Small Firm Factor: Riskiness Evaluation Under A High-Frequency Regime

Rogerio Batista Adelino[1], Leonardo Fernando Cruz Basso[1] & Eli Hadad Jr.[1]

[1] CCSA – Universidade Presbiteriana Mackenzie, Sao Paulo, Brazil

Correspondence: Rogerio Batista Adelino, CCSA – Universidade Presbiteriana Mackenzie, Sao Paulo, Brazil. E-mail: rogerioadelino@gmail.com

Received: May 23, 2024 Accepted: July 17, 2024 Online Published: February 6, 2025

doi:10.5430/ijba.v16n1p1 URL: https://doi.org/10.5430/ijba.v16n1p1

Abstract

Since the advent of multifactor models, the identification of factors that could lead to abnormal returns and their portfolio implications has been the theme of consistent effort and study. Among these factors, the small-cap factor (SMB factor) is one of crucial importance. This paper provides an innovative approach to the evaluation of the SMB factor taking into consideration the time-series comparison of small and large caps under a high-frequency regime, identifying structural differences in terms of stochasticity and presence of jumps. The empirical results confirm the specificity of small caps, understood therefore as a diversifying asset class. Additionally, results fail to confirm abnormal returns in the sample, indicating that factor momentum shall be considered. Moreover, the paper demonstrates a framework for high-frequency statistics usage in investment evaluation.

Keywords: rewarded factors, small firm factor, Brownian Motion, jump, market risk

1. Introduction

Investors and portfolio managers are often concerned about the drivers of return and risk in their portfolios. As a collection of stocks, equity portfolios consist of a combination of specific characteristics related to the particular portfolio components and their combination. Initially, different firms typically embed investment ideas, expected to be profitable under an economic scenario. Moreover, the diversification feature may improve the riskiness profile. Investment strategies cannot, in this sense, be evaluated exclusively in terms of realized returns but rather in terms of the risk exposure utilized to achieve a return.

Measurement and evaluation of risk, then, is a cornerstone in financial literature. Ali & Bashir (2022) perform a bibliometric review of asset pricing that points to this relevance. One of the most important models in this direction, the CAPM, points to a positive relationship between a stock and the market risk, measured by the β coefficient. As idiosyncratic risk is diversified away, investors would retain a certain exposure to the market risk, the only source of risk. In such a model, firms are treated as stand-alone risk components of a portfolio, which is somewhat untrue in reality. Stocks and firms share underlying common risk factors among them, potentially related to specificities of economic sectors, markets in which they operate, supply chain constraints, and many other characteristics. Besides the market risk, portfolio design also implies tilts toward specific factors, highlighting the need to understand these additional exposures.

Blitz & Vidojevic (2019) evaluates factor exposures in the portfolio logic. In this context, portfolios can be understood as a collection of distinct factor exposures, which may drive the capacity of tilting towards a specific directional view. Strategies in this direction, also known as smart betas, should be able to improve the return/risk profile of portfolios.

Fama & French (1992) pioneer this relationship by the inclusion of two additional risk factors, namely the size and the value effects. The first exposure is related to the comparison of the returns of small and large market capitalization firms and the last is relative to the book-to-value ratio. In this way, the authors identify the risk of a stock as a multidimensional vector, with one key element related to the log of the market capitalization of the firm. As a consequence, an investment and portfolio management strategy may be derived, taking into consideration the special features of smaller firm equity, in terms of its risk and return profiles. Hackel, Livnat & Rai (1994) observed, indeed, that financial institutions started offering small-cap investment vehicles as soon as this research was released.

Among the theoretical reasons for such idiosyncratic performance, some articles postulate that the information differential between small and large firms would be the central piece for the distinct investment performance, as in the case of Barry & Brown (1984). In addition to this, another line of research focuses on productivity differentials and funding needs. Works in this line, as in the case of Kim & Burnie (2002), may attribute higher performance volatility to small firms as a result of higher operating and financial leverage and lower access to scale and technology, resulting in higher exposure to economic cycles.

He & Ng (1994) relate the riskiness of the size factor to distress, which could be characterized both by cash flow generation volatility and risk of failure. In addition, the size factor correlates with both yield curve term and credit spreads, indicating high sensitivity of smaller firms to the economic cycle, even considering the potential compensatory effect of a healthier (weaker) economy, with higher (lower) term premium, and the improvement (de-crease) of credit risk, reducing (increasing) credit premium. In another macroeconomic discussion, Everett & Watson (1998) acknowledge that a small firm's mortality is directly related to interest rate dynamics, besides internal causes such as lack of management skills and capital.

Babenko, Boguth & Tserlukevich (2016) identify size risk premium by segregating the effects of rewarded systematic risk, as dictated by the CAPM model, and idiosyncratic cash flow shocks. The research suggests that small firms should have higher required returns in accordance with higher risk.

In terms of empirical research, Hackel, Livnat & Rai (1994) point to abnormal returns using a portfolio of small firms with consistent operating cash flow generation. Switzer (2010) points to an average premium of 2.03% for small-cap equities, with a relevant degree of variability in time. The author, in discussing if this factor is still relevant on the market, observes the fact that the factor Jensen's Alpha regained statistical significance post-year 2000. Moreover, the finding that the performance of small firm equities is dependent on the stage of the economic cycle and country idiosyncrasies may explain such alteration of the size effect in time. On a global analysis over three decades, Hou, Karolyi & Kho (2011) don't favor the size effect, potentially indicating that the factor may not be constant and globally diffused.

Gandhi & Lustic (2015) find evidence of a size effect in the US banking industry. This factor, which is distinct from the size factor in non-financial firms, also plays a role related to riskiness in times of financial distress. The main difference to be taken, according to the authors, is that, in the case of banks, government bailing is more probable in bigger institutions.

In general, the studies on factor investing compare the cross-sectional performance of stocks with subsequent time series modeling to understand the 'average' loading of the specific factor β or contrast the performance of portfolios tilted by a specific factor. These works can be understood as an external approach to the risk factors that seek to identify the existence of the anomaly and eventually propose reasons based on co-movement with other variables or occurrence in time and space.

As a consequence, previous studies on the theme are centered on the fundamental reasons that would lead to performance anomalies, lacking a direct view of the volatility phenomena itself. From a mathematical point of view, risk can be approached as a form of dispersion around a trend or centrality measure. In such way, the riskiness of an asset cannot be properly understood if a time-series methodology is not employed. This fact is becoming even more relevant as the science related to continuous-time econometrics advances. In the current state, volatility can be analyzed in fractions of seconds, as the phenomena really occur in financial markets, and not obscured by data compilation in less frequent time batches.

This understanding may uncover crucial information for the proper treatment of risk factors, such as size, both in the academia and in real-life portfolios. As real investment strategies focus on timing risk factors, according to some economic conditions, the path trailed by this study, focused exclusively on a high-frequency evaluation of the size factor, may imply new and important ground on how to select a specific risk factor. Additionally, this approach may convene a more complete way to understand the interdependence of a risk factor to the general market risk. Previous studies evaluated risk factors as a stand-alone volatility component that is segregated from other factors or market risk. Our approach may indicate a more fluid correlation to the market risk, improving the understanding related to the expectation of the diversification effect brought by a single factor exposure to the portfolio.

