

Article

Are There Dragon Kings in the Stock Market?

Jiong Liu, Mohammadamin Dashti Moghaddam and Rostislav A. Serota * 

Department of Physics, University of Cincinnati, Cincinnati, OH 45221-0011, USA

* Correspondence: serota@ucmail.uc.edu

Abstract: In this study, we undertake a systematic study of historic market volatility spanning roughly five preceding decades. We focus specifically on the time series of the realized volatility (RV) of the S&P500 index and its distribution function. As expected, the largest values of RV coincide with the largest economic upheavals of the period: Savings and Loan Crisis, Tech Bubble, Financial Crisis and Covid Pandemic. We address the question of whether these values belong to one of the three categories: Black Swans (BS), that is, they lie on scale-free, power-law tails of the distribution; Dragon Kings (DK), defined as statistically significant upward deviations from BS; or Negative Dragon Kings (nDK), defined as statistically significant downward deviations from BS. In analyzing the tails of the distribution with $RV > 40$, we observe the appearance of “potential” DK, which eventually terminate in an abrupt plunge to nDK. This phenomenon becomes more pronounced with the increase in the number of days over which the average RV is calculated—here from daily, $n = 1$, to “monthly”, $n = 21$. We fit the entire distribution with a modified Generalized Beta (mGB) distribution function, which terminates at a finite value of the variable but exhibits a long power-law stretch prior to that, as well as a Generalized Beta Prime (GB2) distribution function, which has a power-law tail. We also fit the tails directly with a straight line on a log-log scale. In order to ascertain BS, DK or nDK behavior, all fits include their confidence intervals and p -values are evaluated for the data points to check whether they can come from the respective distributions.

Keywords: black swans; dragon kings; negative dragon kings; generalized beta distribution; volatility



Citation: Liu, J.; Dashti Moghaddam, M.; Serota, R.A. Are There Dragon Kings in the Stock Market? *Foundations* **2024**, *4*, 91–113. <https://doi.org/10.3390/foundations4010008>

Academic Editor: Ricardo Lopez-Ruiz

Received: 2 December 2023

Revised: 18 January 2024

Accepted: 1 February 2024

Published: 8 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Distributions with power-law tails are ubiquitous across multitudes of disciplines [1–3]. Power-law tails are scale-free and as such can meaningfully extend indefinitely, to arbitrarily large values of the variable. On a log-log scale, a power-law tail presents itself as a straight line with a negative slope. Such tails are very difficult to pinpoint since the data become very sparse at large values of the variable. A far more difficult task than that is to be able to locate outliers, should they exist, that deviate substantially from the power-law behavior.

Detection of outliers in itself is a challenging task, regardless of the distribution, and requires various statistical tests [4–6]. One such type of outliers, where the data deviate strongly upward from the above-mentioned straight line, is dubbed “Dragon Kings” and is also encountered across multitudes of disciplines and phenomena [7–10]. In financial markets, power-law tails can lead to what theoretically could be arbitrarily large gains or losses—“Black Swans” in the latter case—while Dragon Kings could indicate even far more distressful events.

While there are some indications of outlier behavior in financial markets [7,11–14], the main goal of this paper is to investigate the realized volatility—a market quantity which is commonly used by market practitioners and economists alike—for the presence of any potential outliers. This is motivated by the fact that intuitively one would suspect that the most catastrophic market downturns and spikes in volatility—Black Monday, Tech Bubble, Financial Crisis and COVID Pandemic—would be plausible candidates for Dragon Kings.

Realized volatility RV is the square root of realized variance, which is defined as follows

$$RV^2 = 100^2 \times \frac{252}{n} \sum_{i=1}^n r_i^2 \tag{1}$$

where

$$\overline{r_n^2} = \frac{1}{n} \sum_{i=1}^n r_i^2 \tag{2}$$

is the average realized variance over n days and

$$r_i = \ln \frac{S_i}{S_{i-1}} \tag{3}$$

are the daily returns, with S_i being the reference (closing) price on day i . This is an annualized value, where 252 represents the number of trading days in a year. In particular, $n = 1$ represent daily returns and $n = 21$, being a typical number of trading days in a month, is useful for evaluating the monthly RV. We point out, however, that in our calculation, n is simply a number of consecutive trading days that can fall on different weeks and months. Specifically, we performed our analysis for $n = 1, 2, 3, 5, 7, 9, 13, 17, 21$. Here, we present results for $n = 1, 7, 21$, which already succinctly illustrate the changes in the RV distribution with n .

Since it is based on actual trades, realized volatility (RV) is the ultimate measure of market volatility, although the latter is more often associated with the implied volatility, most commonly measured by the VIX index [15,16]—the so-called market “fear index”—that tries to predict RV of the S&P500 index for the following month. Its model-independent evaluation [17] is based on option contracts, which are meant to predict future stock prices fluctuations [18]. The question of how well VIX predicts the future realized volatility has been of great interest to researchers [19–22]. Recent results [23,24] show that VIX is only marginally better than past RV in predicting future RV. In particular, it underestimates future low volatility and, most importantly, future high volatility. In fact, while both RV and VIX exhibit scale-free power-law tails, the distribution of the ratio of RV to VIX also has a power-law tail with a relatively small power exponent [23,24], meaning that VIX is incapable of predicting large surges in volatility.

It should be emphasized that RV is agnostic with respect to gains or losses in stock returns. Nonetheless, it has been habitual that large gains and losses occur at around the same time. Here, we wish to address the question of whether the largest values of RV fall on the power-law tail of the RV distribution. As is well known, the largest upheavals in the stock market happened on, and close to, Black Monday, which was a precursor to the Savings and Loan crisis, Tech Bubble, Financial Crisis and COVID-19 pandemic. Plotted on a log-log scale, power-law tails of a distribution show as a straight line. If the largest RV fall on the straight line, they can be classified as Black Swans (BS). If, however, they show statistically significant deviations upward or downward from this straight line, they can be classified as Dragon Kings (DK) [7,8] or negative Dragon Kings (nDK), respectively [5].

The main result of this paper is that the largest values of RV are in fact nDK. We find that daily returns are the closest to the BS behavior. However, with the increase in n , we observe the development of “potential” DK with statistically significant deviations upward from the straight line. This trend terminates with the data points returning to the straight line and then abruptly plunging into nDK territory.

To gain further insight into this phenomenon, we start in Section 2, with the time series of RV from 1970 to 2021, including expanded views of the aforementioned periods of market upheavals. In Section 3 we provide analytical expressions of the two distribution functions used to fit the entire RV distribution: modified Generalized Beta (mGB), which is discussed in great detail in a companion paper [25], and Generalized Beta Prime (GB2), which is essentially a limiting case of mGB and is chosen because it has power-law tails. mGB is chosen because it exhibits a long stretch of power-law dependence before dropping off and terminating at a finite value of the variable, thus mimicking the nDK behavior of RV [25]. Additionally, both mGB and GB2 emerge as steady-state distributions of a

stochastic differential equation for stochastic volatility [25]. In Section 4, we describe fits of RV with mGB and GB2 and provide a detailed description of the tails, specifically in regards to possible DK/nDK. Towards this end, we also use a linear fit (LF) of the tails. For all three fits, we provide confidence intervals [4] and, more importantly, the results of a U-test [5], which evaluates a p -value for the null hypothesis that a data point comes from a fitting distribution [5]. Section 5 is a discussion of results obtained in Sections 4 and 6 summarizes our findings and touches upon possible extensions of this work.

2. Time Series of Realized Volatility

Figure 1 shows the time series of RV for $n = 1$, based on daily returns, $n = 7$, and $n = 21$, where n is the number of days over which the daily RV is averaged in (1) and (2). Only values with $RV > 17$ are shown, and black dots mark values $RV > 10^{1.75} \approx 56$. It is clear that the time series progression with the increase in n is towards a more pronounced amplification of singularly important events, such as periods corresponding to Black Monday, the Financial Crisis and the COVID-19 pandemic—even though the maximum values of RV understandably decrease in the same progression, as the averaging is taken over a larger number of days n . While such progression naturally leads to a question of whether these events might belong to the DK category, we shall see in what follows that they are actually nDK.

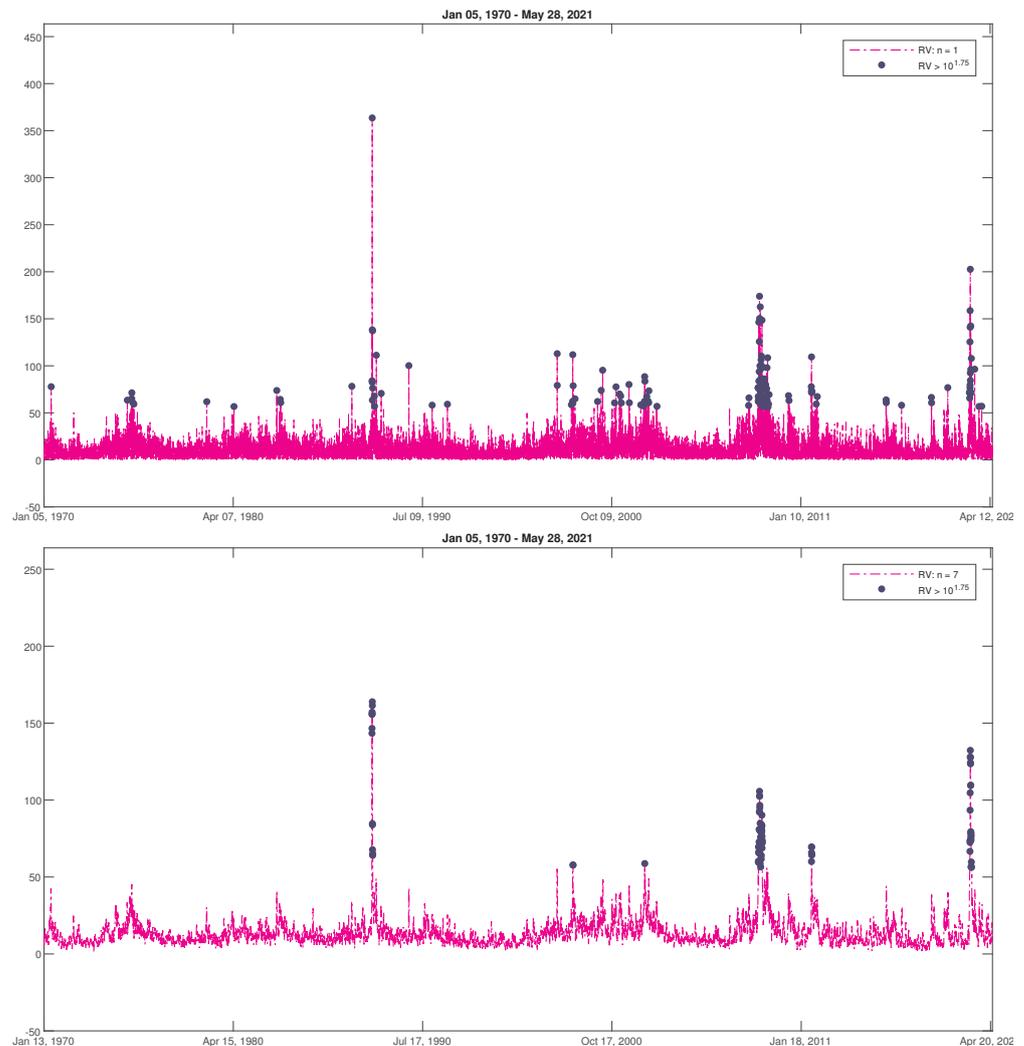


Figure 1. Cont.

