Diverging from the Norm: An Examination of Non-Normality and its Measurement in Asset Returns

Grant Holtes

February 5, 2024

Summary

This paper examines the normality of US equities and fixed income asset-class returns over 104 years. Analysis includes moment computations, hypothesis tests, and visual inspections. Findings indicate large cap equities are closer to normal distribution than small caps, and longer holding periods result in more normal returns across asset classes. While hypothesis tests suggest non-normality, their results are shown to largely depend on sample size. The practical implications of assuming normality are explored, revealing minimal effects on central estimates but significant variations in estimates of low-probability events.

1 Introduction

The return on assets is widely regarded being non-normally distributed, however, assuming that returns are normally distributed is not a rare simplifying assumption. This paper aims to investigate both the normality of returns and the impacts on assuming that returns are normal. Data of asset-class returns from the last 104 years is used, covering US equities and fixed income asset classes.

2 Moments

A simple initial exploration of the distribution of returns is to compute the moments of their sample distribution. This is done for annual returns in Table 1 and for annual log returns in Table 2[1].

	mean	stdev	skew	ex kurtosis
Large-Cap Stocks	0.1216	0.1973	-0.4194	0.0364
Small-Cap Stocks	0.1606	0.3103	0.6306	2.2892
Long-term Corp Bonds	0.064	0.0847	1.1528	2.8442
Long-term Gov Bonds	0.056	0.1025	0.5843	1.6542
Med-term Gov Bonds	0.05	0.0569	1.1093	2.8188
60:40	0.0954	0.1272	-0.364	0.1364

Table 1: Moments of the sample distribution of monthly returns, 1929 to 2023

	mean	stdev	skew	ex kurtosis
Large-Cap Stocks	0.0978	0.1907	-1.0002	1.3356
Small-Cap Stocks	0.1119	0.2817	-0.7325	1.9197
Long-term Corp Bonds	0.059	0.077	0.8025	1.5576
Long-term Gov Bonds	0.0499	0.096	0.0834	1.8679
Med-term Gov Bonds	0.0474	0.0529	0.846	2.0643
60:40	0.0841	0.1208	-0.7503	0.9187

Table 2: Moments of the sample distribution of monthly log returns, 1929 to 2023

A similar analysis can be completed for monthly returns, which provides a larger sample size, with the limitation that the monthly return are being interpreted as monthly returns rather than a twelfth of annual returns [2]. The results for monthly returns are given in Table 3 and for monthly log returns in Table 4.

[1]"60:40" denotes a portfolio of 60% large cap stocks and 40% long term govt bonds, rebalanced with the frequency of return data.

[2]see Holtes, 2024 for a more detailed discussion of this.

	mean	stdev	skew	ex kurtosis
Large-Cap Stocks	0.0096	0.0538	0.3003	9.2369
Small-Cap Stocks	0.0125	0.0809	1.1663	12.7101
Long-term Corp Bonds	0.0051	0.022	0.6157	6.0668
Long-term Gov Bonds	0.0045	0.0253	0.5013	3.9503
Med-term Gov Bonds	0.004	0.0126	0.804	8.495
60:40	0.0076	0.0346	0.2883	7.5102

Table 3: Moments of the sample distribution of monthly returns, 1929 to 2023

	mean	stdev	skew	ex kurtosis
Large-Cap Stocks	0.0082	0.0536	-0.4967	7.6718
Small-Cap Stocks	0.0093	0.0789	-0.1281	7.9834
Long-term Corp Bonds	0.0048	0.0217	0.3734	5.4505
Long-term Gov Bonds	0.0042	0.025	0.293	3.5734
Med-term Gov Bonds	0.0039	0.0125	0.6279	7.5206
60:40	0.007	0.0344	-0.1643	6.501

Table 4: Moments of the sample distribution of monthly log returns, 1929 to 2023

The following conclusions can be made from the moments data:

Large Cap Equities are "more normal": Large cap equities show moments that are closer to a normal distribution than small cap equities.