This paper intends to provide, in this way, an innovative view of the size factor. This research is the first, to our knowledge, that seeks to evaluate the intrinsic characteristics of distinct time series (more specifically, small and large-caps) regarding their degree of continuity and disruption, using the lenses of high-frequency data. The goal is to understand the mechanical differences in the time series of small and large cap indices, and therefore comprehend how the risk manifests itself for different factor loadings. By comparing different Russell indices in high frequency, we depict inside-out features of time series for large and small caps, pointing to their idiosyncratic features and allowing for an analysis of the risk and return profile related to the size effect. The choice for this family of indices is also an important feature of the paper as it emphasizes proper comparison, not influenced by inconsistencies in index rebalancing and re-constitution, to mention a few.

Results obtained clearly differentiate the Russell 2000 Index, the small-cap index scrutinized, from the Russell 1000 Index, the large-cap index, and Russell 3000 Index, the broad market index. Moreover, time series jumps are asymptotically more probable in the Russell 1000 Index, in the sample analyzed. As sample dependency is plausible, the remarkable difference between the small and large caps supports the theoretical approach of considering size as a specific factor exposure. Another important point aligns with the similarities of the large-cap index and the broad market one, implying in low diversification potential of the risk factor unless a specific small-cap vehicle is directly used in the portfolio. At last, all three time series demonstrate the presence of a stochastic component, following a Brownian Motion dynamic.

This article aims to contribute to the theoretical framework of finance and investments in two ways. Initially, it depicts a significant difference in the micro-structure of the time series of the small-cap index, as indicated by a different pattern of jump incidence, when compared to large firm equities. Based on this point, it allows for a distinction between the riskiness pattern of small caps. Also, the research offers a fresh view of the mechanism behind the distinct risk pattern of small caps, from an angle not yet pursued, to the extent of my knowledge. This approach, related to the high-frequency comparison of microstructure risk, is in our view not yet explored in the literature. Additionally, the research also aims to indicate points of attention and action to practitioners in investment management. A broader understanding of risk sources is a crucial tool in evaluating portfolio strategies that may make use of smart beta strategies or explicitly include the evaluation of re-warded risk factors in its composition. At last, this essay brings a practical case study using high frequency, contributing to an analytical path that is still under development.

This paper is structured in the following manner. Session 2 describes different time series models. The session follows a crescendo, in terms of features utilized in the models, culminating with the utilization of features and culminating with a model that allows for discontinuity under high-frequency data. Session 3 discusses the data used in the paper, consisting of 1 minute period time series for the three indices evaluated. Session 4 describes the econometric model used in the research. Session 5 depicts the empirical results achieved and highlights economic implications. The work closes with concluding remarks on Session 6.

2. Theoretical Framework

According to Tsay (2000), the utilization of time series in business and economics can be related to the understanding of the dynamic structure of a process including seasonality and relationship among variables, in which the presence of serial correlation improves regression estimates and allows for forecasting production. In general terms, a time series is produced as a next-period forecast of process-derived mechanics, in which the information set available is used to predict the forward point and an innovation shock that can be associated with the error, noise, or a stochastic process.

Among the many distinctions and classifications potentially used on this subject, this paper shed light on the features included in some models to capture the stochasticity of the movement around univariate time series, which could, under the high-frequency regime, imply in discontinuity of the series. The models presented are not an exhaustive list of time series methodologies existing nor are arranged in a historical timeline. The classification proposed suits the understanding of these models as a set of features included in each that would lead to a specific model design and result interpretation.

Distinct classifications, such as time domain and frequency domain approaches, traditional or Bayesian, and univariate and multivariate, among others, as well as discussions about non-linearity or non-normality and distinct models, can be found in Tsay (2000) and Macaira et al. (2018) but are not addressed in this paper.

2.1 Continuous Stochastic Differential Equation Models

Continuous stochastic models include a random noise element in their construction without representing a disruption in the time series path. In this kind of model, as defined by Evans (2013), the mathematical grounds for forecasting move to stochastic differential equations, as an aleatory element is included in the function, resulting in movement uncertainty in the time series.

Formally stating, Karatzas & Shreve (1998) define such models as a collection of random variables, based on the probability function in the sample space (Ω) and event space (F) and, therefore, measurable under the dimensional distribution of the probability and time, or progressively measurable in terms of its stopping times. Evans (2013) shows that a continuous and differentiable series can be modeled with a smoothing vector b. The equation for this kind of model would be expanded from the one presented in autoregressive models to assume a form of $x_t=b(x_{(t-p)})+B(x_{(t)})\varepsilon_{(t)}$ which considers the inclusion of a random term to deal with the uneven pattern of the curve.

As shown by DeFusco et al. (2007), in its simpler case, a random walk model considers that the best forecast for a specific period would be the value of the variable in the previous period plus a random movement, that cannot be predicted. Such a model has some important features. Initially, it can be seen as a special case of an AR (1) model with an intercept equal to zero and a slope equivalent to one, generating an undefined mean reversion pattern and a variance of errors that is directly proportional to time, and therefore, limiting to infinite as time grows. Based on these facts, normal regression analysis cannot be employed to solve the model (as would be the case for AR models), unless the function is modified through differentiation. An extension of this discussion would reside in the understanding that this process is bounded by a Brownian motion mechanism, as explained by Au, Raj & Thurston (1997).

On a historical note, Lavenda (1985) describes the mathematical mechanism of Brownian motion since the original studies of botanist Robert Brown related to the agitation of particles in a fluid to modern applications such as economics or navigation systems. The concept had crucial relevance in atomic physics and eventually arose into the segregation of a general drift movement and short-term random fluctuations around it. As an example, the Langevin equation, in thermodynamics, decouples movement in a macroscopic world drag force (temperature differentials, for example) and a microscopic fluctuating Brownian force.

According to Evans (2013), as the white noise is transformed to a Brownian motion or Wiener process W(t)=ε(t), a Stochastic Differential Equation is derived from the original Ordinary Differential Equation, as described in the AR model, as it can be noted in Equation (1).

$$\begin{cases} dx_t = b\big(x_{(t)}\big)dt + B(x_{(t)})dW_{(t)} \\ x_{(0)} = x_0 \end{cases} \tag{1}$$

Which may be solved using the stochastic integral approach as depicted in Equation (2).

$$X_t = x_0 + \int_o^t b\big(x_{(s)}\big)\, ds + \int_0^t B\big(x_{(s)}\big)\, dW \tag{2}$$

This equation can be broken into three parts. In addition to the intercept, the second part is a non-aleatory moment related to a drift and the third part is the stochastic movement. In this point, three important issues must be briefly discussed: how the definition of martingales aligns with this model, the application of Wiener process properties to Brownian motion, and the interpretation of white noise.

