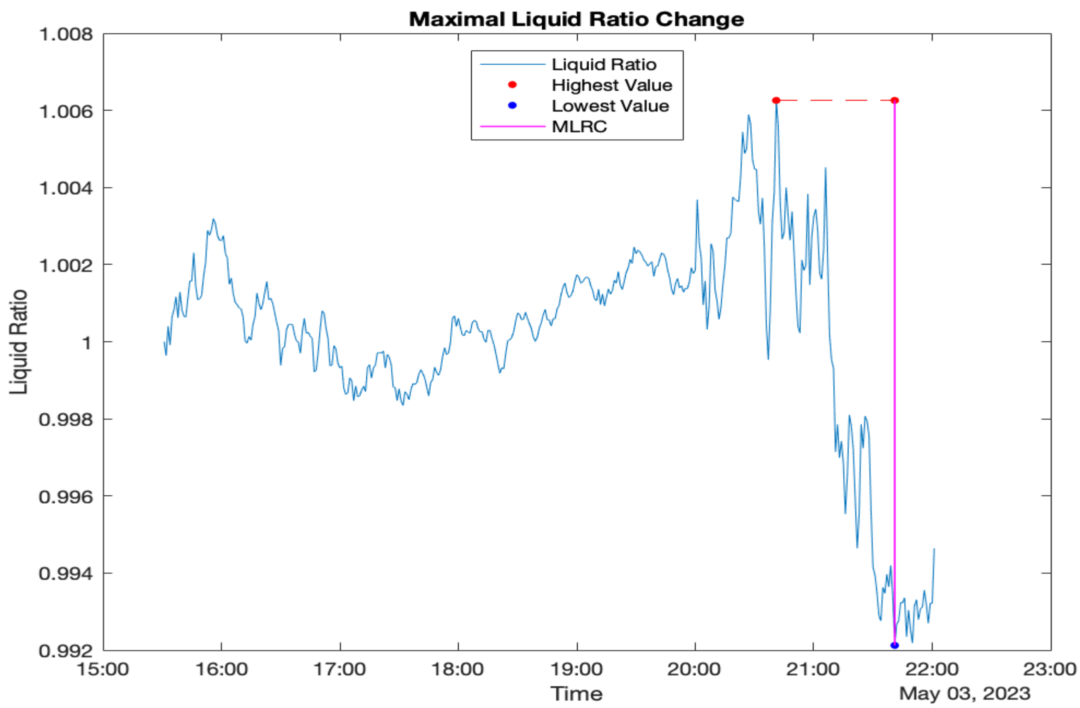


Figure 4: *An illustration for calculation of the Maximal Liquid Ratio Change (MLRC)* $\Delta LR_{\Delta t}^{MAX}$ *in a daily interval.*

Figure 4 shows an example of how the MLRC is calculated. The chart depicts the trend of the Liquid Ratio (blue line) during the working day 03 May 2023 and the corresponding value of MLRC for the specific time interval Δt, given by the difference between the maximum value and the minimum value. For simplicity, the working day was chosen as Δt threshold, but the calculation follows the same procedure regardless of the size of the interval. The MLRC is the analysis variable that provide a way to measure liquidity risk.

The proposed scaling law investigates the possible relationship between the size of this liquidity measure and the size of the time interval in which it occurs, described by the following equation:

$$E(\Delta LR_{\Delta t}^{MAX}) = c(\Delta t)^{\beta}. \tag{2.3}$$

The scaling law takes the same form of that in the first model. To calculate the parameters of the scaling law, the same steps as in model M1 are performed: the logarithms of both terms in Equation (2.3) are taken and the parameters of the scaling law are estimated by linear OLS regression on the newly obtained equation. Also in this case, due to scale invariance property demonstrated by the existence of the scaling law, it is possible to estimate the future values of the liquidity variable.

These two models just introduced are the basic models, based on the simple arithmetic mean in which equal weight is assigned to all data in the range considered. Based on the scaling laws proposed by Qi et al. (2018), it is possible to extend the basic models by replacing the averaging operator with the exponential moving average (EMA). The exponential moving average is used to examine the exponentially declining impact over different time intervals, employing exponentially decreasing weights, which assign the least weight to the oldest liquidity variable and increasing weights for the most recent ones. The EMA responds more quickly to changes in the underlying data and is preferred if is desired a more dynamic view of liquidity trends. As with the basic models, the entire time span T of the sample window is divided into n equidistant sub-intervals $\Delta t_i, (i = 1, \ldots, n)$ and then the EMA is applied instead of the simple mean to the liquidity variable.

So the extended first model, called M1.1 model, is a scaling law with equation:

$$EMA(|\lambda_{\Delta t}|) = c(\Delta t)^{\beta} \tag{1.1.1}$$

where the dependent variable $EMA(|\lambda_{\Delta t}|)$ is calculated as:

$$EMA(|\lambda_{\Delta t}|) = \sum_{i=1}^{n} w_i \cdot (|\lambda_{\Delta t_i}|)$$

with $w_i = \frac{\delta \times (1-\delta)^{(i-1)}}{\sum_{i=1}^{n} \delta \times (1-\delta)^{(i-1)}}$ the weight of the $EMA(|\lambda_{\Delta t}|)$ with property $\sum_{i=1}^{n} w_i = 1$ and $\delta = \frac{2}{n+1}$.

The second extended model (M2.1) is calculated following the same previous procedure but using the Maximal Liquid Ratio Change $\Delta LR_{\Delta t}^{MAX}$ as liquidity variable of interest. The scaling law, therefore, will have equation:

$$EMA(\Delta LR_{\Delta t}^{MAX}) = c(\Delta t)^{\beta}. \tag{2.1.1}$$

where the dependent variable $EMA(\Delta LR_{\Delta t}^{MAX})$ is calculated as in model M1.1 and follows the same properties.

In conclusion, a brief recap of the steps to follow for all the scaling laws is left below:

1. the entire in-sample time span T is divided into n equally-spaced sub-intervals $\Delta t_i, (i = 1, \dots, n)$. The size of the time interval is irrelevant because of the property of the scale invariance, which allows the scaling laws to be valid for any chosen set of thresholds.

2. For each subinterval Δt, we calculate the time series of the liquidity variable of interest, the λ for M1 and M1.1 and the maximal Liquid Ratio change $\Delta LR_{\Delta t}^{MAX}$ for M2 and M2.2, and the mean values obtained at the different subintervals. In this way for each time interval, is obtained the corresponding simple average value (exponential moving average value) of the liquidity variable for basic models (for extended models).

3. To estimate the scaling law parameters, we perform an OLS linear regression on the logarithms of the size of time interval and mean values obtained previously, and check by a log-log plot whether the scaling law exists. If the scaling law exists the plot should show a straight line passing through the regression points (also by looking at the values of R^2, that represent the goodness-of-fit of the regression, it is possible to determine if the scaling law is a good representation of the data. Generally, in all the scaling laws, the values of R^2 are above 0.97, indicating that more than 97% of the variation in the liquidity variables can be explained by the time variable).

4. Once the parameters of the power law have been obtained, given its property of scale invariance, it is possible to estimate the future values of the liquidity variables only by specifying the future time interval for which we want to obtain the estimates. In this study, given the use of high-frequency data and liquidity measures related to market microstructure, the interest is mainly on obtaining future daily risk estimates. The latter can then be used to formulate appropriate trading strategies or to monitor the performance of current trades.

