Deriving Fractional Moments using the Moment Generating Function

Jelle Reisinger

(2780350)

30 June 2025

Bachelor Thesis Econometrics and Data Science

Thesis commission:

PhD candidate Gabriele Mingoli

Dr. yyyy (co-reader)

Abstract

This thesis introduces a novel approach for computing moments of fractional order by combining the moment generating function with techniques from fractional calculus. The accuracy of the method depends on the specific derivative used, and errors tend to increase with larger parameter values of the underlying distribution. By making use of moments of fractional order, statistics such as the standard deviation, skewness and kurtosis are computed in a unique manner to analyze and forecast volatility of stock returns in financial markets. A fractional generalization of downside risk measures is proposed, which can be used to capture both the frequency and magnitude of negative returns. Additionally, moments of fractional order are integrated into regression models, reducing errors and improving prediction performance compared to models using traditional regressors. Whenever closed-form expressions of the moment generating function are unavailable, numerical approximations have been used for the computations, ensuring stable and reliable results.

Keywords: fractional moments, moment generating function, fractional calculus, Caputo-Fabrizio, volatility forecasting, observation-driven regression models

1 Introduction

Statistical moments, defined as the n-th moment of a random variable X via its probability density function, are essential tools for characterizing data and its distribution. The first and second (central) order moments, the mean and variance, respectively provide key insights into the average value and dispersion of a random variable. Higher-order moments offer further information about the shape and symmetry of the distribution. A less commonly discussed class of moments, however, are the fractional moments. The latter are defined in precisely the same manner as the integer moments, but now with $n \in \mathbb{R}$, or even $n \in \mathbb{C}$. From this point onward, when referring to fractional moments, we denote the order by $\alpha \in \mathbb{R}$ instead of n. While these moments may not find as much usage in comparison with the integer moments of a distribution, they can be rather valuable in specific applications.

Fractional moments play a significant role in various fields, including finance, economics, and statistics. One notable application in statistics is their use in approximating integer moments, as described by Novi Inverardi and Tagliani (2024). In such cases, evaluating a 'nearby' existing fractional moment can provide an approximation of the otherwise undefined integer moment (Novi Inverardi and Tagliani, 2024). As an illustration, one can consider the student-t distribution with $\nu = 2$ degrees of freedom. This distribution only has central and raw moments of order k, where $0 < k < \nu$. This implies that its second central moment ($k = 2$) does not exist. One could however consider the central moment of order k, where $k = 1.95$, given that this order of the moment does in fact exist and interpret its value as the variance of the distribution.

A different theoretical application is provided by Mikosch et al. (2013), who analyse fractional moments in the context of stationary solutions to stochastic recurrence equations (SREs). These equations generalize standard econometrics models like Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), which are used to monitor volatility of time series data. In particular, the GARCH(1,1) model can be expressed as an SRE, making the analysis of fractional moments in this context relevant for understanding the long-run behavior of volatility models. Mikosch et al. (2013) analyse the existence and behavior of fractional moments from a recursive perspective and characterizing the distributional properties of financial time series.

In such theoretical studies, it is often important to determine whether a fractional moment exists, even if its exact value is not required. However, in more applied settings, particularly in finance, and engineering, it becomes necessary to explicitly compute fractional moments. This is the main topic we focus on in this thesis: we will not only explore the theoretical relevance of fractional moments, but also analyse the practical importance of computing them.

One important example arises again in the context of financial modeling using GARCH models. These models are commonly applied to financial return series and capture the dynamic volatility that changes over time. The GARCH model achieves this by modelling the volatility based on the returns and variances of previous time periods. It is useful to note that in the context of financial returns, the variable of interest often follows a distribution with heavy tails. As a consequence, not all relevant moments of integer order may exist. In such cases, fractional moments may serve as a viable alternative. Hansen and Tong (2024) propose a method for computing fractional absolute moments of the cumulative return, quantities that would otherwise be undefined using traditional integer-order techniques.

In addition, Hansen and Tong (2024) apply a Heterogenous Autoregressive Gamma Model (HARG) to model intraday stock data. The HARG model assumes the conditional distribution of the variable of interest, say X_t to be of a non-central Gamma distribution. This model is commonly used to model positive-valued time series data (Gourieroux and Jasiak, 2006). In particular, the volatility of the random variable is of interest, which by definition is non-negative.

In this context, X_t is the realized variance. Hansen and Tong (2024) study the conditional moments of the variance process. Specifically, they consider moments of order $\{\frac{1}{2}, \frac{3}{2}, 2\}$ which correspond to the standard deviation, skewness and kurtosis respectively. Furthermore, the moment of order $-\frac{1}{2}$, the inverse volatility has also been computed. In this context, this moment can be used as the inverse of the covariance matrix in order to obtain the Sharpe ratio, which is an index to measure the performance of some investment compared to a risk-free asset (Sharpe, 1994). These fractional moments offer valuable insights into financial return dynamics and are highly relevant for modelling and risk assessment.

Furthermore, computing fractional moments becomes essential in entropy-based density reconstruction. Gzyl et al. (2013) introduce the use of fractional moments as input for maximum entropy methods in the context of insurance risk modelling. In these models, often the probability density function of total ruin, the event where an insurance company's capital becomes negative, is unknown or difficult to compute. The method of maximum entropy has been developed as a way to approximate these unknown density functions by selecting the "least biased" probability density based on a given set of moments. Traditionally, this method relies on moments of integer order. However, Gzyl et al. (2013) demonstrate that fractional moments can be used instead to obtain the unknown ruin distribution function more efficiently and accurately, as is supported by Lin (1992), for real-world applications. Moreover, as suggested before, when it is impossible to take high-order integers moments, which may be required to approximate the unknown density function, moments of fractional order serve as a valuable alternative.

D'Amico et al. (2002) apply a similar approach in the context of option pricing using the Black-Scholes model, a differential equation which can be used to determine the price of an option. They use fractional moments to approximate the underlying density of th asset price and compute related entropic measures. Most importantly, they show that the use of only two fractional moments is sufficient to approximate the entropy and associated distribution at a high accuracy, outperforming the classical integer-moment-based maximum entropy methods both in numerical stability and computational efficiency. Remarkably, using only two fractional moments, the method achieved approximations with three-digit accuracy.

These applications demonstrate that the computation of fractional moments plays a practical and important role in reconstructing risk-relevant distributions via entropy-based methods. This approach is particularly useful in finance and insurance contexts, where the underlying distribution

is often unknown, and entropy-based reconstruction using fractional moments provides a reliable and efficient alternative to traditional techniques such as inverse Laplace transforms (see definition A.0.1).
Beyond finance and risk modelling, computation of fractional moments finds practical applications in fields such as engineering and healthcare. As previously mentioned, Gzyl et al. (2013) used fractional moments combined with the maximum entropy method to predict total ruin. This approach has been extended to the context of estimating lifetime distributions (Gzyl and Mayoral, 2024), a general topic of interest in survival analysis (Clark et al., 2003). Both probability density functions and survival functions were estimated using this technique. As before, the prediction errors remained within three-digit accuracy, further supporting the method's consistency and precision across different fields of research. The same practical advantages of computing moments of fractional order over moments of integer as in Gzyl et al. (2013) still apply.

Other examples include optimizing signal processing and control systems as well as studying the response characteristics of random vibration systems. Wang et al. (2025) have shown that using the concept of fractional moments for the latter leads to higher accuracy and greater numerical stability compared to traditional methods, such as Taylor expansions. Furthermore, in terms of simplicity and efficiency, the method of fractional moments is advantageous, as its computation steps are straightforward and avoid convergence issues, significantly reducing the resources required for computation. Implementing the usage of fractional moments has allowed Wang et al. (2025) to obtain both analytical, as well as numerical solutions to problems within their research field, which again proves its viability in practical applications.

Finally, in the context of non-linear systems, Di Matteo et al. (2014) demonstrate how complex fractional moments can be used to solve differential equations such as the Kolmogorov or Fokker-Planck equation, which are often used in the field of statistical mechanics to model processes such as diffusion, relaxation to equilibrium, and entropy production (Schulten, 2005). By transforming the system into a more agreeable form using Mellin transforms (see definitionA.0.2), Di Matteo et al. (2014) show that solutions can be efficiently computed and later recovered using inverse transformations. A key advantage of this method is that it preserves the entire support of the probability distribution, which traditional integer moment methods fail to achieve.
In this thesis, we propose a novel method to compute moments of fractional order by combining the moment generating function with techniques from fractional calculus. Both analytical and numerical comparisons between the accuracies of the moment generating function in combination with different fractional derivatives will be made.

The remainder of this thesis is structured as follows. Section 2.1 reviews existing methods for computing fractional moments and introduces a novel approach based on combining the moment generating function with fractional calculus. Section 2.2 and section 2.3 lay the theoretical foundation, including the definitions and properties of fractional derivatives and moment generating functions. Section 3 presents numerical results which will be used to compare the accuracy and potential errors of different methods. Finally, in section 4, a practical implementation of this new method in a financial setting will be considered, with a focus on the implications of potential associated errors in real-world applications.

2 Methodology

In this section we present the methodological framework for computing fractional moments using different approaches. We begin by reviewing existing methods before introducing an alternative approach based on fractional calculus. The comparative structure of the analysis, both theoretical and empirical, is also outlined.

2.1 Overview of existing methods and introduction to a novel approach

In this subsection we review the traditional method of computing fractional moments and a modern alternative by Hansen and Tong (2024) based on the complex moment generating function (CMGF). What is more, we also introduce the motivation for using fractional derivatives applied to the moment generating function, which forms the foundation of this thesis' novel approach.

2.1.1 Existing methods of obtaining fractional moments

The formal definition of the statistical moments mentioned in the previous section is as follows: Let X be a random variable.

The α-th moment of X is defined as:

$$\mathbb{E}[X^\alpha] = \begin{cases} \int_{-\infty}^{\infty} (x-c)^\alpha f_X(x)\,dx & \text{if } X \text{ is continuous,} \\ \sum_i (x_i - c_i)^\alpha f_X(x_i) & \text{if } X \text{ is discrete.} \end{cases} \tag{1}$$

In the context of this thesis $\alpha \in \mathbb{R}$. Here $f_X(x)$ is the probability density function if X is a continuous random variable, and the probability mass function if X is a discrete random variable (Feller, 1957).

The traditional method of computing fractional moments is rather straightforward. Similar to integer moments, one simply computes the summation (in the discrete case) or integral (in the continuous case) of $x^\alpha \cdot f_X(x)$. In the context of fractional moments, $\alpha \in \mathbb{R}$ instead of $\mathbb{Z}$ (assuming that negative moments exist). A disadvantage of this method is that for distributions with negative support, moments of fractional order may not be defined. Hansen and Tong (2024) introduce an alternative approach to computing fractional moments using the complex moment generating function (CMGF), which they apply in the context of the aforementioned GARCH models. One of their expressions of importance is given by:

$$\mathbb{E}\,|X-\xi|^\alpha = \frac{\Gamma(\alpha+1)}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-\xi z} M_X(z) + e^{\xi z} M_X(-z)}{z^{\alpha+1}} dt \tag{2}$$

where $z = s + it, s \in \mathbb{N}_+, \xi \in \mathbb{R}$ and α of course the order of the moment.

This formulation extends upon the traditional moment generating function (MGF) but avoids the process of taking derivatives, making it computationally efficient. The inclusion of the Gamma function is logical, as it extends the factorial function to real values, aligning well with the computation of fractional moments. Since this method relies on integration, rather than differentiation, it avoids numerical issues that might arise when computing derivatives, such as obtaining rather great approximation errors. As this method computes absolute moments of fractional order, there will be no problems regarding numerical issues for negative values. The integrand expression is more complicated compared to the integrand of the traditional approach, which may lead to longer run-times.

2.1.2 Obtaining fractional moments by using the moment generating function

While the CMGF method provides a numerically stable and elegant alternative to the traditional method, this thesis explores a different approach: computing fractional moments directly by applying fractional derivatives to the MGF. The MGF is widely used for computing integer moments by differentiation around zero. Extending this approach to fractional orders requires us to take fractional derivatives. Thus, we need to define such fractional derivative operators. These fractional derivatives have a long history and often make use of the aforementioned Gamma function in combination with some integral. This means that, for continuous random variables, where we integrate the MGF, we will have to perform double integration. A consequence might be that obtaining analytical expressions of these moments may not always be possible. A great number of alternative expressions of these fractional derivatives have been created, mostly based on different interpretations of the latter in the field of physics. In this thesis, we will focus on computing the MGF using the Riemann-Liouville derivative, the Caputo-Fabrizio derivative and the Grünwald-Letnikov fractional derivative. Each of these fractional derivatives in combination with the MGF may lead to different moments expressions for the same distribution and same fractional order of the moment. Thus, it is essential to compare each of these definitions with the traditional way of computing fractional moments, to derive their accuracy and conclude which approach is most suitable for fractional moment computation. Similar to the expression of the moments of a random variable, their errors may also be hard to derive analytically, depending on its distribution.

2.1.3 Research design and Methodological steps

To compare the aforementioned fractional derivatives, the research was carried out in several stages. First, a mathematical foundation for the fractional derivatives was established, including a discussion of their respective properties. A brief historical overview of the development of fractional calculus was also provided.

Consequently, basic definitions of statistical moments and the MGF were reviewed, and their properties were re-examined in the context of fractional derivatives. Whenever possible, analytical expressions for the errors associated with each fractional derivative in combination with the MGF were derived. This completed the theoretical part of the research.

Subsequently, numerical values of fractional moments were computed using the different MGF-based methods. When applicable, the corresponding errors relative to theoretical values were analyzed, including summary statistics such as the minimum, maximum, and average error.

Finally, a practical case study was conducted, analyzing the volatility and risk of S&P 500 returns using fractional moments. The new method was evaluated in this context, serving as an empirical benchmark for the proposed methodology. in real-world applications.

2.2 Fractional Calculus

In order to obtain MGF expressions that allow us to compute moments of fractional order, we need to be able to take fractional derivatives. This section is intended to summarise the history and applications of fractional derivatives and lays the foundation of fractional calculus, as well as to provide a number of examples of fractional derivatives.

2.2.1 Overview and applications of fractional derivatives

Although not the main topic of interest of this thesis, it is useful to have some knowledge about the history and applications of fractional derivatives. The study of fractional derivatives has been relevant as early as the year 1695 when the concept of such a derivative was implicitly discussed by Leibniz and Bernoulli (Katugampola, 2014). Since then, numerous definitions of fractional derivatives have been developed. The best-known definition of the fractional derivative is the Riemann-Liouville derivative, its upper derivative of a function $f(x)$ of order α, as mentioned in Kilbas et al. (2006), is denoted as

$$\frac{d^n}{dx^n}\frac{1}{\Gamma(n-\alpha)}\int_a^x (x-t)^{n-\alpha-1}f(t)dt, \text{ where } n=\lceil\alpha\rceil.$$

Michele Caputo (1967) defined a variation on this derivative, where instead of having the integer derivative expression $\frac{d^n}{dx^n}$ in front of the integral, we have $\frac{d^n}{dt^n}$ inside the integral. Due to this adjustment, it is possible to have initial value conditions of (fractional) differential equations expressed as the traditional derivatives of integer-order, which make differential equation problems more intuitive. Other fractional derivatives, such as the Hadamard (1892) and Riesz (1949) derivative, have been defined to take advantage of particular beneficial properties. For example, each of the latter derivatives can be written as a Fourier transformation (see definition A.0.3). As a consequence, the analytical expressions, can often be simplified. A rather unique derivative is the Grünwald-Letnikov derivative which, in contrast to all the aforementioned derivatives, is not based on integral. Instead, it generalizes the difference quotient, $\frac{f(x+h)-f(x)}{h}$, to fractional orders using binomial coefficients (Atici et al., 2021). This variety of definitions emphasizes how dependent fractional derivatives are based on the physical interpretations and practical applications of interest. Beyond their theoretical relevance, fractional derivatives have been of significant importance in various scientific fields since the 19th century. Examples include fractional Fourier transformations, a generalization of the regular Fourier transformations (Missbauer, 2012), fractional diffusion equation models, describing the motion of particles in liquids as a consequence of thermal molecular motions (Einstein, 1905) and the fractional Schrödinger equation, a generalization of the Schrödinger equation, often used in quantum mechanics (Laskin, 2002). Applications of fractional derivatives are less common in the fields of finance and economics, as they are mainly used to describe natural phenomena (Boulaaras et al., 2023). Yet they still offer some great potential. (Symmetric) Levy flights make use of fractional derivatives in order to solve partial differential equations which describe random walk processes in time series (Scalas et al., 2000). The development of fractional derivatives also led to the notion of fractional Brownian motions, a generalization of the Brownian motion (Mandelbrot and Van Ness, 1968). The latter is a continuous-time stochastic process which, similar to Levy flights, may be used to model random walk processes.

2.2.2 Formal definitions of fractional derivatives

In order to obtain the expressions mentioned in 2.1.1, some advanced tools are required. We can find these in the field of fractional calculus. A formal definition of the differintegral operator, which generalizes derivatives and integrals of fractional order, can be found in A. As mentioned in section 2.2.1, quite a number of different definitions have been proposed for computing fractional derivatives. In this thesis, we will focus on some of the more well known fractional derivatives. We will start with the most well-known fractional derivative, which laid the foundation of the study of fractional derivatives as early as in 1832.