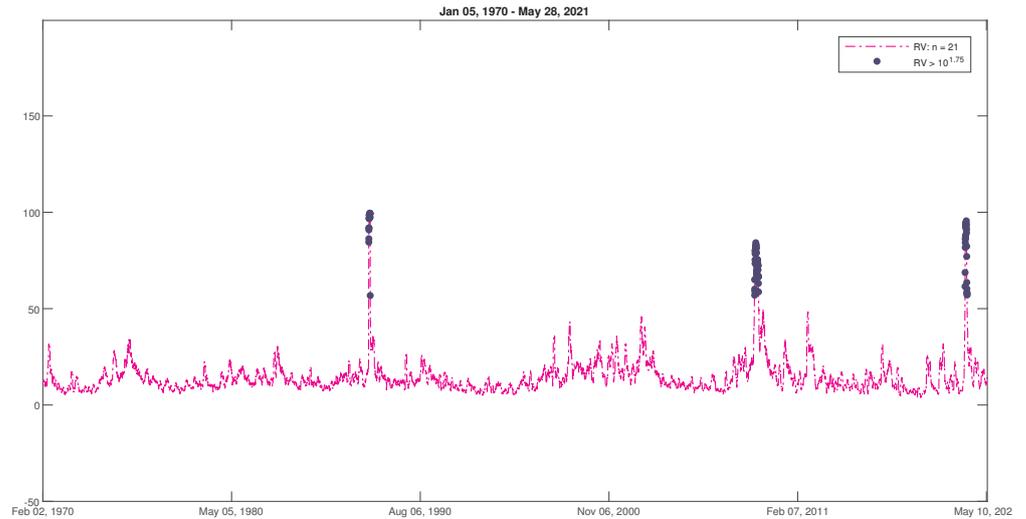


Figure 1. Time series of RV for S&P500, with $RV > 17$ shown. From top to bottom $n = 1, 7, 21$. Black dots indicate $RV > 10^{1.75} \approx 56$.

Figures 2 and 3 give snapshots of the time series in Figure 1 around the largest volatility events: Figure 2 based on daily returns, $n = 1$, and Figure 3 for $n = 7$ and $n = 21$, respectively. Based on Figures 1–3, Black Monday was clearly the most singular volatility event, while Financial Crisis and COVID-19 pandemic were distinguished by more prolonged periods of sustained extraordinarily large RV.

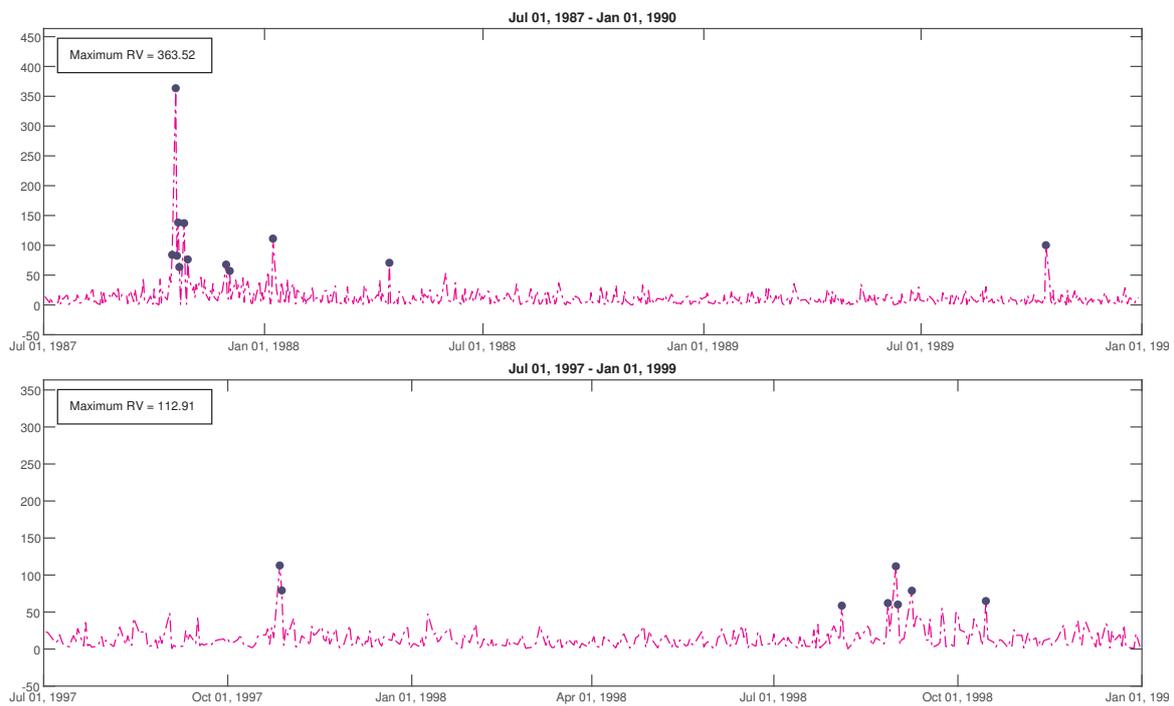


Figure 2. Cont.

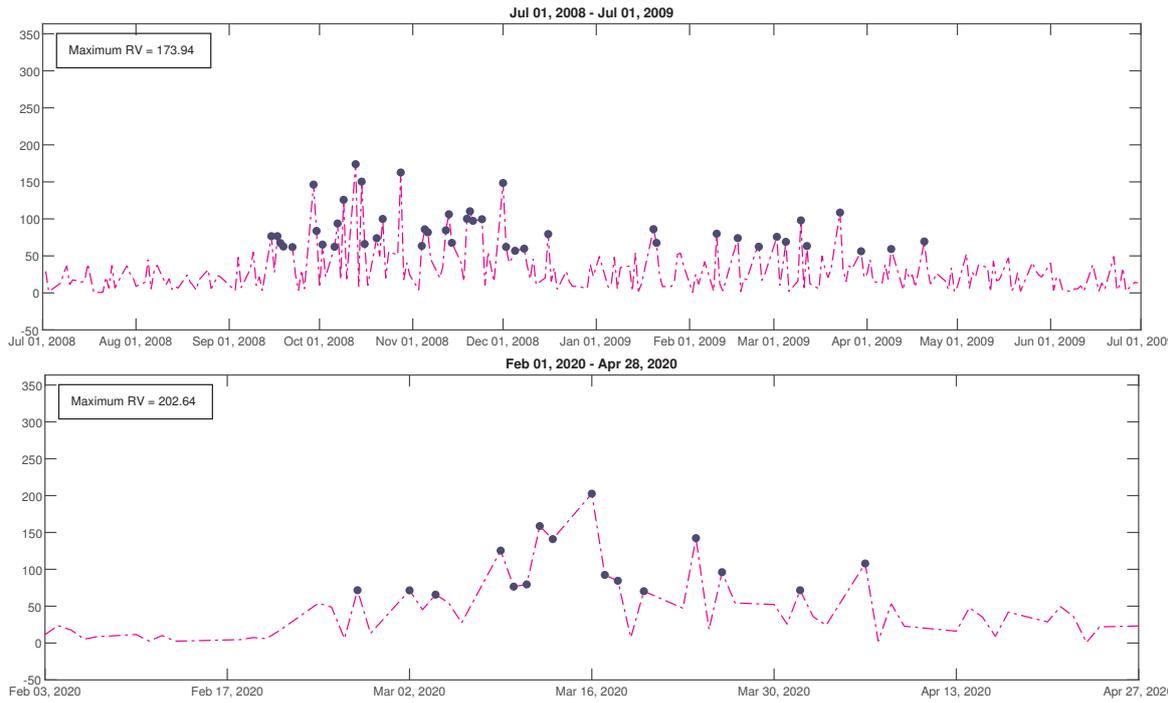


Figure 2. Snapshots of $n = 1$ for S&P500 RV time series in Figure 1 around Black Monday, Tech Bubble, Financial Crisis and COVID-19 pandemic.

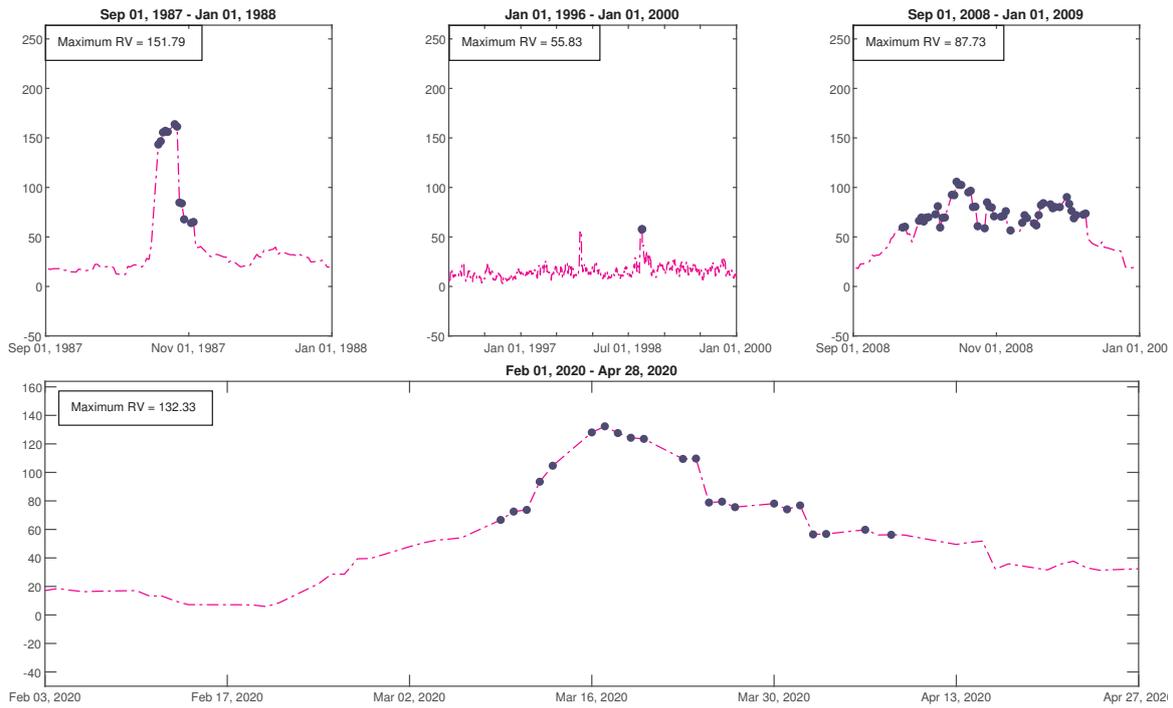


Figure 3. Cont.