3 Empirical Results

The aim of this section is to examine the empirical results obtained through the application of the methodologies described in the previous section.

Section 3.1 presents a detailed description of the high-frequency data used for the analysis, while Section 3.2 focuses on evaluating the validity and performance of the models by comparing the results obtained with the scaling law methods and actual values of liquidity risk measures, choosing the best models based on detailed residual diagnostics. Key findings are illustrated through graphs, tables, and other visual representations to highlight any trends and relationships emerging from the data.

3.1 Data Description

In this empirical analysis, to better understand the relationship between volume and price impact and to assess the liquidity risk of an asset, the NASDAQ Composite Index (*.IXIC*)[18] is examined. It was preferred over others stock market indexes because it is a market capitalization-weighted index, which means that companies with greater capitalization will have a greater impact on the index's performance. Value-weighted indexes are preferred in this type of analysis precisely because the weight of each stock is proportional to its market capitalization, which is the total market value of a company's shares, taking into account both the price and volume of shares on the market. In addition, the NASDAQ Composite Index is composed of stocks that are listed exclusively on the NASDAQ Stock Market, which is home to many technology-related companies and high-growth sectors, leading the index to be more volatile than the others, causing stock price fluctuations to be larger and faster. Finally, the NASDAQ Composite Index was preferred over the NASDAQ-100 because it is more generic by representing the entire NASDAQ stock market.

The initial purpose was to work with tick-by-tick data, but given the difficulty in finding these types of data (especially for past years), the use 1-minute frequency data is preferred, which should not change the analysis and results found. The 1-minute frequency data we used were from 05-May-2022 to 03-May-2023 and included 156760 observations[19]. The peculiarity of high-frequency data is precisely that it is not necessary to go very far back in time to obtain a sufficiently large time series to work on: even a single year generates a massive amount of data.

Once all the necessary data (local date, local time, price[20], trading volume[21]) have been collected, the first step in order to begin the analysis is to clean the sample, taking into

[18]This is the NASDAQ Composite Index ticker that can be downloaded from most of the electronic trading platform.

[19]The 1-minute frequency data are provided by REFINITIV.

[20]The closing price is preferred to the other types of prices because it can be representative of market dynamics.

[21]The trading volume is converted into logarithm to work with more interpretable values. The empirical results of the work does not vary if the actual values are considered.

account only the prices and volumes of orders that were executed when the NASDAQ stock market was open. Given the focus of this analysis on stock indexes rather than exchange rates, which are actively traded in the 24-hour FOREX markets, it is essential to consider the trading hours of the relevant stock market and the impact of holidays when exchanges may be either fully closed or have early closing times. The NASDAQ stock market typically operates during regular business hours, which are from 09:30 a.m. EST (3:30 p.m. Italian time) to 04:00 p.m. EST (10:00 p.m. Italian time) on weekdays, with a significant number of scheduled holidays. Specifically, the U.S. stock exchange remains closed on the following dates:[22]

- New Year's Day (02-Jan 2023)[23]
- Martin Luther King's Day (16-Jan-2023)
- Presidents' Day (20-Feb-2023)
- Good Friday (07-Apr-2023)
- Memorial Day (30-May-2022)
- Juneteenth National Independence Day (20-Jun-2022)[23]
- Independence Day (04-Jul-2022)
- Labor Day (05-Sep-2022)
- Thanksgiving Day (24-Nov-2022)
- Christmas Day (26-Dec 2022)[23]

It is noteworthy to mention instances when the market follows an early closure schedule, as observed on the day following Thanksgiving Day (Nov. 25, 2022). On such day, the market opens only in the morning, with trading concluding at 01:00 p.m. EST, consequently, it will not be considered in the analysis.

Upon completing the data cleaning process, the empirical analysis may begin. It will be carried out on 248 trading days. each characterized by a trading duration of 6 hours and 30 minutes (391 observations per day) for a total of 96968 total annual observations.

Figure 5 plots the daily prices and the daily volumes[24] for the NASDAQ Composite Index from the 05-May-2022 to 03-May-2023. Looking at the graph, it is interesting to note that trading volumes show peaks in the middle of each quarterly month. This phenomenon can be attributed to the release of quarterly financial reports by technology companies listed on the NASDAQ market, exerting substantial influence on trading activities.

[22]The brackets show the dates in our sample.

[23]If the festivity falls on the weekend, the market closes on the following Monday.

[24]The daily price represents the closing price observed in the last minute of each trading day, whereas daily volume encompasses the cumulative trading volumes recorded throughout the respective day.

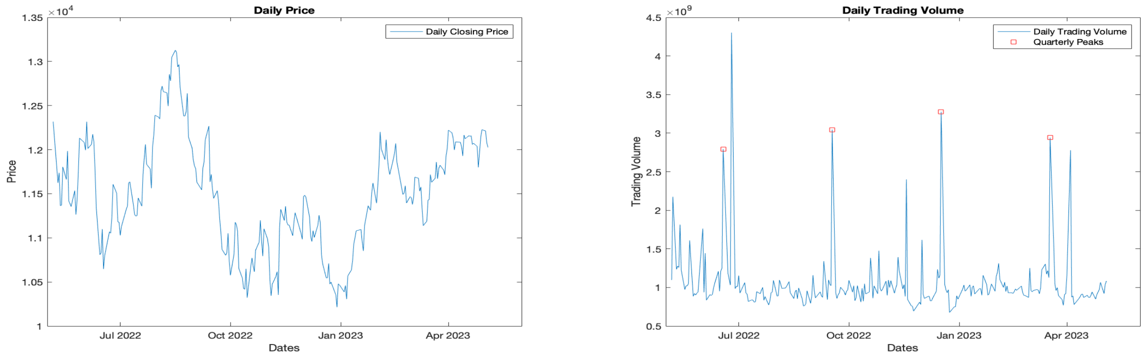


Figure 5: *The daily prices (left) and the daily volumes (right) of the NASDAQ Composite Index in 2022-2023.*

To evaluate the validity and performance of the models, we divided the sample into two sub-samples: the in-sample window, which refers to the data used to estimate the parameters and fit the model, and the out-of-sample window, the data used to evaluate the performance of the model by comparing the obtained estimates with the actual values. Following Qi et al. (2018), who show that to obtain good risk prediction performance it is sufficient to consider 1 month in-sample data, 20 working days (which corresponds to an actual month if holidays are excluded) will be considered as in-sample dataset in the analysis. The out-of-sample evaluation period, therefore, will consist of the remaining 228 working days.

Considering the scale invariance property of the scaling law method, any length of the time variable can be chosen to estimate the liquidity variable. However, given the focus on liquidity measures primarily applied in intraday trading, it is considered more appropriate to concentrate on daily risk forecasts. The in-sample windows, in fact, progress by one day, leading to the re-estimation of the scaling laws parameters on a daily basis using updated in-sample 1-minute datasets. In other words, tomorrow's estimates are obtained from the previous month.

Instead of presenting 912 OLS regressions models on the log-transformation of the scaling laws, Table 1 summarizes the average estimated scaling law parameters and the goodness-of-fit, reported as R^2, for all the power laws.[25]

Scaling Law Models	Intercept	Slope	R^2
Model 1	4.0651	-0.7937	0.9745
Model 1.1	4.1567	-0.8222	0.9785
Model 2	-7.8815	0.5413	0.9970
Model 2.1	-7.8613	0.5391	0.9967

Table 1: *The average estimated scaling law parameters.*

[25]See Appendix A for more details.