Definition 2.2.1. The left-side Riemann-Liouville fractional derivative of order α, proposed by Joseph Liouville (1832), is defined as:

$$D^{\alpha}_{a+}f(x) = \frac{d^n}{dx^n}D_x^{-(n-\alpha)}f(x) = \frac{d^n}{dx^n}\frac{1}{\Gamma(n-\alpha)}\int_a^x (x-t)^{n-\alpha-1}f(t)dt \tag{3}$$

Here, $n = \lceil \alpha \rceil$, the ceiling function and $\Gamma(.)$ is the Gamma function (see definition A.0.7).

Note that we just defined the left-side Riemann-Liouville fractional derivative, suggesting that there also exists a right side derivative. In the case of the latter, we would evaluate the associated integral the other way around. Namely,

$$D^{\alpha}_{b-}f(x) = \frac{d^n}{dx^n}\frac{1}{\Gamma(n-\alpha)}\int_x^b (x-t)^{n-\alpha-1}f(t)dt.$$

We intend on using the left-side derivative, as is supported by Tarasov (2023). The reason is, due to the fact that many functions in probability theory, most importantly the cumulative distribution function, are defined as an integral from some constant to x.

Remark 2.2.1. For values $\alpha \in \mathbb{N}_+, n = \lceil \alpha \rceil = \alpha$, so $\Gamma(n-\alpha) = \Gamma(0)$, which is undefined. Thus for $\alpha \in \mathbb{N}_+$, we define: $D^\alpha f(x) = \frac{d^\alpha}{dx^\alpha} f(x)$, which is simply the regular expression for derivatives of integer order.

A modification of the Riemann-Liouville derivative is the Caputo-Fabrizio derivative, which is defined as follows:

Definition 2.2.2. The Caputo-Fabrizio fractional derivative of order $\alpha, \alpha \in [0,1)$, as introduced by Caputo and Fabrizio (2015), is defined as:

$$D^\alpha f(x) = \frac{1}{1-\alpha} \int_a^x \exp\left(\frac{-\alpha}{1-\alpha}(x-t)\right) f'(t)dt \tag{4}$$

With $a \in [-\infty, x)$. The Caputo-Fabrizio derivative is always defined as the integral from some constant to the variable x. This is another reason for choosing to work with the left-side Riemann-Liouville integral. In this way, it will be more straightforward to compare the two integrals. What is more, note that for the Caputo-Fabrizio derivative, the order $\alpha \in [0,1)$. This does not mean, however, that one can only compute fractional derivatives of order 1 or lower. There exists a rather convenient property of the differintegral operator which allows one to combine orders of derivatives, which will be discussed in a moment.

Finally, we will define the Grünwald-Letnikov derivative, which is defined as follows:

Definition 2.2.3. The Grünwald-Letnikov fractional derivative of order α, as discussed in Zhmakin (2022), is defined as:

$$D^\alpha f(x) = \lim_{h\to 0} \frac{1}{h^\alpha} \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} f(x-kh) \tag{5}$$

Where $\binom{\alpha}{k}$ is the binomial coefficient, with $0 \leq k \leq \alpha$ (see definition A.0.8).

It is immediately clear, observing the summation symbol instead of the integral, that this derivative behaves quite differently from the two derivatives defined above. The Grünwald-Letnikov derivative is an extension on derivatives based of the concept of finite differences (Flajolet and Sedgewick, 1995).

We will consider a number of properties which come in useful when working with fractional derivatives.

Proposition 2.2.1. *The fractional derivatives above adhere to the following properties:*

(i) Linearity: Let $f(x), g(x)$ be functions and $a, b, x \in \mathbb{R}$. Then we have that $D^\alpha (af(x) + bg(x)) = a \cdot D^\alpha f(x) + b \cdot D^\alpha g(x)$.

(ii) $D^\alpha f(x) = f(x)$, for $\alpha = 0$

(iii) For sufficiently smooth functions f, we have that $D^{\alpha+\beta} f(x) = D^\alpha(D^\beta f(x)) = D^\beta(D^\alpha f(x))$, with $\alpha, \beta \in \mathbb{R}$. Note that, for definition 2.2.2, this property only holds for $\beta \in \mathbb{N}, \alpha \in [0,1)$.

The proofs of these of the properties stated in this proposition can be found in Appendix B.1. Most of these proofs have been provided by myself, while some other proofs, which are outside the scope of thesis are based on other papers. Note that the third property is especially useful for the Caputo-Fabrizio derivative. This property allows one to take fractional derivatives of an order greater than 1 by comparing fractional derivatives and regular integer derivatives.

We will now provide two explicit examples of these fractional derivatives. For simplicity, we will let $a = 0$:

Example 2.2.1. (i) We consider the Riemann-Liouville derivative of order $\frac{3}{2}$ for some constant $c \in \mathbb{R}$:

$$D_{RL}^{\frac{3}{2}}(c) = \frac{d^2}{dx^2}\frac{1}{\Gamma(2-\frac{3}{2})}\int_0^x c(x-t)^{2-\frac{3}{2}-1}dt$$
$$= \frac{d^2}{dx^2}\frac{c}{\sqrt{\pi}}\int_0^x (x-t)^{-\frac{1}{2}}dt = \frac{d^2}{dx^2}\frac{-2c}{\sqrt{\pi}}\cdot\sqrt{x-t}\Big|_0^x$$
$$= \frac{d^2}{dx^2}\frac{2c\sqrt{x}}{\sqrt{\pi}} = \frac{-c}{2x\sqrt{\pi x}} \neq 0.$$

As stated, the fractional derivative of a constant is not equal to zero when using the Riemann-Liouville derivative. This is also the case for the Grünwald-Letnikov derivative, but not for the Caputo-Fabrizio derivative.

(ii) We compute the semi-derivative of $\frac{x}{2}$ using the Caputo-Fabrizio derivative:

$$D_{CF}^{\frac{1}{2}}\left(\frac{x}{2}\right) = \frac{1}{1-\frac{1}{2}}\int_0^x \exp\left(\frac{-\frac{1}{2}}{1-\frac{1}{2}}(x-t)\right)\cdot\frac{1}{2}dt = \int_0^x \exp(t-x)dt$$
$$= \exp(t-x)\Big|_0^x = 1-\exp(-x).$$

For $x \geq 0$, this expression is equal to the CDF of the exponential distribution with $\lambda = 1$. A remarkable result.

An explicit example of the Grünwald-Letnikov derivative has not been included. This is as it is rather difficult to obtain analytical expressions for this derivative. Thus, later on in this thesis, when computing fractional moments and their potentially associated errors, the main focus for the Grünwald-Letnikov derivative will be on its numerical computations.

Remark 2.2.2. As shortly mentioned in the introduction, it is possible to generalize the order of derivatives even further, extending α to be in $\mathbb{C}$ instead of $\alpha \in \mathbb{R}$. This means that, when combining such derivatives with the MGF, we will obtain complex moments. Since statistical moments of complex order do almost not find any usage in applications, as they lack interpretability, they are not the main focus of this research. For the interested reader, Love (1971) has done some impressive research on the fundamentals of derivatives of complex order. The obtained expressions are somewhat similar to those of the fractional derivatives which have been discussed above.

2.3 The Moment Generating Function of fractional order

Having defined the techniques required to compute fractional derivatives, we now try to combine them with the MGF. Before proceeding, we will briefly review the formal definition of a moment in statistics.

2.3.1 Moments

Recall the definition of statistical moments as in definition 1 of section 2.1.1. If $c = \mu_x$, where μ_x denotes the expected value of X, then our higher moments are called central moments. In the context of this research, we will focus on the case $c = 0$, corresponding to raw moments of a random variable X. This choice has been made as the MGF, which we will soon define, only computes raw moments of higher order. A moment of order α is said to exist, if $\mathbb{E}[X^\alpha] < \infty$.

2.3.2 The Moment Generating Function

We now formally introduce the moment Generating Function, one of the most significant subjects of this thesis.

Definition 2.3.1. The moment generating function of a variable X, is defined as

$$M_X(t) = \mathbb{E}[e^{tX}]$$

provided that $\mathbb{E}[e^{tX}] < \infty$, for all $t \in (-h, h)$, which contains 0, for some $h > 0$

Remark 2.3.1. Deriving the expression $M_X(t) = \mathbb{E}[e^{tX}]$ is typically straightforward. Generally, it simply requires a handful of steps of analytic evaluation. This procedure is not that interesting nor relevant to this research. Thus, when making use of expressions of the MGF, we will simply refer to the distribution table in Appendix D.

We will state the theorem which makes the MGF so useful. This theorem allows us to compute moments of higher order by taking derivatives of the given order instead of integrals.

Theorem 2.3.1. *If $M_X(t)$ exists on some interval $(-h, h)$, as defined before, we have that:*

$$\mathbb{E}[X^n] = M_X^{(n)}(0), \textit{ for } n \in \mathbb{N}$$

We introduce the following well-known properties for the MGF $M_X(t)$:

Proposition 2.3.1. *For X, Y random variables, we have that:*

(i) $M_X^{(0)}(t) = \mathbb{E}[e^{0X}] = 1$. *This property can be used to confirm that a given function is a valid probability density function (i.e., integrates to one).*

(ii) Location scale-transform. Assuming $M_X(t)$ exists, for constants $\mu, \sigma \in \mathbb{R}$, we have that:

$$M_{\mu+\sigma X}(t) = e^{\mu t} \cdot M_X(\sigma t)$$

(iii) If $X \perp Y$, then $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.

The proofs of the latter can be found in Appendix B.2.

There are several additional topics closely related to the MGF, including Fourier transforms, Laplace transforms, Wick rotations, and characteristic functions. While these subjects are relevant to the theoretical foundation of the MGF , they fall outside the scope of this thesis and will therefore not be discussed. Readers interested in exploring these concepts further may find Kolmogorov and Fomin (1999) to be a valuable resource.

2.3.3 Computing moments of negative order using the Moment Generating Function

In specific cases, as were mentioned in section 1, moments of negative order can be of interest to characterize the data. Such an example was the moment of order $-\frac{1}{2}$, which in some contexts may be used to obtain the Sharpe Ratio. Thus, it is useful to understand how to compute moments of such orders. We will consider the continuous case:

$$\mathbb{E}[X^{-n}] = \int_{-\infty}^{\infty} x^{-n} f_X(x)dx = \int_{-\infty}^{\infty} \left(\frac{1}{x}\right)^n f_X(x)dx.$$

We immediately observe a rather obvious problem. This integral is undefined at $x = 0$ and diverges in a neighbourhood around zero. Khuri and Casella (2002) have stated the following corollary for the existence of a moment with negative first order:

Corollary 2.3.1. *If $f_X(x)$ is a continuous pdf defined on $(-\infty, \infty)$, and if*

$$\lim_{x \to 0} \frac{f_X(x)}{|x|^\alpha} < \infty$$

for $\alpha > 0$, then

$$\mathbb{E}[X^{-1}] \text{ exists.}$$

Most common distribution functions do not adhere to this corollary, however, the Gamma function does (see example B.2).

If such a moment of negative order exists, we should be able to obtain it using the MGF. In the 20-th century, Cressie et al. (1981) have published the following remarkable theorem:

Theorem 2.3.2. *Assuming the negative n-th raw moment exists, the negative n-th raw moment can be computed as follows:*

$$\mathbb{E}[X^{-n}] = \frac{1}{\Gamma(n)} \int_0^\infty t^{n-1} M_X(-t) dt$$

where n is a positive integer.

The proof of this Theorem can be found in Appendix B.2.

Since this is an extension on the regular functions of the MGF, this technique is of interest for this thesis. Thus, it will be shortly be discussed. We compute the first inverse moment of the Gamma distribution, by making use of the latter theorem for the MGF.

Example 2.3.1. Let

$$f_X(x) \sim \Gamma(\alpha, \lambda) = \frac{x^{\alpha-1} e^{-\lambda x} \lambda^\alpha}{\Gamma(\alpha)}, M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$$

$$\mathbb{E}[X^{-1}] = \frac{1}{\Gamma(1)} \int_0^\infty t^{(1-1)} \left(\frac{\lambda}{\lambda - (-t)}\right)^\alpha dt = \int_0^\infty \left(\frac{\lambda}{\lambda + t}\right)^\alpha dt$$

$$= \lambda^\alpha \int_0^\infty (\lambda + t)^{-\alpha} dt, \text{ Let } u = \lambda + t, \frac{du}{dt} = 1, dt = du :$$

$$\lambda^\alpha \int_0^\infty u^{-\alpha} du = \lambda^\alpha \frac{u^{1-\alpha}}{1-\alpha}\Big|_0^\infty = \lambda^\alpha \frac{(\lambda + t)^{1-\alpha}}{1-\alpha}\Big|_0^\infty$$

$$= \lambda^\alpha \left(0 - \frac{\lambda^{1-\alpha}}{1-\alpha}\right) = \frac{-\lambda}{1-\alpha} = \frac{\lambda}{\alpha - 1}.$$

Which corresponds with our result from example B.2.

2.3.4 Extending the Moment Generating Function to fractional order

Now that we have introduced the definitions of the MGF and discussed a number of relevant properties, we will combine these with the techniques developed in section 2.2. Therefore, we can at last obtain moments of fractional order using the MGF.

To avoid confusion regarding what fractional derivative is being used in combination with the MGF, we will from now on, work with the following notation:

Definition 2.3.2. We define the MGF of order $\alpha \in \mathbb{R}$ by ${}_{RL}M_X^{(\alpha)}, {}_{CF}M_X^{(\alpha)}, {}_{GL}M_X^{(\alpha)}$ for the MGF in combination with the Riemann-Liouville, Caputo-Fabrizio and Grünwald-Letnikov fractional derivative respectively.

Remark 2.3.2. The three properties mentioned in proposition 2.3.1 still hold for the MGF of fractional order. This is the case as the first property makes makes use of the derivative of order 0. Which has been defined to just be the original function itself as stated in the second property of proposition2.2.1. The other two properties do not involve any derivatives of any order. Thus, they are generally applicable to the MGF, regardless of its order or kind of derivative.

Before stating any results about the accuracy of these new MGF expressions, we first consider an explicit example to illustrate the interaction between fractional derivatives and the MGF.

Example 2.3.2. (i) We let $f_X(x_i) \sim Bernoulli(p)$, with $\mathbb{P}(X = 1) = p$ and with associated MGF expression: $M_X(t) = (1 - p) + p \cdot \exp(t)$, now we consider

$$_{CF}M_X^{(\frac{1}{2})} = \frac{1}{1 - \frac{1}{2}} \int_{-h}^{t} \exp\left(\frac{-\frac{1}{2}}{1 - \frac{1}{2}}(t - s)\right) M_X'(s)ds.$$

In this case, the domain of $M_X(t) = (-\infty, \infty)$, so we let $-h = -\infty$, and $M_X'(s) = p \cdot \exp(s)$, thus we obtain:

$$_{CF}M_X^{(\frac{1}{2})} = 2\int_{-\infty}^{t} \exp\left((s - t)\right) \cdot (p \cdot \exp(s))ds.$$

$$= 2p \cdot \exp(-t) \int_{-\infty}^{t} \exp(2s)ds$$

$$= p \cdot \exp(-t) \left(\exp(2s)\Big|_{-\infty}^{t}\right) = p \cdot \exp(t)$$

Now, all that is left to do is set $t = 0$ and we obtain that $_{CF}M_X^{(\frac{1}{2})} = p$.

(ii) Computing $\mathbb{E}[X^{\frac{1}{2}}]$ in the traditional fashion, we obtain:

$$\mathbb{E}[X^{\frac{1}{2}}] = \sum_{x} x^{\frac{1}{2}}\mathbb{P}(X = x) = 0^{\frac{1}{2}} \cdot \mathbb{P}(X = 0) + 1^{\frac{1}{2}} \cdot \mathbb{P}(X = 1)$$

$$= 0 \cdot (1 - p) + 1 \cdot p = p$$

It is amazing and maybe even somewhat surprising that both expressions obtain the same result. Indeed, this result is actually more of a coincidence. It is important to note that this agreement of results is coincidental and specific to the chosen distribution. Namely, all raw higher moments of a Bernoulli random variable are p. If we had taken any other distribution in combination with a moment of fractional order, it becomes highly likely that the MGF returns a different value compared to the traditional method of computing moments. What is more, if we were to take

$$_{RL}M_X^{(\frac{1}{2})} = \frac{d}{dt}\frac{1}{\sqrt{\pi}} \int_{-h}^{t} (t - s)^{\frac{-1}{2}} f(s)ds$$

we obtain an integral which may diverge based on the choice of $-h$. These observations lead to the following theorem.

Theorem 2.3.3. *Consider the three MGF's as defined in definition 2.3.2. Assume* $_{RL}M_X^{(\alpha)}$ *and* $_{GL}M_X^{(\alpha)}$ *are well defined on some open interval* $(-h, h)$*, then the MGF expressions* $_{RL}M_X^{(\alpha)}$ *and* $_{GL}M_X^{(\alpha)}$ *accurately obtain raw moments of order* $\alpha \in \mathbb{R}$

The proof can be found in Appendix B.2.
Unfortunately, this result does not hold for ${}_{CF}M_X^{(\alpha)}$, which leads to the following theorem.

