

Modelling skewness and elongation in financial returns: the case of exchange-traded funds

Sanjiv Jaggia and Alison Kelly-Hawke

Recent studies have documented the importance of asymmetry and tail-fatness of returns on portfolio-choice, asset-pricing, value-at-risk and option-valuation models. This article explores the nature of skewness and elongation in daily Exchange-traded Fund (ETF) return distributions using g , h and $(g \times h)$ distributions. These exploratory data analytic techniques of Tukey (1977) reveal patterns that are hidden from a cursory glance at conventional measures for skewness and elongation. The g , h and $(g \times h)$ distributions provide parameter estimates that indicate substantial variation in skewness and elongation for individual ETFs; nonetheless, some trends are discovered when the funds are grouped by fund size and style of investing. Monte Carlo simulations suggest that these exploratory techniques are able to capture patterns found in commonly used Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family of models.

1. Introduction

This article explores the patterns of daily Exchange-traded Fund (ETF) return distributions. ETFs are mutual funds that trade like stocks. They are structured like index mutual funds; that is, a particular ETF contains a collection of stocks that typically track an index, like the Dow Jones Industrial Average or the S&P 500 stock index. An ETF thus combines the valuation feature of a mutual fund with the tradability feature of a closed-end fund. The advantage of examining the properties of ETFs over a particular index arises because ETFs tend to be more isolated from market microstructure noise, such

as nonsynchronous trading, as compared to an index. In addition to analysing measures of return and risk for various ETFs, a major goal of this article is to evaluate thoroughly the higher moments of skewness and kurtosis. In particular, this article explores the nature of skewness and elongation in daily ETF return distributions using g , h and $(g \times h)$ distributions.

Numerous studies have documented that the form of the distribution of returns is a crucial assumption for mean–variance portfolio theory, theoretical models of capital asset prices and the prices of contingent claims.¹ Statistical inference also relies heavily on distributional assumptions. If these assumptions are violated, there are resulting implications for portfolio analysis.

Harvey and Siddique (2000a) show how conditional skewness explains a significant part of the variation in returns even when factors based on size and book-to-market value are added to the asset pricing model. Dittmar (2002) incorporates skewness and kurtosis to the asset pricing model and his results indicate that nonlinearities substantially improve upon the model's ability to describe a cross section of returns.

Hansen (1994) introduces the generalized Student's t -distribution to model innovations of a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model. This two parameter distribution is asymmetric and allows excess kurtosis, which can also be used to model time-varying conditional higher moments (Jondeau and Rockinger, 2003). Christoffersen *et al.* (2006) develop a model of stock returns that allows for skewness as well as conditional heteroskedasticity and a leverage effect. They then introduce an option pricing formula consistent with this model. Using S&P 500 stock options over the period 2 January 1990 through 31 December 1992, in-sample and up to 10 weeks out-of-sample performance of their model achieves a better fit than standard GARCH models.

Most empirical studies calculate skewness and kurtosis as an average and find that stock market returns have negative skewness and severe excess kurtosis. Given the presence of outliers, however, conventional measures of skewness and kurtosis may be quite inadequate in capturing the true behaviour of financial returns. Kim and White (2004) show how a single outlier can dramatically influence conventional measures. They conclude that one must look beyond conventional measures of skewness and kurtosis to gain insight into market returns behaviour.

In order to analyse ETF returns, this article follows the exploratory data analytic techniques first suggested by Tukey (1977) and later applied to a housing allowance demand experiment by Hoaglin (1985). These techniques are simple to compute and allow large flexibility and robustness in their fitting. Badrinath and Chatterjee (1988) apply this technique to daily and monthly returns on the Center for Research in Security Prices (CRSP) equal-weighted and value-weighted market portfolios covering the period July 1962 through December 1985. They conclude that the distribution of the market portfolio is adequately explained as a skewed and elongated distribution. In subsequent work with daily common-stock-return distributions, Badrinath and Chatterjee (1991) find substantial variation in the parameter estimates for skewness and elongation for individual firms, but discover some trends across industry groups and firm sizes.