3.2 Forecasting Performance Results

Two different criteria are used to evaluate model performance. In the case of the first M1 model (and its extension M1.1), the estimates obtained through the scaling law method were compared with the actual daily λ value. Again, in order to assess the actual validity of the model, it was preferred the comparison with the absolute value of daily λ, taking into account only the magnitude of the price change without looking at the direction. This measure allows us to know how much the trading volume impacts the price change for each business day and represents the variable of interest to estimate. Daily Kyle's Lambda λ were calculated using the same formula (Eq.(1.1)) by fixing the time interval Δt to the trading day. The time series of daily λ is stationary, as can be seen by running an Augmented Dickey-Fuller Test[26] that allows us to study stationarity by testing the null hypothesis that a unit root is present in the time series sample. Working with stationary time series is very important because it allows us to simplify the predictability of future values, improving the efficiency and consistency of the models over time.

For the scaling law, a "rolling window" framework was used to estimate the model parameters. The scaling law parameters are estimated by a simple linear OLS regression with a monthly in-sample data window that is moved forward one day at a time[27], each time re-estimating the model parameters. In this way, daily risk estimates are obtained using the previous month's data and then compared with the corresponding actual value. The graphical results of daily risk forecasts using the M1 model are shown in Figure 6.

The light blue line represents the absolute value of λ for each trading day. The series moves in a range from a minimum of $\lambda = 0.0013$ to a maximum of $\lambda = 2.4585$ and seems to alternate between periods of high volatility, characterized by high jumps, and periods of settling in which λ values are more contained. Moreover, the highest peaks seem to occur always in the last trading days of the month.

The risk forecasts derived from the scaling laws are depicted by the red line (basic M1 model) and the black line (extended M1.1 model), with the blue line representing the monthly moving average of daily λ (simply the average of previous monthly values). Graphically, the models exhibit good prediction of the dynamic fluctuations in risk and volatility, yet they falter in forecasting extreme movements (both spikes and losses). Nevertheless, when comparing the risk forecasts with the moving average, it becomes evident that they closely align with the overarching dynamic trend of λ. This suggests that the models' estimations successfully anticipate the average price impact trend. Despite the notable daily λ volatility and occurrences of extreme fluctuations, the shape of the risk forecast curves follow the real trend of λ values.

Regarding the second M2 model, to evaluate the model performance the estimate obtained through the scaling law (2.3) are compared with the daily maximum fall of the

[26] Dickey and Fuller (1979)

[27] In this case, one-day rolling window means large volumes of 1-minute data

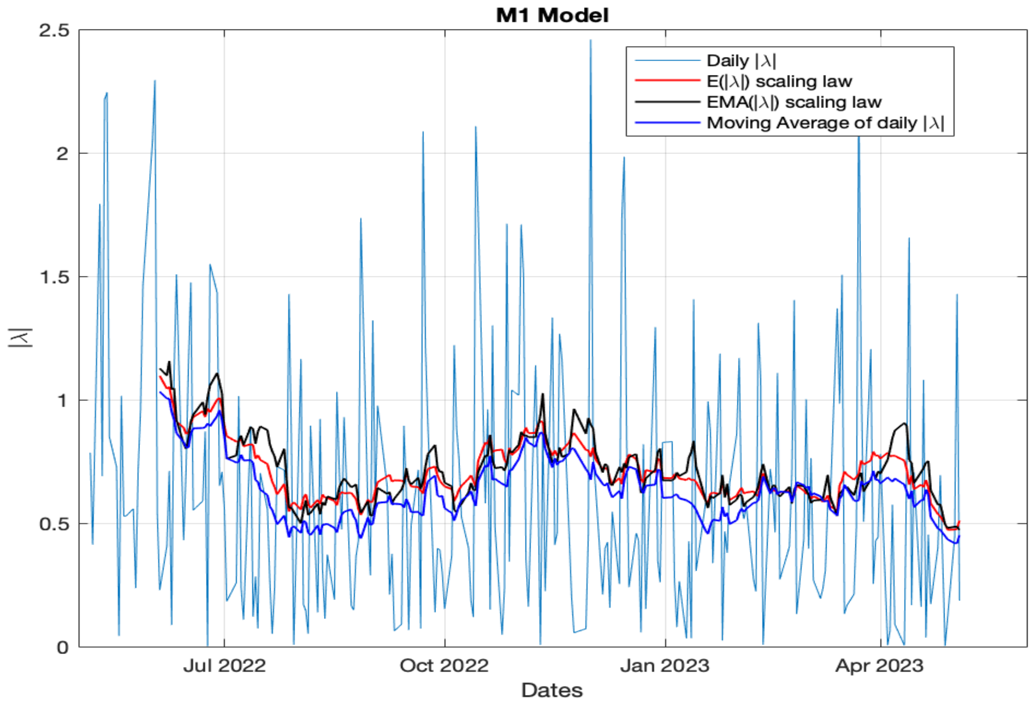


Figure 6: *The scaling laws forecasting results of basic and extended M1 model using NASDAQ Composite Index (out-of-sample period is from 03 Jun 2022 to 03 May 2023).*

$P/VWAP$, i.e. the biggest negative change of daily Liquid Ratio (LR). In this case, the goal of the model is not to study the effective future value of the Ratio, but to study the maximum amplitude of it, in order to understand how far the price can deviate from its benchmark value $VWAP$ and analyse the relationship.

The expected value for the maximum change in the Liquid Ratio (MLRC) over a daily horizon can be interpreted as a measure of the stock's liquidity risk for the next day. It represents the maximum amount of fluctuation that is expected for the $P/VWAP$ ratio during the trading day. Fluctuations in this liquidity measure may be due to various factors and market conditions, mainly occurring when there are changes in trading volumes. Since $VWAP$ is a volume-weighted measure, a significant change in volume can greatly affect its value, leading to fluctuations in the LR measure. So, also this measure allows us to consider the relationship that exists between price impact and trading volume.

Also for this model is used the same calculation procedure. First, the 1-minute $P/VWAP$ ratio was calculated for each trading day using the formula given by Eq.(2.1). Subsequently, for each business day was calculated the biggest negative change of the ratio in order to consider only the greatest drops of the ratio. Usually, rapid falls in the $P/VWAP$ are due to an increase in the volume trading, which, by increasing the value of $VWAP$, inversely impacts the ratio. A large drop in it may be a symptom of bearish pressure and market weakness, due to the fact that the stock is facing difficulties in maintaining previous price levels leading to increased selling by investors, or a correction of a previous rally, a situation