Theorem 2.3.4. *Consider the three MGF's as defined in definition 2.3.2. Assume ${}_{CF}M_X^{(\alpha)}$ is well defined on some open interval $(-h, h)$, then the MGF ${}_{CF}M_X^{(\alpha)}$ inaccurately approximates moments of order $\alpha \in \mathbb{R}$ with approximation error given by*

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} x^\alpha f_X(x)dx - \int_{-\infty}^{\infty} \frac{x^{n+1}}{(1-\beta)x+\beta} f_X(x)dx & \text{if } X \text{ is continuous,} \\ \displaystyle\sum_i \left(x_i^\alpha - \frac{x_i^{n+1}}{(1-\beta)x_i+\beta} \right) f_X(x_i) & \text{if } X \text{ is discrete.} \end{cases}$$

with $\alpha \in \mathbb{R}, \beta = \alpha - n$ and $n = \lfloor \alpha \rfloor$.

The proof can be found in Appendix B.2.

Remark 2.3.3. In the case when $\alpha \in \mathbb{N}$, we have that $n = \alpha$, and thus $\beta = 0$, therefore, ${}_{CF}M_X^{(\alpha)}$ is accurate for integer orders. This is an expected result, as the MGF for integer moments is accurate and from section 2.2 we know that $D_{CF}^{\alpha+\beta} f(x) = D_{CF}^\alpha(D_{CF}^\beta f(x))$, with $\alpha \in \mathbb{N}, \beta \in [0, 1)$. In this case, let $\beta = 0$. So we get $D_{CF}^{\alpha+0} f(x) = D_{CF}^\alpha f(x)$ which is just a regular derivative of integer order.

3 Error Analysis

We now focus on computing the newly obtained MGF expressions numerically. To highlight the errors of the Caputo-Fabrizio MGF, moments of various orders with different parameter configurations will be computed. However, before computing these errors involving explicit distributions, we first focus on the expression

$$E(x, \alpha) = x^\alpha - \frac{x^{n+1}}{(1-\beta)x+\beta}$$

which is of great importance in order to fully comprehend the errors discussed in section 3.3

3.1 Analysing a particular error term of interest of the Caputo-Fabrizio MGF

We explicitly focus on the term

$$E(x, \alpha) = x^\alpha - \frac{x^{n+1}}{(1-\beta)x+\beta} \tag{6}$$

for simplicity. The expression we analyse, denoted $E(x, \alpha)$ is not the same as the expression obtained in theorem 2.3.4. Yet, it is the part of the expression that involves the order term α and is thus of great interest. What is more, considering the result obtained in theorem 2.3.4, it is easy to see that ${}_{CF}M_X^{(\alpha)}$ is accurate $\iff x^\alpha = \dfrac{x^{n+1}}{(1-\beta)x+\beta}$. Thus, we consider the function $E(x, \alpha)$, to be the function of their differences and analyse when this function equals zero. Plotting this expression for different orders of α, we obtain the following figure:

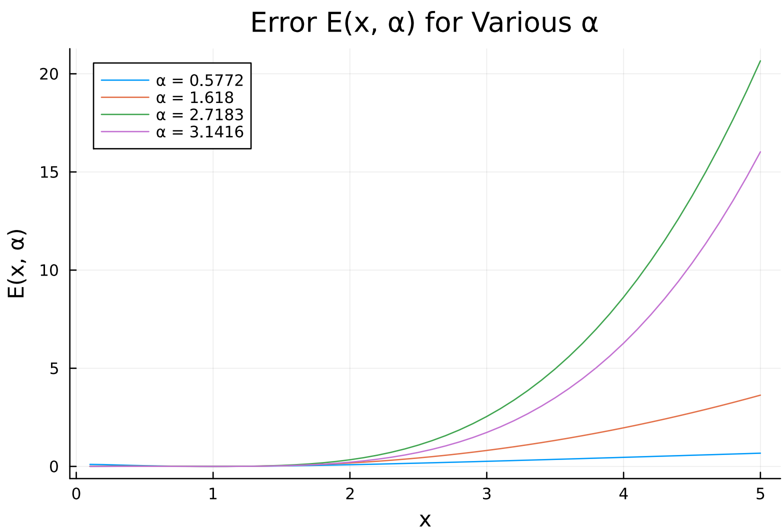

Figure 1: Error of Caputo-Fabrizio MGF

It can be observed from figure 1 above that, in general for greater orders α, the error tends to increase. However, the greatest order of α does not provide the greatest error. Instead it is the second greatest order of α that does so. This may be explained by the fact that for values of α, being exactly in the middle of two integers, the value of $(1-\beta)x+\beta$ will be the greatest. As a result, the fraction of error term 6 becomes small, thus the error tends to increase.

This hypothesis is indeed confirmed in the following figure, which focuses on the growth rate of the error, for a fixed value of x and an increasing order α.

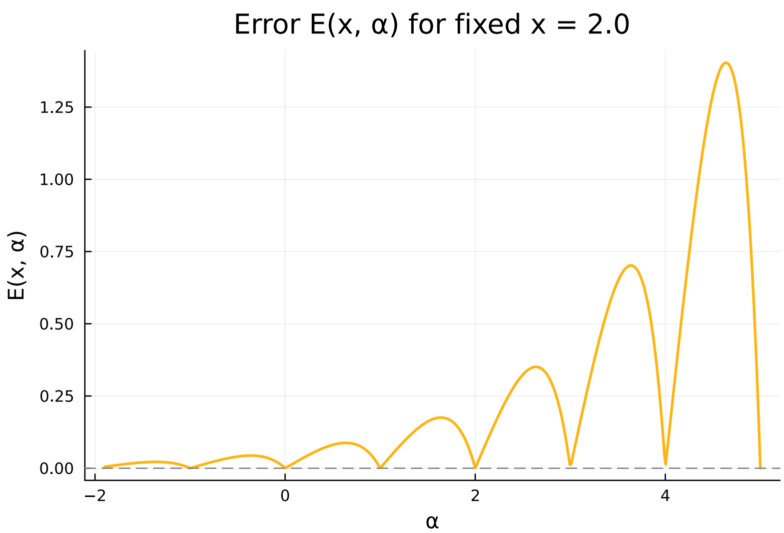

Figure 2: Error of Caputo-Fabrizio MGF, for fixed x

Figure 2 indeed seems to indicate that for some $\alpha \in (a,b)$, where $a,b \in \mathbb{Z}$ are consecutive integers that $E(x,\alpha)$ attains it maximum value for some α 'slightly greater' than $\frac{a+b}{2}$. Figure 2 also reveals that the approximation error is considerably smaller for negative fractional moments compared to the positive ones. This suggests that, when negative moments are well-defined, the Caputo-Fabrizio MGF may still offer reasonable accuracy, despite its known limitations.

Note that within each subsequent integer interval, the function $E(x,\alpha)$ appears to be concave. Thus, as is supported by figure 2, it is only possible to obtain local maxima. Ideally, one would aim to analytically determine the optimal order $\hat{\alpha}$ such that minimizes $E(x, \hat{\alpha})$. However, due to the concavity of $E(x, \alpha)$, such analytical minimization is not feasible. Figure 2 seems to suggest that the order of α near integer orders yields smaller approximation errors, which makes sense intuitively. However, the latter is not rigorous evidence. Therefore, in the following section, we will analyse the numerical approximation errors of the Caputo-Fabrizio MGF by computing moments of fractional order for different distributions and various orders α.

3.2 Practical Issues

Computing fractional moments of a specific distribution poses some difficulties. In the first place, we want to compute the moment of fractional order α for a random variable X. Official packages to compute raw moments of fractional order do not exist yet, thus these methods will have to be created first. In order to accomplish the desired result, we directly compute $\int_{-\infty}^{\infty} x^\alpha f_X(x)dx$. If the support of the random variable is infinite, this integral is often not numerically well defined or takes too long to compute. Thus, in such cases we use a bounded support $(-100, 100)$ which provides a sufficiently accurate approximation. What is more, for many fractional orders α in combination with negative values of x, x^α is of complex form. If a distribution has a support which contains negative values, it is possible to compute the complex or absolute moment of fractional order, which avoids any numerical issues. This approach, however, lacks interpretability, as the intuition behind absolute and complex moments is harder to grasp.

It now remains to compute the fractional moment using the Caputo-Fabrizio MGF. Unfortunately, it is rather difficult to compute fractional derivatives of an MGF expression numerically. The reason is, that programming packages often take (fractional) derivatives on a certain point x rather than over the entire function. Thus, we simply use the result obtained in theorem 2.3.4. That is, instead of the moment of order α, the Caputo-Fabrizio MGF computes the moment of order $n+1$ divided by some variables dependent on α. More explicitly, we compute ${}_{CF}M_X^{(\alpha)}(0) = \displaystyle\int_{-\infty}^{\infty} \frac{x^{n+1}}{(1-\beta)x+\beta} f_X(x)dx$. Throughout the rest of the thesis, this is the expression that is being computed whenever values of the Caputo-Fabrizio MGF are discussed. This integral faces the same problems as the aforementioned integral including x^α. What is more, for specific values of x and β, the denominator tends to go towards zero, which is numerically unstable. Thus, in such cases, we add a small $\epsilon > 0$ in order to avoid this issue.

3.3 Accuracy Analysis

For this accuracy analysis we will compute fractional moments based on different parameter values and orders of α. We set $\alpha \in [0, 5]$ with steps of 0.01. This yields 500 different fractional moments per parameter-configuration. We consider three distributions, namely the Exponential, Normal and Poisson distribution. This selection allows us to cover both continuous and discrete distributions, and distributions with negative and positive supports.

3.3.1 Computing fractional moments for various distributions

Before analysing the errors of the Caputo-Fabrizio MGF, we first compare the values obtained by each method. Specifically, we compute:

- **Theoretical values:** $\displaystyle\int_{-\infty}^{\infty} x^\alpha f_X(x)dx$, which has been shown to coincide with the Riemann-Liouville and Grünwald-Letnikov MGFs.
- **Caputo-Fabrizio MGF:** $\displaystyle\int_{-\infty}^{\infty} \frac{x^{n+1}}{(1-\beta)x+\beta} f_X(x)dx$

- **Complex moment generating function (CMGF):** defined as in equation 2 in section 2.1.1, which computation is given in algorithm 1. In this case, $\xi = 0$.

- **Empirical values:** $\mathbb{E}[|X|^{\alpha}] \approx \dfrac{1}{n}\sum_{i=1}^{n} |x_i|^{\alpha}$. Where $\{x_i\}_{i=1}^{n}$ are simulated samples and $n = 1.000.000$.

For the Poisson, Exponential and Normal distribution, we obtain the following figures:

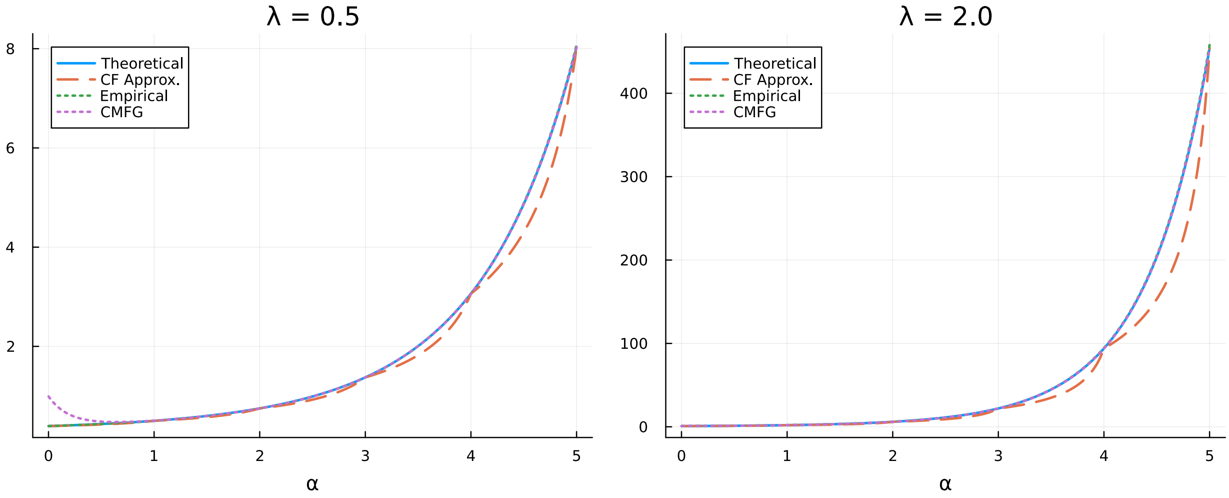

Figure 3: Comparison of different fractional moments for the Poisson Distribution

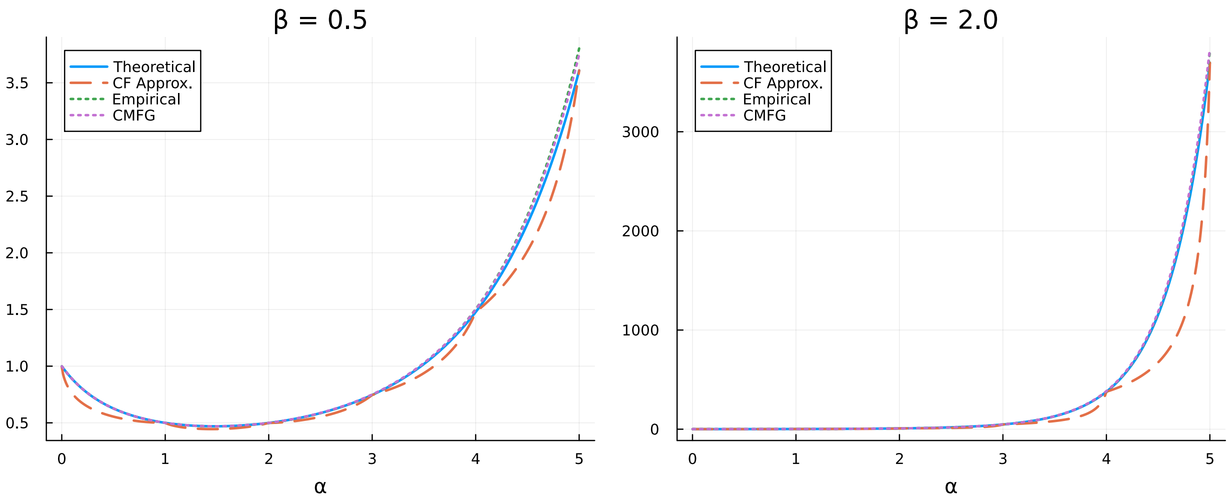

Figure 4: Comparison of different fractional moments for the Exponential Distribution

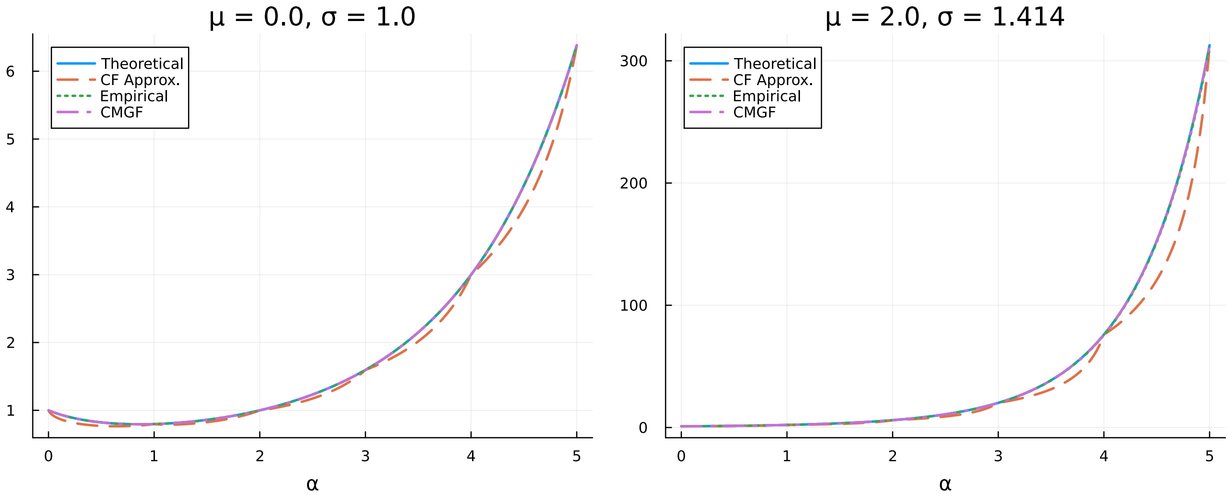

Figure 5: Comparison of different fractional moments for the Normal Distribution

The figures above show fractional moment values for various orders α and different parameter values as computed using the Caputo-Fabrizio MGF, the two accurate MGFs, the CMGF and the empirical method. The results are consistent across all three distributions. The values of the accurate MGFs (blue line), the CMGF (dashed purple line) and empirical values (dashed green-line) almost completely agree, confirming the accuracy and usability of these methods. The values obtained by the Caputo-Fabrizio MGF are always smaller or equal to the values of the accurate MGF expressions. This observation is in line with the result that the errors are always non-negative. Moreover, for all three distributions, an increase in the parameter values, leads to substantially larger fractional moment values. Correspondingly, the errors of the Caputo-Fabrizio MGF also increase, confirming that the approximation error grows with the distribution parameters.