Starting with the notion that martingales are random processes in which the knowledge of previous moments, or stopping times, cannot provide information for future movements of the variable, Karatzas & Shreve (1998) underline that the standard example of such process is a one-dimensional Brownian motion, that also could be directly related to the continuous time version of symmetric random walk. This is, in fact, the model explained in this session, meaning that it refers to a martingale.

In addition to that, as an important application in the field of finance, a sub-martingale is an interesting delimitation of that concept as it embeds a non-negative return expectation, of $E(x_t\|F)\geq x_s$ for $0\leq s\leq t\leq\infty$. The sub-martingale is subjected, in this way, to Doob-Meyer decomposition as a bounded martingale and an increasing trend, which is the base for the discrete-time stochastic integral.

Karatzas & Shreve (1998) consider an alternative for building a Brownian motion from interpolation, when $\{B_t, F_t, t\geq 0\}$, with $0\leq s\leq t\leq\infty$ and values $B_s = x$ and $B_t = z$, has a mean $(x + z)/2$ and variance $(t-s)/4$. The authors highlight this approach as a continuation of the one known as the Wiener process, where $E(W_t) = 0$ and $\sigma^2(W_t) = t$, under a normal probability distribution $w_t(x) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/(2t)}$. Hall (1969) demonstrates that the study of martingales and certain sub-martingales can be analyzed using Wiener process and drift, in a sense that they would share the same law. In this way, the application of Brownian motion to random walks is theoretically sound.

Following this point, Karatzas & Shreve (1998) contextualize white noise as the dW, in terms of “statistical communication theory” (KARATZAS & SHREVE, 1998, page 395). This term is used later in subsequent models and can be interpreted in terms of the randomness implied by time.

Brownian motion’s logic is based on the probability that a certain function path will be encompassed under a certain interval at a specific point in time. The Brownian motion follows a Normal Distribution where the expected mean is zero, indicating an unbiased path W(t)=N(0,t).

At this point, we evaluate some of the aspects related to the estimation of probability density function as its implications on the Brownian motion. As described by Parzen (1962), this problem is similar to the estimation of spectral density function on a stationary time series, so the methods used may be similar to the estimation of a maximum likelihood parameter. This may be particularly important to the objective of this paper as the central problem resides in the forecast of time series with the spectrum of variation defined by a probability distribution function derived from a specific density.

2.2 Jumps, Martingales, “Broken Curves”: Models for High-Frequency Data

The models presented until now assume that the time series is continuous. This is true even in the case where a stochastic element is considered, such as presented in Equation (3). Ait-Sahalia & Jacod (2012) state that the asset price must follow a semi-martingale assumption to avoid arbitrage opportunities. As analyzed, the drift and the Brownian motion elements correspond to this assumption in Equation (3). On the other hand, the authors work with an additional component in their model, opening the possibility of curve discontinuation due to the presence of jumps.

Makinen et al. (2019) define jumps as large price movements in microstructure that will extrapolate the explainable capacity of the Brownian motion. As it creates a relevant discontinuity on the price path, the return potentially obtained in the jump is abnormal to the continuous innovation process. From another point of view, jumps impose a high degree of realized volatility, which may spark volatility spikes in the traditional time series as volatility clusters.

The evaluation of such models is becoming more feasible due to the onset of high-frequency data, pushed by developments in computing and information technologies. From the information availability standpoint, open and close positions, bid-ask spreads, volumes, and other relevant stock market information can be provided in the blink of a second to practitioners and scholars. From a trading point of view, Makinen et al. (2019) indicate that the technological advancement of systems and algorithms is enabled by enhanced computing and colocation schemes, and may manage a great number of market orders in a split of milliseconds.

Goodhart & O’Hara (1997) highlight this development on the potential for a continuous observation of transactions, not bounded by discrete time frames. This potentially new perspective may impose crucial changes on econometric models as the process may become time-varying.

The reason for this point is that the analytical limit of this data is related to the possibility of recording each transaction as a separate time moment. In such cases, defined as ultra-high-frequency, Engle (2000) defines a model in which the information embedded in the transaction is a random variable but also its timing is. For this strongly granular approach, many moments in time will not register a single transaction, meaning that its occurrence per se implies informational interest. Moreover, as more moments are filled with transactions, the author identifies the volume as the reciprocal of duration between the trades, which is negatively correlated with moment volatility.

In addition to this point, Goodhart & O’Hara (1997) associate market movement with the number of transactions, in a way that the high-frequency data series can still be seen as a martingale but not a Markovian, meaning that the enchainment of prices does influence the next stages of the time series. In such a way, the market is a dynamic construct with changing sentiments of buyers and sellers being computed continuously, with some implications. Examples of these effects are that, as news is processed, the market may show a peculiar movement that traditional econometric models will fail to perceive or, as critical technical points are reached, volatility may be severely impacted until confirmation or rejection of the signal is clear.

In this way, a high-frequency periodicity allows for a distinctive pattern in which it may be broken in such moments of abnormal transactions. In more formal terms, previous findings of volatility clustering can be translated in terms of the number of transactions and the potential presence of jumps in the time series.

The described inclusion, in the previous session, of a Brownian motion into time series such as a random walk implies that part of the movement is aleatorily defined under a normality assumption with variance as a function of time. This randomness, on the other hand, doesn’t disrupt the continuous character of the time sample path, as explained by Ait-Sahalia & Jacod (2012).

The model developed by the authors is depicted in Equation (3). In this case, besides the continuous random element, it also incorporates discontinuous jumps, which would be apprehensible when returns are being analyzed on a high-frequency pattern.

$$x_t = x_0 + \int_0^t b_s \, d_s + \int_0^t \sigma_s \, dw_s + JUMPS \tag{3}$$

This model raised a strong interest in analyzing markets. Pan (2002) acknowledges the presence of both stochastic and jump components in stock returns and that the risk of jumps is priced into options, and therefore into implied volatility, especially in moments of higher spot volatility. Makinen et al. (2019) develop a jump prediction methodology based on the asymmetry of limit orders book. Hu et al. (2019) identify the jumps as a source of market volatility, in an environment where a jump in stock, especially in the financial sector, may contaminate other firms.

It is straightforward that Equation (3) is the same as Equation (2), plus the jump component, which marks the discontinuity feature of the model. The application of the model proposed by Ait-Sahalia & Jacod (2012) seeks to decide which are the components to include and their magnitude, besides outlining the economic implications of these models. As an example, such discontinued models may change current views on normal contingent claim valuation and optimal portfolio choices, as they would be dependent on price dynamics. Moreover, many high-frequency trading strategies rely on specific components of the model, based on small profit with a “better than 50” chance that is repeated several times over very small time intervals. By using ultra-high-frequency, there’s a limitation related to data sampling that is available, reducing the capacity of looking into long series or illiquid assets. In addition, market microstructure noise is always a concern.