Mills (1995) also uses exploratory data techniques to examine the distribution of daily returns of three London Stock Exchange indices over the period 1986 to 1992. He too concludes that returns are both skewed and extremely kurtotic. However, he finds that the deregulation of the stock exchange in October 1986 and the run-up and aftermath of 'Black Monday' (the market crash of 19 October 1987) alter the shape of the return distributions quite dramatically. Dutta and Babbel (2005) use the exploratory data analysis on 3-month London Interbank Offered Rate (LIBOR) data as implied by its option prices. They find that the implied distribution is modelled more accurately by the g , h and $(g \times h)$ distributions as compared to other commonly used distributions.

The contributions of this article are threefold: (1) to explore the nature of skewness and elongation in daily ETF return distributions using g , h and $(g \times h)$ distributions proposed by Tukey, (2) to search for patterns of skewness and elongation over different classes of ETFs (categorized by fund size and style of investing) that may enable investors to make more careful decisions in their portfolio selection and (3) to use Monte Carlo simulations to analyse how well these exploratory techniques capture patterns found in commonly used GARCH family of models.

The rest of the article is organized as follows. Section II summarizes elements of the g , h and $(g \times h)$ distributions and the estimation procedures. Section III provides descriptive statistics of the sample data and presents the results. Section IV uses simulation based on GARCH models to analyse the sampling behaviour of the g , h and $(g \times h)$ estimators. Section V concludes and discusses implications for portfolio diversification models and portfolio selection.

II. The g , h and $(g \times h)$ Distributions

Skewness and the g distribution

The skewness of a distribution is judged in terms of its departure from symmetry. In this application, the random variable X is defined as the daily return on an ETF and Z is a standard normal random variable such that

$$X = A + BY_g(Z) \quad (1a)$$

where

$$Y_g(Z) = \frac{[\exp(gZ) - 1]}{g} \quad (1b)$$

The parameters A and B refer to the location and scale of X , respectively. The function $Y_g(Z)$ is said to have the g -distribution where the parameter g controls the amount and direction of skewness; thus, a value of $g=0$ corresponds to no skewness.

In order to estimate g , we implement the approach suggested by Tukey (1977); that is, we let x_p , x_{1-p} , z_p and z_{1-p} represent the p -th and $(1-p)$ th percentiles of the random variable X and a standard normal random variable Z , respectively, where $p < 0.5$. Rewriting Equation 1a for the p -th and $(1-p)$ th percentiles gives

$$x_p = A + B/g^*[\exp(gz_p) - 1] \quad (2a)$$

and

$$x_{1-p} = A + B/g^*[\exp(gz_{1-p}) - 1] \quad (2b)$$

Noting that $z_p = -z_{1-p}$ and $x_{0.5} = A + B/g^*[\exp(g*0) - 1] = A$, it follows from dividing Equation 2b from Equation 2a and solving for g yields:

$$g_p = \left(\frac{-1}{z_p}\right) \ln \left[\frac{x_{1-p} - x_{0.5}}{x_{0.5} - x_p} \right] \quad (3)$$

Thus, g_p measures skewness in terms of the logarithm of the relative distances of the $(1-p)$ th and the p -th percentiles from the median. Multiple estimates of g may be obtained from Equation 3 by using selected values of p . These estimates provide a summary of how skewness changes across the sample and informative plots. If the estimates of the g_p 's are more or less constant, then the median of these values can be used as an estimate of the overall skewness. In addition, the variation in these estimates provides information on the stability of the median estimate of g .

In some cases, the median may provide a good estimate of g , but the power of this particular methodology stems from being able to focus on different percentiles of the distribution. The typical approach to choosing the percentiles is to use letter values; the sequence of percentiles is chosen such that $p=1/2$ corresponds to M (median), $p=1/4$ corresponds to F (fourths), $p=1/8$ corresponds to E (eighths) and so on, such that $p=1/16$, $1/32$, $1/64$, $1/128$, $1/256$, etc. are D, C, B, A, Z, etc. By definition, the letter values pay more attention to the tails of the distribution than the middle since the tail area is repeatedly being halved.

If the data are symmetric, the median is the point of symmetry and each pair of letter values must be symmetrically placed about the median. That is, the lowest fourth of the distribution will be as far below

the median as the upper fourth is from the median. A simple way to check on symmetry is to define a set of midsummaries, one for each pair of letter values. The midsummary is the average of the two letter values (upper and lower) or $1/2[x_p + x_{1-p}]$. (The distance between the upper and lower values for any letter value, $[x_p - x_{1-p}]$, is called the *letter spread* and the positive distance between the median and any letter value is called the *half-spread*.) In a perfectly symmetric distribution, all midsummaries would be equal to the median. If the data were skewed to the right, the midsummaries would increase as they came from the letter values further into the tails. For data skewed to the left, they would decrease. If apparent skewness is due to one or two stray values, only the most extreme letter values and midsummaries would be affected.