We obtain the following table, including the run-time (in seconds) of these different methods:

Method	**Poisson**	**Exponential**	**Normal**
Theoretical	0.159	0.045	0.060
CF Approx.	0.161	0.035	0.029
CMGF	1.264	0.181	0.832
Empirical	49.993	66.076	64.815

Table 1: Run-time of different methods

The run-time values reported in table 1 are averages over 1000 executions for the theoretical and Caputo-Fabrizio MGFs, 100 executions for the CMGF and 10 executions for the empirical method. Unsurprisingly, the empirical method is the least computationally efficient method, due to the requirement of one million simulated samples. The theoretical and Caputo-Fabrizio MGF methods have comparable run-times, both being highly efficient. The CMGF method is also fast, but noticeably slower than the former two, likely due to the increased complexity of the integrand. The results of the Poisson distribution somewhat differ from those of the Normal and Exponential distribution. This discrepancy may come from the discrete nature of the Poisson distribution, which requires its moments to be computed in a slightly different numerical manner.

We now focus on the errors of the Caputo-Fabrizio MGF. Specifically, for each of distribution we compute the difference between the theoretical fractional moments: $\int_{-\infty}^{\infty} x^{\alpha} f_X(x)dx$ and those computed by the Caputo-Fabrizio MGF: $\int_{-\infty}^{\infty} \frac{x^{n+1}}{(1-\beta)x+\beta} f_X(x)dx$ for various orders α. This error expression corresponds to the analytical approximation error given in theorem 2.3.4. The observant reader might notice that all of the errors in the upcoming sections are non-negative and may incorrectly conclude that these are the absolute or squared errors. This is not the case. Since the expression for the pointwise error function $E(x, \alpha)$ as in 6 is always non-negative, it follows that $x^\alpha \geq \frac{x^{n+1}}{(1-\beta)x+\beta}$ and thus $\int_{-\infty}^{\infty} x^\alpha f_X(x)dx \geq \int_{-\infty}^{\infty} \frac{x^{n+1}}{(1-\beta)x+\beta} f_X(x)dx$ for any distribution. Thus the approximation error from theorem 2.3.4 is guaranteed to be non-negative.

3.3.2 Poisson Distribution

First, we let $X \sim Poisson(\lambda)$, with the probability mass function defined as in Appendix D. The Poisson distribution has positive support, thus numerical issues will be avoided. It is not possible to obtain moments of negative order for the Poisson distribution.

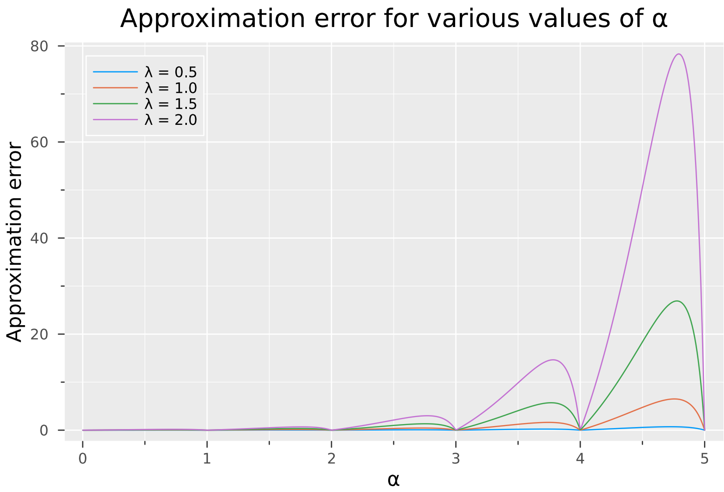

Figure 6: Approximation error for the Poisson Distribution

The trend of the errors in figure 6 is rather similar to the trend displayed in figure 2. That is, the errors behave concavely between each interval of subsequent integers. The plots of the error trends for the Exponential distribution is also very similar. Thus, this plot will not be included. Figure 6 offers us some additional informational. Namely, the approximation errors increase with the parameter values of the underlying distribution function, in this case the Poisson distribution.

λ	**minimum**	**maximum**	**mean**	**standard deviation**	**skewness**	c_v
0.5	0.0	0.715	0.128	0.191	1.915	1.494
1.0	0.0	6.5	1.021	1.713	2.075	1.679
1.5	0.0	26.904	3.895	7.001	2.176	1.797
2.0	0.0	78.34	10.687	20.143	2.251	1.885

Table 2: Poisson Distribution - Approximation Error Statistics

The statistics in table 2 further support this observation. Both the average and maximum of the errors increase, as the parameter λ grows. Moreover, the standard deviation of the errors also increases and the distribution of the errors tends to become more right-skewed as λ increases. For all values of λ, the minimum error is zero, which occurs when $\alpha \in \mathbb{N}$. In the last column, the coefficient of variation has been reported, which has been given by the formula $\frac{\sigma}{\mu}$, where μ and σ denote the sample mean and sample standard deviation respectively (Hendricks and Robey, 1936). This metric of measurement allows us to objectively compare the variability of the errors of each of the parameter configurations. The lower c_v, the lower the relative variability of the errors. This low variability is something we strive for, as it allows us to describe our data with greater confidence. Once more, this statistic also increases with λ. For $\lambda = 0.5$ we may still be able to obtain reliable results. In contrast, the other configurations are likely to be too unreliable in practice, especially for orders of $\alpha > 3$.

3.3.3 Exponential Distribution

We now consider $X \sim Exponential(\beta)$, with PDF as defined in Appendix D. The support of the Exponential distribution is $[0, \infty)$, thus the integration process will avoid most numerical issues. Moreover, all relevant moments of the Exponential distribution are raw moments, which are precisely the moments we are interested in.

We obtain the following associated table with some core statistics:

β	**minimum**	**maximum**	**mean**	**standard deviation**	**skewness**	c_v
0.5	0.0	0.329	0.074	0.083	1.762	1.12
1.0	0.0	22.119	2.907	5.612	2.314	1.93
1.5	0.0	225.731	25.477	54.91	2.513	2.155
2.0	0.0	1128.26	115.489	265.154	2.655	2.296

Table 3: Exponential Distribution - Approximation Error Statistics

The results in table 3 align with the conclusions derived from table 2. The greater the parameter values and order of α, the greater the (average) approximation error. An interesting observation is that for $\beta = 1.5$ the average approximation error exceeds the maximum approximation error for $\beta = 1.0$. For all values of β, the skewness is positive, implying that the distribution of the approximation errors is right-skewed. Thus, clearly the distribution of the errors is not symmetric and therefore certainly not a Normal distribution. From table 3, it is clear that the coefficient of variation increases as the parameter β increases. For $\beta = 0.5$ all result are reasonable. The Caputo-Fabrizio MGF in combination with the Exponential(0.5) could still be viable in practice. For Exponential(1), this conclusion may depend on the maximum order of α.

3.3.4 Normal Distribution

We now consider $X \sim \mathcal{N}(\mu, \sigma)$, with PDF as defined in Appendix D. This distribution is of particular interest as it has two parameters. Thus, we can analyse the differences of effect of both parameters on the approximation error. Therefore, while raw fractional moments of the Normal distribution are not commonly used in practice, we can still obtain valuable insights from these computations. In order to avoid numerical issues, absolute values have been used to compute the fractional moments.

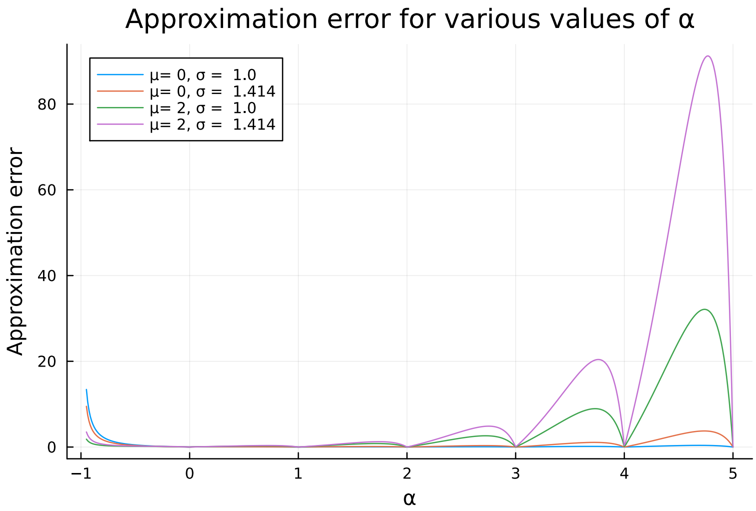

Figure 7: Approximation error for the Normal Distribution

As illustrated in figure 7, the approximation error increases with μ and σ. This result is consistent with the result obtained for the other two distributions. Moreover, this figure also displays the errors for order $\alpha \in (-0.95, 0)$. Interestingly, for negative orders of α, the effect of the parameter values seems to be the reverse. Larger values of μ and σ are associated with smaller approximation errors.

We obtain the following associated table with some core statistics:

μ	σ	**minimum**	**maximum**	**mean**	**standard deviation**	**skewness**	c_v
0.0	1.0	0.0	13.439	0.298	1.07	7.898	3.591
0.0	1.414	0.0	9.505	0.7	1.141	2.824	1.629
2.0	1.0	0.0	32.123	4.517	8.055	2.269	1.783
2.0	1.414	0.0	91.231	11.411	22.404	2.428	1.963

Table 4: Normal Distribution - Approximation Error Statistics

The statistics in table 4 support the visual conclusions drawn from figure 7. The average of the approximation error for $\mu = 0$ is acceptable. However, the size of its coefficient of variation compared to the other parameter configurations stands out. Indeed, configuration $(\mu, \sigma) = (0, 1)$ and configuration $(\mu, \sigma) = (0, \sqrt{2})$ have a rather similar standard deviation, however, the latter has approximately twice the average error. Except for this case, c_v again tends to increase as the parameter values increase, suggesting higher relative variability in approximation errors. Similar to the other distributions, for all parameter-configurations, the distribution of the errors is right-skewed. Thus, the approximation errors of a Normal distribution are not Normally distributed themselves! This result is confirmed in figure 8. Even in the histogram of the parameter-configuration with the greatest parameter values, more than 50% of the approximation errors cluster around zero, with only a small number of observations falling in the range $(40, 80)$. As before, the average errors and variability for large parameters may be too great to be reliable depending on the context of the application.

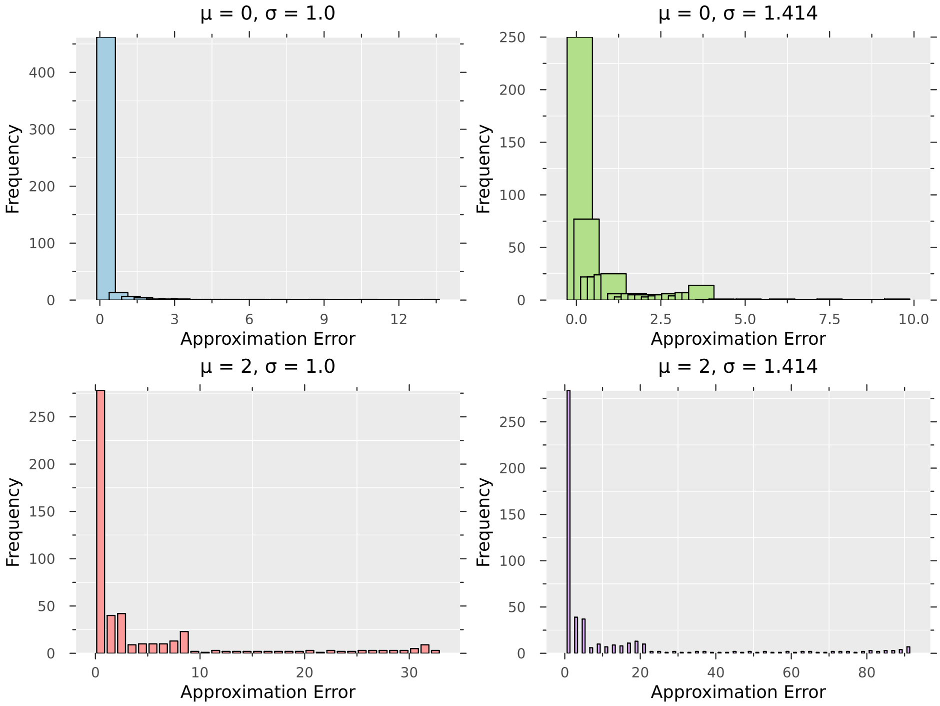

Figure 8: Error Histogram for the Normal Distribution

3.4 General results

Based on the three distributions analysed, we arrive at the following general conclusions:

- As the parameter value of the underlying distribution increases, both the average and maximum approximation error tend to increase.
- For all distributions considered, higher parameter values, lead to greater standard deviations in the approximation errors.
- Furthermore, the coefficient of variation also increases with larger parameter values, indicating that the standard deviation increases at a higher rate than the mean - a stronger result than the increase in the standard deviation alone.
- For all considered distributions and parameter values, the distribution of the approximation errors is right-skewed. Thus, we can conclude that the latter does not follow a Normal, or any other symmetric, distribution.

These findings raise important considerations regarding the reliability of the Caputo-Fabrizio MGF in practice. While high parameter values often result in a large coefficient of variation, suggesting that the quality of the approximation behaves inconsistently. Figure 8 offers a different perspective. Namely, in the case of the Normal distribution, at least 50% of the errors are clustered around zero. Most importantly, in this section, we are considering a sample of approximation errors of 500 fractional moments. When applying moments of fractional order in practice such as in Hansen and Tong (2024) or Gzyl et al. (2013), one usually computes no more than five or ten moments. Considering the latter, metrics such as the standard deviation and coefficient of variation may not always be relevant or decisive when evaluating the usability of the Caputo-Fabrizio MGF. Therefore, rejecting a parameter configuration solely based on its high variability may be too strict. In general, as long as the fractional order of the moment remains below 3 or is close to an integer, the Caputo-Fabrizio MGF tends to yield sufficiently accurate results. This supports its viability in practical cases, such as in Hansen and Tong (2024), where only fractional moments of relatively low order are of interest.

4 Practical Case

To demonstrate the usage of the Moment Generating Function for computing moments of fractional order, we consider a dataset containing the weekly S&P 500 index from the second week of January 2000 up until the last week of December 2024, provided by Wharton Research Data Services (WRDS) (2025).

4.1 First look at the data

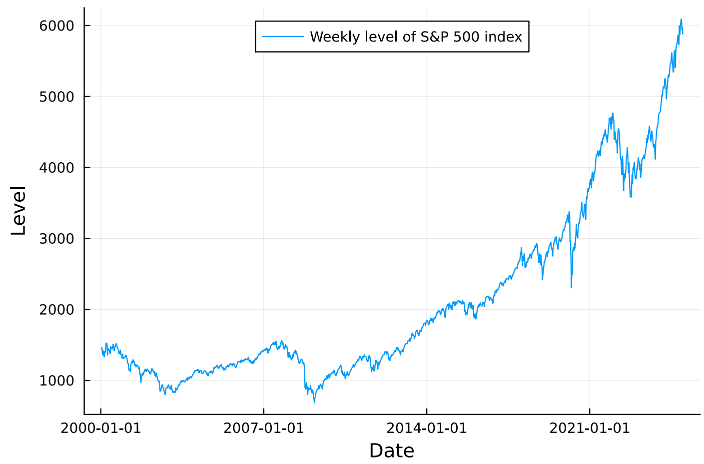

Figure 9: S&P 500 level over time

It is clear from figure 9 that the S&P 500 index is non-stationary, as the mean of the process is not constant. We prefer to work with data that is stationary, thus we consider the log-returns (variations in stock-price) as shown in figure 10. These values have been scaled by a factor of 100 to express them as percentages

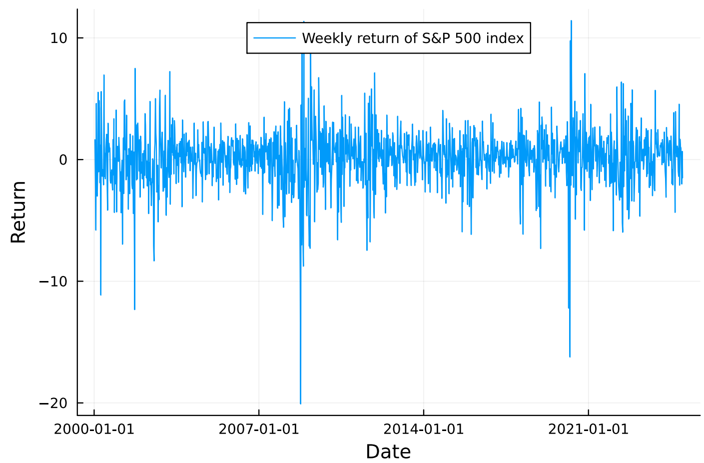

Figure 10: S&P 500 log-returns over time

The log-returns series appears to be more suitable for volatility analysis compared to the data displayed in figure 9. The mean of the log-returns seems to be stationary (lying around zero), the variance of the log-returns also seems to remain constant over time. Additionally, there are no signs of seasonality present. Notable spikes in the log-returns around 2008 and 2021 likely correspond the global financial crisis and COVID pandemic respectively. In times of (financial) uncertainty, it is more likely that returns may take more extreme values, making them less predictable.