3. Data

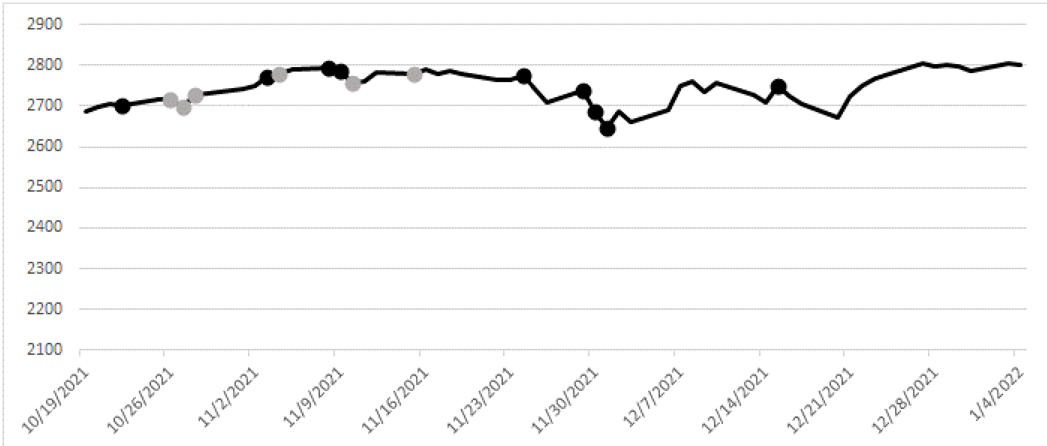
The data utilized in this research comprises 54 trading days from Oct/10th/2021 to Jan/4th/2022. The Russell 1000 Index, Russel 2000 Index, and the Russell 3000 Index were collected based on a 1-minute periodicity, exclusive to regular operating times of stock exchanges, 9:30 AM to 4:00 PM. In all data points, the closing index is recorded, with the exception of 9:30 AM when both the closing and opening prices are recorded. The last point collected is the closing of 3:59 PM. Returns are calculated using continuous compounding from one minute to another, excluding the first minute of trading each day, as suggested by Ait-Sahalia & Jacod (2012) as a result of the distinct volatility profile of that first minute.

Each trading day in the sample contains 389 points, except for Black Friday on Nov/26th/2021, where the market operated until 1:00 PM. Based on this, the database consists of 20,826 points for each of the three indices.

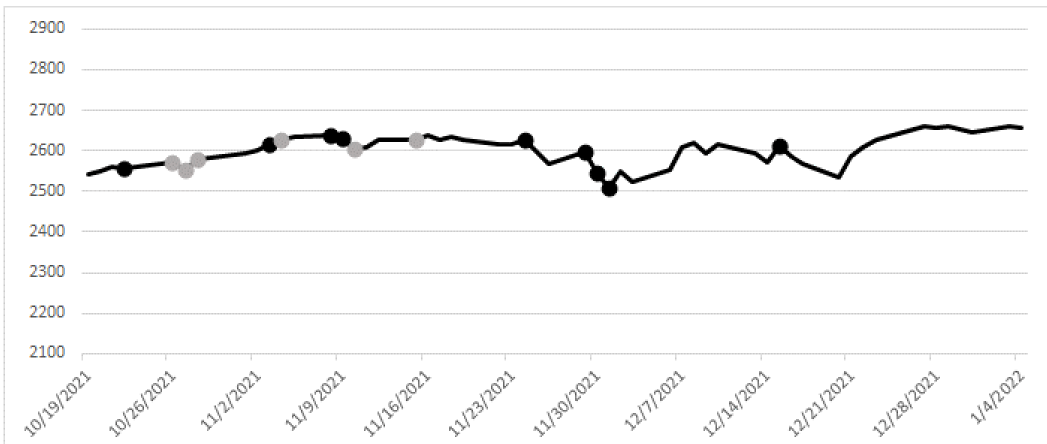
The Russell family of indices is one of the most important standards in the market, with more than 9 trillion USD of assets directly benchmarked. They are designed to evaluate broad market performance and allow for specific views of market subsets, in terms of size and style (FTSE, 2021). Moreover, the cohesive methodology allows for direct comparison among those indices without external problems, such as reconstitution and rebalancing. The entire family derives from an index that contains the 4,000 largest public companies in the US, from which the Russell 3000 Index is generated as a gauge for the broad market, including 3,000 firms. Both Russell 1000 and Russell 2000 Indices are subgroups of the Russell 3000 Index, where the largest 1,000 companies are included in the first as a measurement of the large-cap segment, and the smallest 2,000 firms are included in the last as a gauge of the small-cap segment.

During the period, the performance of small caps largely lagged large caps, with a period return of 0.05% against 5.32%. In addition, a visual inspection of Figure 1 unveils a higher degree of volatility associated with the first, in the time frame. In numeric terms, the daily standard deviation of the small-cap index is 1.40% while the same statistic for the large-cap is 0.89% within the sample. Another important point depicted in the chart is related to relevant points in terms of volatility events. During the period, there were 9 dates related to interest rate decisions (Nov/3rd/2021, and Dec/15th/2021), Fed meeting minutes publication (Nov/24th/2021), and general speeches of Fed Chairman J. Powell. All these events are marked in black. Also, 8 days were selected as the most important earnings-reporting days in the period, ranging from 217 to 660 companies reporting on the day. Those days are depicted in gray.

Panel A: Russel 3000 Index with Selected Dates



Panel B: Russel 1000 Index with Selected Dates



Panel C: Russel 2000 Index with Selected Dates

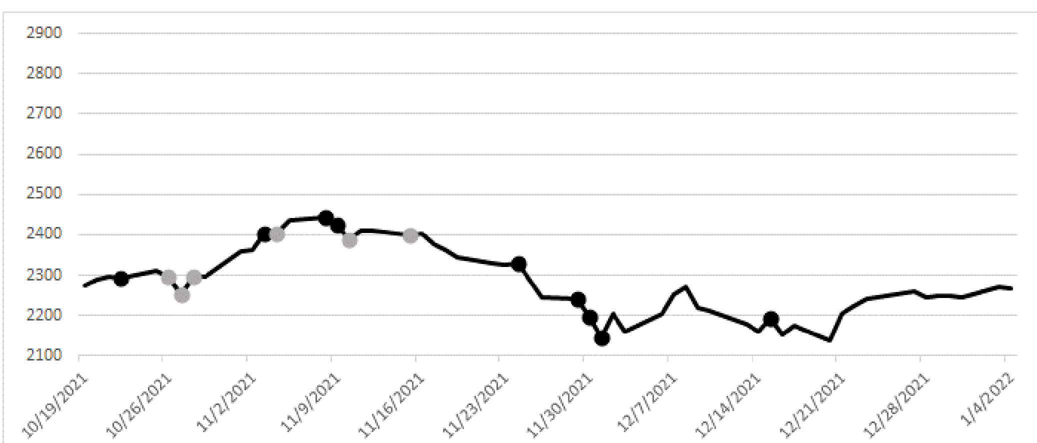


Figure 1. Time Series of Russell Indices

Source: Model Calculations

The chart demonstrates that volatility in the indices is not exclusive to the occurrence of events. By evaluating abrupt changes in the time series, it is possible to understand that some volatility clusters coincide with event days but this is not a necessary condition. It seems, though, that Fed days and busy announcement days are important for the market. In addition, the change in slope on the Russell 2000 Index has a more pronounced effect than on the large-cap segment.