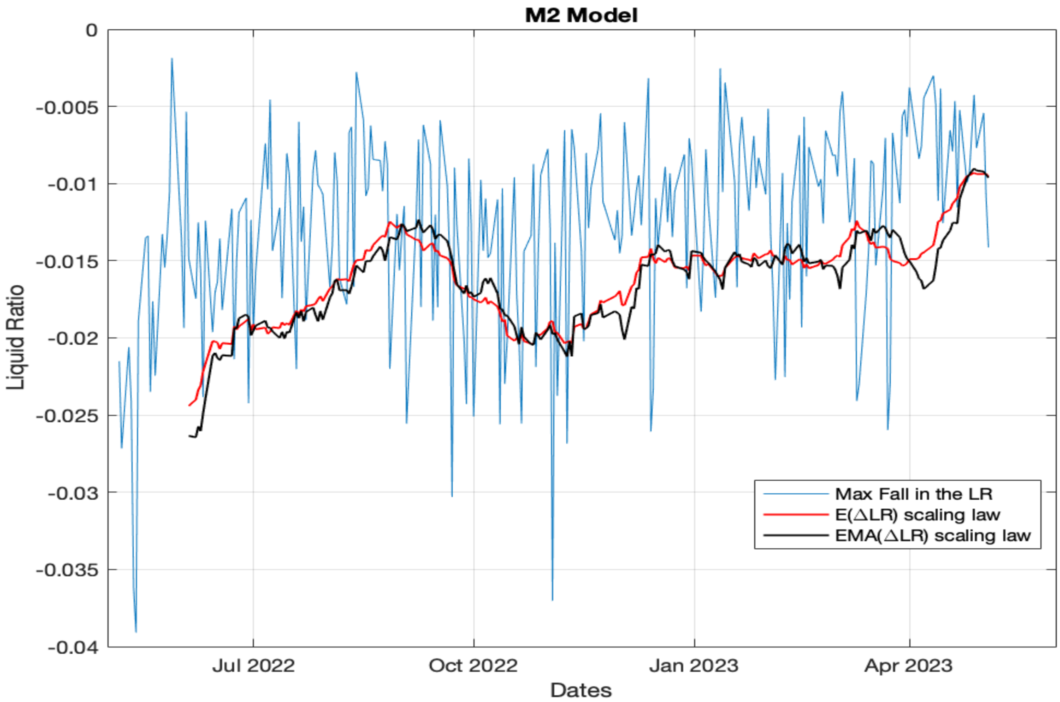
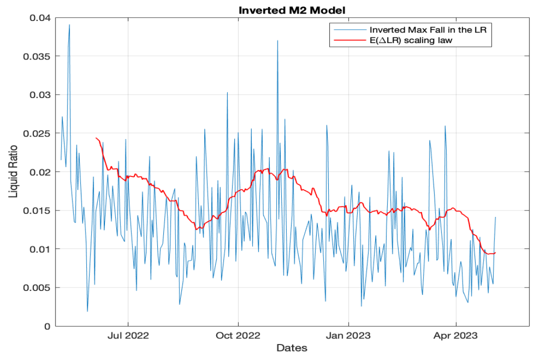


Figure 7: *The scaling laws forecasting results of basic and extended M2 model using NASDAQ Composite Index (out-of-sample period is from 03 Jun 2022 to 03 May 2023).*

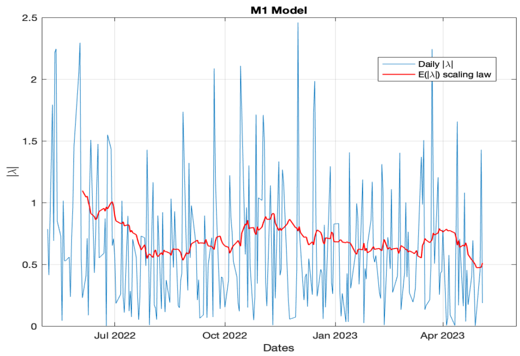
in which the ratio was initially above 1 and then fell sharply below 1, due to a slowdown in prices after they had been rising above the volume-weighted average of trading for a certain time. Once calculated the daily maximum fall, as for the M1 model, a "rolling window" framework was used to estimate the model parameters of the scaling laws. The forecasting performance in the scaling law M2 method is shown in the Figure 7.

In this case the light blue line represents the maximum fall of the LR over a daily horizon. Also in this case the series is stationary[28], but seems to be less volatile than the λ series. The alternation of period with high volatility and period with lower volatility is more evident: for an initial period the maximum fall shrinks until the end of August 2022 and then have some big drops from September 2022 to November 2022 and reduce again after it. The forecasted daily Maximal Liquid Ratio Change (MLRC) using the basic model M2 and the extendend model M2.1 are represented respectively by the red line and the black line. Graphically, the dynamic trend of the maximum fall is well represented by the forecasts of the scaling laws, in particular, in contrast to the M1 model, they underestimate the maximum fall and fail to predict extreme movements only in the period when the biggest drops are concentrated (from September 2022 to November 2022). In general, however, the forecast curve is very similar to that of the actual values, especially during periods characterized by smaller and less volatile maximum drops, implying that, on average, the model effectively predicts the Liquid Ratio trend when market conditions are stable.

[28] By running an Augmented Dickey-Fuller test by Dickey and Fuller (1979)



(a) Inverted M2 model



(b) M1 model

Figure 8: *The relationship between trading volume and price impact explained by the scaling law models.*

One peculiarity that can be seen from both graphs is that, by inverting the graph of the M2 model, the shape of the forecast curves follows a fairly similar trend. Figure 8 depicts this characteristic: initially, the curves decrease, indicating that during the summer months the markets slow down and trading volume has less impact on price[29], and then rise sharply in the period from September 2022 to November 2022 and reassess thereafter taking more or less constant values in the period from January 2023 to April 2023. This feature illustrates the functionality of both models in capturing the relationship between trading volume and price impact, consistently evaluating liquidity risk. This capability enables the combination of both measures to yield a more comprehensive and holistic perspective on market liquidity.[30]

Relying solely on a single graphical representation makes it challenging to discern which of the scaling law models exhibits superior performance in forecasting risk. To compare the models, it is necessary to look at the residuals, calculated as the difference between actual and predicted values of the liquidity variable.

Scaling Law Models	Mean	Std	Skewness	Kurtosis
Model 1	-0.0956	0.5100	1.0019	3.7888
Model 1.1	-0.1056	0.5116	0.9383	3.6481
Model 2	-0.0038	0.0058	0.9272	3.9081
Model 2.1	-0.0040	0.0060	0.8598	3.7036

Table 2: *Sample moments of the forecast errors for the scaling law models in the out-of-sample period (03 Jun 2022 - 03 May 2023).*

Table 2 summarize the descriptive statistics for the residuals of all the scaling law

[29] Jacobsen and Visaltanachoti (2009)

[30] The same results are obtained if extended models are considered.

models. Before commenting on the results obtained, it is necessary to proceed with the diagnostic operations to determine whether the models used are correctly specified or not.

A first look at the residuals plot denotes how, for all models, the error series are stationary with neither trend nor seasonal components.[31] To properly study the correctness of the models it is necessary to look at the ACF of the residuals.[32] The autocorrelation function (ACF) measures the linear dependence between values of the process at different lags, and indicate the amplitude and the length of the memory of the process. By simply looking at the correlogram of the residuals, that gives information about the existing ACFs in the observed sample, it is possible to notice that all the bars corresponding to the estimated autocorrelation are inside the blue lines (the interval $(-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}})$) and can be considered null. Values of the autocorrelation function observed for the residuals inside the interval suggest that the estimated autocorrelation of them may be due to randomness, meaning that the prediction errors do not depend significantly on previous observations over time. This is a desirable condition in statistical models and time series analyses, as it indicates that the model is able to capture variation in the data without leaving predictable or structured residuals. To look deeply at the appropriateness of the model a Ljung-Box test[33] was carried out on the residuals of the model to assess the null hypothesis that a series of residuals exhibits no autocorrelation for a fixed number of lags (the residuals are independently distributed), against the alternative that some autocorrelation coefficient is nonzero (the residuals exhibit serial correlation). The resulting p-values are all above the 5% threshold meaning that the residuals of the models are independent, since the null hypothesis cannot be rejected, concluding that the models are correctly specified.