The following core statistics of the log-returns, as displayed in figure 10, have been computed:

Mean	**Variance**	**Skewness**	**Kurtosis**	**P-value**
0.108	6.17	-0.878	7.225	1e-99

Table 5: Summary Statistics log-returns of S&P 500

Indeed, as anticipated, the (unconditional) mean seems to be around zero. Moreover, the high value of the kurtosis in table 5 indicates that the log-returns of the S&P 500 are not normally distributed, as this value is far greater than 3. This result is expected in the context of financial markets, due to frequent occurrence of extreme values. The normal distribution, known for its thin tails, underestimates the probability of such extremes. The p-value included in table 5 is the associated p-value of the augmented Dickey-Fuller unit root test, with null hypothesis: the log-returns of the s&P 500 stock index is a unit root process. The p-value is so small that we reject the hypothesis at any conventional significance level. Therefore, we formally conclude that the weekly log-returns of the S&P 500 index are (weakly) stationary.

4.2 Modelling volatility of the log-returns

It is well known that the stock prices, as well as the mean of log-returns, cannot be predicted. Therefore, we focus on the volatility of the log-returns instead. We implement a widely used approach, namely the Generalized Conditional Heteroskedasticity model of order (p, q), denoted GARCH(p, q), as proposed by Bollerslev (1986) with observation equation

$$r_t = \sigma_t \cdot \epsilon_t \tag{7}$$

and updating equation:

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \cdot \sigma_{t-i}^2 + \sum_{i=1}^{q} \alpha_i \cdot r_{t-i}^2 \tag{8}$$

This model has been selected as figure 11 shows that a substantial number of lags exhibit autocorrelations exceeding the 95% confidence interval (given by the red dashed line). This suggests that autocorrelation function does not decay exponentially and in such cases the GARCH model is often suitable (Tsay, 2010).

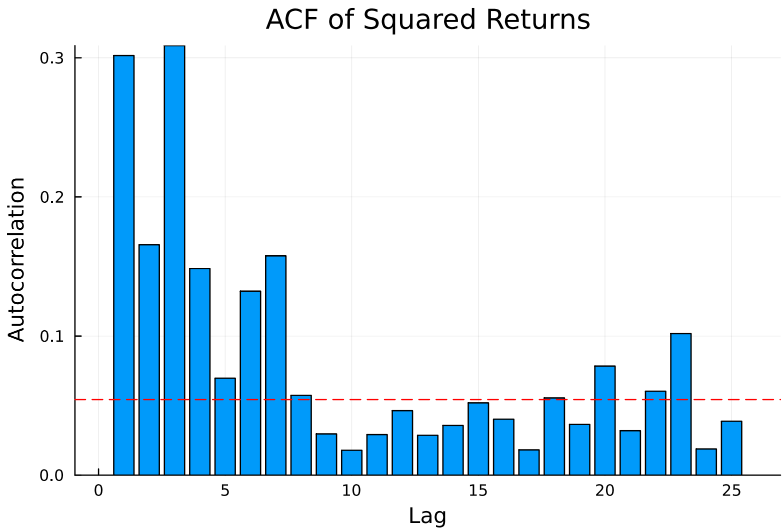

Figure 11: ACF of Squared log-returns

Note that the first index at $t = 0$ has been removed as $corr(r_i, r_j)$ is always equal to 1 for $i = j$.

The selection of the optimal GARCH(p,q) model has been based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The model having both the lowest AIC and BIC value is to be preferred.

Model	**AIC**	**BIC**
GARCH(1,1)	5699	5714
GARCH(1,2)	5700	5721
GARCH(1,3)	5697	5723
GARCH(2,1)	5700	5721
GARCH(2,2)	5702	5728
GARCH(2,3)	5699	5730
GARCH(3,1)	5702	5727
GARCH(3,2)	5703	5734
GARCH(3,3)	5699	5735

Table 6: AIC and BIC for GARCH models

From table 6 we observe that the GARCH$(1, 3)$ model achieves the lowest AIC value, while the GARCH$(1, 1)$ has the lowest BIC value. Both models also obtain relatively low values for the other criterion, indicating that they are good candidates for modelling volatility. For simplicity, we proceed with the GARCH$(1, 1)$ model as that means we will only need to estimate three parameters instead of five, leading to more straightforward interpretability. The parameters have been estimated using maximum likelihood estimation.

Parameter	Estimate	Standard error
ω	0.37	0.084
α	0.224	0.030
β	0.726	0.031

Table 7: Parameter Estimates of GARCH(1,1)

We have that, $\hat{\alpha} + \hat{\beta} < 1$ confirming that the log-returns represent a weakly stationary white noise sequence. Moreover, the standard errors of the estimates are relatively small, indicating high precision.

4.3 Using fractional moments to analyse the log-returns

Now that we have described the data using core statistics, fitted appropriate models and estimated their parameters, we finally consider the usage of moments of fractional order to enhance the analysis.

4.3.1 Computing conditional absolute moments of fractional order

Following Hansen and Tong (2024), we compute the conditional expectation of the absolute variance over a given period H, specifically

$$\mathbb{E}[|X_{T+H}|^{\gamma} \mid \mathcal{F}_{\mathcal{T}}] \tag{9}$$

here $X_{T+H} = Var(R_{T,H}) = \mathbb{E}[R_{T,H}^2] = \mathbb{E}[\sum_{t=T+1}^{T+H} r_t^2] = \sum_{t=T+1}^{T+H} \mathbb{E}[r_t^2]$. Here r_t denotes the log-return at time t and $\mathcal{F}_{\mathcal{T}}$ represents the natural filtration, which is the sigma-algebra generated by the process up to time T (Lowther, 2009). In other words, $\mathcal{F}_{\mathcal{T}}$ contains all information about log-returns observed up to time T. Note that $Var(R_{T,H})$ is indeed equal to $\mathbb{E}[R_{T,H}^2]$ as $\mathbb{E}[R_{T,H}]^2 = \mathbb{E}[\sum_{t=T+1}^{T+H} r_t]^2 = \left(\sum_{t=T+1}^{T+H} \mathbb{E}[r_t]\right)^2$. We assume that $r_t \sim \mathcal{N}(0, \sigma_t^2)$ and since the autocorrelation function of log-returns is zero, we have that each r_t is independent of the observations at different time indices. So it follows that: $\left(\sum_{t=T+1}^{T+H} \mathbb{E}[r_t]\right)^2 = \left(\sum_{t=T+1}^{T+H} 0\right)^2 = 0$. To avoid confusion, we denote the order of the fractional moment by γ rather than α which is already used as a parameter in the GARCH(1,1) model. We compute moments of order $\gamma \in \{-0.5, 0.5, 1.5, 2\}$ as suggested by Hansen and Tong (2024). According to the latter the orders of γ correspond to the following moments:

- $\gamma = -0.5$: Inverse of the volatility of the returns. This value may be used for Sharpe ratio forecasting. Namely it may serve as the inverse of the covariance matrix in the context of portfolio investments. In such a case, we would have to consider a multivariate distribution between the different assets and thus extend the MGF for fractional moments to a multivariate case. This is not the focus of this thesis, thus the Sharpe ratio will not be computed in this manner.

- $\gamma = 0.5$: Conditional volatility or standard deviation of returns

- $\gamma = 1.5$: Conditional skewness (of the absolute value)

- $\gamma = 2$: Conditional kurtosis

These fractional methods are computed using three different methods.

- **Empirical Simulation:** We consider the empirical value of the moments, as defined in section 3.3, obtained by performing 100.000 Monte Carlo simulations of the distribution of the variance (see algorithm 3) followed by averaging the results to obtain the empirical estimates.
- **Integral method:** We compute the standard integral, namely $\int_{-\infty}^{\infty} x^\gamma f_X(x)\,dx$, where $f_X(x)$ is the PDF of $X_{T,H}$. This PDF is acquired by performing Kernel Density Estimation on the values of $X_{T,H}$.
- **Caputo-Fabrizio MGF:** We also compute the expectations of fractional order using the inaccurate method of the Caputo-Fabrizio MGF. This integral also makes use of the same PDF obtained by Kernel Density Estimation.

In this case, the CMGF as in Hansen and Tong (2024) is not computed. An explicit expression of the MGF of the distribution of the variance is required. Such an expression is not available in this context. We compute these moments for different time horizons, namely $H \in \{4, 8, 16\}$. That is, we compute the expectation of the absolute variance of order γ for one month, two months and four months in the future, respectively. This allows us to evaluate whether the time horizon affects the performance of the Caputo-Fabrizio method. The full procedure is detailed in algorithm 2.

We obtain the following table:

H	order	empirical	standard	CF
4	-0.5	0.25	0.255	0.138
4	0.5	4.948	4.938	1.855
4	1.5	201.653	201.763	56.748
4	2.0	1568.7	1569.69	1569.69
8	-0.5	0.164	0.164	0.06
8	0.5	7.098	7.095	1.94
8	1.5	552.323	552.142	114.85
8	2.0	5909.13	5905.96	5905.96
16	-0.5	0.113	0.113	0.028
16	0.5	10.073	10.066	1.971
16	1.5	1550.63	1547.5	230.747
16	2.0	23559.0	23482.6	23482.6

Table 8: Conditional expectations for various orders

The results from table 8 can be interpreted as follows:

- $\gamma = -0.5$: The inverse of the volatility decreases with H, implying that the volatility increases as the forecast horizon extends. All values are positive, due to use of absolute moments.
- $\gamma = 0.5$: The standard deviation of the returns indeed increase when the value of H increases. This is an expected result as, in a short or medium-term context, greater time horizons lead to more uncertainty. This value will likely decrease for much greater values of H, as $\mathbb{E}[|X_{T+H}|^{0.5} \mid \mathcal{F}_T]$ will converge to the square root of the unconditional variance.
- $\gamma = 1.5$: The conditional skewness greatly increases when the value of H increases, which implies that cumulative returns have heavy tails. Even though the skewness is computed using absolute values, which may limit its interpretability, the increasing magnitude may still suggest growing asymmetry in the cumulative returns.

- $\gamma = 2$: The conditional kurtosis vastly increases as the value of H increases. For all values of H, the kurtosis is considerably greater than 3, implying that it is incorrect to assume that the conditional absolute cumulative returns are normally distributed. As with skewness, the interpretability is limited due to the use of absolute values.

For all values of H and non-integer values of γ the empirical results and theoretical integral results are closely aligned. Similarly, for all values of H and $\gamma = 2$, the results of the integral method and Caputo-Fabrizio MGF are the same, as the latter is accurate for integer derivatives of the MGF. As indicated in section 3, the Caputo-Fabrizio MGF greatly under-estimates the values of fractional moments. These errors increase with H, implying that the method increasingly underestimates volatility - and hence risk - for longer horizons. We therefore conclude that the Caputo-Fabrizio MGF is unreliable for computing expectations of fractional order in forecasting applications. This conclusion is consistent with the results of section 3.3, where the size of the errors grew as the distribution parameters increased. Figure 12 emphasizes the increment of the errors of the Caputo-Fabrizio MGF.

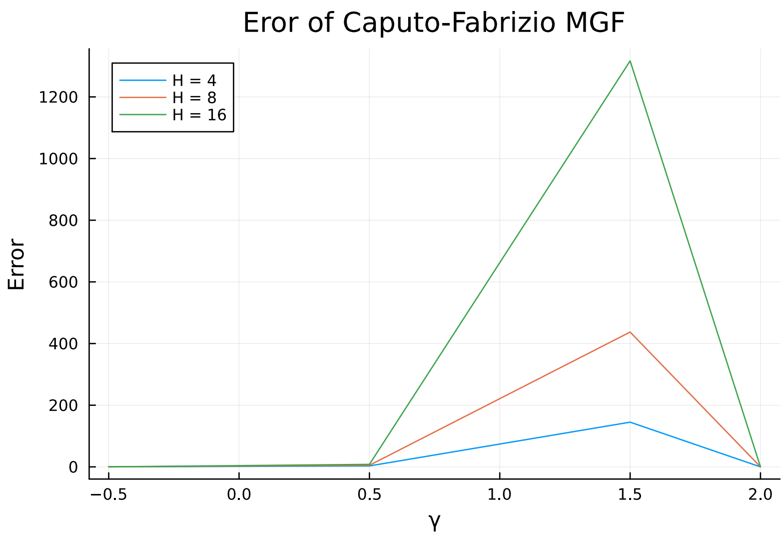

Figure 12: Error of Caputo-Fabrizio MGF for various values of H and γ

Note that figure 12 only includes discrete values of $\gamma \in \{-0.5, 0.5, 1.5, 2\}$. Therefore, the plotted functions may give a misleading impression of continuity. For instance, the error at $\gamma = 1$ should be exactly zero, yet the graph figure suggests otherwise due to the interpolation between points.

4.3.2 More details regarding computation and interpretability of fractional moments

An observant reader might wonder why we compute the expectation of $|X_{T+H}|$ instead of simply computing the expectation of X_{T+H}. We have that $X_{T+H} = Var(R_{T,H}) \geq 0$ so why are these absolute values necessary? The answer is twofold. When we are computing moments of negative order, i.e. for $\gamma = -0.5$ we are computing values of a function that behaves similar to $\frac{1}{\sqrt{X_{T+H}}}$. It is quite clear that this computation will suffer from numerical issues for variances in a neighbourhood of zero. The second reason lies in the manner of how we compute the standard and CF-MGF moments. As aforementioned, there exists no simple analytical expression of the MGF in this context (Hansen and Tong (2024) managed to find one, but considered a different model to monitor volatility, which is outside of the scope of this thesis). Thus we need to work with a PDF of X_{T+H}. To obtain this density, we use Kernel Density Estimation (KDE), which may occasionally produce negative values, especially around the tails of the distribution, where there are less data points. The KDE uses the same data points that were simulated for the empirical moments. As a result, the integral method and Caputo-Fabrizio MGF method lose their advantage of efficient run-time over the empirical method. Computing negative values of fractional order is not well-defined, thus in this case, we have to include absolute values as well. As seen in table 8, this may result in values that are less intuitive or interpretable. We could also have considered to give a complex form to X_{T+H} allowing for negative values. However, these results are most likely even less intuitive. Another option would have been to square all orders to ensure that the integral is well defined for all values of the PDF. This would lead to all orders being integer orders again, which defeats the purpose of considering moments of fractional orders in the first place.

4.3.3 Analyzing portfolio risks using Value-at-Risk and fractional orders of lower partial moments

Next, we consider two different measures of risk, namely the Value-at-Risk (VaR) measure and the lower partial moment (LPM). Where the former is defined as

$$\alpha\text{-VaR} = z_\alpha \cdot \sigma_{t+H} \tag{10}$$

where z_α is the quantile of level α of the standard normal distribution (Holton, 2013). The LPM is defined as follows:

$$LPM_n(\tau) = \int_{-\infty}^{\tau} (\tau - x)^n dF_X(x) \tag{11}$$

where $\tau = \alpha$-VaR and $F_X(x)$ is the distribution of a random variable X (Wojt, 2009). In our case X is the absolute cumulative return. The intuition of the VaR is as follows: given some portfolio with a weekly $\alpha -$ VaR of some percentage p, there is a probability of $\alpha\%$ that the value of the portfolio will fall by more than $p\%$ of its value in one week. The interpretation of the LPM, which is usually defined for $n \in \mathbb{N}$ is as follows. The LPM of order $n = 1$, is called the expected shortfall, which captures the expected downside deviation below the target value, τ. If the LPM is of order $n = 0$, it represent the probability of performing under the target value, τ (Sorino, 2001). We extend the order of the LPM by letting $n \in \mathbb{R}$, focusing on $n \leq 1$, as proposed by Fishburn (1977). According to the latter, the LPM of order $0 < n < 1$ emphasizes the frequency of downside incomes. The following figure depicts the Value-at-Risk with $\alpha = 5\%$, along with the LPM of order $n \in \{0, 1, 0.5\}$, where $LPM_{0.5}(\tau)$ has been computed using both the accurate and inaccurate expressions of the fractional MGFs. The risks are based on the forecasts of the absolute cumulative returns up until H = 26, so half a year in advance.