In order to analyse skewness graphically, a plot of the sample upper values against the lower values should form a line with slope equal to -1 if the returns are symmetric about the median. A numerical estimate of the slope can be obtained by regressing sample upper values against sample lower values. Testing would then reveal whether or not the slope is significantly different from -1 .

Elongation and the h distribution

Elongation refers to the stretch of the tails of a distribution. A more elongated distribution gives greater probability to outcomes that are quite notably more extreme. Since there is no natural standard as symmetry is for skewness, a common practice is to use the Gaussian distribution as the standard when measuring elongation. In this section, we analyse elongation in the presence of symmetry. The random variables X and Z are those previously defined such that

$$X = A + BY_h(Z) \quad (4a)$$

where

$$Y_h(Z) = Z \exp\left(\frac{hZ^2}{2}\right) \quad (4b)$$

The function $Y_h(Z)$ is said to have the h distribution where the parameter h measures the elongation (or kurtosis) of X . If $h=0$, there is no elongation relative to the Gaussian distribution; for $h > 0$ or $h < 0$ the distribution exhibits thicker or thinner tails than the Gaussian distribution, respectively.

Analogous to the procedures used earlier, an estimate of h is obtained by first rewriting Equation 4a for the p th and $(1-p)$ th percentiles

of X , noting that $z_p = -z_{1-p}$, and subtracting x_{1-p} from x_p . This process yields

$$\frac{[x_p - x_{1-p}]}{2z_p} = B \exp\left(\frac{hz_p^2}{2}\right) \quad (5)$$

The numerator on the left-hand side of Equation 5 is the letter spread while the denominator measures the corresponding distance (letter spread) for a unit normal random variable. This value is defined as the pseudosigma, or p -sigma, and it measures the extent to which a distribution is more elongated than the Gaussian distribution. That is, a value of p -sigma greater than one implies a distribution with thicker tails than the Gaussian distribution. An estimate of h is obtained by regressing $\ln(p\text{-sigma})$ against $z_p^2/2$ for selected percentiles.

Skewness, elongation and the $(g \times h)$ distribution

Since skewness may induce elongation, or both may exist in a distribution, a joint assessment is necessary. Here, the $(g \times h)$ distribution is obtained by multiplying the g and h distributions. Now the random variables X and Z are such that

$$X = A + BY_{g,h}(Z) \quad (6a)$$

where

$$Y_{g,h}(Z) = \left(\frac{1}{g}\right) \{ \exp(gZ) - 1 \} \exp\left(\frac{hZ^2}{2}\right) \quad (6b)$$

In order to estimate h conditionally on g , or h^* , we rework Equations 6a and 6b as done earlier to arrive at

$$\frac{g(x_p - x_{1-p})}{(\exp(gz_p) - \exp(-gz_p))} = B \exp\left(\frac{hz_p^2}{2}\right) \quad (7)$$

The left-hand side of Equation 7 is called the Corrected Full Spread (CFS), and an estimate of h^* can be obtained by regressing $\ln(\text{CFS})$ on $z_p^2/2$.

III. Descriptive Statistics and Results

Data and descriptive statistics

The data for the sample are the time series of the daily adjusted closing price (adjusted for dividends and

splits) on an ETF that was continuously traded from 1 January 2003 to 31 December 2007. We calculate daily returns for each ETF as logarithmic price changes, that is, $\ln(p_t) - \ln(p_{t-1})$.² The source of the data is provided by <http://finance.yahoo.com>. The total number of ETFs in the sample is 112 and for each ETF there are 1258 days of data. For detailed analysis, the sample is partitioned into six groups. Funds are first classified by fund size: subgroups are (a) small funds (23) and (b) large funds (89). The size divisions reflect those used in the Morningstar investment style box. Given fund size, funds are then classified by style of investing: subgroups are (a) value, (b) blend and (c) growth. The appendix provides a detailed explanation of how the funds are grouped with respect to size and the style of investing.