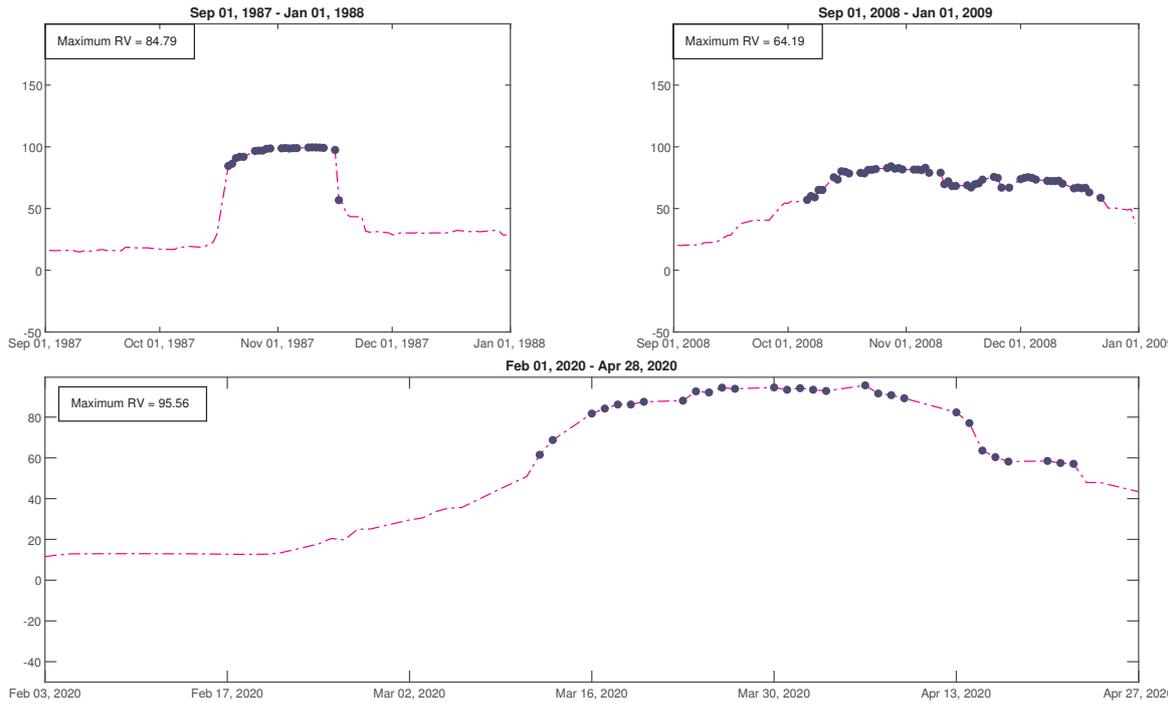


Figure 3. Snapshots of $n = 7$ (top two) and $n = 21$ (bottom two) for S&P500 RV time series in Figure 1 for the same periods as in Figure 2.

3. Generalized Beta Distribution Function

A companion paper [25] discusses in great detail the modified Generalized Beta distribution function (mGB) used here to fit the distributions of RV. A generalization of the traditional [26] GB can be written, in a slightly modified form relative to that of [25], as follows:

$$f_{GB}(x; \alpha, \beta_1, \beta_2, p, q) = \frac{\alpha \left(1 + \left(\frac{\beta_2}{\beta_1}\right)^\alpha\right)^p \left(\frac{x}{\beta_2}\right)^{\alpha p - 1} \left(1 + \left(\frac{x}{\beta_2}\right)^\alpha\right)^{-p - q} \left(1 - \left(\frac{x}{\beta_1}\right)^\alpha\right)^{q - 1}}{\beta_2 B(p, q)}, \tag{4}$$

where β_1 and β_2 are scale parameters and α , p and q are shape parameters, all positive; $B(p, q)$ is the beta function and $x \leq \beta_1$. Although it has a concise and transparent form, it does not come out as a solution of a stochastic differential equation (SDE) [27–29], which is desirable for the purpose of modeling behavior of quantities, such as stochastic volatility, which is important for the understanding of RV [30].

The probability density function (PDF) of mGB, which comes out as a solution of an SDE (with minor caveats explained in [25]) and which is used here to model the RV distribution, can be written as

$$f_{mGB}(x; \alpha, \beta_1, \beta_2, p, q) = \frac{\alpha(p + q) \left(1 + \left(\frac{\beta_2}{\beta_1}\right)^\alpha\right)^{p + 1} \left(\frac{x}{\beta_2}\right)^{\alpha p - 1} \left(1 + \left(\frac{x}{\beta_2}\right)^\alpha\right)^{-p - q - 1} \left(1 - \left(\frac{x}{\beta_1}\right)^\alpha\right)^{q - 1}}{\beta_2 B(p, q) \left(q + \left(\frac{\beta_2}{\beta_1}\right)^\alpha (p + q)\right)}, \tag{5}$$

The cumulative distribution function (CDF) and complementary CDF (CCDF) of mGB are given, respectively, by

$$F_{mGB}(x; \alpha, \beta_1, \beta_2, p, q) = I\left(\frac{\left(\frac{x}{\beta_1}\right)^\alpha + \left(\frac{x}{\beta_2}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}; p, q\right) + \frac{1}{B(p, q) \left(q + \left(\frac{\beta_2}{\beta_1}\right)^\alpha (p + q)\right)} \left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}\right)^q \left(\frac{\left(1 + \left(\frac{\beta_2}{\beta_1}\right)^\alpha\right) \left(\frac{x}{\beta_2}\right)^\alpha}{\left(1 + \left(\frac{x}{\beta_2}\right)^\alpha\right)}\right)^p, \tag{6}$$

and

$$1 - F_{mGB}(x; \alpha, \beta_1, \beta_2, p, q) = I\left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}; q, p\right) - \frac{1}{B(p, q) \left(q + \left(\frac{\beta_2}{\beta_1}\right)^\alpha (p + q)\right)} \left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}\right)^q \left(\frac{\left(1 + \left(\frac{\beta_2}{\beta_1}\right)^\alpha\right) \left(\frac{x}{\beta_2}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}\right)^p, \tag{7}$$

where the first term in (6) and (7) represent, respectively, CDF and CCDF of GB (whose PDF is given by (4)), while $I(y; p, q) = B(y; p, q) / B(p, q)$ and $B(y; p, q)$ are, respectively, the regularized and incomplete beta functions [31].

In what follows, we will be specifically interested in the $\beta_2 \ll \beta_1$ circumstance, since, for $\beta_2 \ll x \ll \beta_1$, GB and mGB exhibit a power-law dependence,

$$f_{mGB} \propto \left(\frac{x}{\beta_2}\right)^{-\alpha(q+1)-1}, \quad 1 - F_{mGB} \propto \left(\frac{x}{\beta_2}\right)^{-\alpha(q+1)}; \quad f_{GB} \propto \left(\frac{x}{\beta_2}\right)^{-\alpha q - 1}, \quad 1 - F_{GB} \propto \left(\frac{x}{\beta_2}\right)^{-\alpha q}. \tag{8}$$

In the limit of $\beta_1 \rightarrow \infty$, mGB and GB become, respectively, mGB2 and GB2 (the latter is also known as Generalized Beta Prime) and are given by [25]

$$f_{mGB2}(x; \alpha, \beta_2, p, q) = \frac{\alpha(p + q) \left(\frac{x}{\beta_2}\right)^{\alpha p - 1} \left(1 + \left(\frac{x}{\beta_2}\right)^\alpha\right)^{-p - q - 1}}{q \beta_2 B(p, q)}, \tag{9}$$

and

$$f_{GB2}(x; \alpha, \beta_1, \beta_2, p, q) = \frac{\alpha \left(\frac{x}{\beta_2}\right)^{\alpha p - 1} \left(1 + \left(\frac{x}{\beta_2}\right)^\alpha\right)^{-p - q}}{\beta_2 B(p, q)}. \tag{10}$$

Unlike mGB and GB, for whom the power-law dependences in (8) eventually terminate at β_1 , mGB2 and GB2 will sustain these power-law dependences indefinitely.

Below, we will use (7) to fit CCDF of distributions of RV. As explained in [25], mGB2 and GB2 are equivalent, since q and p are independently defined at this level of the GB family of distributions and q can be shifted by unity in the definition of mGB2/GB2. Consequently, we choose a more familiar CCDF of GB2

$$1 - F_{GB2}(x; \alpha, \beta_1, \beta_2, p, q) = I\left(\frac{1}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}; q, p\right) \tag{11}$$

to fit the CCDF of the RV data. Insofar as the main difference between mGB and GB is concerned, it is their behavior near β_1 in the present context [25]. Namely,

$$1 - F_{GB} \approx \frac{1}{q B(p, q)} \left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}\right)^q \approx \frac{1}{q B(p, q)} \left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{\beta_1}{\beta_2}\right)^\alpha}\right)^q, \tag{12}$$

and

$$1 - F_{mGB} \approx \frac{1 + \frac{p}{q}}{q B(p, q)} \left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{x}{\beta_2}\right)^\alpha}\right)^q \left(\frac{\beta_2}{\beta_1}\right)^\alpha \approx \frac{1 + \frac{p}{q}}{q B(p, q)} \left(\frac{1 - \left(\frac{x}{\beta_1}\right)^\alpha}{1 + \left(\frac{\beta_1}{\beta_2}\right)^\alpha}\right)^q \left(\frac{\beta_2}{\beta_1}\right)^\alpha, \tag{13}$$

that is, $1 - F_{mGB}$ drops off to zero (F_{mGB} saturates to unity) faster than $1 - F_{GB}$ due to the factor $\left(\frac{\beta_2}{\beta_1}\right)^\alpha$. This feature accounts for a better fit via mGB versus GB, which may be due to the fact that mGB emerges from a physically motivated stochastic model [25].

4. Fitting Distribution of Realized Volatility

4.1. Methodology

We fit CCDF of the full RV distribution—for the entire time span discussed in Section 2—using mGB (7) and GB2 (11). We use Bayesian fitting since the latter (and

the gradient descent) consistently provide better fits of the sparsely populated tails. The fits are shown on the log-log scale in Figures 4–13, together with the linear fit (LF) of the tails with $RV > 40$. LF excludes the end points, as prescribed in [5], that visually may be nDK candidates. (In order to mimic LF, we also excluded those points in GB2 fits, which has a minimal effect on GB2 fits, including the slope and KS statistic). To make the progression of the fits as a function of n clearer, we included results for $n = 5$ and $n = 17$, in addition to $n = 1, 7, 21$, which we used in Section 2. Confidence intervals (CI) were evaluated per [4], via the inversion of the binomial distribution. p -values were evaluated in the framework of the U-test, which is discussed in [5] and is based on order statistics (a wide variety of tests statistics is discussed in [6]):

$$p(x_{k,m}) = 1 - B(F(x_{k,m}); k, m - k + 1), \tag{14}$$

where $x_{k,m}$ is the k 's member of numbers between 1 and m ordered by increasing the magnitude (RV values in this case), and $F(x_{k,m})$ is the assumed CDF (here, mGB, GB2 and LF). p -values are evaluated in order to test the null hypothesis H_0 : all observations of the sample are generated by the same fitting distribution. The p -value (14) is defined as a probability of exceeding the observed value $x_{k,m}$ under the null hypothesis. If among the p -values there are some small values— ≤ 0.05 here—then those observed values are identified as DK with probability $1 - p$. Conversely, large p -values— ≥ 0.95 here—are identified as nDK with the probability p [5].

4.2. Results

For each number of days n in Equations (1) and (2), from top to bottom, Figures 4–13 are organized as follows:

- Full data CDF fit with mGB and GB2 and LF of the tails;
- Same as above shown for $RV > 40$;
- p -values of all three fits for $RV > 40$, with $p < 0.05$ indicating DK and $p > 0.95$ nDK;
- LF with its CI;
- GB2 fit with its CI;
- mGB fit with its CI.

In the CI plots, upward pointing triangles indicate p -values consistent with DK, while downward pointing triangles indicate p -values consistent with nDK [5]. Figure 14 shows LF for $n = 7, 17, 21$, where the last 10% of the range of values were excluded, that is, values greater than $0.9 \max\{RV\}$, as opposed to excluding points visually, as in Figures 4–13.

Figure 15 shows LF and GB2 slopes and the Kolmogorov–Smirnov (KS) statistic for GB2 and mGB as a function of n ; the horizontal line with the table value of KS statistic for our sample size [32] is shown for guidance only since mGB and GB2 here are distributions with estimated parameters. Tables 1–3 list, respectively, parameters of the mGB and GB fits, their standard errors (SE) and linear slope values of GB2 and LF in Figure 15.