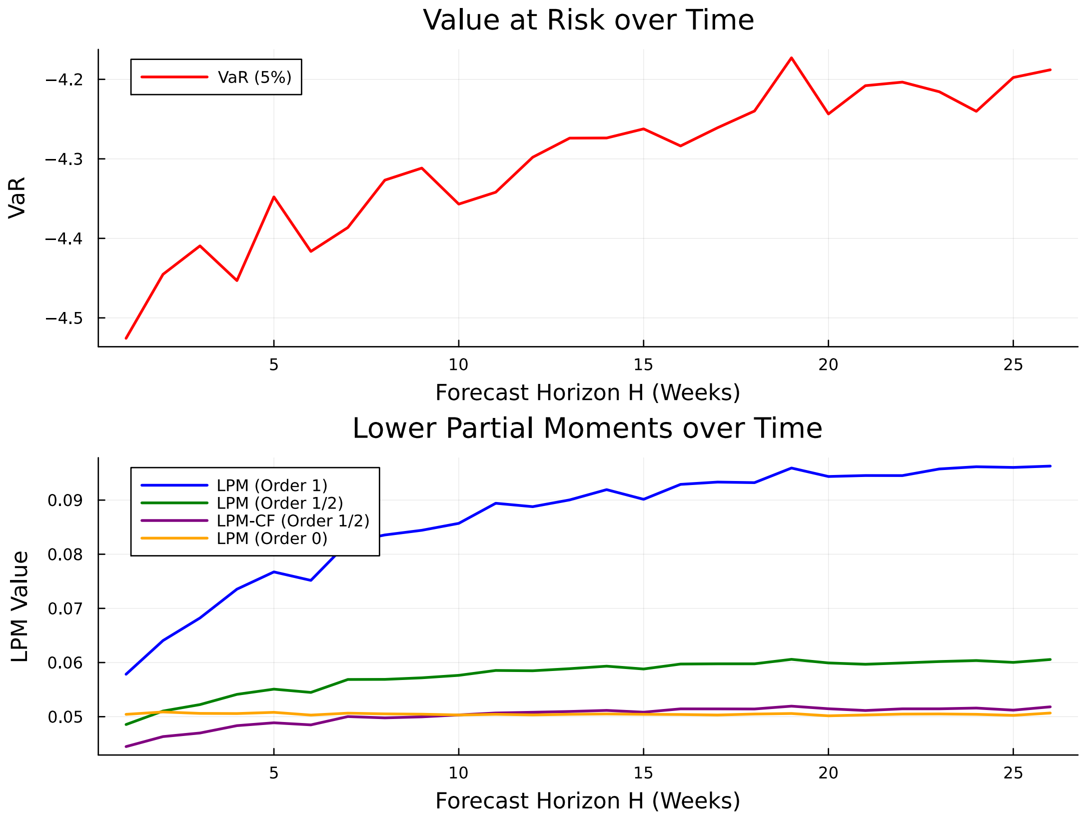

Figure 13: Comparison of VaR and LPM for various orders of n

- From figure 13 we observe that as H increases, the VaR quantile decreases in absolute magnitude, indicating that extreme losses become less severe relative to shorter horizons. This might seem counterintuitive at first, as stock returns may seem less predictable over larger periods of time. The observation, however, actually aligns with the theoretical expectations. That is, $\lim_{h \to \infty} \sigma_T^2(h) = \omega/(1 - \alpha - \beta)$. So as H increases, $\sigma_T^2(h)$ will get closer to its unconditional variance and the function will become less steep.

- As H increases, $LPM_1(\tau)$ tends to increase slightly. This suggests that, while the expected downside loss below the threshold τ may slightly increase over time, it does so at a decreasing rate. The flattening of the $LPM_1(\tau)$ function reflects the convergence of the aforementioned unconditional variance which is in line with the theory of GARCH models.

- As H increases, $LPM_0(\tau)$ seems to follow the same trends as the other orders of $LPM_n(\tau)$. The latter suggests that the downside risk grows marginally as time increases but eventually becomes constant. This is in line with the fact that return distributions converge to their long-run unconditional distribution as time increases (Hamilton, 1994).

- $LPM_{0.5}(\tau)$ shows a gradual increase, as H increases, after which it eventually flattens. This implies that the frequency and magnitude of downside deviations slowly grow over time and eventually reaching a maximum. The latter reflects the diminishing risk over long periods (Sortino and van der Meer, 1991).

- As concluded before, fractional moments computed by the Caputo-Fabrizio MGF, inaccurately underestimated. All values of $LPM_{0.5}(\tau)$ computed via the Caputo-Fabrizio MGF are consistently lower than those computed using direct integration, once more confirming its systematic underestimation of downside risk. The Caputo-Fabrizio MGF approach does not fully capture the tail behaviour of the return distribution, especially for small values H where it even obtains smaller values compared to $LPM_0(\tau)$.

4.3.4 Modelling volatility using an observation-driven regression model

In this final section, we try forecasting volatility using a different model. Namely, we consider an observation-driven regression model, a special case of the Generalized Autoregressive Score framework (Creal et al., 2013). This model has observation equation:

$$y_t = \beta_t x_t + \epsilon_t \tag{12}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ and updating equation:

$$\beta_t = \omega + \phi\beta_{t-1} + \alpha(y_{t-1} - \beta_{t-1}x_{t-1})x_{t-1} \tag{13}$$

This model has the advantage of allowing β_t to evolve over time, in contrast to models like GARCH(1,1) which impose a constant parameter structure. Moreover, it is known that GARCH models are sensitive to extreme events. Therefore, it may be useful to consider a different method of obtaining future volatility. The time-varying parameters are especially useful when working with financial return series that contain extreme values or structural changes. A fixed parameter may be overly influenced by outliers, whereas a dynamic β_t allows the model to adapt more flexibly to such deviations without negatively affecting prediction accuracy. In our application, the goal is to predict the variance of cumulative returns over the next four weeks ($t + 4$), making use of all the information available at time t. Thus we consider the model: $y_{t+4} = \beta_t x_t + \epsilon_{t+4}$ We will consider two different versions of this model.

- Standard model: we let $x_t = |r_t|$ represent the realized absolute return at time t, as in Taylor (2007).
- Extended model: we let x_t represent the conditional expectation of the absolute variance over two weeks of order 0.5. That is $x_t = \mathbb{E}[|X_{t+2}|^{0.5} \mid \mathcal{F}_t]$, with X_{t+2} defined as before. In this case we have that $x_t = \mathbb{E}[|X_{t+2}|^{0.5} \mid \mathcal{F}_t] = \left(\sum_{s=t+1}^{t+2} \mathbb{E}\left[r_s^2 \mid \mathcal{F}_t\right]\right)^{0.5}$. The square root of the variance of the absolute returns over a period of two weeks. This choice is motivated by section 4.3.1 where this value was interpreted as the standard deviation of the absolute returns over a period of two weeks. It is common practice to use past volatility to predict volatility (Poon and Granger, 2003). In this case, we try to predict the variance of the cumulative returns over a month using the standard deviation of the cumulative returns over 2 weeks.

The decision of $H = 2$ has been made as to form a compromise between short-term volatility noise ($H = 1$) and greater forecast uncertainty of a longer horizon ($H = 3$). This choice is in line with findings of Andersen and Bollerslev (1998), which show that such forecasts are informative for longer-term volatility prediction. Since realized (absolute) returns have been proven to be of interest for forecasting volatility (Taylor, 2007), $H = 0$ for the regressor in the standard model. The goal of this specific case study is to compare the regression performance between a model with and without a moment of fractional order serving as regressor. To maintain the focus on this objective and avoid confusion from the various different definitions of fractional moments, we will only consider the fractional moments computed using the expression of the accurate MGFs. This choice ensures an objective comparison. If we were to use the Caputo-Fabrizio MGF for the extended model, it would most likely perform poorly as the fractional moments, which serve as the forecasted standard deviations at time $t = 2$, are simply incorrect.

The dataset has been split as follows: 80% of the data has been used for the training-set and the remaining 20% functions as the test-set. Around 50% of the training-set has been reserved for estimating the parameters of the GARCH(1,1) model. We already estimated the parameters of a GARCH(1,1) model, but the estimates of those parameters were based on the entire dataset. In this case, we need parameter estimates for each different time t to obtain the most accurate computations of $\mathbb{E}[|X_{t+2}|^{0.5} \mid \mathcal{F}_t]$. We still make use of the parameters we estimated in table 7 as we let $\sigma_0^2 = \hat{\omega}/(1 - \hat{\alpha} - \hat{\beta})$ be the initial value of σ_t^2. We compute $\mathbb{E}[|X_{t+2}|^{0.5} \mid \mathcal{F}_t]$ for all t-values in our selected window. The computation is rather similar to those of section 4.3.1, with only a slight modification in the code (see algorithm 4). After performing maximum likelihood estimation, we obtain the following estimates:

Model	$\hat{\omega}$	$\hat{\phi}$	$\hat{\alpha}$	$\hat{\sigma^2}$
Standard model	0.066	0.78	0.007	0.839
Extended model	0.08	0.812	0.219	0.683

Table 9: Parameter Estimates of the two different regression models

The two models yield similar estimates for ω and ϕ. The most notable differences appear in the parameter estimates of α and σ^2. The higher value of $\hat{\alpha}$ implies that the updates to β_t of the extended model are more responsive to the product of the past error term and regressor compared to the standard model, namely $\epsilon_{t-1} \cdot x_{t-1}$. Meanwhile, the smaller value in the extended model of $\hat{\sigma^2}$ of $\epsilon_{t-1} \sim \mathcal{N}(0, 0.683)$ compared to $\epsilon_{t-1} \sim \mathcal{N}(0, 0.839)$ indicates reduced variance in the residuals, suggesting a better fit of the model.

The following plot displays the path of β_t for the standard and extended model, that is, it shows the evolution of the value of β_t over the training period (2010-2022).

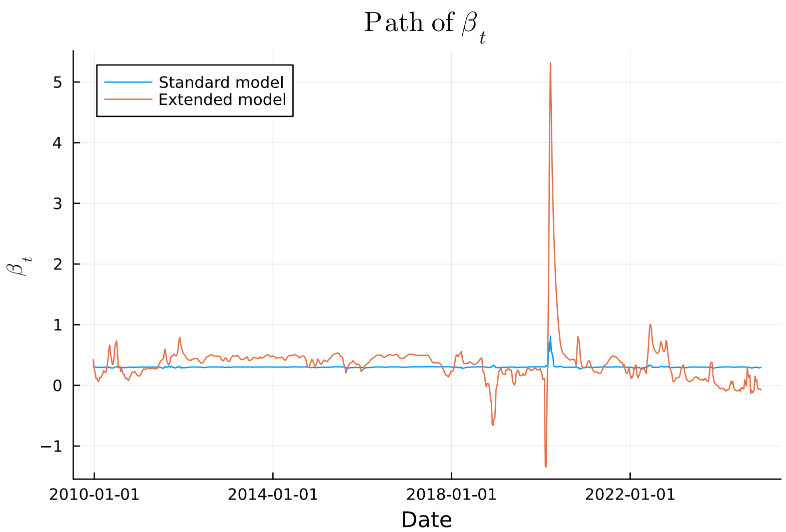

Figure 14: Path of β_t for the two different models

In the standard model β_t remains relatively stable, clustering around a value of 0.5, except for a spike around 2020 (COVID pandemic). In contrast, the extended model produces a more volatile β_t fluctuating between values around -1 and 5. This is possibly due to how β_t is updated in the extended model. Namely β_{t+2} is dependent on $\mathbb{E}[|X_{t+1}|^{0.5} \mid \mathcal{F}_t]$. We have seen in section 4.3.1 how great the values of these fractional moments can get. Therefore, as the value of $\mathbb{E}[|X_{t+1}|^{0.5} \mid \mathcal{F}_t]$ explodes, so will the value of β_t. The spike during the COVID pandemic perfectly illustrates this behaviour, where greater uncertainty drove up the value of the fractional moment and thus the value of β_t.

The table below includes performance measures of the standard model and extended model respectively.

Model	**MSE**	**MAE**	**RMSE**	R^2
Standard model	0.275	0.387	0.524	0.023
Extended model	0.207	0.349	0.454	0.267

Table 10: Performance measures of the two different regression models

The values of the mean squared error, mean absolute error and root mean squared error of the extended model are all smaller compared to those of the standard model, indicating that the extended model is preferred. It also achieves a higher value for R^2. Note that R^2 is not always a reliable performance measure for time series data. This follows from the fact that in a time series context, there exists autocorrelation between the residuals, thus violating the assumption that residuals should be independent of each other. Using R^2 in the context of time series often results in low values of R^2, indicating poor predictive performance, while this may not necessarily be the case. Yet, in table 10 we observe that the associated R^2 value of the extended model is roughly ten times the value of the associated R^2 value of the basic model. This suggests that the regressor x_t in the extended model has a much greater correlation with the variation in y_t compared to the regressor x_t in the standard model. That is, the regressor x_t in the extended model explains a greater portion of the variance in y_t. Thus, in this context, the performance measure R^2 is still found to be useful. The following figure depicts the actual variance along with the variance predicted by the two different models:

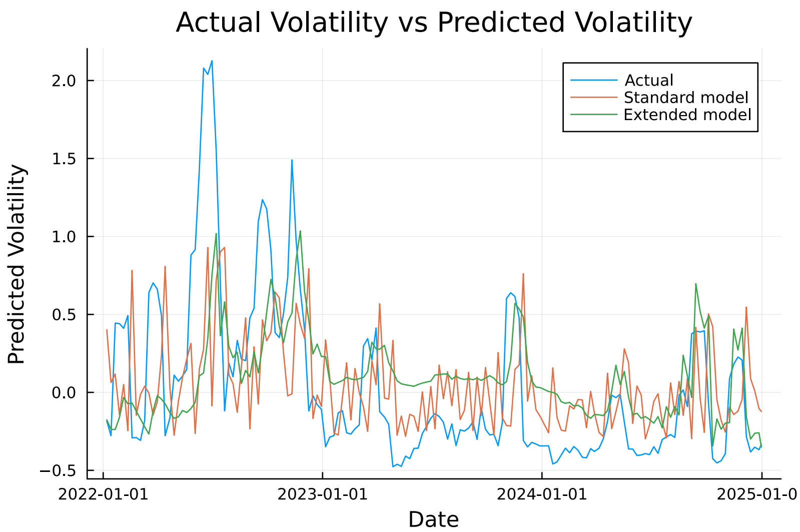

Figure 15: Actual volatility vs Predicted volatility

From figure 15 we observe that the actual variance attains more extreme values compared to the predictions from both models. While the β_t value of the extended model was more flexible, as shown in figure 14, it is the predicted volatility of the standard model that exhibits more fluctuation. The predicted volatility path of the extended model is much smoother, similar to the actual variances. This might explain the significant difference in performance measures between the two models.

Incorporating an expression involving fractional moments as a regressor leads to an improved performance in predictions of volatility. The extended model not only obtains lower prediction errors, but also better captures the structure of the data. This comes with an increase in volatility of the coefficient path, which may be either an advantage or disadvantage depending on the specific application.

5 Conclusion

This thesis has explored a relatively new method for computing moments of fractional order. This novel approach included combining the moment generating function with various different fractional derivatives. It was shown that the moment generating function in combination with either the Riemann-Liouville fractional derivative or the Grünwald-Letnikov fractional derivative leads to accurate computations of fractional moments. In contrast, the moment generating function in combination with the Caputo-Fabrizio fractional derivative leads to a consistent under-estimation of moments of fractional order. It was shown that in general the mean and maximum error, the standard deviation, skewness and the coefficient of variation of the error increase as the parameter values of the underlying distribution grow. Nevertheless, the Caputo-Fabrizio moment generating function remains accurate for integer orders and in general, most of the errors are clustered around zero.

The practical relevance of moments of fractional order was demonstrated through the application of volatility forecasting in financial markets. These moments were used to derive informative statistics for future stock returns and volatility. The values obtained using the Riemann-Liouville and Grünwald-Letnikov moment generating functions aligned closely with the empirical values, thus confirming their effectiveness in real world applications.

Additionally, the concept of lower partial moments was extended to fractional orders, which represent the frequency and magnitude of downside deviations. In both applications, the Caputo-Fabrizio moment generating function was shown to under-estimate the moments of fractional order. In particular when the forecasting horizon increased. This suggests that the Caputo-Fabrizio moment generating function is not well suited for financial risk analysis, as it tends to understate volatility and associated risk.

Moreover, moments of fractional order were implemented as regressors in an observation-driven regression model. This approach allowed for a more flexible β_t compared to the β_t of the standard model, which led to reduced errors and improved forecasting accuracy compared to a standard model. In the absence of closed-form expressions of the moment generating function in this setting, it was required to make use of traditional numerical integration. To ensure numerical stability, absolute values had to be used, which leads to the computation of absolute moments of fractional order. While this is an effective solution, it limits the interpretability of the results, compared to unrestricted fractional moments. This limitation could be avoided if closed form expressions of the moment generating function are available.

Future research could include exploring moment generating functions in combinations with different fractional derivatives as those proposed by Hadamard (1892), Riesz (1949) and Marchaud (1927). It would be valuable to assess which combinations obtain accurate values of moments of fractional order and whether a general theorem regarding the accuracy of the moment generating function in combination with fractional derivatives can be established. Another topic of interest would be to consider computing fractional moments of multivariate distributions as briefly mentioned in Hansen and Tong (2024). This would allow for computing fractional cross-moments to explore dependencies between different random variables. On a practical level, developing Julia packages capable of performing symbolic differentiation similar to SymPy in Python (Meurer et al., 2017) would reduce reliance on manual derivation. Currently, the lack of such tools in Julia requires all derivative expressions to be computed by hand before use, making computation an inefficient and painful process.

Acknowledgements

I would like to thank my supervisor Gabriele Mingoli for the great guidance and clear and timely communication throughout the process. Most importantly, I am especially grateful that he gave me the freedom to explore a somewhat abstract and theoretically topic that potentially lies outside the conventional scope of econometrics topics for the bachelor thesis.

A Appendix: Relevant Functions and Identities

Definition A.0.1. Let $F(s) = \mathcal{L}(f) = \int_0^\infty e^{-st} f(t)dt$ be the Laplace transform of f. Then $f = \mathcal{L}^{-1}(f)$ is said to be the inverse Laplace transform of F. The Laplace transform can be used to convert differential equations to much simpler algebraic equations (Thiagarajan, 2022).

Definition A.0.2. The Mellin transform of some function $f(t)$ is defined as $\mathcal{M}\{f(t)\}(s) = \int_0^\infty t^{s-1} f(t)dt$. The Mellin transform is often used in computer science for the analysis of algorithms (Flajolet and Sedgewick, 1996).