Table 1 reports summary statistics and conventional measures of skewness and excess kurtosis³ for the data set as a whole as well as each of the six subgroups. The average daily return and SD for the entire data set are 0.062 and 1.138%, respectively. The entire data set and all subgroups display skewness that is, on average, negative; however, the skewness coefficient over the entire sample ranges from -0.65 to 0.78 . Not a single fund in the *Small value* or *Small blend* subgroup has a positive skewness coefficient. In addition, the average skewness coefficient for the *Small value* subgroup yields the most negative value, while the corresponding statistic for the *Large growth* subgroup generates the least negative statistic; interestingly, these two subgroups have relatively low excess kurtosis.

A return distribution with positive skewness has frequent small losses and a few extreme gains, while a return distribution with negative skewness has frequent small gains and a few extreme losses. Investors are likely to be attracted by positive skewness because the mean return falls above the median (see Elton *et al.*, 2003; Reilly and Brown, 2003, and the references therein). Relative to the mean return, positive skewness amounts to a limited, though frequent, downside compared with a somewhat unlimited, but less frequent, upside. Harvey and Siddique (2000b) show that an investor may be willing to accept negative expected return in the presence of large positive skewness.

The entire sample and all subgroups are, on average, more elongated than the Gaussian distribution.

² We have also replicated this study using market adjusted returns computed as $\ln(p_t/p_{t-1}) - \ln(m_t/m_{t-1})$, where m_t represents the market (S&P 500) price at time t . The results are broadly similar to those using unadjusted returns. These results are not included in the article for the sake of brevity.

³ Excess kurtosis is defined as kurtosis -3 ; for the Gaussian distribution, kurtosis equals 3. A positive (negative) value of excess kurtosis implies a distribution that is simultaneously more (less) peaked and has fatter (thinner) tails than the Gaussian distribution.

Table 1. Daily return characteristics of the sample of ETFs

	Mean	SD	Skewness (min, max)	Excess kurtosis	N
Entire sample	0.062	1.138	-0.18 (-0.65, 0.78)	2.40	112
<i>Small value</i>	0.062	1.072	-0.41 (-0.57, -0.21)	1.74	9
<i>Small blend</i>	0.058	0.980	-0.26 (-0.30, -0.22)	0.55	6
<i>Small growth</i>	0.061	1.363	-0.12 (-0.39, 0.26)	1.41	9
<i>Large value</i>	0.064	1.083	-0.22 (-0.65, 0.21)	2.40	34
<i>Large blend</i>	0.067	1.117	-0.16 (-0.55, 0.78)	3.91	29
<i>Large growth</i>	0.053	1.219	-0.07 (-0.42, 0.25)	1.69	25

Notes: All returns are in percent per day.
Excess kurtosis = Kurtosis - 3.

Most notably, the *Large value* and the *Large blend* subgroups, the subgroups containing the most ETFs, have the highest average excess kurtosis coefficients. Such return distributions have a greater percentage of extremely large deviations from the mean return. Most investors would perceive a greater chance of extremely large deviations from the mean as increasing risk. In fact, not a single ETF in the sample has negative excess kurtosis which would imply thinner tails than the Gaussian distribution. However, the *Small blend* subgroup has an average kurtosis coefficient of 0.55. Further analysis will reveal that some of the funds from this subgroup have elongation comparable to that of the Gaussian distribution.

Assuming for a moment that the means and the SDs are not significantly different from one another for the various subgroups, an investor may be inclined to avoid *Small value* and *Small blend* ETFs since all funds in these subgroups lack a positive skewness statistic. Further, an investor contemplating purchasing an ETF in the other four subgroups might focus solely on the fund within each subgroup with the maximum positive skewness and the smallest kurtosis statistic. However, these conventional measures of skewness and kurtosis do not reveal anything about the behaviour of skewness and kurtosis across the different tails of the distribution. Moreover, it is not possible to isolate any patterns from the data using only one measure. A more thorough analysis of the data will reveal patterns that are hidden from a cursory glance at the conventional measures for skewness and kurtosis. In many instances, an investor equipped with this detailed information might make a radically different decision on which ETF to either purchase or avoid.

Results from applying the *g*- and *h*-distributions

For expositional purposes, the ETFs with the minimum and maximum skewness coefficients from each subgroup are selected for analysis.