Table 1. All estimated parameters of mGB and GB2 fits of S&P500 RV distribution.

n	Parameters of mGB($\alpha, \beta_1, \beta_2, p, q$)	Parameters of GB2(α, β, p, q)
1	(1.5500, 399.9009, 27.4233, 0.6519, 1.7828)	(1.5289, 27.4156, 0.6545, 2.7780)
5	(2.3708, 200.5519, 10.7210, 1.7255, 0.4456)	(2.3749, 10.7480, 1.7205, 1.4409)
7	(2.4744, 180.8711, 9.7136, 2.1430, 0.3848)	(2.4681, 9.7169, 2.1371, 1.3796)
17	(2.3026, 120.110, 6.3561, 6.1121, 0.5403)	(2.2797, 6.3471, 6.0946, 1.5334)
21	(2.4016, 104.9925, 6.3853, 6.3429, 0.5043)	(2.3850, 6.2193, 6.5117, 1.4457)

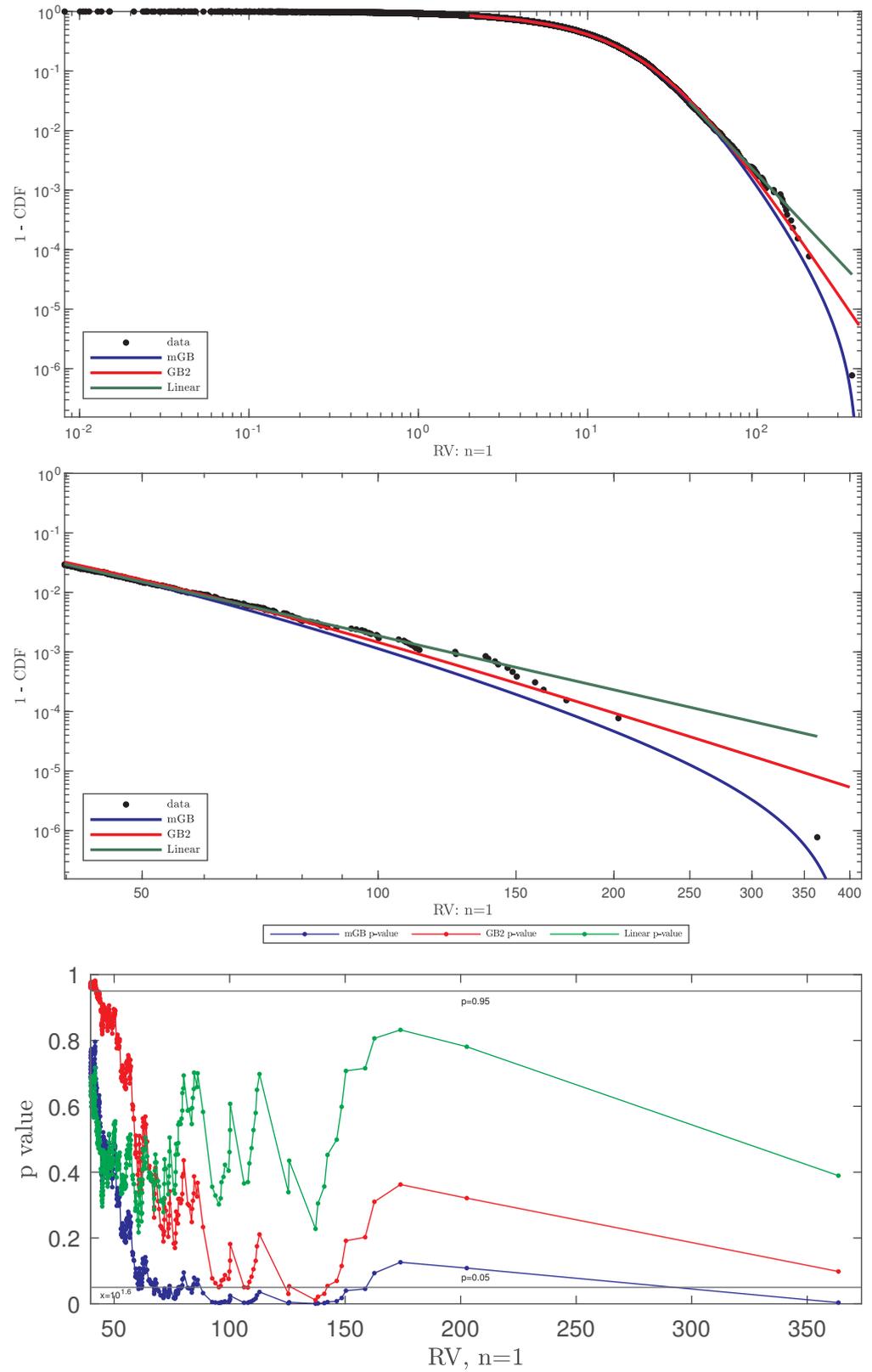


Figure 4. Linear, GB2 and mGB fits of S&P500 RV (top), tail $RV > 40$ (middle), and p -values (bottom) for $n = 1$.

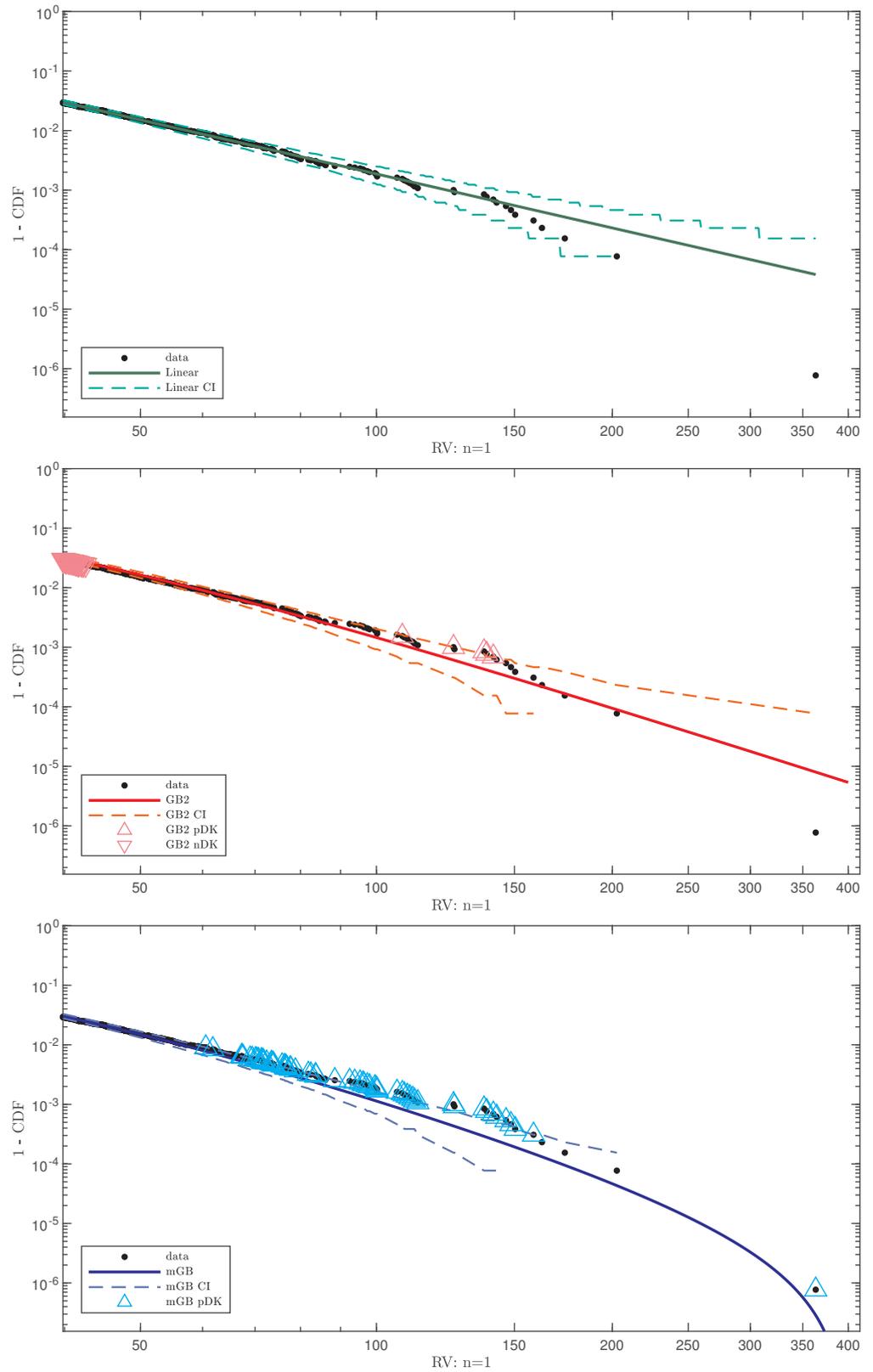


Figure 5. Same as the middle Figure 4 (tail parts of fits of S&P500 RV for $n = 1$) with the respective CI, “potential” DK (up triangles) and nDK (down triangles).

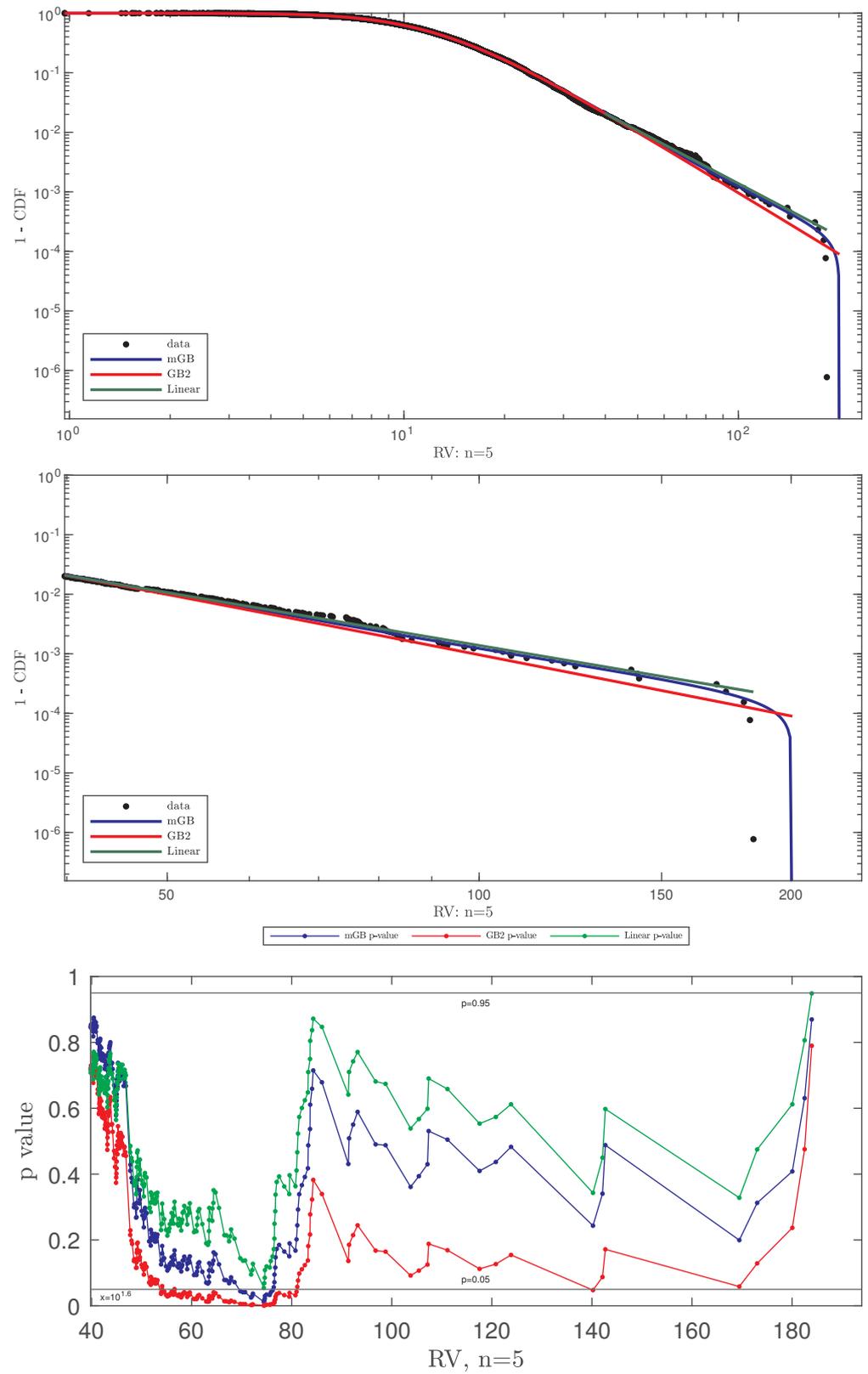


Figure 6. Linear, GB2 and mGB fits of S&P500 RV (top), tail $\text{RV} > 40$ (middle), and p -values (bottom) for $n = 5$.