Now it is possible to compare the models by looking at the descriptive statistics of the residuals (Table 2). Since the residuals' mean is relatively low for all the scaling laws (a positive results because it means that the errors are on average close to zero), the sample moments to analyse for the comparison of all risk models are the skewness and the kurtosis, with particular attention given to the standard deviation of the errors relative to the first model M1 and its extension. M1 and M1.1 models, in fact, present high values of standard deviation due to the fact that they fail to predict well the extreme movements of the λ series. However, by considering the residuals with respect tho the moving average of daily λ (the blue line in Figure 6), it can be seen that the standard deviation values drop considerably[34], keeping the other descriptive statistics similar. This feature further reinforces the hypothesis that the models are able to predict the average price impact trend well, while failing to account for extreme changes.

Instead, by looking at the skewness and kurtosis of the residuals, it is possible to compare models to see which type, basic or extended model, shows better performance in forecasting

[31]See Figure 10 in Appendix B for more details.
[32]See Figure 11 in Appendix B for more details.
[33]Ljung and Box (1978)
[34]The Std values drop respectively to 0.0342 and 0.0643 for M1 and M1.1 model.

risk. First of all, from Table 2, it can be noticed that all of the residual distributions are leptokurtic (with kurtosis values hovering around 3.7, slightly higher than those of a normal distribution) and show evidence of positive skewness (with values above 0.8).[35] These results suggest that the errors are influenced by outliers leading to a skewed and slightly ”stretched” distribution. Therefore, it is necessary to identify these extreme values and check whether, by excluding them from the analysis, the distribution of the residuals approximates that of a Standard Normal. To verify the presence of outliers, it is sufficient to look at the box plot of the distributions, which provides a general overview of the residuals and its quartiles.[36] Once the outliers are identified, it is possible to check whether the error distributions better approximate that of a Standard Normal, when extreme values are excluded from the distribution, by looking at the Q-Q plot of the new residuals and at the descriptive statistics of skewness and kurtosis.[37]

Scaling Law Models	Skewness	Kurtosis
Model 1	0.6682	2.7288
Model 1.1	0.6225	2.7210
Model 2	0.5768	2.9183
Model 2.1	0.5136	2.8366

Table 3: *Skewness and kurtosis of the residuals for the scaling law models excluding the outliers.*

Table 3 illustrates the level of skewness and kurtosis obtained by removing the outliers from the distribution. The exclusion of outliers reduced the kurtosis value, bringing it closer to that of a Standard Normal distribution, and skewness for all model specifications, reducing the thickness of the tails but keeping the right tail more pronounced than the left tail. Now, looking at Table 2 and Table 3, it can be observed that the level of skewness and kurtosis are lower in the case of the extended models. In this analysis, significant positive or negative skewness implies asymmetry of the errors' distribution, corresponding to under- or overestimation of risk. The same is valid for kurtosis: lower level of kurtosis implies lower probability of underestimation or overestimation of risk.[38] According to these considerations, therefore, it is possible to conclude that scaling law models based on the Exponential Moving Average (EMA) show better performance in forecasting risk than basic models, although the results obtained by simple arithmetic mean are not so different.

[35]See Figure 12 in Appendix B for more details.
[36]See Figure 13 in Appendix B for more details.
[37]See Figure 14 in Appendix B for more details.
[38]Qi et al. (2018)

4 Conclusions

This study aims to be part of the financial risk management framework by proposing forecasting models based on scaling laws. With the advent of high-frequency data, it has become necessary to develop new risk measurement and forecasting methodologies that take into account the dynamic nature of financial markets. The scaling laws method does not consider assumptions about the distributions of returns, incorporating all available and relevant information at all different time scales, only through the use of one month of high-frequency data.

In this empirical analysis, new forecasting models have been proposed to incorporate the relationship between volume and price impact in order to assess the liquidity risk of an asset. The goal of the work is to find a scaling law that relates the main liquidity measures of the market microstructure framework with different time intervals to evaluate and forecast liquidity risk at each time horizon, due to the scale invariance property of scaling laws. The concept of scale invariance is the basis of this work because it allows to relate two variables in the form of a power law, which has a similar distribution for each magnitude of the variables. Indeed, in this case, if the assumption of scale invariance is acceptable, it is possible to know the value of liquidity risk at each interval precisely because of the scaling law models that relate the two quantities.

The measures of liquidity analyzed in this study are Kyle's Lambda (λ), introduced by Kyle (1985), and the $P/VWAP$ ratio, called Liquid Ratio, which assess liquidity as price impact, specifically indicating how much the price of an asset can vary in response to a given amount of trading volume. Under the assumption that scale invariance is applicable, the existence of two scaling law models (a basic model and an extended model) have been demonstrated, both relating the time variable, denoted as the size of the time interval Δt, to each of the liquidity variables. The basic models (M1 and M2) consider the simple arithmetic mean of the liquidity variables values found at different time intervals, while the extendend models (M1.1 and M2.1) consider the exponential moving average (EMA) in order to incorporate the exponentially declining impact over the different time intervals. The analysis was conducted with a 1-minute frequency dataset of the NASDAQ Composite Index from 05-May-2022 to 03-May-2023, considering only data when the NASDAQ market was open.

To evaluate the validity and performance of the models, the estimates of liquidity risk, obtained using one month of data, were compared with the actual values of the liquidity measures analysed. The results show that, for the scaling law method, one month in-sample data are sufficient to obtain good risk prediction performance, specifically the results obtained from the extended scaling laws, based on the exponential moving average (EMA), are slightly better than those from the basic models, although both of them exhibit a good representation of the dynamic trend of liquidity variables. Residual diagnostics shows some positive skewness and kurtosis in the residual distributions for all model specifications, even when outliers are excluded from the distribution. This may mean that the models slightly overestimate liquidity risk, resulting in predicted values that are slightly higher than actual values. To accurately predict the relationship between trading volume and price impact, it is necessary to take into account a number of factors that are difficult to calculate, such as the presence of information asymmetries in the market or events that can greatly influence the behavior of market participants. Generally, the models developed in this study predict well the dynamic trends in market liquidity, especially when volatility is more concentrated and market conditions are stable. When aware of the practical implications of overestimating risk, which can lead to disregarding the execution of some orders that are considered unsuitable, and the presence of outliers that can affect future estimates, these scaling law models can be valuable tools in the field of financial risk management.

In conclusion, this empirical study is intended to be a starting point for the analysis of new liquidity risk management methodologies that take into account financial market fluctuations and the relationship between trading volume and price changes, through the use of high-frequency data that provide short-term market information to measure and predict risk and volatility over longer time horizons.

References

Amihud, Y. (2002) Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31–56.

Bangia, A., Diebold, F., Schuermann, T., and Stroughair, J. (1999) *Modeling liquidity risk with implications for traditional market risk measurement and management* (Papers No. 99-06). Wharton School Center for Financial Institutions, University of Pennsylvania.

Berkowitz, S. A., Logue, D. E., and Noser Jr, E. A. (1988) The total cost of transactions on the nyse. *The Journal of Finance*, *43*(1), 97–112.

Black, F. (1971) Toward a fully automated stock exchange, part *I*. *Financial Analysts Journal*, *27*(4), 28–35.

Brennan, M., and Subrahmanyam, A. (1996) Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, *41*(3), 441–464.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009) Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703.

Dickey, D. A., and Fuller, W. A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431.

Ekren, I., Mostowski, B., and Žitković, G. (2022) Kyle's model with stochastic liquidity.