Definition A.0.3. The Fourier transform of some function $f(x)$ is defined as $\hat{f}(\xi) = \int_{-\infty}^\infty f(x)e^{-2\pi i\xi x}dx$, $\xi \in \mathbb{R}$. The Fourier transform can be used to simplify differential equations and finds applications in signal and image processing (Fourier, 1822).

Definition A.0.4. Let D be the differential operator, such that $Df(x) = \frac{d}{dx}f(x)$. Then, as in Samko et al. (1993), the fractional derivative of order α is defined as

$$D^\alpha f(x) = \frac{d^\alpha}{dx^\alpha} f(x)$$

In this definition, α can be any real number. When taking regular derivatives, $\alpha \in \mathbb{N}$. In most of our cases, we are interested in the instance where $\alpha \in \mathbb{R}_+$. It is also possible to study derivatives of negative order, which can be used to obtain moments of negative order of a random variable, provided that such an order exists. A derivative of negative order is simply an integral of positive order. This is defined as follows:

Definition A.0.5. Let I be the integral operator, such that $If(x) = \int f(x)dx$. Then by Cauchy (1823) the fractional integral of order α is defined as

$$(I^\alpha f)(x) = \frac{1}{(\alpha-1)!}\int (x-t)^{\alpha-1} f(t)dt$$

Combining the previous two definitions, we obtain the following, more general, definition, as in Oldham and Spanier (1974).

Definition A.0.6. The differintegral operator is defined as

$$R^\alpha f(x) = \begin{cases} I^{|\alpha|} f(x) & \text{if } \alpha < 0 \\ D^\alpha f(x) & \text{if } \alpha > 0 \\ f(x) & \text{if } \alpha = 0 \end{cases}, \text{ with } \alpha \in \mathbb{R}. \tag{14}$$

We define the (Euler-)Gamma function as follows:

Definition A.0.7. Let $z \in \mathbb{C}$, for $\Re(z) > 0$, we have the following: $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$

The Gamma function can be seen as an extension of the factorial function, for non-integers. This function is defined for complex numbers and all there subsets (so also real numbers), as long as the condition above holds. For positive integers values z, we have the following identity: $\Gamma(z) = (z-1)!$ Other important identities, not necessarily for z an integer, are:

- $\Gamma(z+1) = z\Gamma(z)$

- $\Gamma(2) = \Gamma(1) = 1$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Definition A.0.8. For $0 \leq k \leq n$, the Binomial Coefficient is defined as follows:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{\Gamma(n+1)}{\Gamma(k+1) \cdot \Gamma(n-k+1)}$$

where $n, k \in \mathbb{N}$ and $\Gamma(.)$ as defined in A.0.7.

Definition A.0.9. Vandermonde's identity: for non-negative integers, k, l, m, n, we have that

$$\sum_{k=0}^{l} \binom{m}{k} \cdot \binom{n}{l-k} = \binom{m+n}{l}$$

A modification on the latter identity has been called the Chu-Vandermonde identity. This is the same identity, but it his been proven that the identities still hold for complex values m, n as long as l is a positive integer (Askey, 1975).

Definition A.0.10. The binomial series is a generalization of the binomial formula, namely:

$$(1+x)^{\alpha} = \sum_{k=0}^{\infty} \binom{\alpha}{k} \cdot x^k$$

For the interchange-ability of derivatives and integrals and sums, we can apply the following two theorems:

Theorem A.0.1. *Leibnitz's Rule (as stated in Flanders (1973)): Let $f(x,\theta), a(\theta), b(\theta)$ be differentiable with respect to θ, then we have that:*

$$\frac{d}{d\theta} \int_{a(\theta)}^{b(\theta)} f(x,\theta) dx = f(b(\theta),\theta) \frac{db(\theta)}{d\theta} - f(a(\theta),\theta) \frac{da(\theta)}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial f(x,\theta)}{\partial \theta} dx.$$

For the special case, where $a(\theta), b(\theta)$ are constant we have that:

$$\frac{d}{d\theta} \int_{a}^{b} f(x,\theta) dx = \int_{a}^{b} \frac{\partial f(x,\theta)}{\partial d\theta}.$$

For the interchange-ability of derivatives and summations, the following theorem has been given by Casella and Berger (2002):

Theorem A.0.2. *Suppose that the series $\sum_{x=0}^{\infty} h(\theta, x)$ converges for all θ in an interval (a,b) of real numbers and*

(i) $\frac{\partial h(\theta,x)}{\partial \theta}$ *is continuous for all x*

(ii) $\sum_{x=0}^{\infty} \frac{\partial h(\theta,x)}{\partial \theta}$ *converges uniformly on every closed bounded sub-interval of (a,b)*

Then:

$$\frac{d}{d\theta} \left(\sum_{x=0}^{\infty} h(\theta, x) \right) = \sum_{x=0}^{\infty} \frac{\partial h(\theta, x)}{\partial \theta}$$

B Appendix: Proofs of section 2

B.1 Proofs section 2.2

Proof. Proof of 2.2.1

(i) We will prove for the Riemann-Liouville derivative, the proof for the Caputo-Fabrizio derivative is very similar and the Grünwald-Letnikov derivative is a direct consequence of the linearity of the sum.

$$D^{\alpha}(af(x)+bg(x)) = \frac{d^n}{dx^n}\frac{1}{\Gamma(n-\alpha)}\int_0^x (x-t)^{n-\alpha-1}\left(af(t)+bg(t)\right)dt$$

$$= \frac{d^n}{dx^n}\left(\frac{a}{\Gamma(n-\alpha)}\int_0^x (x-t)^{n-\alpha-1}f(t)dt + \frac{b}{\Gamma(n-\alpha)}\int_0^x (x-t)^{n-\alpha-1}g(t)dt\right)$$

Where we simply split the integral and put the constants in front.

$$= \frac{d^n}{dx^n}\frac{a}{\Gamma(n-\alpha)}\int_0^x (x-t)^{n-\alpha-1}f(t)dt + \frac{d^n}{dx^n}\frac{b}{\Gamma(n-\alpha)}\int_0^x (x-t)^{n-\alpha-1}g(t)dt$$

As the regular derivative operator is linear.

$$= aD^{\alpha}f(x) + bD^{\alpha}g(x)$$

(ii) Intuitively, this makes perfect sense, as the 0-th derivative is just no derivative, thus simply the function $f(x)$. But for these derivatives, a little bit more effort is required to prove this rather obvious fact.
For the Grünwald-Letnikov derivative we get:

$$D^0 f(x) = \lim_{h\to 0}\frac{1}{h^0}\sum_{k=0}^{\infty}(-1)^k\binom{0}{k}f(x-kh) = \lim_{h\to 0}\frac{1}{1}\sum_{k=0}^{\infty}(-1)^k\frac{0!}{k!(0-k)!}f(x-kh).$$

The factorial identity of the binomial coefficient only holds for $0 \leq k \leq \alpha$. Since $\alpha = 0$ and k is always a positive integer lesser or equal to $\alpha, k = 0$. Thus, we get:

$$= \lim_{h\to 0}\sum_{k=0}^{\infty}(-1)^0\frac{0!}{0!(0-0)!}f(x-0h) = \lim_{h\to 0} f(x-0h) = f(x).$$

For the Caputo-Fabrizio derivative, we obtain the following:

$$D^0 f(x) = \frac{1}{1-0}\int_0^x \exp\left(\frac{0}{1-0}(x-t)\right)f'(t)dt = \int_0^x f'(t)dt = f(x).$$

(By the first fundamental theorem of calculus).
Finally, for the Riemann-Liouville derivative, we can simply make use of 2.2.1 to note that in this context $\alpha = 0$ is included in the natural integers. So $D^{\alpha} = \frac{d^{\alpha}}{dx^{\alpha}}f(x) = \frac{d^0}{dx^0}f(x) = f(x)$.

(iii) The proof for the Riemann-Liouville derivative is given by Koning (2015). And the proof for the Caputo-Fabrizio derivative is given by Losada and Nieto (2015). For the Grünwald-Letnikov derivative we get:

$$D^{\alpha}(D^{\beta}f(x)) = \lim_{h\to 0}\frac{1}{h^{\alpha}}\sum_{k=0}^{\infty}(-1)^k\binom{\alpha}{k}\left(\frac{1}{h^{\beta}}\sum_{l=0}^{\infty}(-1)^l\binom{\beta}{l}f(x-lh-kh)\right)$$

$$= \lim_{h\to 0}\frac{1}{h^{\alpha+\beta}}\sum_{k=0}^{\infty}(-1)^k\binom{\alpha}{k}\sum_{l=0}^{\infty}(-1)^l\binom{\beta}{l}f(x-(k+l)h).$$

We substitute $m = k + l$ to deal with the double sums:

$$\lim_{h\to 0}\frac{1}{h^{\alpha+\beta}}\sum_{m=0}^{\infty}f(x-mh)\sum_{k=0}^{m}(-1)^k(-1)^{m-k}\binom{\alpha}{k}\binom{\beta}{m-k}$$

Now we make use of identify A.0.9 to obtain:

$$= \lim_{h\to 0}\frac{1}{h^{\alpha+\beta}}\sum_{m=0}^{\infty}(-1)^m\binom{\alpha+\beta}{m}f(x-mh) = D^{\alpha+\beta}f(x).$$

It can be shown in a similar way that the latter expression is equal to $D^{\beta}(D^{\alpha}f(x))$.

□

B.2 Proofs section 2.3

Proof. Example: let $f_X(x) \sim \Gamma(\alpha,\lambda) = \dfrac{x^{\alpha-1}e^{-\lambda x}\lambda^{\alpha}}{\Gamma(\alpha)}$. This PDF is defined on $(0,\infty)$. So the function is not defined on $\mathbb{R}$. This, however, is not a problem, as we can just evaluate the right limit. Since the Gamma function already uses α as a parameter, we will evaluate $\dfrac{f_X(x)}{|x|^{\beta}}$, with $\beta > 0$:

$$\lim_{x\to 0_+}\frac{f_X(x)}{|x|^{\beta}} = \lim_{x\to 0_+}\frac{x^{\alpha-1-\beta}e^{-\lambda x}\lambda^{\alpha}}{\Gamma(\alpha)} = \lim_{x\to 0_+}\frac{x^{\alpha-(1+\beta)}e^{-\lambda x}\lambda^{\alpha}}{\Gamma(\alpha)}.$$

Thus for $\alpha \geq \beta+1, \lim_{x\to 0_+}\frac{f_X(x)}{|x|^{\beta}} < \infty$. Therefore, the first negative moment of the Gamma distribution should exist.

We compute the first negative moment:

$$\mathbb{E}[X^{-1}] = \int_0^{\infty}x^{-1}\frac{x^{\alpha-1}e^{-\lambda x}\lambda^{\alpha}}{\Gamma(\alpha)}dx = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}x^{\alpha-2}e^{-\lambda x}dx$$

Using the substitution $u = \lambda x, \dfrac{du}{dx} = \lambda, dx = \dfrac{du}{\lambda}$, we get:

$$= \frac{\lambda^{\alpha}}{\Gamma(\alpha)}\int_0^{\infty}\left(\frac{u}{\lambda}\right)^{\alpha-2}e^{-u}du = \frac{\lambda^{\alpha}}{\Gamma(\alpha)\lambda^{\alpha-1}}\int_0^{\infty}\left(\frac{u}{\lambda}\right)^{\alpha-2}e^{-u}du.$$

This integral is equal to $\Gamma(\alpha-1)$ (See definition A.0.7). So we get:

$$\mathbb{E}[X^{-1}] = \frac{\lambda^{\alpha}\Gamma(\alpha-1)}{\Gamma(\alpha)\lambda^{\alpha-1}} = \frac{\lambda\Gamma(\alpha-1)}{(\alpha-1)\Gamma(\alpha-1)} = \frac{\lambda}{(\alpha-1)}.$$

Thus, for $\alpha \neq 1, \mathbb{E}[X^{-1}] = \frac{\lambda}{(\alpha-1)}$. Fortunately, this is always the case, since we had just derived that the integral only converges when $\alpha \geq \beta + 1$, with $\beta > 0$. In other words, $\alpha > 1$. So this holds. □

Proof. Proof of Proposition 2.3.1

$$M_X^{(n)}(t) = \frac{d^n}{dt^n} \int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d^n}{dt^n} e^{tx} f_X(x)dx$$

(We can interchange differentiation and integration since all partial derivatives of $e^{tx}f(x)$ are continuous and the absolute value of the integral converges, as we assume the n-th moment exists, see Theorem A.0.1).

$$= \int_{-\infty}^{\infty} x^n e^{tx} f_X(x)dx, \text{evaluating at } t = 0 := \int_{-\infty}^{\infty} x^n e^{0x} f_X(x)dx$$

$$= \int_{-\infty}^{\infty} x^n f_X(x)dx = \mathbb{E}[X^n]$$

The proof for the case that $f_X(x)$ is discrete is very similar. In that case, one would have to change the order of the derivative and summation, which has been justified in Theorem A.0.2. □

Proof. Proof of Proposition 2.3.1

(i) This is trivial. For X a continuous random variable, we get:

$$M_X^{(0)}(t) = \int_{-\infty}^{\infty} x^0 \cdot e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} 1 \cdot e^{0x} f_X(x)dx = \int_{-\infty}^{\infty} f_X(x)dx.$$

Assuming that $f(x)$ is a PDF, this integrates to 1 by definition. If this integral is not equal to 1, this implies that $f(x)$ is not a PDF. The proof for the discrete case is the same but with a summation instead of an integral sign.

(ii)
$$M_{\mu+\sigma X}(t) = \mathbb{E}[e^{(\mu+\sigma X)t}] = \mathbb{E}[e^{\mu t} \cdot e^{\sigma X t}].$$

Since this is the expectation of x, every term that is not dependent on x can be taken out of the summation:

$$= e^{\mu t} \cdot \mathbb{E}[e^{\sigma X t}] = e^{\mu t} \cdot M_X(\sigma t)$$

(iii)
$$M_{X+Y}(t) = \mathbb{E}[e^{(X+Y)t}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{(x+y)t} f_{X,Y}(x,y)dxdy$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{xt} \cdot e^{yt} f_{X,Y}(x,y)dxdy.$$

$f_{X,Y}(x,y)$ is the joint pdf for X, Y. But we know that the latter is equal to $f_X(x) \cdot f_Y(y)$, if X, Y are independent. Thus we get:

$$= \left(\int_{-\infty}^{\infty} e^{xt} f_X(x)dx \right) \left(\int_{-\infty}^{\infty} e^{yt} f_Y(y)dy \right) = M_X(t) \cdot M_Y(t).$$

□

Proof. Proof of Theorem 2.3.2: Suppose for the moment that X is a positive random variable. Since $x \cdot f_X(x)$ is integrable for $x > 0$, we have:

$$\mathbb{E}(X) = \int_0^{\infty} x \, dF(x) = \int_0^{\infty} \int_{-\infty}^{0} e^{tx} \, dt \, dF(x).$$

$$= \int_{-\infty}^{0} \int_{0}^{\infty} e^{tx}\, dt\, dF(x) = \int_{0}^{\infty} M_X(-t)\, dt.$$

The interchange of the order of integration is subject to $\mathbb{E}(e^{-tX})$ being integrable from $t = 0$ to $t = \infty$.

Finally, by substituting X^{-1} for X, we find:

$$\mathbb{E}(X^{-1}) = \int_{0}^{\infty} M_X^{-1}(-t)\, dt.$$

There are two natural ways to generalize (1) to $\mathbb{E}(X^{-1})$; one way gives:

$$\mathbb{E}(X^{-n}) = \int_{0}^{\infty} \int_{0}^{\infty} \cdots \int_{0}^{\infty} M_X(-t_n)\, dt_n \cdots dt_2 dt_1, \tag{2}$$

while the second way gives:

$$\mathbb{E}(X^{-n}) = \frac{1}{\Gamma(n)} \int_{0}^{\infty} t^{n-1} M_X(-t)\, dt.$$

(Cressie et al., 1981) □

Proof. Proof Theorem 2.3.3. We consider X to be a continuous variable. The three MGF expressions are accurate if $\mathbb{E}[X^\alpha] - M_X^{(\alpha)}(0) = 0$, in other words, if

$$\int_{-\infty}^{\infty} (x^\alpha - D^\alpha e^{tx}) \cdot f_X(x) dx = 0 \iff x^\alpha = D^\alpha e^{tx}$$

where D^α the differential operator of order α. Note that we are taking derivatives w.r.t. t, not x. The proof will be shown for the Grünwald-left derivative.