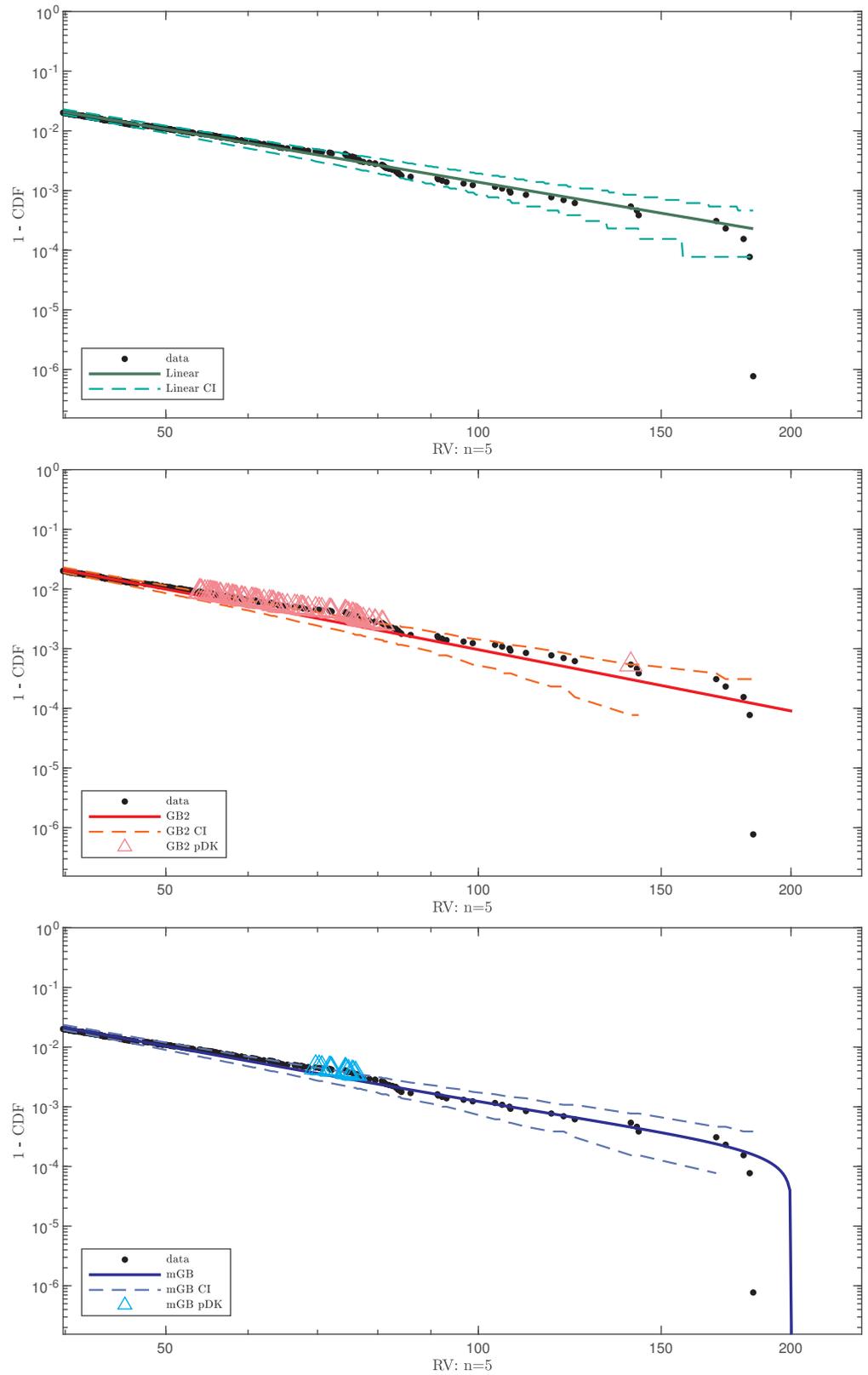


Figure 7. Same as the middle Figure 6 (tail parts of fits of S&P500 RV for $n = 5$) with the respective CI, “potential” DK (up triangles) and nDK (down triangles).

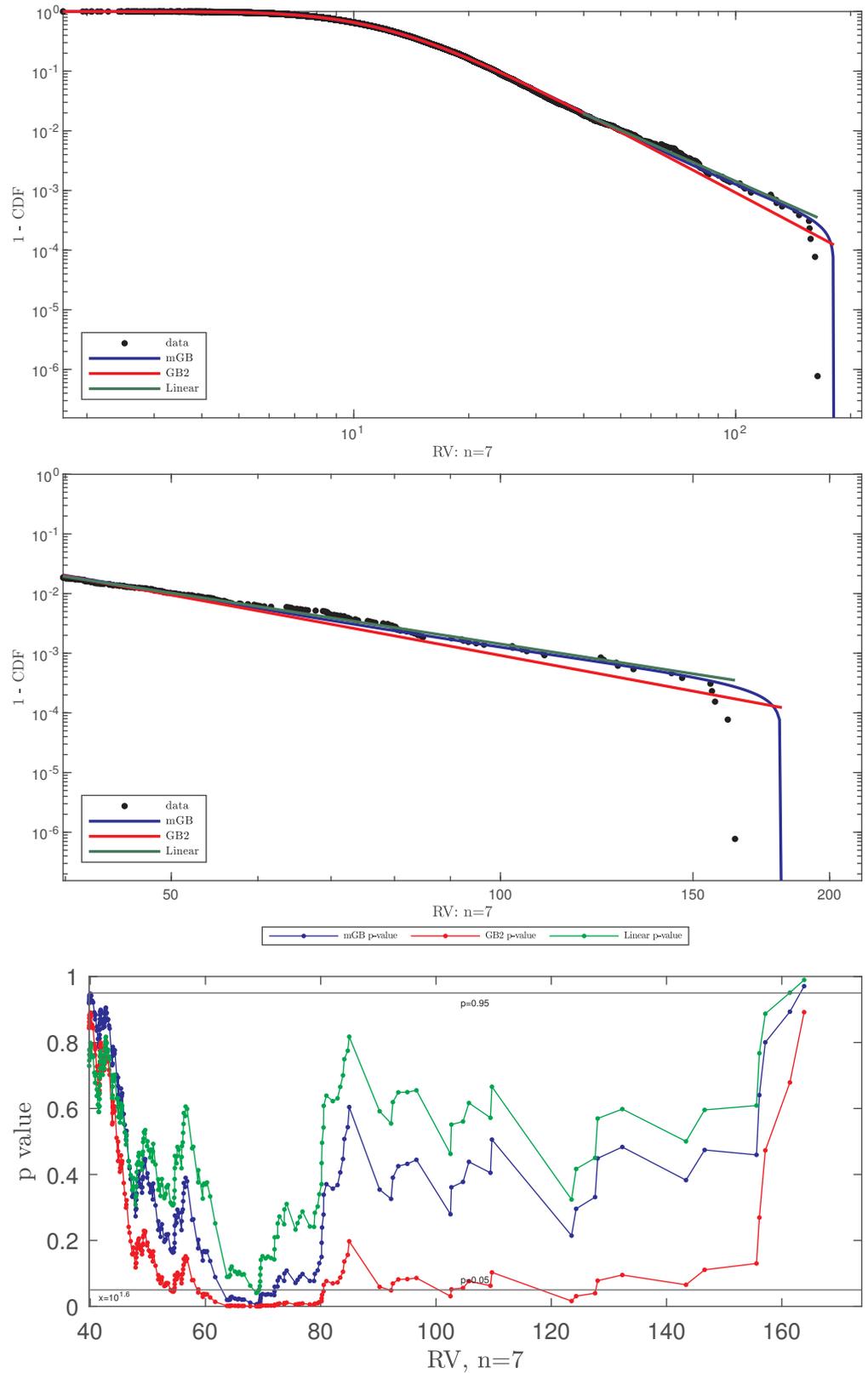


Figure 8. Linear, GB2 and mGB fits of S&P500 RV (top), tail $RV > 40$ (middle), and p -values (bottom) for $n = 7$.

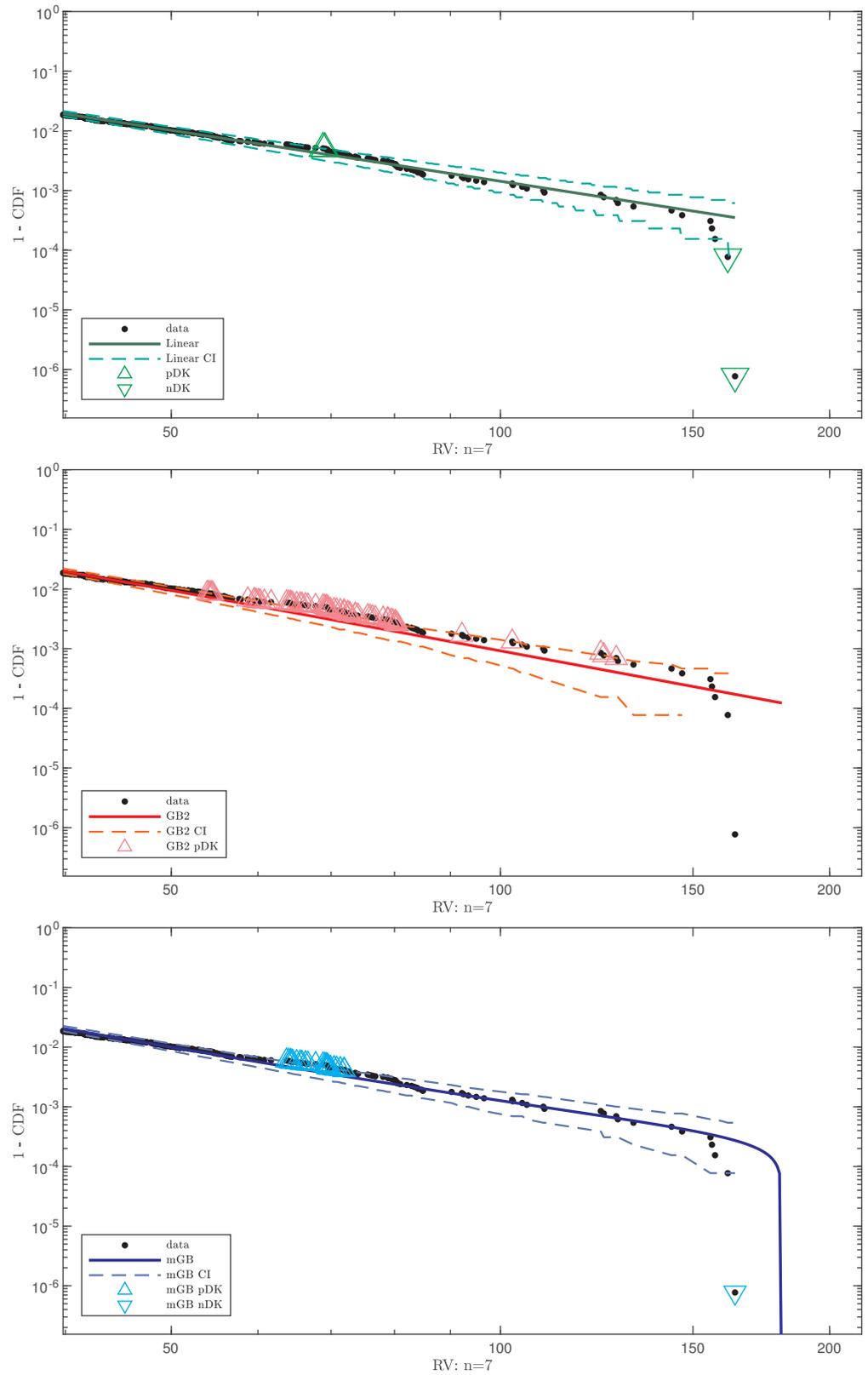


Figure 9. Same as the middle Figure 8 (tail parts of fits of S&P500 RV for $n = 7$) with the respective CI, “potential” DK (up triangles) and nDK (down triangles).