Emna, R., and Chokri, M. (2014) Measuring Liquidity Risk in an Emerging Market: Liquidity Adjusted Value at Risk Approach for High Frequency Data. *International Journal of Economics and Financial Issues*, *4*(1), 40–53.

Glattfelder, J. B., Dupuis, A., and Olsen, R. B. (2011) Patterns in high-frequency fx data: Discovery of 12 empirical scaling laws. *Quantitative Finance*, *11*(4), 599–614.

Glosten, L. R., and Harris, L. E. (1988) Estimating the components of the bid/ask spread. *Journal of Financial Economics*, *21*(1), 123–142.

Harris, L. (2002) *Trading and exchanges: Market microstructure for practitioners.* Oxford University Press.

Hasbrouck, J. (1991) Measuring the information content of stock trades. *Journal of Finance*, *46*(1), 179–207.

Jacobsen, B., and Visaltanachoti, N. (2009) The halloween effect in u.s. sectors. *The Financial Review*, *44*(3), 437–459.

Kyle, A. S. (1985) Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.

Ljung, G. M., and Box, G. E. (1978) On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.

Madhavan, A. (2002) Vwap strategies. *Trading*, *1*, 32–39.

Matz, L., and Neu, P. (2007) *Liquidity risk measurement and management: A practitioner's guide to global best practices.* John Wiley & Sons (Asia) Pte Ltd.

Muller, U. A., Dacorogna, M., Olsen, R., Pictet, O. V., Schwarz, M., and Morgenegg, C. (1990) Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking & Finance*, *14*(6), 1189–1208.

O'Hara, M. (1995) *Market microstructure theory.* Blackwell.

O'Hara, M. (2015) High frequency market microstructure. *Journal of Financial Economics*, *116*(2), 257–270.

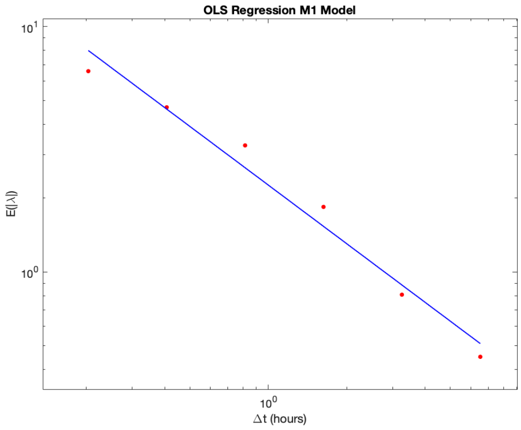
Qi, J. (2011) *Risk measurement with high-frequency data: Value-at-risk and scaling law methods* (Doctoral dissertation). University of Essex.

Qi, J., Yi, L., and Chen, Y. (2018) Forecasting market risk using ultra-high-frequency data and scaling laws. *Quantitative Finance*, *18*(12), 2085–2099.

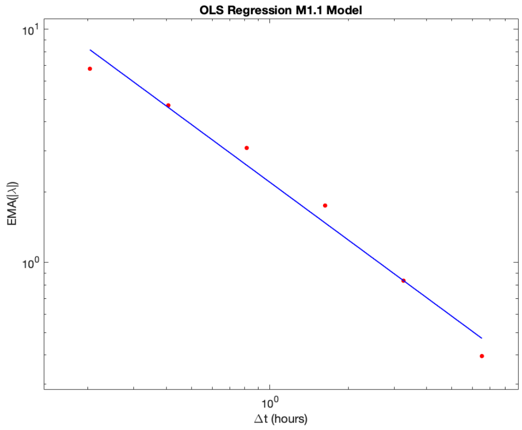
Russell, J. R., and Engle, R. F. (2010) Chapter 7 - analysis of high-frequency data. In Y. AÏT-SAHALIA and L. P. HANSEN (Eds.), *Handbook of financial econometrics: Tools and techniques* (pp. 383–426). North-Holland.

Appendix

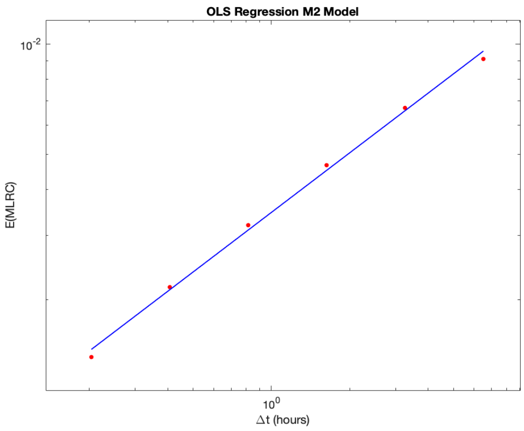
A OLS Regression Models



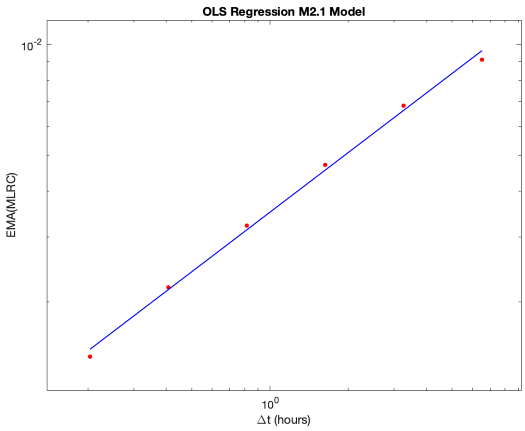
(a) M1 Model



(b) M1.1 Model



(c) M2 Model

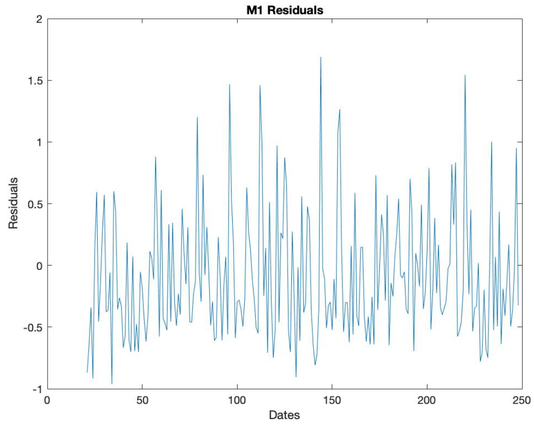


(d) M2.1 Model

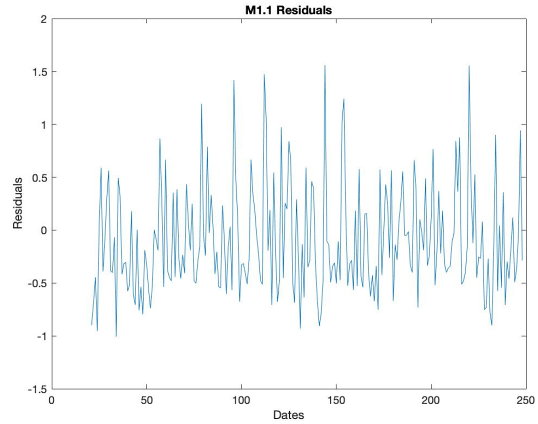
Figure 9: *An example of the scaling law regression lines for all the model specifications using NASDAQ Composite Index at 1-minute frequency from 03 April 2023 15:30:00 to 02 May 2023 22:00:00 (Italian Time).*

Figure 9 depicts the OLS regression models applied to estimate the scaling law parameters using last monthly observations. The graph shows that the estimated regression line is a good approximation of the data, confirming the existence of scaling law for the variables. All the models specifications present a R^2 close to 0.99.