Longer time period returns are "more normal": Excess kurtosis is lower for annual returns than monthly returns, across asset classes. This is confirmed in Figure 1 and Figure 8 (appendix), where kurtosis is computed for a range of holding periods. This shows that both returns and log returns on equities appear to converge towards normality as the holding period increases. Similar conclusions can be drawn on skew, with skew tending towards zero as holding period increases, as in Figure 2. This conclusion is limited by the length of the data history and the high serial correlation between returns on overlapping holding periods. It is worth noting that returns on bonds appear to converge to a distribution with thinner tails than a normal distribution after 15 years.

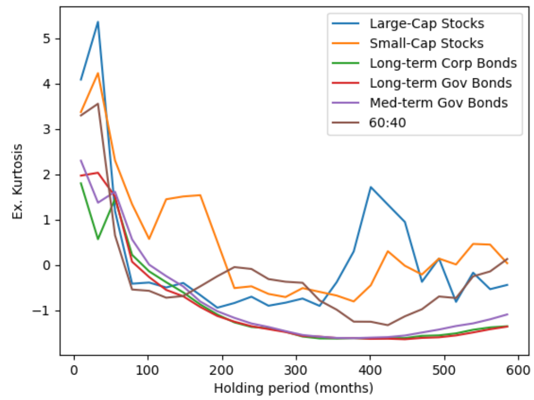

Figure 1: Kurtosis of returns by holding period and asset class. Similar results are seen for log returns, see appendix: Figure 8.

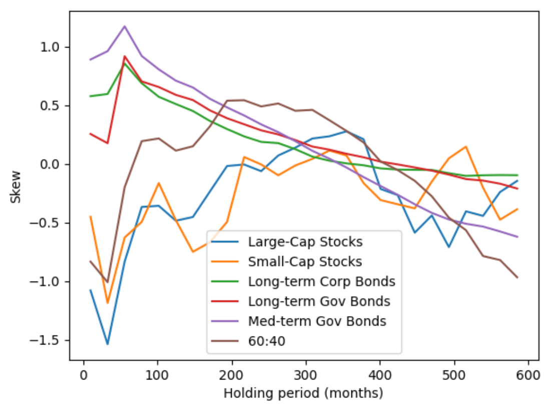

Figure 2: Skew of returns by holding period and asset class.

3 Hypothesis tests

The analysis of the moments of the distribution can complimented by a range of goodness-of-fit hypothesis tests to test the null hypothesis that an observed dataset is drawn from a normal distribution. The results of these tests are interpreted as:

$p < critical value$**:** The null hypothesis is rejected and there is evidence that the data was not drawn from a normal distribution.

$p >= critical value$**:** The null hypothesis is not rejected and we cannot conclude that the data was not drawn from a normal distribution. This should not be interpreted to imply that the data is drawn from a normal distribution.

The results from a range of tests are given for annual returns in Table 5 and for monthly returns in Table 6. As detailed by Wah, 2011, the Shapiro-Wilk test has the highest power for a given significance level, followed by the Anderson-Darling test.

	n	Shapiro-Wilk	Kolmogorov-Smirnov	Anderson-Darling
Large-Cap Stocks	98	Null not rejected (0.2292)	Not normal (0.0)	Null not rejected (0.15)
Small-Cap Stocks	98	Not normal (0.0135)	Not normal (0.0)	Null not rejected (0.15)
Long-term Corp Bonds	96	Not normal (0.0002)	Not normal (0.0)	Not normal (0.025)
Long-term Gov Bonds	98	Not normal (0.0025)	Not normal (0.0)	Not normal (0.01)
Med-term Gov Bonds	98	Not normal (0.0001)	Not normal (0.0)	Not normal (0.01)
60:40	98	Null not rejected (0.3622)	Not normal (0.0)	Null not rejected (0.15)

Table 5: Hypothesis tests on the normality of annual returns, 1929 to 2023. A threshold p value of 0.05 is used.