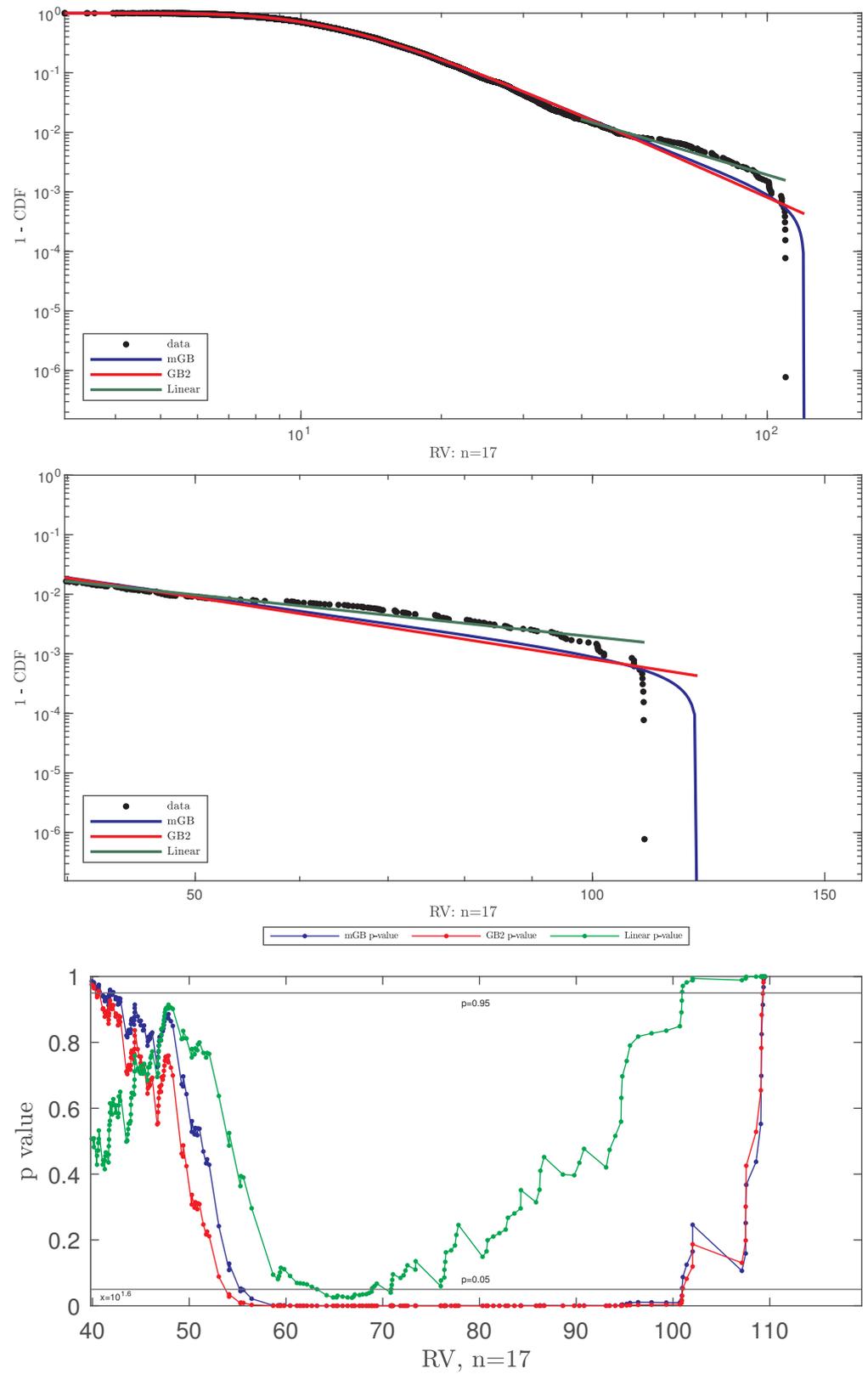


Figure 10. Linear, GB2 and mGB fits of S&P500 RV (top), tail $RV > 40$ (middle), and p -values (bottom) for $n = 17$.

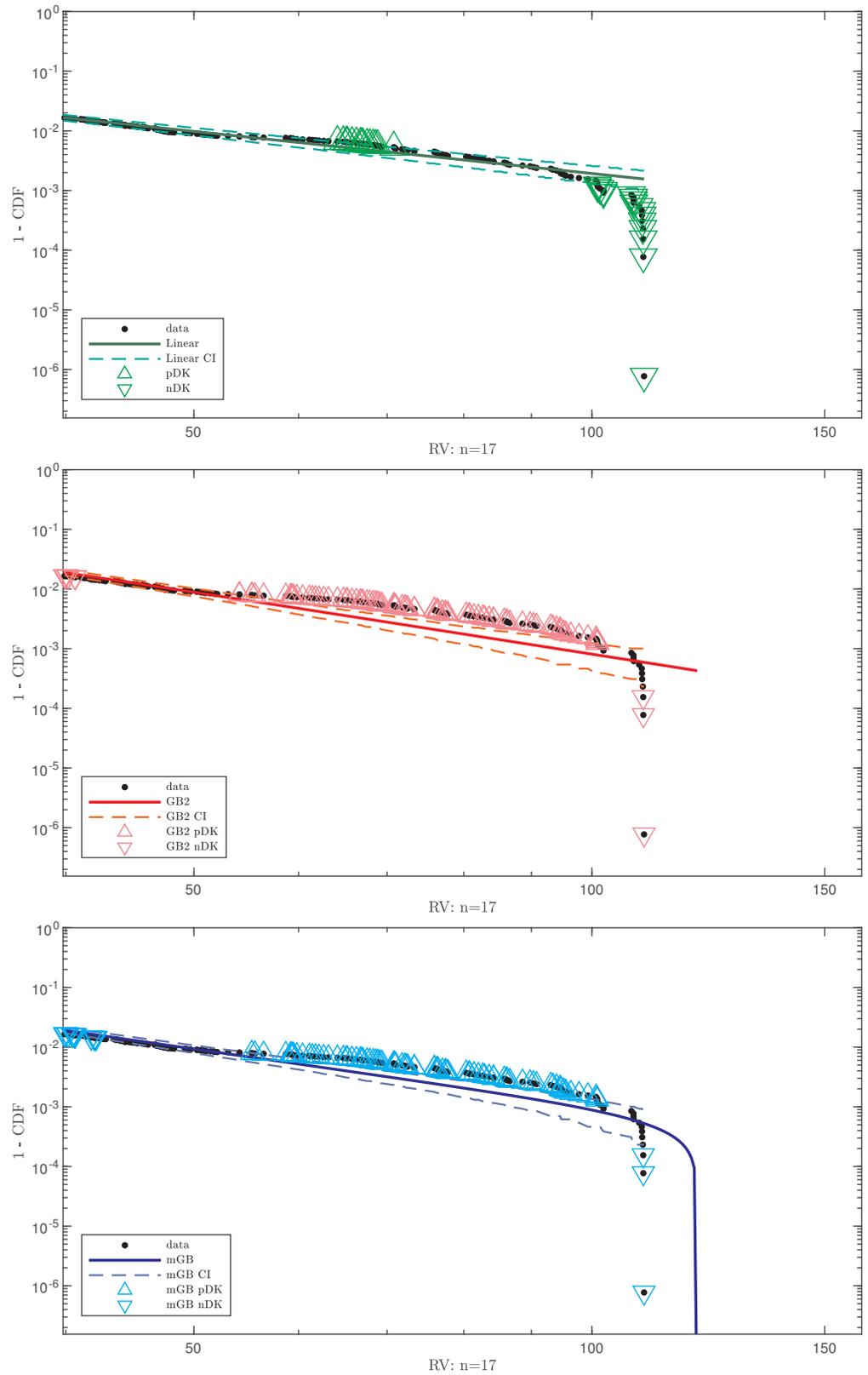


Figure 11. Same as the middle Figure 10 (tail parts of fits of S&P500 RV for $n = 17$) with the respective CI, “potential” DK (up triangles) and nDK (down triangles).

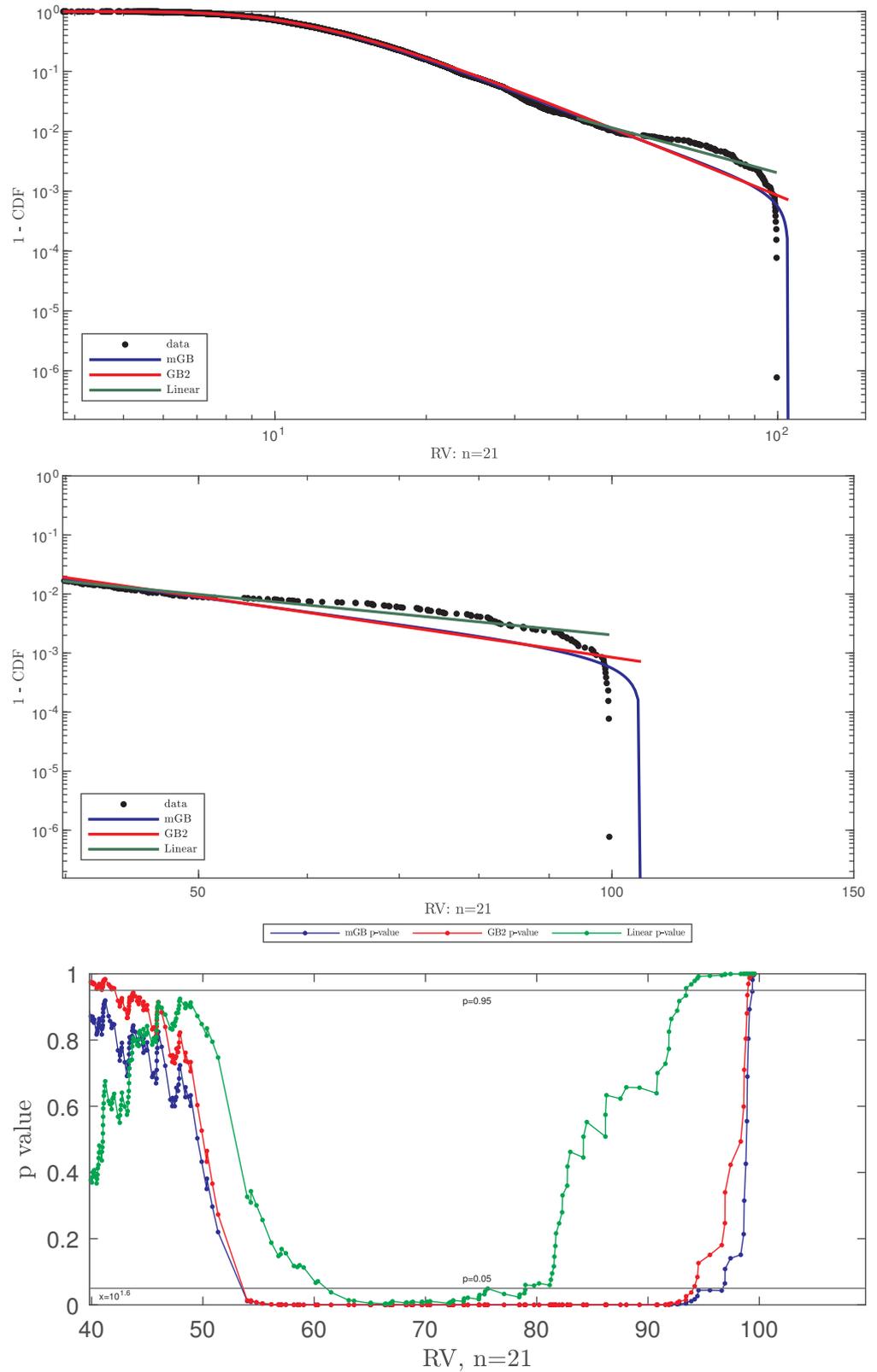


Figure 12. Linear, GB2 and mGB fits of S&P500 RV (top), tail $RV > 40$ (middle), and p -values (bottom) for $n = 21$.