Another important finding is related to the greater adherence of the Russell 3000 Index to the Russell 1000 Index. This point implies that, from an index composition point of view, an investment in a broad market index is somewhat similar to a large-cap position, and that small caps, then, can be seen as a distinct set, with potential benefits in terms of diversification, as prescribed by Modern Portfolio Theory. To evaluate this aspect, Table 1 provides basic descriptive statistics for the indices, with Panel A focusing on daily frequency and Panel B on minute frequency.

In order to test for the correlations among the indices, the test developed by Lawley (1963) is used. In both cases, the Lawley Chi2 test yields p-values of 0.0000. We, therefore, reject the null hypothesis of symmetric correlation and that correlations are not equal to each other. This result confirms the different nature of small caps, as pointed out by the literature, in terms of their risk and return profiles.

Table 1. Descriptive Statistics of Russell Indices

Panel A: Descriptive Statistics – Daily Frequency

	Russell 3000	Russell 1000	Russell 2000
Initial Date	Oct/18/2021	Oct/18/2021	Oct/18/2021
Final Date	Jan/04/2022	Jan/04/2022	Jan/04/2022
Sample Size	54	54	54
Average Return	0.0009004	0.0009603	8.41e-06
Median Return	0.0015858	0.0016299	0.000272
Standard Deviation	0.0091624	0.0089614	0.014046
Skewness	-0.2264384	-0.2120141	-0.1035943
Kurtosis	3.233298	3.2383	2.950903
Correlations			
Russel 3000	1.0000		
Russell 1000	0.9987	1.0000	
Russell 2000	0.8777	0.8522	1.0000

Panel B: Descriptive Statistics – Minute Frequency

	Russell 3000	Russell 1000	Russell 2000
Initial Date	Oct/19/2021 Open	Oct/19/2021 Open	Oct/19/2021 Open
Final Date	Jan/04/2020 Close	Jan/04/2020 Close	Jan/04/2020 Close
Sample Size	20,826	20,826	20,826
Average Return	3.38e-08	1.22e-07	-1.99e-06
Median Return	5.50e-06	1.22e-07	-1.99e-06
Standard Deviation	0.0002838	0.0002861	0.0003267
Skewness	-0.1910511	-0.199024	0.1555979
Kurtosis	11.95432	11.9699	11.66066
Correlations			
Russel 3000	1.0000		
Russell 1000	0.9922	1.0000	
Russell 2000	0.7579	0.7228	1.0000

As for the distribution, the minute-wise distribution is remarkably different in terms of the fat-tail situation, in all three cases. This point aligns with the findings of Ahadzie & Jeyasreedharan (2020) related to the direct relationship between the increase in sample frequency and the realized Skewness and Kurtosis. On the other hand, there is not a strong distinction in the asymmetry for Russell 3000 and Russell 1000, as all indices show negative skewness both in minute and daily frequencies. On the Russell 2000, the signal inverts when the frequency is changed, indicating an idiosyncrasy in the high-frequency pattern for the series.

4. Model and Measuring Devices

In order to understand the specific features of each of the Russell family indices, time series will be subjected to tests under the high-frequency regime. This will allow for comparison among the indices, with potential economic conclusions arising. The research is based on the methodology of realized power variations, as proposed by Ait-Sahalia & Jacod (2012). The model consists of the calculation of test statistics on the realized power variation variable (B), as depicted in Equation (4), with three controls being adjusted to isolate the specific components of the time series.

$$B(p, u_n, \Delta n) = \sum_{i=1}^{T/\Delta n} |\Delta_i^n x|^p \, 1_{\{|\Delta_i^n x| \leq u_n\}} \quad (4)$$

The first adjustment is related to the power (p), which function is related to the isolation of the continuous or the jump components in the series. When the value of p = 2, the effect of both jumps and continuous components of the series are balanced. This is the result, as explained by Ahadzie & Jeyasreedharan (2020), of the convergence of realized variance to the sample variance in a way that higher frequency sampling may imply more efficient estimates for the variance.

On the other hand, the limits of the third and fourth moments converge to the sum of cubic and quartic jumps, respectively. In this way, an increase in the power variable will highlight the effects of the jumps in the series and disregard the continuous diffusion process.

The second adjustment variable is related to the truncation mechanism (u_n), as a form to isolate specific jump sizes. This variable is important in segregating large and small jumps, which allows for the evaluation of its occurrence, in terms of being finite or infinite. At last, the sampling frequency (Δn) is a key element in identifying the limiting behavior of the realized power variation, by comparing the highest available frequency with a larger periodicity.

From the analytical methodology of Ait-Sahalia & Jacod (2012), two test statistics are calculated to identify the presence of Brownian Motion and jump processes in the time series. The first test (sw), used to identify if the series contains the stochastic component, is based on Equation (5), where a truncation level is chosen according to a certain number of standard deviations. As the frequency Δn tends to zero, meaning that a ‘higher frequency’ is selected, the asymptotic behavior of the standard deviation is σ→0, and the jump component is eliminated to capture the continuous part of the time series. Additionally, the power (p) is set at levels inferior to 2, also to highlight the continuous element.

$$sw(p, u_n, k, \Delta_n) = \frac{B(p, u_n, \Delta_n)}{B(p, u_n, k\Delta_n)} \quad (5)$$

In this case, the null hypothesis is defined in the direction of the existence of the Brownian Motion element in the time series, so this defines the first hypothesis of this research. The result of sw converges to 1 in the case where no stochasticity is perceived in the time series and approaches k^(1-p⁄2) when Brownian Motion is present. In this case, the parameter k implies the frequency interpolation element.

The second test statistic (SJ) assesses the presence of jumps in the time series. Equation (6) depicts the test, also comparing realized power variations in different frequency patterns mediated by the variable k. Moreover, this test is calculated based on the untruncated form B(p, ∞, Δn) with a power higher than 2. As the test SJ tends to one in its limit, the presence of jumps is detected, while as it tends to zero based on a decay rate of kp/2-1, no jumps are present and the series is based on a continuous Brownian process.

$$s_J(p, \infty, \Delta_n) = \frac{B(p, \infty, k\Delta_n)}{B(p, \infty, \Delta_n)} \quad (6)$$

The null hypothesis of the test statistic is related to the presence of jumps in the time series. Accordingly, the second hypothesis of the research states that there will be jumps in the price path of the Russell indices. Moreover, the measurement will focus on potential differences among the indices, that would lead to distinction in the microstructure of them, and therefore differentiate from an inside-out point of view of the volatility dynamics of the series.