Table 2 reports the Sharpe ratio, conventional measures of skewness and kurtosis, as well as the *g*, *h* and *h** statistics for each of these ETFs. In general, funds with a negative (positive) conventional skewness coefficient yield negative (positive) median *g* values. There are funds, however, for which the relative magnitude of the robust measure given by the median *g* value is not consistent with its conventional skewness coefficient. For instance, BB2 Internet HOLDERS and Biotech HOLDERS have similar skewness coefficients with values of 0.26 and 0.25, respectively; however, the median *g* value of Biotech HOLDERS (0.059) is more than three times that of BB2 Internet HOLDERS' value (0.019). Moreover, BLDRS Developed Markets 100 ADR Index generates the highest skewness coefficient in the entire sample, yet its median *g* value of 0.032 is by no means the highest value. BLDRS Emerging Market 50 ADR has a slightly more negative skewness coefficient when compared to the value for iShares MSCI Austria Index, -0.42 versus -0.39, respectively, but its median *g* value is more than three and a half times less negative than that for iShares MSCI Austria Index, -0.016 versus -0.058, respectively. These results point to the fact that simply analysing the conventional measures for skewness and kurtosis may be misleading.

It is interesting to note that those ETFs with negative skewness tend to have the higher Sharpe ratios. Even though the Sharpe ratio does not include skewness and/or kurtosis explicitly, studies have shown that these higher moments are inherently priced. For example, Leland (1999) develops a model of market returns to show that investors seem to outperform the market if they are willing to accept negatively skewed returns.

Arguably, it appears as if ETF managers are in some way 'selling risk' in order to maintain good Sharpe ratios. Similarly, managers following a strategy of limiting downside risk are incorrectly underrated.

Table 2. Higher moments of the sample of ETFs

	Sharpe ratio ^a	Skewness	Kurtosis	Median <i>g</i>	<i>h</i>	<i>h</i> [*]
<i>Small value</i>						
Ishares Cohen&Steers Rty Majors	0.048	-0.57	2.54	-0.102 ^b	0.053 ^c	0.049 ^d
Ishares Russell 2000 Value Index	0.038	-0.21	0.73	-0.048 ^b	0.023 ^c	0.022 ^d
<i>Small blend</i>						
MidCap SPDRs	0.047	-0.30	0.42	-0.075 ^b	0.005	0.003
Vanguard Ext'd Mkt Ind VIPERs	0.055	-0.22	0.46	-0.037	0.001	0.000
<i>Small growth</i>						
Ishares MSCI Austria Index	0.087	-0.39	2.37	-0.058 ^b	0.070 ^c	-0.037 ^d
BB2 Internet HOLDERS	0.009	0.26	4.19	0.019	NA ^e	NA
<i>Large value</i>						
StreetTRACKS D.J. STOXX 50	0.048	-0.65	5.19	-0.059 ^b	0.101 ^c	0.099 ^d
Telecomhldrs HOLDERS	0.032	0.21	5.13	0.055 ^b	0.117 ^c	0.116 ^d
<i>Large blend</i>						
Ishares MSCI Australia Index	0.072	-0.55	2.10	-0.079 ^b	0.056 ^c	0.054 ^d
BLDRS Asia 50 ADR Index	0.038	0.78	36.55	0.032 ^b	0.289 ^c	0.289 ^d
<i>Large growth</i>						
BLDRS Emerging Market 50 ADR	0.059	0.42	5.85	-0.016 ^b	0.119 ^c	0.119 ^d
Biotech HOLDERS	0.036	0.25	2.74	0.059 ^b	0.085 ^c	0.056 ^d

Notes: ^aThe Sharpe ratio is calculated assuming a risk-free annual interest rate of 4%.

^bWhen upper values are regressed against lower values for the relevant subgroup, the slope is significantly different from -1 at the 5% significance level – indicating that the distribution is not symmetric.

^cWhen $\ln(p\text{-sigma})$ is regressed against $z_p^2/2$, the estimate of *h* (the slope) is significantly different from 0 at the 5% significance level – indicating elongation that deviates from the Gaussian distribution.

^dWhen $\ln(\text{CFS})$ is regressed against $z_p^2/2$, the estimate of *h*^{*} (the slope) is significantly different from 0 at the 5% significance level – indicating elongation that deviates from the Gaussian distribution.

^eThe values for *h* and *h*^{*} are not available (NA) for this fund due to the extraordinary number of 0% return values in the sample.