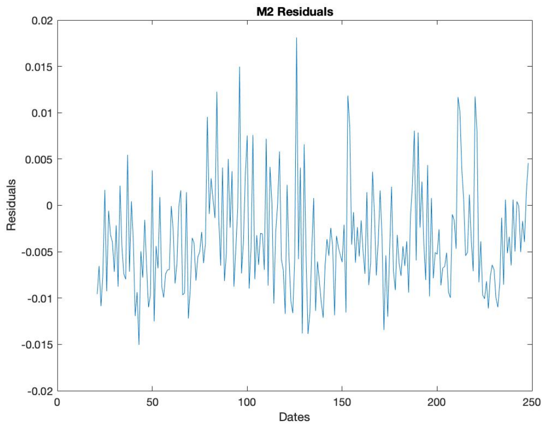
B Residuals diagnostics



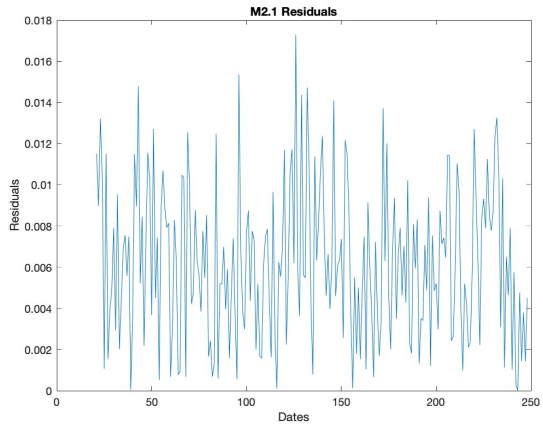
(a) M1 Residuals



(b) M1.1 Residuals



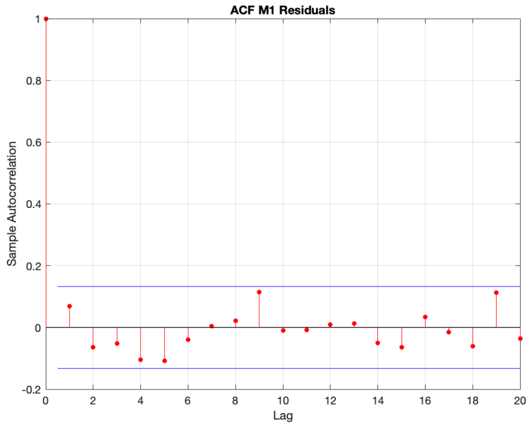
(c) M2 Residuals



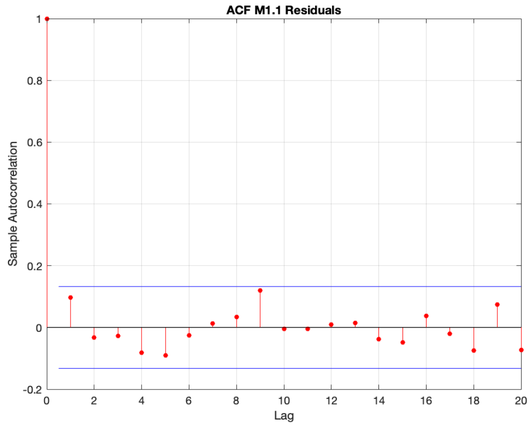
(d) M2.1 Residuals

Figure 10: *Plot of the residuals.*

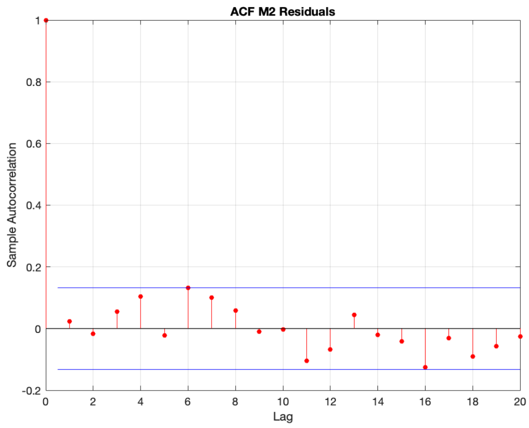
Figure 10 shows the residuals for all the scaling law models. The p-values of the Augmented Dickey-Fuller test are all below the 5% threshold for all the scaling models, rejecting the null hypothesis that a unit root is present in the time series.



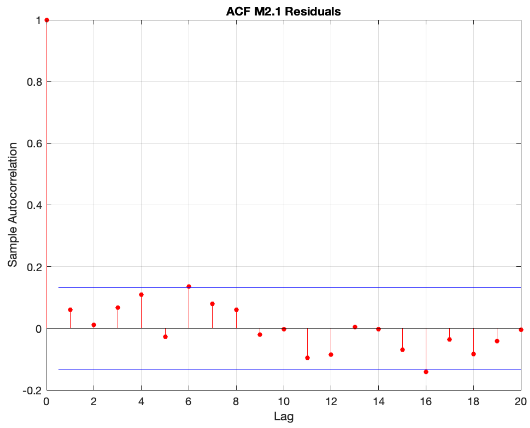
(a) M1 Residuals



(b) M1.1 Residuals



(c) M2 Residuals



(d) M2.1 Residuals

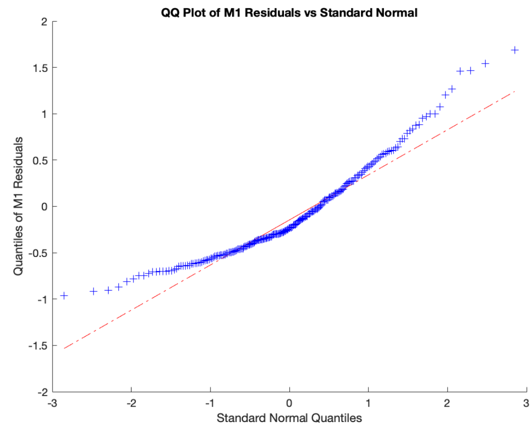
Figure 11: *ACF of the residuals.*

Figure 11 shows the ACF of the residuals for all the scaling law models. The p-values of the Ljung-Box test are shown in the following Table 4.

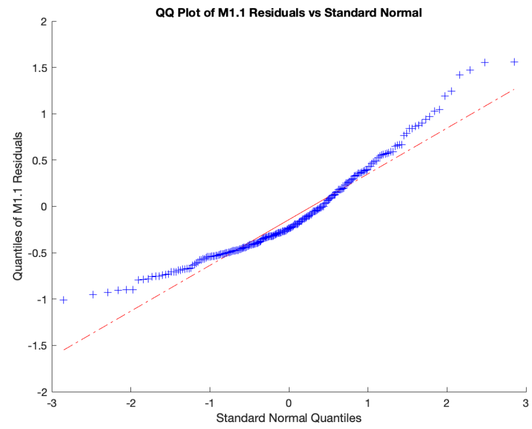
Scaling Law Models	P-value
Model 1	0.1585
Model 1.1	0.3015
Model 2	0.6177
Model 2.1	0.4318

Table 4: *P-values of the Ljung-Box test with 5 lags.*

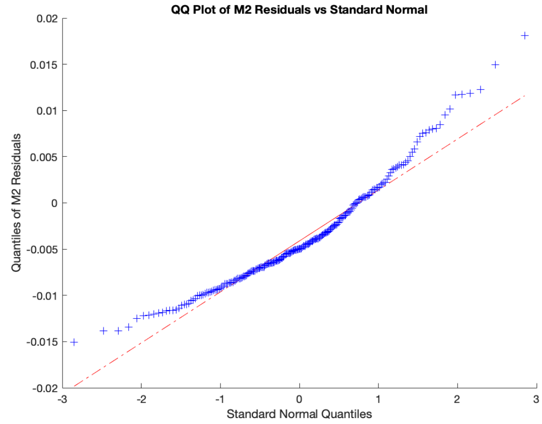
For all model specifications, the null hypothesis of independence of residuals for 5 lags cannot be rejected, since the p-values are all above the 5% threshold.