5. Empirical Results

The empirical result depicts the presence of Brownian Motion on the three Russell indices. For the calculation of sw, the variable power (p) is set between 1 and 1.75 with a 0.25 increment, the truncation varies from 5 to 10 standard deviations under the Δn regime with an increment of 1 standard deviation, the frequency interpolation (k) is set at 2 and 3, and the periodicity Δn is set on 1 and 2 minutes. For each index, then, 96 estimates of sw are calculated and presented in the histograms in Figure 2. Such parameters are the same used by Ait-Sahalia & Jacod (2012).

In all cases, the calculation of sw fails to reject the null hypothesis that Brownian Motion is present. Moreover, noise is not a dominating structure as potentially the result of the periodicity (Δn) used. In the case of the Russell 2000 Index, part of the histogram is positioned at a sw below the value of 1, indicating a non-stochastic component in the time series. Evaluating the scatterplot matrix, it is remarkable that the linear relationship between large caps and the broad index is less pronounced for the small caps segment.

In regards to the presence of jumps, empirical results are calculated using values for the power (p) variable between 3 and 6 with increments of 0.25, frequency interpolation as 2 and 3, and frequency (Δn) defined as 1 and 2 minutes. Under these conditions, a sample of 52 estimates of SJ is calculated for each index.

The null hypothesis that jumps are present is rejected for all three indices, as portrayed in Figure 3. This result is potentially the effect of the periodicity of 1 and 2 minutes. In high-frequency works, as in the case of Ait-Sahalia & Jacod (2012), frequencies collected range from 5 seconds to 2 minutes, and therefore, the data available for this analysis is located on the top-notch of these high-frequency works. Moreover, the authors point out that SJ is found near the value of one in higher frequencies and increases as the periodicity becomes less granular.

Most important, on the other hand, is the remarkable difference between the Russell 2000 Index and the other Russell Indices family members. Evidence from Figure 3.3 delineates a longer tail for the Russell 2000 Index in comparison with its large-cap counterpart and the broad market. This finding is relevant in three dimensions. On an operational dimension, the result must be analyzed as presented, in terms of its difference from the other indices. In this case, it implies that the time series behavior of small caps is distinct from large caps, resulting in the de-coupling of risk and return profiles and validating the theoretical landscape of rewarded factors.

Panel A (Top-Left): SW Histogram for Russel 1000 Index

Panel B (Top-Right): SW Histogram for Russel 3000 Index

Panel C (Bottom-Left): SW Histogram for Russel 2000 Index

Panel D (Bottom-Right): Scatterplot Matrix of SW for Russell Indices

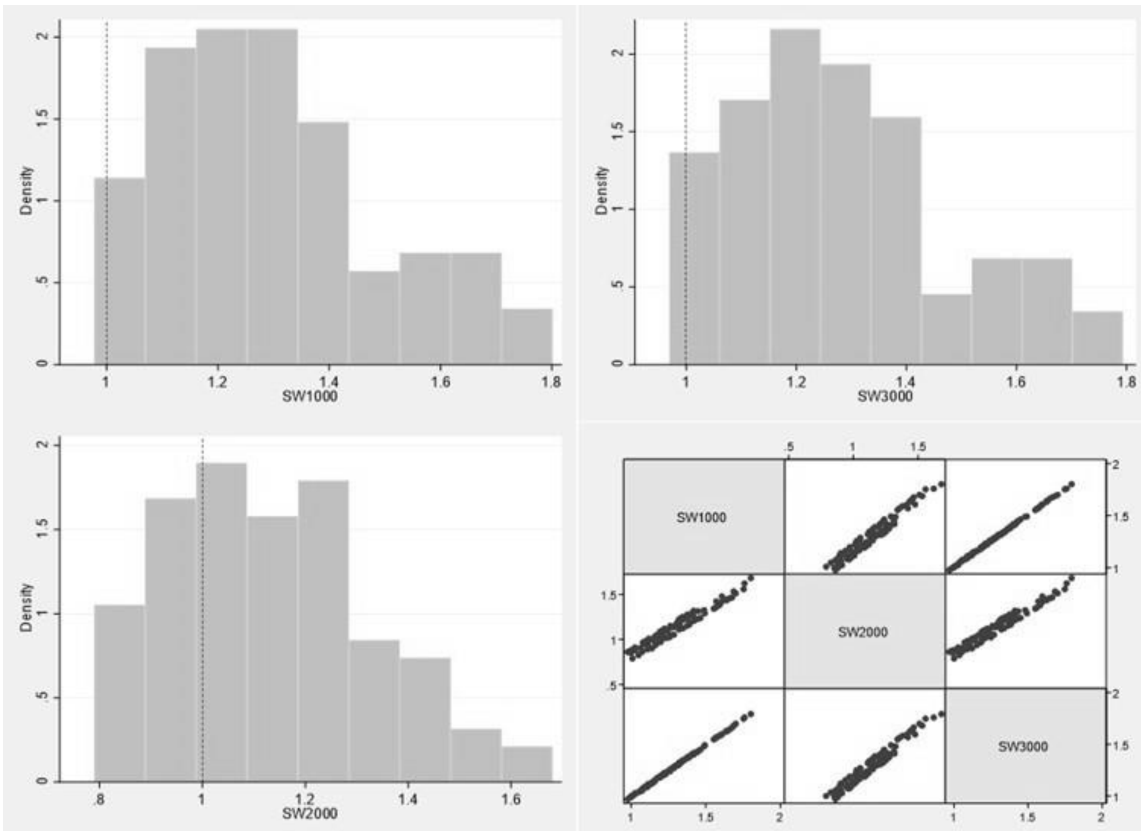


Figure 2. Histograms for SW (Presence of Brownian Motion)

Source: Model Calculations

On a sample dimension, this result would imply a lower probability of time series jumps for small caps. The point is that this finding may be sample dependent, more remarkably considered in light of smaller returns of small caps in the period. This point is remarkably important as it may imply that small caps may add diversification benefits, when included in portfolios, but not necessarily active return. In other words, excess return that may arise by rewarded factors may be understood as dynamic and transitory.

On the time series functioning dimension, this result imposes, constrained by the sample, two conclusions. The first would be related to the lower probability of jumps in the Russell 2000 Index and the second to a lagged asymptotic trend to jump pattern, as more granular data is introduced. On the other hand, the observed volatility of the Russell 2000 Index is higher than the Russell 1000 or 3000 Indices. In this sense, the research decouples the presence of volatility and jumps, implying that volatility may well arise from the continuous and stochastic element of the time series.