For a more detailed analysis of the *g*- and *h*-distributions, we focus on the two funds within the *Large blend* subgroup. Within this subgroup, the ETF with the most negative skewness is the ishares MSCI Australia Fund (symbol EWA) with a skewness statistic of -0.55, while the ETF with the most positive skewness is the BLDRS Asia 50 ADR Index Fund (symbol ADRD) with a skewness statistic of 0.78. What is particularly striking about ADRD is the kurtosis coefficient of 36.55, a value approximately 15 times the average kurtosis coefficient value in the entire sample. Tables 3 and 4 present sample upper and lower letter values as well as midsummaries for 13 percentiles for these two ETFs.

One should recall that for a symmetric distribution, a plot of the upper values against the lower values would form a line with a slope equal to -1 – implying letter values that are equidistant from the median. For EWA, when the upper letter values are regressed against the lower values, the slope has a value of -0.87. Further, this value is statistically different from the value of -1 at the 5% significance level. The results indicate a rather substantial departure from symmetry – a result reinforced once the *g* values

are analysed. For ADRD, a regression of the upper values against the lower values reveals a slope of -1.14 that also is statistically different from -1 at the 5% level; however, positive skewness is implied here since |-1.14| is greater than |-1.00|, and an inspection of the midsummaries (Table 4) indicates values that eventually increase as one moves further into the tails.

In order to further capture the behaviour of skewness in returns, *g* values are estimated for different letter values according to Equation 3 and are presented in Tables 3 and 4. In general, a series that exhibits constant values for its estimates of *g* tends to have a simple pattern of skewness; the lognormal distribution is such an example. For EWA, the median *g* value is -0.079 and 12 of the 13 *g* values are negative; however, there is considerable variation in the magnitude of the values. An examination of ADRD reveals a median *g* value of 0.032, however four of the 13 values are negative. The patterns of skewness in these series are far from simple and suggests that they cannot be adequately explained by skewness coefficients of -0.55 for ADRD and 0.78 for EWA.

Table 3. iShares MSCI Australia Index Fund, 1 January 2003–31 December 2007

Letter values for EWA						
Lower (1)	Upper (2)	Z_p (3)	Midsummary (4)	g (5)	p -sigma (6)	Corrected p -sigma (7)
0.162	0.162	0	0.162	0		
−0.538	0.896	−0.674	0.179	0.071	1.063	1.063
−1.186	1.408	−1.150	0.111	−0.069	1.127	1.126
−1.790	1.854	−1.534	0.032	−0.093	1.188	1.185
−2.594	2.307	−1.863	−0.144	−0.135	1.316	1.311
−3.437	2.678	−2.154	−0.380	−0.166	1.420	1.413
−4.039	3.166	−2.418	−0.437	−0.139	1.490	1.481
−4.379	3.498	−2.660	−0.440	−0.116	1.481	1.470
−4.895	3.825	−2.886	−0.535	−0.112	1.511	1.498
−5.443	4.546	−3.097	−0.448	−0.079	1.613	1.596
−5.716	5.229	−3.297	−0.243	−0.045	1.660	1.641
−5.858	5.636	−3.487	−0.111	−0.027	1.648	1.627
−5.930	5.840	−3.668	−0.045	−0.019	1.604	1.582
−5.965	5.942	−3.842	−0.012	−0.015	1.550	1.526

Note: Columns 1, 2, 4, 6 and 7 are presented in terms of percentage returns. See text for definition of terms.

Table 4. BDRS Asia 50 ADR Index Fund, 1 January 2003–31 December 2007

Letter values for ADRD						
Lower (1)	Upper (2)	Z (3)	Midsummary (4)	g (5)	p -sigma (6)	Corrected p -sigma (7)
0		0	0	0		
−0.409	0.579	−0.674	0.085	0.516	0.733	0.732
−0.945	1.097	−1.15	0.076	0.13	0.888	0.887
−1.554	1.611	−1.534	0.029	0.024	1.032	1.031
−2.278	2.264	−1.863	−0.007	−0.003	1.219	1.219
−3.178	3.137	−2.154	−0.021	−0.006	1.466	1.465
−4.917	4.858	−2.418	−0.029	−0.005	2.022	2.020
−8.808	11.532	−2.660	1.362	0.101	3.823	3.819
−12.346	12.166	−2.886	−0.090	−0.005	4.247	4.241
−14.055	14.422	−3.097	0.183	0.008	4.597	4.590
−14.892	16.529	−3.297	0.818	0.032	4.765	4.756
−15.219	17.701	−3.487	1.241	0.043	4.720	4.711
−15.383	18.287	−3.668	1.452	0.047	4.589	4.579
−15.465	18.580	−3.842	1.558	0.048	4.431	4.420

Note: Columns 1, 2, 4, 6 and 7 are presented in terms of percentage returns. See text for definition of terms.