	n	Shapiro-Wilk	Kolmogorov-Smirnov	Anderson-Darling
Large-Cap Stocks	1175	Not normal (0.0)	Not normal (0.0)	Not normal (0.01)
Small-Cap Stocks	1176	Not normal (0.0)	Not normal (0.0)	Not normal (0.01)
Long-term Corp Bonds	1155	Not normal (0.0)	Not normal (0.0)	Not normal (0.01)
Long-term Gov Bonds	1176	Not normal (0.0)	Not normal (0.0)	Not normal (0.01)
Med-term Gov Bonds	1176	Not normal (0.0)	Not normal (0.0)	Not normal (0.01)
60:40	1175	Not normal (0.0)	Not normal (0.0)	Not normal (0.01)

Table 6: Hypothesis tests on the normality of monthly returns, 1929 to 2023. A threshold p value of 0.05 is used.

These results compliment the previous conclusion that longer holding times result in more normal returns, with the p-values being higher for the tests on annual returns than on monthly returns.

3.1 Are hypothesis tests useful?

A common criticism of distribution tests is that the result of the test is not particular useful in measuring how well a normal distribution fits the observed data. Assuming that the data generating process is not drawing from a perfect normal distribution, the results of the test is more commonly determined by sample size rather than distribution shape, with a large enough sample being almost certain to reject the null hypothesis [3].

This is observed in the results, with the monthly returns tests overwhelming rejecting the null at a high confidence, while the annual returns tests generally had higher p values and did not reject the null in some cases.

This can be further confirmed by calculating the p-value for a range of different sample sizes, bootstrapping by resampling from the monthly returns dataset. Figure 3 shows that the median p-value of the tests on the bootstrapped samples is above a critical value of 0.05 until a sample size of 50, and 95% of the samples were below the critical value from a sample size of 250.

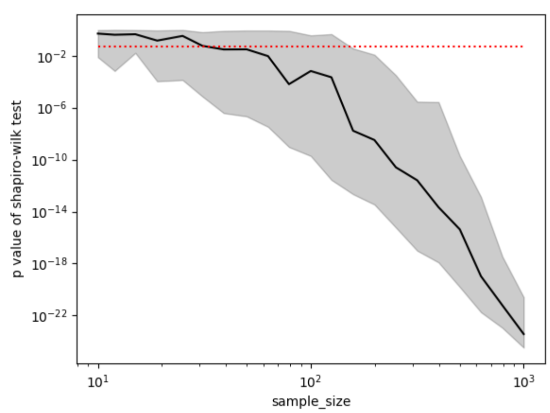

Figure 3: p-value on a Shapiro-Wilk test on monthly large-cap returns for a range of sample sizes. 80% confidence interval based on bootstrapped tests. Both axis in log scale.

4 Visual analysis

A qualitative alternative to hypothesis testing is the visual inspection of the distribution. While this seems unsophisticated, it can provide a useful tool to spot where and how distributions differ. The two plots that are used are the cumulative distribution function (CDF) and Q-Q plot.

An example of this can be seen using large-cap stock returns, with the CDF given in Figure 4 and the Q-Q plot given in Figure 5. These plots can be compared against the monthly return data, again for large-cap stocks in Figure 6 and Figure 7. Overall the below conclusions can be made:

[3] This is analysis was inspired by the work of Allen Downey, blog: Probably Overthinking It. See https://allendowney.blogspot.com/2013/08/are-my-data-normal.html

The normal distribution fits annual returns well: While there is a lot of noise in the plots, this is generally random deviations around the fitted normal distribution rather than a systematic deviation, as is observed in the monthly returns.

Fat tails are more observable in monthly returns: The higher kurtosis of the monthly returns is observable in the "S-shaped" deviation from the line $x = y$ in the Q-Q plot and as a steepening of the CDF compared to the fitted normal.