Panel A (Top-Left): SJ Histogram for Russel 1000 Index

Panel B (Top-Right): SJ Histogram for Russel 3000 Index

Panel C (Bottom-Left): SJ Histogram for Russel 2000 Index

Panel D (Bottom-Right): Scatterplot Matrix of SJ for Russell Indices

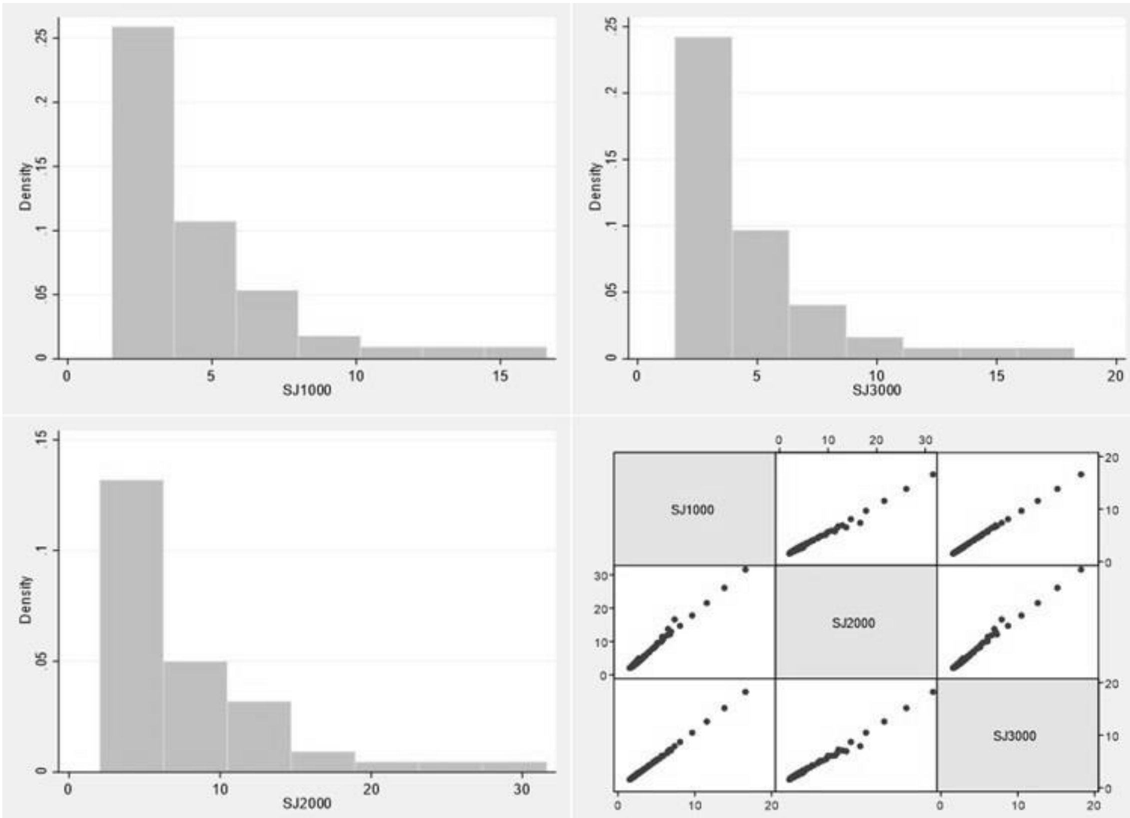


Figure 3. Histograms for SW (Presence of Jumps)

Source: Model Calculations

In order to emphasize the distinct pattern of the small caps, an additional evaluation may be significant. The differences in the SJ statistic among the distinct Russell indices are regressed on the power coefficient (p). In this direction, it may be possible to understand how distinct is the microstructure of each segment about each other, and therefore understand potential contrasts in the riskiness and potential of diversification. Table 2 depicts these results.

Initially, all models depicted statistical significance for the parameter power and constant. In terms of the power, based on the idea that a higher power coefficient highlights the jump component of the time series, the relevance of the coefficient becomes crucial in defining the relative propensity of a series to jump, in comparison to the other. In this way, this analysis embeds a risk pattern contrast arising from the path disruption in the high-frequency time series.

The evaluation of the three equations highlights the different characteristics of small caps. To comprehend this statement, it may be important to notice that the coefficient calculated for the power coefficient on the difference between large caps and the broad index is -0.1823. In this sense, the test statistics difference, which starts close to a zero level considering that SJ is only applicable at power 3 and above, depicts a higher chance of jumps in the Russell 1000 Index, which grows at a lower degree than the Russell 3000 Index. On the other hand, the degree of separation between those two series is still low implying that, despite the fact that the large-cap is more prone to jumps, the risk pattern of both indices is somewhat equivalent.

In a different direction, the evaluation of the Russell 2000 Index leaves no doubt about a distinct risk pattern for small caps.

Table 2. Regression Results in the Difference on Russell Indices

	Difference Russell 2000 and 1000	**Difference Russell 1000 and 3000**	**Difference Russell 2000 and 3000**
Observations	52	52	52
Coefficients and P-Values			
Power	2.3124	-0.1823	2.1301
	(0.000)	(0.000)	(0.000)
Constant	-7.0963	0.5767	-6.5197
	(0.000)	(0.003)	(0.000)
R-Squared	0.4882	0.2930	0.5051

In the above equations, as much as the difference approaches zero at the lower economic significance of the power, the coefficients of 2.3124 and 2.1301 determine a consistently lower propensity to jump than its peers. As much as this result may appear counterintuitive, it calls for sample dependency. The interesting implication of this point relates to the fact that a position in small caps represents a risk modification of a portfolio, as its riskiness pattern is distinct from large caps, but not necessarily in terms of enlarging expected returns. The success of such investments will depend on the relative momentum of small caps in relation to large caps, which in the last case depends on the allocation skill in terms of timing.

6. Conclusions

Empirical results confirm the existence of the specificity of the size factor. As a time series that evolves from a different pattern, in terms of its microstructure, the riskiness of the small-cap segment differentiates itself from the broad market and large caps. As analyzed, this conclusion is not aligned with traditional views that advocate for abnormal returns for this asset class. More specifically, it points to the fact that there is a diversification benefit relative to small caps but the presence of higher-than-market returns will depend on timing. In this sense, the study corroborates with studies and strategies such as Smart Betas.

As an example, active managers may understand the appropriateness of tilting portfolios toward small caps, as active returns may be dependent on market momentum for the asset class and the broad market.

This paper seeks to explore the applications of high-frequency econometrics. In this way, it paves the way for a series of studies that may rely on this sort of data applied to material market questions. For example, additional studies can be done using higher frequency data, exploring other factors, and confirming previous results obtained for the US in international markets. Another important indication for future research is related to the conditions in which jumps occur, such as liquidity or market breadth. At last, a broader understanding of the extent the absolute returns can be explained by the presence of jumps also may be an important line of research.

In such a way, this study open a promising field to the evaluation of risk factors under a market microstructure approach. Future research opportunities could related to the evaluation of other risk factors rather than size, the analysis of higher time frequencies, and the changing behavior of time-series among these different time patterns, and the evaluation of directional dynamics of different exposures, potentially in correlation with specific market conditions. Besides that, this methodological approach may be suitable for microstructure understanding of event-driven anomalies, such as IPOs or earnings surprises.