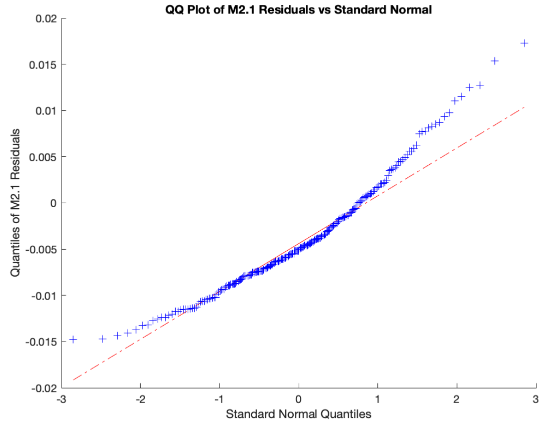
(a) M1 Residuals



(b) M1.1 Residuals



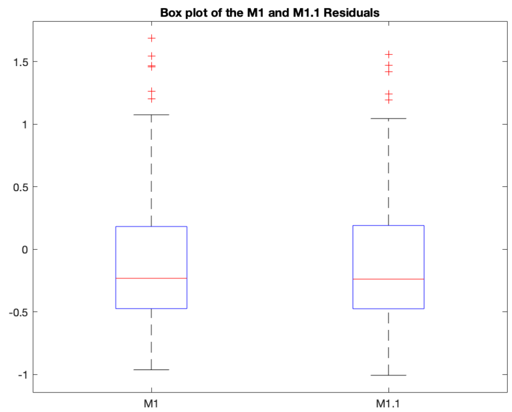
(c) M2 Residuals



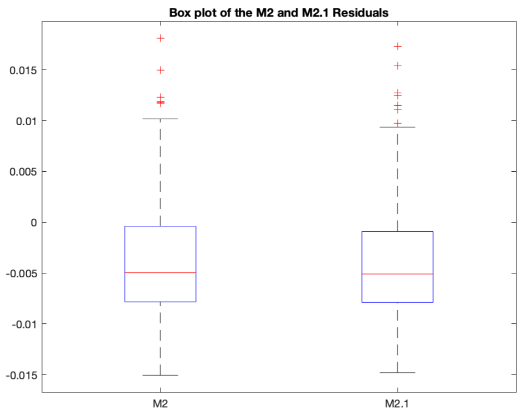
(d) M2.1 Residuals

Figure 12: *QQ plots of the residuals distributions vs Standard Normal distribution($\mathcal{N}$).*

Figure 12 shows the QQ plots of the residuals for all the scaling law models. The evidence of skewness and kurtosis is underscored by looking at these graphs, due to the presence of outliers in both tails of the distributions. All of them are leptokurtic and positively skewed.



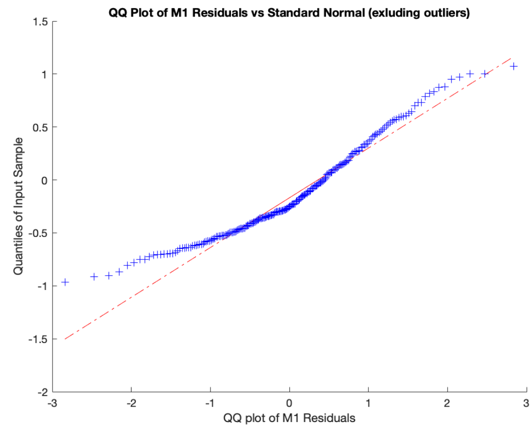
(a) M1 & M1.1 Residuals



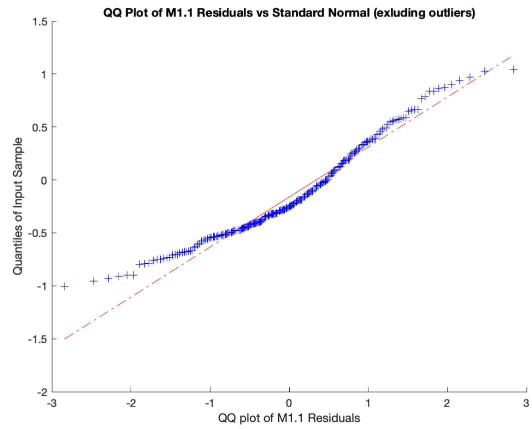
(b) M2 & M2.1 Residuals

Figure 13: *Box plots of residuals distributions.*

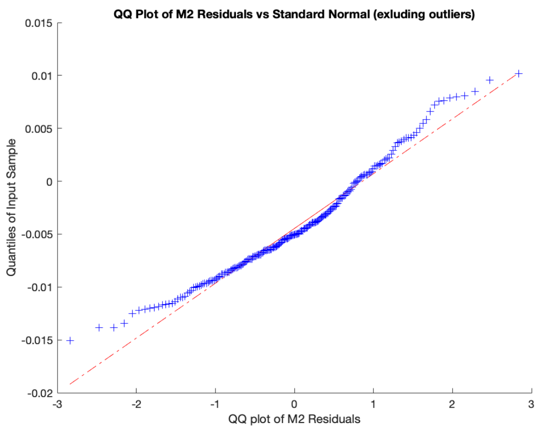
Figure 13 shows the box plots of the residuals for all the scaling law models. By looking at the graphs it is possible to notice that the outliers are all on the right tail of the distribution, due to the fact that the residuals are positve skewed.



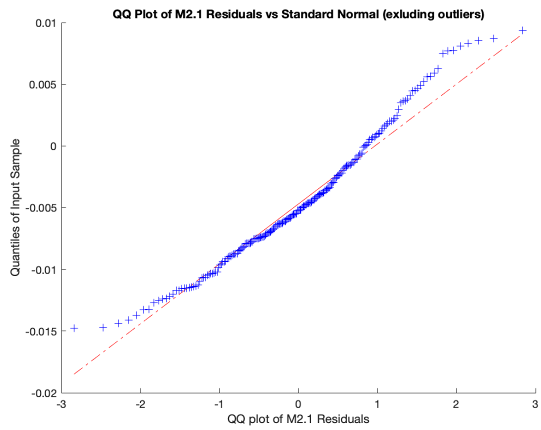
(a) M1 Residuals (excluding outliers)



(b) M1.1 Residuals (excluding outliers)



(c) M2 Residuals (excluding outliers)



(d) M2.1 Residuals (excluding outliers)

Figure 14: *QQ plots of residuals distributions vs Standard Normal distribution($\mathcal{N}$) (exluding outliers).*

Figure 14 shows the QQ plots of the residuals for all the scaling law models excluding the outliers. The distributions of the residuals without the outliers seem to better approximate that of a Standard Normal, especially in the right tail, although some kurtosis and skewness remain.