In order to determine whether the return series has thicker or thinner tails than the Gaussian distribution, p -sigma estimates are calculated for different letter values using Equation 5 and are presented in Tables 3 and 4. A p -sigma value of 1 implies neutral elongation or that of the Gaussian distribution. An inspection of Table 3 reveals p -sigma estimates for EWA that appear greater than 1, suggesting fatter tails than the Gaussian. According to Table 4, the p -sigma estimates for ADRD are even greater than those obtained for EWA. An estimate of h is obtained

for each ETF by regressing $\ln(p\text{-sigma})$ against $z_p^2/2$ for the selected letter values. The h estimates for EWA and ADRD are 0.056 and 0.289, respectively. Further testing reveals that the estimates of h for both funds are significantly different from 0 at the 5% significance level, reflecting fatter tails than the Gaussian – a common result with financial return data.

In the above analysis of elongation, an implicit assumption of symmetry was maintained. Since skewness may induce elongation, a joint assessment

is necessary. Using median g estimates and Equation 7, corrected p -sigma estimates are calculated and are presented in Tables 3 and 4. Finally, estimates of h^* are obtained by regressing $\ln(\text{CFS})$ on $z_p^2/2$ for the selected letter values. The h^* estimates for EWA and ADRD are 0.054 and 0.289, respectively, where both estimates are statistically significant at the 5% level.

Within the *Large blend* subgroup, judging EWA based on its conventional skewness statistic of -0.55 and kurtosis statistic of 2.10 would appear to be misguided. Exploratory data techniques for this fund reveal a small negative median g estimate as well as an h estimate that is significantly different from 0. Further, since EWA's g estimates are quite variable across percentiles even the median estimate is not representative of its overall skewness. Similar findings are found once the other subgroups are analysed. For example, within the *Small blend* subgroup, the Vanguard extended market index VIPER fund has one of the higher Sharpe ratios, but a negative skewness coefficient of -0.22 and a kurtosis coefficient of 0.46 . Further, this fund has a median g estimate of -0.037 and h and h^* estimates of 0.001 and 0.000 , respectively. When upper values are regressed against lower values the slope is not statistically different from -1 – indicating a distribution that is more or less symmetric. In addition, the estimates of h and h^* are not significantly different from 0 – reflecting elongation comparable to the Gaussian distribution. An investor who avoids this fund on the basis of its relatively large negative skewness coefficient and kurtosis coefficient of 0.46 might be making a poor decision.

IV. Simulation Based on GARCH Models

The GARCH models are used in modelling financial time series that exhibit time-varying volatility clustering, for example, periods of swings followed by periods of relative calm. These models have further been extended to include time-varying conditional higher moments. As mentioned earlier, the explanatory data analysis of Tukey is attractive for its computational ease. It is also considered robust to complex patterns of skewness and kurtosis in distributions. In this section we study how well this exploratory analysis captures the patterns found in commonly used GARCH family of models where the higher moments may or may not vary over time. In particular, we simulate data based on such models to analyse the sampling behaviour of the g , h and h^* estimators.

Consider continuously compounded returns $r_t = 100 * \ln(P_t/P_{t-1})$ for $t = 1, 2, \dots, T$, where $r_t = \mu_t + \varepsilon_t$. The GARCH(1, 1) models the residual of a times series regression as $\varepsilon_t = \sigma_t z_t$, where z_t is i.i.d. with $E(z_t) = 0$, $\text{Var}(z_t) = 1$ and conditional volatility is specified as $\sigma_t^2 = a_0 + b_0 \varepsilon_{t-1}^2 + c_0 \sigma_{t-1}^2$. We use 1000 observations to simulate the GARCH models; in fact, 1050 observations are considered but the first 50 observations are discarded to remove any influence from initial values. The choice of parameters is consistent with the available evidence on market returns. In particular, we use $\mu = 0.019$, $\sigma_0 = 0.956$ along with $a_0 = 0.06$, $b_0 = 0.05$ and $c_0 = 0.90$ for the analysis.