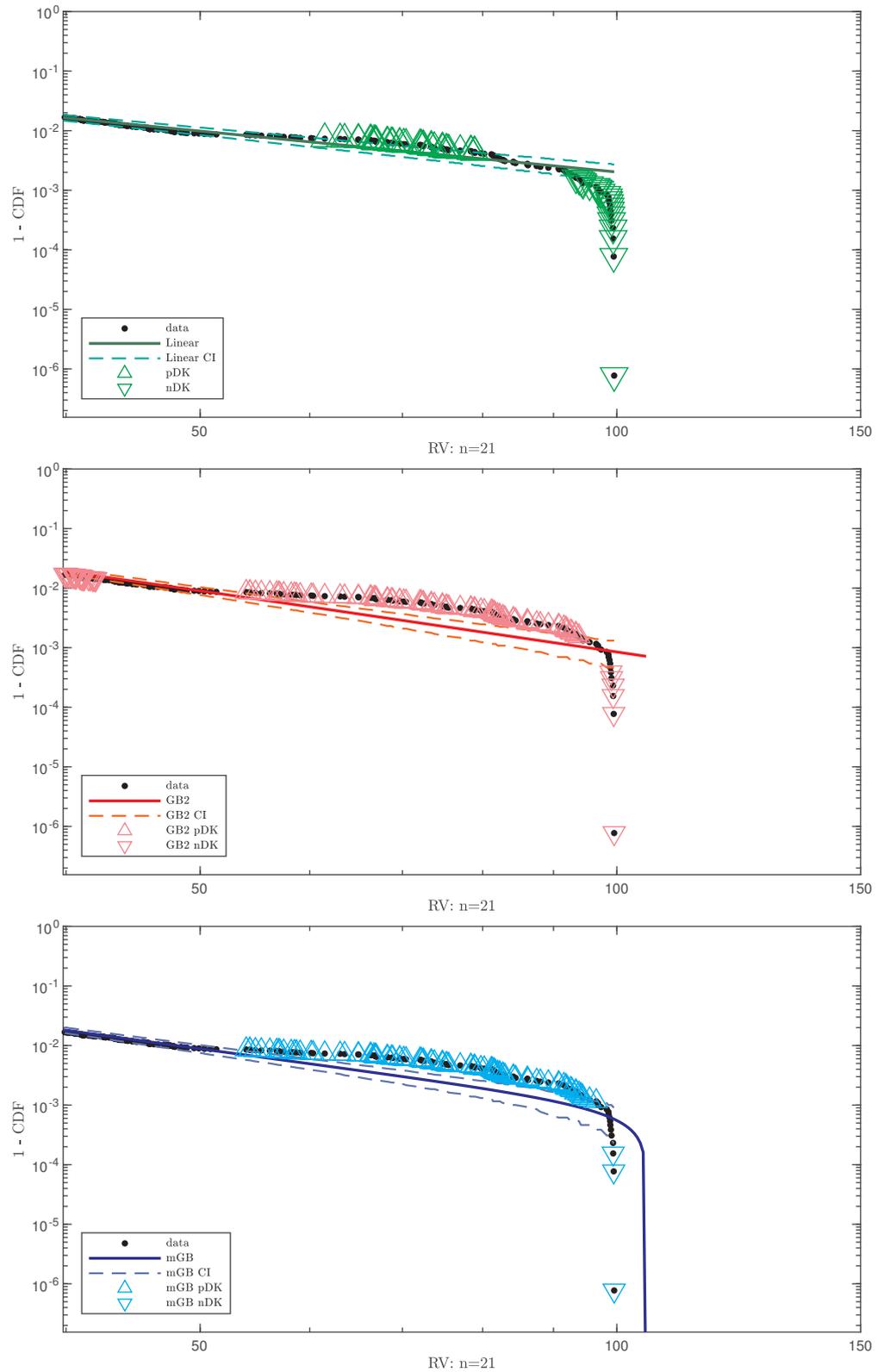


Figure 13. Same as the middle Figure 12 (tail parts of fits of S&P500 RV for $n = 21$) with the respective CI, “potential” DK (up triangles) and nDK (down triangles).

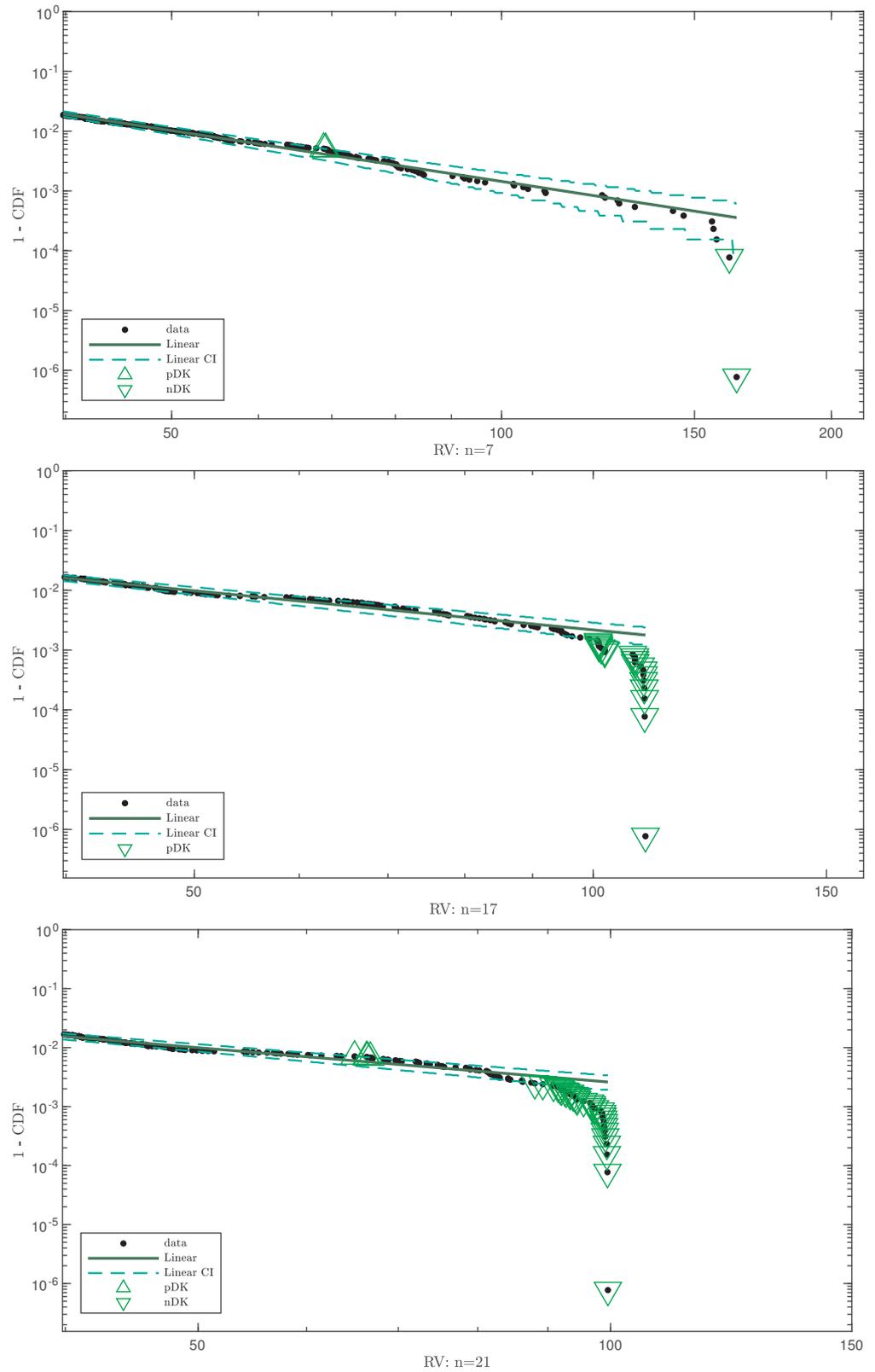


Figure 14. From top to bottom, LF of the tail of S&P500 RV, $RV > 40$, for $n = 7, 17, 21$ with end values $> 0.9 \max\{RV\}$ excluded from LF.

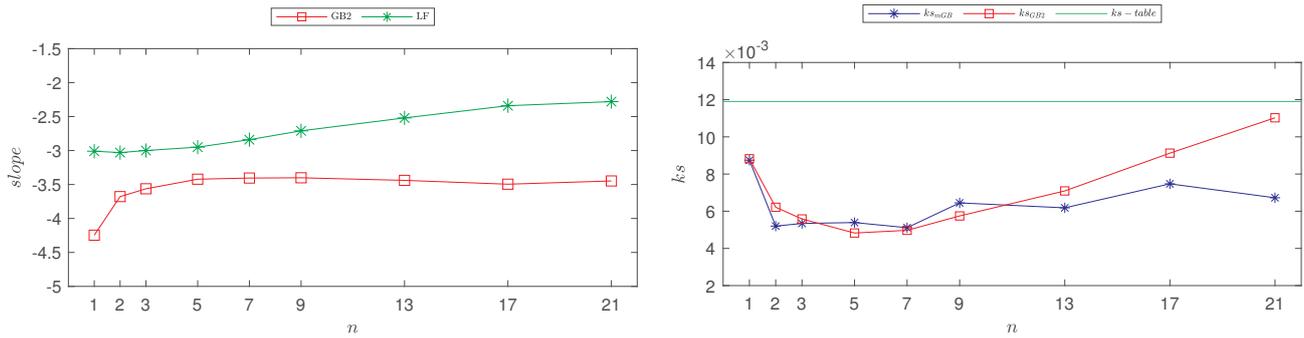


Figure 15. As a function of n : linear slope values of linear (stars) and GB2 (squares) fits (also listed in Table 3)—(left); KS statistic values indicating relative goodness of fit of mGB (stars) and GB2 (squares)—(right).