Similar results are seen for annual log-returns of large-cap equities, see appendix Figure 9. For small-cap equities, the fatter tails in annual returns is seen as a steeper CDF in Figure 10 and a slight S-shaped Q-Q plot in Figure 11

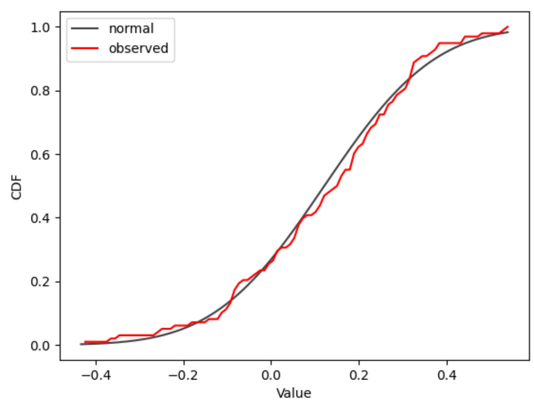

Figure 4: CDF of annual US large-cap stock returns, fitted normal distribution. 1929 to 2023.

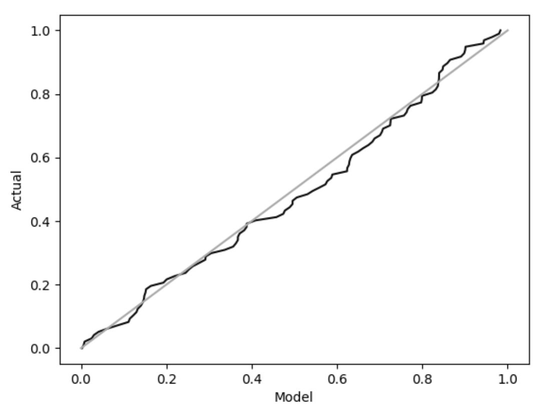

Figure 5: Q-Q of annual US large-cap stock returns, fitted normal distribution. 1929 to 2023.

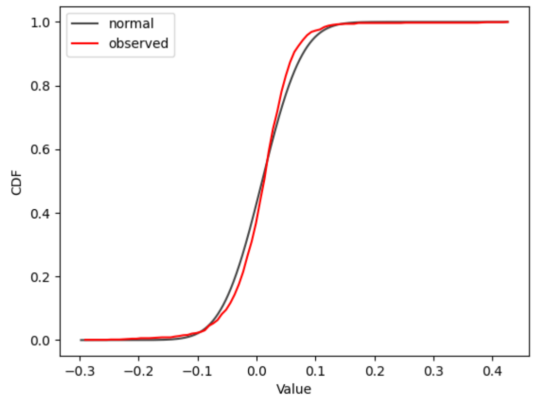

Figure 6: CDF of monthly US large-cap stock returns, fitted normal distribution. 1929 to 2023.

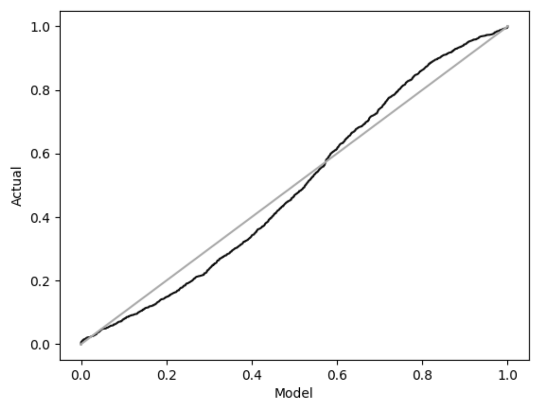

Figure 7: Q-Q of monthly US large-cap stock returns, fitted normal distribution. 1929 to 2023.

5 Impacts of a normal assumption

A more practical test is to measure the impact of a normal returns distribution assumption on the results of a forecast or analysis. In order to measure the size of the impact a baseline is required to compare against, which for simplicity is a Gaussian kernel fitted on the historical return distribution. Two use cases for the assumed distribution are considered:

Stochastic forecasts: The return distribution is used to generate returns in a Monte-Carlo portfolio simulation.

CVAR: The return distribution is used to compute conditional value at risk (CVAR).

5.1 Stochastic forecasts

The results for the stochastic forecast test are given in Table 7. This shows that both the mean and median tend to be slightly higher for the simulations drawing from the normal distribution, but this effect is minimal and likely insignificant when compared to uncertainty regarding predicted returns. Similar results are observed when using monthly data, as in Table 10.