(i)

$$\begin{aligned} {}_{GL}M_X^{(\alpha)} = D_{GL}^\alpha(e^{tx}) &= \lim_{h \to 0} \frac{1}{h^\alpha} \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} \exp(x(t - kh)) \\ &= \exp(xt) \lim_{h \to 0} \frac{1}{h^\alpha} \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} \exp(-xh)^k \\ &= \exp(xt) \lim_{h \to 0} \frac{1}{h^\alpha} (1 - \exp(-xh))^\alpha \end{aligned}$$

(where we used the binomial series identity defined in A.0.10). Now we use the Taylor expansion of $\exp(-xh) = 1 - xh + \frac{(xh)^2}{2!} - \frac{(xh)^3}{3!} + ... - ....$ Since $h \to 0$, we get: $\exp(-xh) = 1 - xh + \mathcal{O}(h^2)$ (the rest of the terms are negligible). Thus we get:

$$\exp(xt) \lim_{h \to 0} \frac{1}{h^\alpha} (1 - (1 - xh + \mathcal{O}(n)))^\alpha = \exp(xt) \lim_{h \to 0} \frac{1}{h^\alpha} (h(x + \mathcal{O}(h)))^\alpha$$

$$= \exp(xt) \lim_{h \to 0} \frac{1}{h^\alpha} (h^\alpha((x + \mathcal{O}(h)))^\alpha) = \exp(xt) \lim_{h \to 0} (x + \mathcal{O}(h))^\alpha$$

$$= \exp(xt) \cdot x^\alpha.$$

Finally, we take a value of t around 0 and obtain ${}_{GL}M_X^{(\alpha)}(0) = x^\alpha$

(ii) We consider the Riemann-Liouville derivative:

$$D^{\alpha}_{RL}(e^{tx}) = \frac{d^n}{dt^n}\frac{1}{\Gamma(n-\alpha)}\int_{-\infty}^{t}(t-s)^{n-\alpha-1}\cdot e^{sx}ds$$

Koning (2015) has shown that this derivative is equal to $x^{\alpha}e^{tx}$. Thus, if we take a value of t around 0, we get: ${}_{RL}M_X^{(\alpha)}(0) = x^{\alpha}$

The proof also holds for the case when X is a discrete random variable. $\square$

Proof. Proof of Theorem 2.3.4. We compute

$${}_{CF}M_X^{(\alpha)} = D^{\alpha}_{CF}(e^{tx}) = \frac{d^n}{dt^n}\frac{1}{1-\beta}\int_{-\infty}^{t}\exp\left(\frac{-\beta}{1-\beta}(t-s)\right)x\exp(xs)ds$$

$$= \frac{d^n}{dt^n}\frac{x}{1-\beta}\exp\left(\frac{-\beta t}{1-\beta}\right)\int_{-\infty}^{t}\exp\left(s\left(\frac{\beta}{1-\beta}+x\right)\right)ds \text{ where } \beta = \alpha - n \text{ and } n = \lfloor\alpha\rfloor.$$

Now let $u = \dfrac{\beta}{1-\beta} + x$, so we get:

$$\frac{d^n}{dt^n}\frac{x}{1-\beta}\exp\left(\frac{-\beta t}{1-\beta}\right)\int_{-\infty}^{t}\exp(us)ds = \frac{d^n}{dt^n}\frac{x}{1-\beta}\exp\left(\frac{-\beta t}{1-\beta}\right)\frac{1}{u}exp(us)\Big|_{-\infty}^{t}$$

$$= \frac{d^n}{dt^n}\frac{x}{1-\beta}\exp\left(\frac{-\beta t}{1-\beta}\right)\frac{1}{u}\exp(ut) = \frac{d^n}{dt^n}\frac{x\exp(xt)}{(1-\beta)x+\beta}$$

Now we apply the derivative of the above expression, with respect to t, n times:

$$\frac{d^n}{dt^n}\frac{x\exp(xt)}{(1-\beta)x+\beta} = \frac{x^{n+1}\exp(xt)}{(1-\beta)x+\beta}$$

Now we take a value of t around 0 and obtain:

$${}_{CF}M_X^{(\alpha)} = D^{\alpha}_{CF}(e^{tx}) = \frac{x^{n+1}}{(1-\beta)x+\beta}.$$

The latter expression is not equal to x^{α}, thus the error of ${}_{CF}M_X^{(\alpha)}$ is:

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty}x^{\alpha}f_X(x)dx - \int_{-\infty}^{\infty}\frac{x^{n+1}}{(1-\beta)x+\beta}f_X(x)dx & \text{if } X \text{ is continuous,} \\ \displaystyle\sum_i\left(x_i^{\alpha} - \frac{x_i^{n+1}}{(1-\beta)x_i+\beta}\right)f_X(x_i) & \text{if } X \text{ is discrete.} \end{cases}$$

$\square$

C Appendix: Algorithms

Algorithm 1 Compute Fractional Moment via CMGF Method

Require: Order r, location parameter ξ, scale parameter s, distribution name `dist_name`, distribution parameters `parameters`, tolerance `tol`

Ensure: Approximate value of the fractional moment

Define integrand function:
  For $t \in [0, \infty)$:
    $z \leftarrow s + it$
    $Mz \leftarrow \texttt{MGF\_DIST}(z, \texttt{parameters}, \texttt{dist\_name})$
    $M_{-z} \leftarrow \texttt{MGF\_DIST}(-z, \texttt{parameters}, \texttt{dist\_name})$
    numerator $\leftarrow e^{-\xi z} \cdot Mz + e^{\xi z} \cdot M_{-z}$
    integrand$(t) \leftarrow \Re\left(\frac{\text{numerator}}{z^{r+1}}\right)$
Compute integral: integral $\leftarrow \int_0^{100}$ integrand$(t)\, dt$ using `quadgk` with relative tolerance `tol`
return $\frac{\Gamma(r+1)}{\pi} \cdot$ integral

Algorithm 2 Analyze Fractional Moments

Require: Parameters: ω, α, β, list of horizons: H_values, list of orders: $orders$, number of iterations: n_sim

Ensure: DataFrame of conditional moments for each H and order

Initialize empty results DataFrame
for each H in H_values **do**
  $R_H^2 \leftarrow$ SimulateConditionalVariance$(\omega, \alpha, \beta, H, n_sim)$
  $pdf\_R \leftarrow$ KDE distribution from R_H^2
  for each $order$ in $orders$ **do**
    $empirical \leftarrow$ mean$(|R_H^2|^{order})$
    $standard \leftarrow$ FractionalMoment(pdf_R, $order$, use_abs=true)
    $CF \leftarrow$ FractionalMomentCF(pdf_R, $order$, use_abs=true)
    Append $(H, order, empirical, standard, CF)$ to results
  end for
end for
return results

Algorithm 3 Simulate Conditional Variance

Require: Parameters: ω, α, β, list of horizons: H_values, number of iterations: n_sim
Ensure: Array R_H^2 of accumulated conditional variances

Initialize R_H^2 as a zero array of length n_sim
$\sigma_0^2 \leftarrow \omega/(1-\alpha-\beta)$
for $i = 1$ to n_sim **do**
    $\sigma^2 \leftarrow \sigma_0^2$
    $variance\_R_H \leftarrow 0$
    for $h = 1$ to H **do**
        $\sigma \leftarrow \sqrt{\sigma^2}$
        $r \sim \mathcal{N}(0, \sigma)$
        $variance\_R_H \leftarrow variance\_R_H + r^2$
        $\sigma^2 \leftarrow \omega + \alpha r^2 + \beta \sigma^2$
    end for
    $R_H^2[i] \leftarrow variance\_R_H$
end for
return R_H^2

Algorithm 4 Compute z from returns using GARCH and fractional moments

Require: Parameters: $H \leftarrow 2$, $n_{\text{sim}} \leftarrow$ 100,000, window $\leftarrow$ 520, target_length $\leftarrow$ `nothing`, returns, fractional order γ
Ensure: Computed array `z`

$n \leftarrow$ length(`returns`)
max_index $\leftarrow n - H$
if target_length $=$ `nothing` **then**
    range $\leftarrow$ window : max_index
else
    range $\leftarrow$ (max_index $-$ target_length $+ 1$) : max_index
end if
Initialize empty array z
for each t in `range` **do**
    $r_{\text{window}} \leftarrow$ `returns`$[(t -$ window $+ 1) : t]$
    Fit GARCH(1,1) model to r_{window}
    Extract parameters ω, α, β from model
    $R_H^2 \leftarrow$ simulate_conditional_variance$(\omega, \alpha, \beta, H, n_{\text{sim}})$
    distribution$_R \leftarrow$ kde(R_H^2)
    pdf$_R \leftarrow$ KDEDistribution(distribution$_R$)
    moment $\leftarrow$ fractional_moment(pdf$_R, \gamma$; use_abs $=$ true)
    Append `moment` to `z`
end for
return z

D Appendix: Table of Common Distributions

Distribution	**PDF / PMF**	**MGF** $M_X(t)$	**Restrictions**
Bernoulli(p)	$f(x) = p^x(1-p)^{1-x}$	$M_X(t) = (1-p) + pe^t$	$0 \leq p \leq 1,\ x \in \{0,1\}$
Binomial(n,p)	$f(x) = \binom{n}{x} p^x(1-p)^{n-x}$	$M_X(t) = (1-p+pe^t)^n$	$n \in \mathbb{N},\ 0 \leq p \leq 1$
Poisson(λ)	$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$	$M_X(t) = \exp(\lambda(e^t - 1))$	$\lambda > 0,\ x \geq 0$
Exponential(β)	$f(x) = \frac{1}{\beta} e^{-x/\beta}$	$M_X(t) = \frac{1}{1-\beta t},\ t < 1/\beta$	$\beta > 0,\ x > 0$
Gamma(α, β)	$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$	$M_X(t) = (1-\beta t)^{-\alpha},\ t < 1/\beta$	$\alpha, \beta > 0,\ x > 0$
$\mathcal{N}(\mu, \sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$	$\sigma > 0$

Table 11: Common distributions and their moment generating functions

References

Andersen, T. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*(4), 885–905.

Askey, R. (1975). *Orthogonal Polynomials and Special Functions*. Society for Industrial and Applied Mathematics.

Atici, F. M., S. Chang, and J. Jonnalagadda (2021). Grünwald-letnikov fractional operators: from past to present. *Fractional Differential Calculus 11*(1), 147–159.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Boulaaras, S., R. Jan, and V.-T. Pham (2023). Recent advancement of fractional calculus and its applications in physical systems. *The European Physical Journal Special Topics 232*(14), 2347–2350.

Caputo, M. (1967, 11). Linear models of dissipation whose q is almost frequency independent—ii. *Geophysical Journal International 13*(5), 529–539.

Caputo, M. and M. Fabrizio (2015, 04). A new definition of fractional derivative without singular kernel. *Prog Fract Differ Appl 1*, 73–85.

Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

Cauchy, A.-L. (1823). *Summary of the lessons given at the Royal Polytechnic School on infinitesimal calculus, Volume I.* GTextes1. Paris: Impr. royale, Debure frères. Contains a table of contents, searchable online.

Clark, T. G., M. J. Bradburn, S. B. Love, and D. G. Altman (2003). Survival analysis part i: Basic concepts and first analyses. *British Journal of Cancer 89*(2), 232–238.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Cressie, N., A. S. Davis, J. L. Folks, and G. E. Policello (1981). The moment-generating function and negative integer moments. *The American Statistician 35*(3), 148–150.

Di Matteo, A., M. Di Paola, and A. Pirrotta (2014). Probabilistic characterization of nonlinear systems under poisson white noise via complex fractional moments. *Nonlinear Dynamics 77*(3), 729–738.

D'Amico, M., G. Fusai, and A. Tagliani (2002, May). Valuation of exotic options using moments. *Operational Research 2*(2), 157–186.

Einstein, A. (1905). On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Annalen der Physik*, 549–560.

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons. Volumes I–II published between 1957 and 1971.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns. *The American Economic Review 67*(2), 116–126.

Flajolet, P. and R. Sedgewick (1995). Mellin transforms and asymptotics: Finite differences and rice's integrals. *Theoretical Computer Science 144*(1), 101–124.

Flajolet, P. and R. Sedgewick (1996). The Average Case Analysis of Algorithms: Mellin Transform Asymptotics. Research Report RR-2956, INRIA.

Flanders, H. (1973). Differentiation under the integral sign. *The American Mathematical Monthly 80*(6), 615–627.

Fourier, J.-B.-J. (1822). *Théorie analytique de la chaleur*. Paris: F. Didot. 2 unnumbered leaves of plates: illustrations; 26 cm.

Gourieroux, C. and J. Jasiak (2006, 03). Autoregressive gamma process. *Journal of Forecasting 25*, 129–152.

Gzyl, H. and S. Mayoral (2024). Determining the lifetime distribution using fractional moments with maximum entropy. *Heliyon 10*(15), e35250.

Gzyl, H., P.-L. Novi-Inverardi, and A. Tagliani (2013). Determination of the probability of ultimate ruin by maximum entropy applied to fractional moments. *Insurance: Mathematics and Economics 53*(2), 457–463.

Hadamard, J. (1892). Essai sur l'étude des fonctions données par leur développement de taylor. *Journal de Mathématiques Pures et Appliquées 8*, 101–186.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, P. and C. Tong (2024, 10). Fractional moments by the moment-generating function.

Hendricks, W. A. and K. W. Robey (1936). The sampling distribution of the coefficient of variation. *The Annals of Mathematical Statistics 7*(3), 129–132.

Holton, G. A. (2013). *Value-at-Risk: Theory and Practice* (Second ed.). Self-published. Published online.

Joseph Liouville, École polytechnique (Palaiseau, E. (1832). Journal de l'École polytechnique. Accessed: 2025-03-30.

Katugampola, U. N. (2014). A new approach to generalized fractional derivatives.

Khuri, A. and G. Casella (2002). The existence of the first negative moment revisited. *The American Statistician 56*(1), 44–47.

Kilbas, A. A., H. M. Srivastava, and J. J. Trujillo (2006). Theory and applications of fractional differential equations. In *Theory and Applications of Fractional Differential Equations*.

Kolmogorov, A. N. and S. V. Fomin (1999). *Elements of the Theory of Functions and Functional Analysis*. Dover Publications. Translated by Leo F. Boron.

Koning, D. (2015). *Fractional Calculus*. Ph. D. thesis, University of Groningen.

Laskin, N. (2002, Nov). Fractional schrödinger equation. *Phys. Rev. E 66*, 056108.

Lin, G. D. (1992). Characterizations of distributions via moments.

Losada, J. and J. Nieto (2015, 04). Properties of a new fractional derivative without singular kernel. *Prog Fract Differ Appl 1*, 87–92.

Love, E. R. (1971). Fractional derivatives of imaginary order. *Journal of the London Mathematical Society s2-3*(2), 241–259.

Lowther, G. (2009). Filtrations and adapted processes. `https://almostsuremath.com/2009/11/08/filtrations-and-adapted-processes/`. Accessed: 2025-06-04.

Mandelbrot, B. B. and J. W. Van Ness (1968). Fractional brownian motions, fractional noises and applications. *SIAM Review 10*(4), 422–437.

Marchaud, A. (1927). Sur les dérivées et sur les différences des fonctions de variables réelles. *Numdam Thèses de l'entre-deux-guerres 78*, 98.

Meurer, A., C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. L. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, and F. Johansson (2017). Sympy: symbolic computing in python. *PeerJ Computer Science 3*, e103.

Mikosch, T., G. Samorodnitsky, and L. Tafakori (2013). Fractional moments of solutions to stochastic recurrence equations. *Journal of Applied Probability 50*(4), 969–982.

Missbauer, A. (2012). Gabor frames and the fractional fourier transform.

Novi Inverardi, P. L. and A. Tagliani (2024). Probability distributions approximation via fractional moments and maximum entropy: Theoretical and computational aspects. *Axioms 13*(1).

Oldham, K. B. and J. Spanier (1974). *The Fractional Calculus: Theory and Applications of Differentiation and Integration to Arbitrary Order*, Volume 111 of *Mathematics in Science and Engineering*. New York and London: Academic Press.

Poon, S.-H. and C. W. Granger (2003, June). Forecasting volatility in financial markets: A review. *Journal of Economic Literature 41*(2), 478–539.

Riesz, M. (1949). L'intégrale de riemann-liouville et le problème de cauchy. *Acta Mathematica 81*(1), 1–222.

Samko, S. G., A. A. Kilbas, and O. I. Marichev (1993). *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach Science Publishers.

Scalas, E., R. Gorenflo, and F. Mainardi (2000). Fractional calculus and continuous-time finance. *Physica A: Statistical Mechanics and its Applications 284*(1), 376–384.

Schulten, K. (2005). Phys 550: Non-equilibrium statistical mechanics, lecture notes. `https://www.ks.uiuc.edu/Services/Class/PHYS550/LectureNotes.html`. Spring 2005, University of Illinois at Urbana-Champaign.

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management 21*(1), 49–58.

Sorino, F. A. (2001). Appendix - the forsey-sortino model tutorial. In F. A. Sortino and S. E. Satchell (Eds.), *Managing Downside Risk in Financial Markets*, Quantitative Finance, pp. 245–252. Oxford: Butterworth-Heinemann.

Sortino, F. A. and R. van der Meer (1991). Downside risk. *The Journal of Portfolio Management 17*(4), 27–31.

Tarasov, V. E. (2023). Fractional probability theory of arbitrary order. *Fractal and Fractional 7*(2).

Taylor, S. J. (2007). *Modelling Financial Time Series* (2nd ed.). WORLD SCIENTIFIC.

Thiagarajan, S. (2022). Theory of laplace transforms and their applications. *University of Chicago*. Definition, examples, and applications of the Laplace transformation.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). John Wiley & Sons.

Wang, Y., R. Tani, and K. Uchida (2025). A numerical solution method for the fractional moment problem within engineering. *Engineering Computations*.

Wharton Research Data Services (WRDS) (2025). S&p 500 index data. `https://wrds-www.wharton.upenn.edu/`. Accessed via WRDS. Data provided by Standard & Poor's.

Wojt, A. (2009). Portfolio selection and lower partial moments. Accessed: 2025-06-06.

Zhmakin, A. (2022, 12). A compact introduction to fractional calculus.