Table 2. Standard error of all estimated parameters of mGB and GB2 fits of S&P500 RV distribution.

n	SE of Parameters of mGB ($\alpha, \beta_1, \beta_2, p, q$)	SE of Parameters of GB2 (α, β, p, q)
1	(0.2049, 5.6537, 0.3884, 0.1196, 0.3171)	(0.0079, 0.0018, 0.0018, 0.0026)
5	(0.0690, 0.0127, 0.0152, 0.0175, 0.0922)	(0.0018, 0.0004, 0.0004, 0.0006)
7	(0.1193, 0.0246, 0.0348, 0.0246, 0.0281)	(0.0021, 0.0005, 0.0005, 0.0007)
17	(0.0614, 0.0220, 0.0150, 0.0288, 0.0325)	(0.0055, 0.0013, 0.0013, 0.0018)
21	(0.6300, 0.2807, 0.0114, 0.3419, 0.2451)	(0.0085, 0.0020, 0.0020, 0.0028)

Table 3. Slopes in Figure 15 of GB2 tail ($-\alpha q$) and of LF.

n	GB2	LF
1	-4.25	-3.01
5	-3.42	-2.95
7	-3.41	-2.84
17	-3.50	-2.34
21	-3.45	-2.28

5. Discussion

While the standard search for Dragon Kings involves performing a linear fit of the tails of the distribution combined with statistical tests, such as U-test [5] and evaluation of confidence intervals [4], here, we tried to broaden our analysis by also fitting the entire distribution—of realized volatility of S&P500 index in this case—using mGB (7) and GB2 (11), the two members of the Generalized Beta family of distributions [25,26]. As explained in the paragraph that follows (7), the central feature of mGB is that, after exhibiting a long power-law dependence, it eventually terminates at a finite value of the variable. GB2, on the other hand, has a power-law tail that extends mGB’s power-law dependence to infinity.

The key to understanding the results of fits in Section 4 is the analysis of the structure of RV used by the markets—a square root of realized variance (1). At its core is the average of the consecutive daily realized variances (2). Distribution of daily realized variance can be modeled using a duo of stochastic differential equations—for stock returns and stochastic volatility—which produces distributions of daily variance such as mGB [25] and GB2 [30]. Via a simple change of variable, daily RV would then follow the same distributions but with renormalized parameters.

Even assuming the knowledge of the distribution of the daily realized variance, finding the distribution of the averages constitutes a daunting task. To begin with, using convolution to evaluate the distribution of a sum of just two such complex distributions as mGB and GB2 is already not amenable to an analytical evaluation. To complicate things further, the consecutive daily RV cannot be treated as an independent identically

distributed variable (i.i.d.), due to the correlations that persist on average for up to roughly 5–7 days [24].

With the above in mind, we first address Figures 4–13. According to Figures 4 and 5, daily RV appears to be the closest to being commensurate with the Black Swan behavior, as both LF and GB2 approximate the tail of the distribution better than mGB and LF does not point to the existence of either DK, $p < 0.05$, or nDK, $p > 0.95$. The $n = 1$ behavior undergoes a dramatic change with the increase in n , as observed in Figures 6–13, where we observe that, first, the “potential” DK, $p < 0.05$, develop at the earlier portions of the tails, only to terminate in nDK at the tail ends.

Generally speaking, the existence of the large number of “potential” DK in the tail of the distribution indicates that the distribution is not describing the tail adequately. This becomes pronounced for large n for all three fits—LF, GB2 and mGB—although less so for LF, which also does not exhibit “potential” DK for small n . However, if we adopt a different procedure for LF, whereby instead of visually excluding nDK candidates at the tail end we exclude the values whose RV is greater than 0.9 of the maximum RV, we observe in Figure 14 that it has little effect on LF for small n but all but eliminates “potential” DK.

For large n we also observe that mGB approximates the tail end better than GB2—consistent with smaller KS values in Figure 15 and the smaller number of nDK. However, neither approximates the preceding portion of the tail well, as indicated by the “potential” DK. This has to do with the fact that neither of the distributions appear as a solution of a first-principle model describing the average RV. Finally, in the first plot in Figure 15, we observe that after roughly 5–7 days, the slope of the GB2 tail saturates, consistent with the correlation range of daily RV [24]. The slope of LF, on the other hand, increases with n . However, neither is consistent with a naive assumption of the distribution having the same slope as that of the daily RV.

6. Conclusions

We showed that for daily returns, the distribution of realized volatility likely has a power-law tail, consistent with the Black Swan behavior. The averaged realized volatility develops a strong negative Dragon King signature, as the number of days involved in the averaging of daily realized variances increases. The breadth and strength of the S&P500 index analyzed here may be a contributing factor in suppressing the runaway power-law Black Swan behavior.

It should be emphasized, however, that there exists a plethora of other quantities describing markets, some of which may exhibit Black Swan or Dragon King behavior—to wit, the study of outliers in [11–14]. For that matter, given that market crashes are often characterized by their longevity, accumulated (multi-day) stock returns and associated realized volatility may be more appropriate quantities to study [33,34]. On the other hand, the realized volatility studied here is widely used both by economists and market practitioners, and it provides a good retrospective picture of market calamities.

We also attempted to fit the entire distributions using modified Generalized Beta and Generalized Beta Prime distributions. While the former qualitatively mimics termination of the empirical distributions of realized volatility, goodness of fit—which is in itself a complicated subject of future research—remains an open question.

A natural extension of this work, and specifically of the application of the Generalized Beta family of distributions, would be, aside from the aforementioned accumulated returns and volatility, the analysis of separate gains and losses of stock returns vis-a-vis realized volatility, as well as of other large data sets that may exhibit power-law tails and outliers, such as income distributions and associated quantities such as house prices and house price indices [35]. It would also be of great interest to compare these distributions across different markets and geographical areas.

Author Contributions: J.L. and M.D.M. performed all numerical calculations, R.A.S. was a lead on the analytical part with J.L.’s active participation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: We obtained S&P500 data at Yahoo! Finance. Our datasets are available upon request.

Acknowledgments: We used Wolfram Mathematica in a subset of analytical calculations and MathWorks Matlab for much of the numerical work.

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Clauset, A.; Shalizi, C.R.; Newman, M.E. Power-Law Distributions in Empirical Data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
2. Gabaix, X. Power laws in economics and finance. *Annu. Rev. Econ.* **2009**, *1*, 255–294. [CrossRef]
3. Saichev, A.; Malevergne, Y.; Sornette, D. *Theory of Zipf's Law and Beyond*; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin/Heidelberg, Germany, 2009.
4. Janczura, J.; Weron, R. Black swans or dragon-kings? A simple test for deviations from the power law. *Eur. Phys. J. Spec. Top.* **2012**, *205*, 79–93. [CrossRef]
5. Pisarenko, V.F.; Sornette, D. Robust statistical tests of Dragon-Kings beyond power law distribution. *Eur. Phys. J. Spec. Top.* **2012**, *205*, 95–115. [CrossRef]
6. Wheatley, S.; Sornette, D. Multiple Outlier Detection in Samples with Exponential Pareto Tails: Redeeming the Inward Approach Detecting Dragon Kings. *arXiv* **2015**, arXiv:1507.08689.
7. Sornette, D. Dragon-Kings, Black Swans and the Prediction of Crises. *Int. J. Terraspace Sci. Eng.* **2009**, *2*, 1–18. [CrossRef]
8. Sornette, D.; Ouillon, G. Dragon-kings: Mechanisms, statistical methods and empirical evidence. *Eur. Phys. J. Spec. Top.* **2012**, *205*, 1–26. [CrossRef]
9. Golosovsky, M.; Solomon, S. Runaway events dominate the heavy tail of citation distributions. *Eur. Phys. J. Spec. Top.* **2012**, *205*, 303–311. [CrossRef]
10. Medina, J.A. Extreme reaction times determine fluctuation scaling in human color vision. *Phys. A* **2016**, *461*, 125–132. [CrossRef]
11. Johansen, A.; Sornette, D. Stock market crashes are outliers. *Eur. Phys. J. B* **1998**, *1*, 141–143. [CrossRef]
12. Johansen, A.; Sornette, D. Large Stock Market Price Drawdowns Are Outliers. *J. Risk* **2001**, *4*, 69–110. [CrossRef]
13. Sornette, D.; Johansen, A. Significance of log-periodic precursors to financial crashes. *Quant. Financ.* **2001**, *1*, 452–471. [CrossRef]
14. Filimonov, V.; Sornette, D. Power law scaling and “Dragon-Kings” in distributions of intraday financial drawdowns. *Chaos Solitons Fractals* **2015**, *74*, 27–45. [CrossRef]
15. CBOE VIX Index. Available online: https://www.cboe.com/tradable_products/vix/ (accessed on 1 January 2024).
16. VIX Options and Futures Historical Data. Available online: <http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data> (accessed on 1 January 2024).
17. Demeterfi, K.; Derman, E.; Kamal, M.; Zou, J. A guide to volatility and variance swaps. *J. Deriv.* **1999**, *6*, 9–32. [CrossRef]
18. The CBOE Volatility Index—VIX. Previous Location of VIX White Paper, Appears to Have Been Since Removed. 2003. Available online: <https://www.cboe.com/micro/vix/vixwhite.pdf> (accessed on 1 January 2024).
19. Christensen, B.J.; Prabhala, N.R. The Relation Between Implied and Realized Volatility. *J. Financ. Econ.* **1998**, *50*, 125–150. [CrossRef]
20. Vodenska, I.; Chambers, W.J. Understanding the Relationship between VIX and the S&P 500 Index Volatility. In Proceedings of the 26th Australasian Finance and Banking Conference, Sydney, Australia, 17–19 December 2013.
21. Kownatzki, C. How good is the VIX as a predictor of market risk? *J. Account. Financ.* **2016**, *16*, 39–60.
22. Russon, M.D.; Vakili, A.F. On the non-linear relationship between VIX and realized SP500 volatility. *Invest. Manag. Financ. Innov.* **2017**, *14*, 200–206. [CrossRef]
23. Dashti Moghaddam, M.; Liu, Z.; Serota, R.A. Distributions of Historic Market Data—Implied and Realized Volatility. *Appl. Econ. Financ.* **2019**, *6*, 104–130. [CrossRef]
24. Dashti Moghaddam, M.; Liu, J.; Serota, R.A. Implied and Realized Volatility: A Study of Distributions and Distribution of Difference. *Int. J. Financ. Econ.* **2021**, *26*, 2581–2594. [CrossRef]
25. Liu, J.; Serota, R.A. Rethinking Generalized Beta family of distributions. *Eur. Phys. J. B* **2023**, *96*, 24. [CrossRef]
26. McDonald, J.B.; Xu, Y.J. A generalization of the beta distribution with applications. *J. Econom.* **1996**, *66*, 133–152. [CrossRef]
27. Risken, H. *The Fokker-Planck Equation*; Springer: Berlin/Heidelberg, Germany, 1996.
28. Hertzler, G. “Classical” Probability Distributions for Stochastic Dynamic Models. In Proceedings of the 47th Annual Conference of the Australian Agricultural and Resource Economics Society, Fremantle, Australia, 12–14 February 2003.
29. Jacobs, K. *Stochastic Processes for Physicists*; Cambridge University Press: Cambridge, UK, 2010.
30. Dashti Moghaddam, M.; Serota, R.A. Combined Multiplicative-Heston Model for Stochastic Volatility. *Phys. A Stat. Mech. Its Appl.* **2021**, *561*, 125263. [CrossRef]
31. NIST Digital Library of Mathematical Functions. Available online: <https://dlmf.nist.gov> (accessed on 6 February 2024).
32. Massey, F.J. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1985**, *80*, 954–958. [CrossRef]
33. Liu, Z.; Dashti Moghaddam, M.; Serota, R.A. Distributions of Historic Market Data—Stock Returns. *Eur. Phys. J. B* **2019**, *92*, 60. [CrossRef]

-
34. Dashti Moghaddam, M.; Liu, Z.; Serota, R.A. Distributions of historic market data: Relaxation and correlations. *Eur. Phys. J. B* **2021**, *94*, 83. [[CrossRef](#)]
 35. Liu, J.; Farahani, H.; Serota, R.A. Exploring distributions of housing prices and housing prices index. *arXiv* **2023**, arXiv:2312.14325.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.