Notable the 95th percentile of simulations of equity returns showed much higher returns when the normal distribution was used, likely due to the lack of accounting for the negative skew in the sample distribution. As is suggested by Figure 2, this is likely to be less of an issue for longer forecast periods.

	mean	median	5th quantile	95th quantile
Large-Cap Stocks	-0.0378	-0.0026	0.0238	0.1423
Small-Cap Stocks	0.1574	0.3241	0.0795	0.3309
Long-term Corp Bonds	0.0383	0.0476	-0.0184	0.0188
Long-term Gov Bonds	-0.0059	0.0125	-0.0023	-0.1033
Med-term Gov Bonds	0.0004	0.0111	-0.0343	-0.0241
60:40	0.0153	-0.0005	0.0529	0.1143

Table 7: Comparison of 10Y stochastic forecasts of returns, using fitted normal distribution and fitted Gaussian kernel. Values are the difference in forecast statistics (normal less kernel). Based on annual US data, 1929 to 2023.

5.2 CVAR

As CVAR calculations concentrate on the tails of the probability distribution, the results are naturally more sensitive to the choice of distribution specification. As is expected with the high kurtosis of the monthly returns, the normal distribution underestimates the tail risk in the results in Table 8.[4]

	normal distribution	observed	gaussian kernel
Large-Cap Stocks	-0.1095	-0.1514	-0.1509
Small-Cap Stocks	-0.1704	-0.2161	-0.2161
Long-term Corp Bonds	-0.0443	-0.0556	-0.0548
Long-term Gov Bonds	-0.052	-0.0618	-0.0616
Med-term Gov Bonds	-0.0231	-0.0301	-0.03
60:40	-0.07	-0.0931	-0.0924

Table 8: Comparison of monthly return CVAR (at the 2.5th percentile), using fitted normal distribution and fitted Gaussian kernel. Based on monthly US data, 1929 to 2023.

The results for annual return CVAR, Table 9, show that the normal distribution underestimates tail risk for equities and the 60:40 portfolio, but overestimates tail risk for bonds.

[4] In all cases the Gaussian kernel and observed CVAR are naturally very similar, given that the Gaussian kernel is simply a continuous smoothing of the observed CDF.

	normal distribution	observed	gaussian kernel
Large-Cap Stocks	-0.2791	-0.3546	-0.3343
Small-Cap Stocks	-0.4453	-0.4932	-0.469
Long-term Corp Bonds	-0.1033	-0.0736	-0.0719
Long-term Gov Bonds	-0.1524	-0.156	-0.1443
Med-term Gov Bonds	-0.0658	-0.0506	-0.0476
60:40	-0.1636	-0.2112	-0.1962

Table 9: Comparison of annual return CVAR (at the 2.5th percentile), using fitted normal distribution and fitted Gaussian kernel. Based on annual US data, 1929 to 2023.

6 Conclusion

The following conclusions can be made from the observations in this paper:

Returns are measurably non-normal: Analysis of the moments of the returns distribution and hypothesis tests on the goodness-of-fit against a normal distribution suggest that returns are unlikely to be normal.

Returns are more normal at longer holding periods: Excess kurtosis appears to decrease and skew tends towards zero with holding period for all asset types.

Sample size is a major limitation: Goodness-of-fit hypothesis tests are seemingly dictated by sample size, rather than being a reliable measure for the normality of returns.

A normal assumption has an impact when estimating low-probability events: The impacts section demonstrates that a normal assumption does not have a large impact on central estimates, but can have a large impact on estimates of low-probability events such as CVAR calculations or estimating the 95th percentile of 10Y returns.

References

Holtes, G. (2024). Return volatility estimates: A review and practical analysis. *Independent*.

Wah, N. R. Y. B. (2011). Power comparisons of shapiro–wilk, kolmogorov–smirnov, lilliefors and anderson–darling tests. *Journal of Statistical Modeling and Analytics*.