This research has limitations regarding its proposed feature of pointing to volatility idiosyncrasies instead of a directional strategy. Further studies may deep this methodology towards this objective. To this extent, broader tests must be conducted to evaluate sample dependency. We have been able to depict the riskiness peculiarity of the factor size. To extend into a directional approach, an objective that is beyond the scope of this research, on the other hand, different samples should be employed.

Acknowledgments

We greatly appreciate the valuable contributions of research professors at our university.

Authors' contributions

All authors read and approved the final manuscript. The three authors contributed equally to the study.

Funding

This article was produced with our own resources.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of Sciedu Press.

The journal and publisher adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Ahadzie, R., & Jeyasreedharan, N. (2020). Effects on Intervaling on High Frequency Realized Higher Order Moments. *Quantitative Finance, 20*(7), 1169-1184. https://doi.org/10.1080/14697688.2020.1725100

Ait-Sahalia, Y., & Jacod, J. (2012). Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High-Frequency Data. *Journal of Economic Literature, 50*(4), 1007-1050. https://doi.org/10.1257/jel.50.4.1007

Ali, A., & Bashir, H. (2022). Bibliometric Study on Asset Pricing. *Qualitative Research in Financial Markets, 14*(3), 433-460. https://doi.org/10.1108/QRFM-07-2020-0114

Au, K., Raj, M., & Thurston, D. (1997). An Intuitive Explanation of Brownian Motion as a Limit of Random Walk. *Journal of Financial Education, 23*, 91-94

Babenko, I., Boguth, O., & Tserlukevich, Y. (2016). Idiosyncratic Cash Flows and Systematic Risk. *The Journal of Finance, 71*(1), 425-455. https://doi.org/10.1111/jofi.12280

Barry, C., & Brown, S. (1984). Differential Information and the Small Firm Effect. *Journal of Financial Economics, 13*(2), 283-294. https://doi.org/10.1016/0304-405X(84)90026-6

Blitz, D., & Vidojevic, M. (2019). The Characteristics of Factor Investing. *The Journal of Portfolio Management, 45*(3), 69-86. https://doi.org/10.3905/jpm.2019.45.3.069

DeFusco, R., McLeavey, D., Pinto, R., & Runkle, D. (2007). *Quantitative Investment Analysis.* John Wiley & Sons.

Didericksen, D., Kokoszka, P., & Zhang, X. (2012). Empirical Properties of Forecasts with the Functional Autoregressive Model. *Computational Statistics, 27*, 285-298. https://doi.org/10.1007/s00180-011-0256-2

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 40*(2), 987-1008.

Engle, R. (2000). The Econometrics of Ultra-High-Frequency Data. *Econometrica, 68*(1), 1-22. https://doi.org/10.1111/1468-0262.00091

Evans, L. (2013). An Introduction to Stochastic Differential Equations. *American Mathematical Society*.

Everett, J., & Watson, J. (1998). Small Business Failure and External Risk Factors. *Small Business Economics, 11*(4), 371-390. https://doi.org/10.1023/A:1008065527282

Fama, E., & French, K. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance, 47*(2), 427-465. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

FTSE. (2021). Russell US Equity Indices: Construction and Methodology. Retrieved from https://research.ftserussell.com/products/downloads/Russell-US-indexes.pdf

Gandhi, P., & Lustig, H. (2015). Size Anomalies in US Bank Stock Returns. *The Journal of Finance, 70*(2), 733-768. https://doi.org/10.1111/jofi.12235

Goodhart, C., & O'Hara, M. (1997). High-Frequency Data in Financial Markets: Issues and Applications. *Journal of Empirical Finance, 4*(2), 73-114. https://doi.org/10.1016/S0927-5398(97)00003-0

Hackel, K., Livnat, J., & Rai, A. (1994). The Free Cash Flow / Small Cap Anomaly. *Financial Analysts Journal, 50*(5), 33-42. https://doi.org/10.2469/faj.v50.n5.33

Hall, W. (1969). Embedding Submartingales in Wiener Processes with Drift, with Applications to Sequential Analysis. *Journal of Applied Probability, 6*(3), 612-632.

He, J., Ng, L. (1994). Economic Forces, Fundamental Variables, and Equity Returns. *The Journal of Business, 67*(7), 599-609. https://doi.org/10.1086/296648

Hou, K., Karolyi, A., & Kho, B. (2011). What Factors Drive Global Stock Returns?. *The Review of Financial Studies, 24*(8), 2527-2574. https://doi.org/10.1093/rfs/hhr013

Hu, S., Gu, Z., Wang, Y., & Zhang, X. (2019). An Analysis of the Clustering Effect of a Jump Risk Complex Network in the Chinese Stock Market. *Physica A Statistical Mechanics and its Applications, 523*, 622-630. https://doi.org/10.1016/j.physa.2019.01.114

Karatzas, I., & Shreve, S. (1998). *Brownian Motion and Stochastic Calculus*. Springer.

Kim, M., & Burnie, D. (2002). The Firm Size Effect and the Economic Cycle. *The Journal of Financial Research, 25*(1), 111-124. https://doi.org/10.1111/1475-6803.00007

Lavenda, B. (1985). Brownian Motion. *Scientific American, 252*(2), 70-85.

Lawley, D. (1963). On Testing a Set of Correlations Coefficients for Equality. *The Annals of Mathematical Statistics, 34*(1), 149-151. https://doi.org/10.1214/aoms/1177704249

Macaira, P., Thome, A., Oliveira, F., & Ferrer, A. (2018). Time Series Analysis with Explanatory Variables: a Systematic Literature Review. *Environmental Modelling and Software, 107*, 199-209. https://doi.org/10.1016/j.envsoft.2018.06.004

Makinen, Y., Kanniaken, J., Gabbouj, M., & Iosifidis, A. (2019). Forecasting Jump Arrivals in Stock Prices: New Attention-Based Network Architecture Using Limit Order Book Data. *Quantitative Finance, 19*(2), 2033-2050. https://doi.org/10.1080/14697688.2019.1634277

Pan, J. (2002). The Jump Risk Premia Implicit in Options: Evidence from an Integrated Time Series Study. *Journal of Financial Economics, 63*(1), 3-50. https://doi.org/10.1016/S0304-405X(01)00088-5

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematic Statistics, 33*(3), 1065-1076. Retrieved from https://www.jstor.org/stable/2237880

Switzer, L. (2010). The Behavior of Small Cap Vs. Large-cap Stocks in Recessions and Recoveries: Empirical Evidence for the United States and Canada. *North American Journal of Economics and Finance, 21*(3), 332-346. https://doi.org/10.1016/j.najef.2010.10.002

Tsay, R. (2000). Time Series and Forecasting: Brief History and Future Research. *Journal of American Statistical Association, 95*(450), 638-643. https://doi.org/10.1080/01621459.2000.10474241