7 Appendix

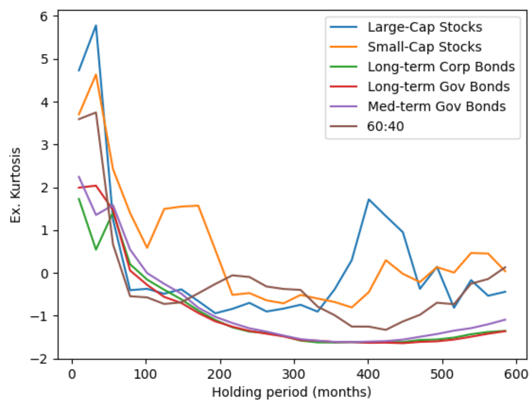

Figure 8: Kurtosis of log returns by holding period and asset class.

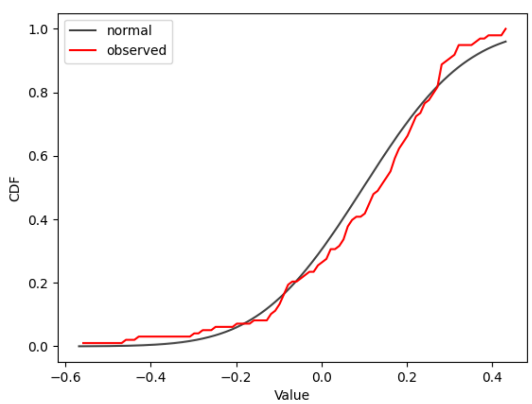

Figure 9: CDF of annual US large-cap stock log returns, fitted normal distribution. 1929 to 2023.

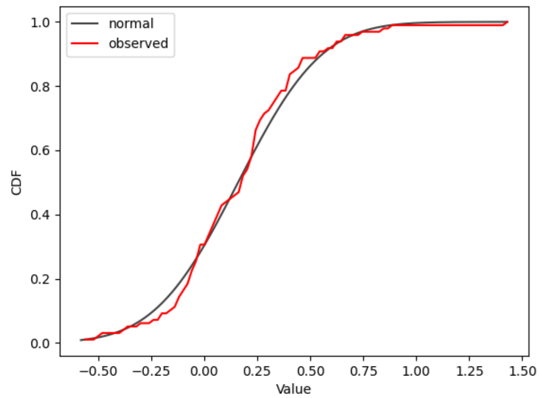

Figure 10: CDF of annual US small-cap stock returns, fitted normal distribution. 1929 to 2023.

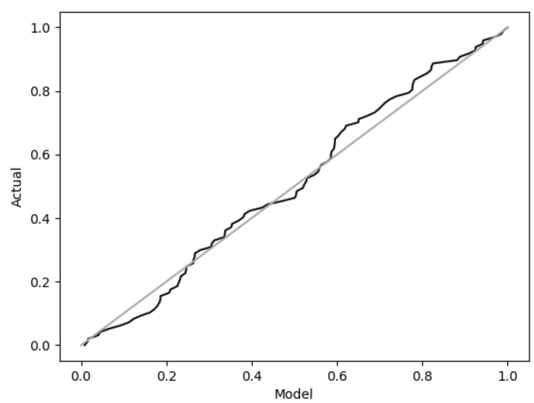

Figure 11: Q-Q of annual US small-cap stock returns, fitted normal distribution. 1929 to 2023.

	mean	median	5th quantile	95th quantile
Large-Cap Stocks	0.0024	0.0012	0.0192	-0.0067
Small-Cap Stocks	-0.0222	-0.0226	-0.017	0.0081
Long-term Corp Bonds	-0.0016	-0.0008	0.0008	-0.014
Long-term Gov Bonds	0.0034	0.006	0.0151	-0.0082
Med-term Gov Bonds	0.0005	0.0021	0.0074	-0.0104
60:40	0.0001	-0.0024	0.0225	0.0065

Table 10: Comparison of 10 month stochastic forecasts of returns, using fitted normal distribution and fitted Gaussian kernel. Based on monthly US data, 1929 to 2023.