The above parameter values are used to simulate the GARCH models where the residuals are drawn from the Gaussian, Student's t , and generalized Student's t -distributions. Although the Student's t -distribution allows for variations in the tail thickness, it is considered restrictive since it is not consistent with a stylized fact that stock market returns are skewed. The generalized Student's t -distribution offers flexibility in that it not only allows excess kurtosis (as in the standard Student's t -distribution) but also skewness (Hansen, 1994). The two parameter density function of this distribution that is normalized to have zero mean and unit variance is

$$g(z|\eta, \lambda) = \begin{cases} bc \left(1 + \frac{1}{\eta-2} \left(\frac{bz+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & z < -a/b \\ bc \left(1 + \frac{1}{\eta-2} \left(\frac{bz+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2} & z \geq -a/b \end{cases} \quad (8)$$

where $2 < \eta < \infty$ and $-1 < \lambda < 1$. The constants are given by

$$a = 4\lambda c \left(\frac{\eta-2}{\eta-1}\right), \quad b^2 = 1 + 3\lambda^2 - a^2 \text{ and} \\ c = \frac{\Gamma((\eta+1)/2)}{\sqrt{\pi(\eta-2)}\Gamma(\eta/2)} \quad (9)$$

This generalized distribution allows positive ($\lambda > 0$) as well as negative ($\lambda < 0$) skewness. Further it specializes to the Student's t -distribution when $\lambda = 0$ and to the Gaussian distribution for $\lambda = 0$, $\eta \rightarrow \infty$. For simulations, we use the degrees of freedom parameter $\eta = 6$ to allow for excess kurtosis. For departure from symmetry, we use $\lambda = -0.25$ for negative skewness and $\lambda = 0.25$ for positive skewness.

Kim and White (2004) use Monte Carlo simulations to demonstrate that the conventional measures of skewness and kurtosis are very sensitive to outliers since they are based on the sample mean which is known to be an inadequate measure

Table 5. Simulation analysis of the g , h and $(g \times h)$ distributions. GARCH models with constant higher moments simulated with the Gaussian and Student's t -distributions

Statistics	Gaussian			Student's t ($df = \eta = 6$)		
	g	h	h^*	g	h	h^*
Median (Mean)	0.000 (0.000)	-0.019 (-0.019)	-0.020 (-0.019)	0.000 (0.000)	0.058 (0.061)	0.057 (0.060)
$P_{0.5}(P_{99.5})$	-0.075 (0.078)	-0.051 (0.020)	-0.051 (0.019)	-0.122 (0.132)	-0.001 (0.162)	-0.001 (0.161)
$P_{2.5}(P_{97.5})$	-0.057 (0.059)	-0.043 (0.010)	-0.043 (0.010)	-0.093 (0.093)	0.011 (0.129)	0.010 (0.128)
$P_5(P_{95})$	-0.048 (0.049)	-0.040 (0.005)	-0.040 (0.005)	-0.078 (0.077)	0.017 (0.114)	0.017 (0.114)
$P_{25}(P_{75})$	-0.019 (0.019)	-0.028 (-0.010)	-0.028 (-0.010)	-0.032 (0.032)	0.040 (0.078)	0.039 (0.077)

Notes: Results are based on 5000 replications using a random sample of 1000 observations drawn from the various distributions. An allowance is made for outliers. P_x denotes the x -th percentile value of the bootstrap distributions of g , h and h^* .

Table 6. Simulation analysis of the g , h and $(g \times h)$ distributions. GARCH models with constant higher moments simulated with the generalized Student's t -distribution

Statistics	$\eta = 6; \lambda = -0.25$			$\eta = 6; \lambda = 0.25$		
	g	h	h^*	g	h	h^*
Median (Mean)	-0.167 (-0.167)	0.056 (0.060)	0.047 (0.050)	0.167 (0.167)	0.056 (0.060)	0.047 (0.050)
$P_{0.5}(P_{99.5})$	-0.290 (-0.041)	-0.002 (0.166)	-0.010 (0.152)	0.037 (0.293)	-0.003 (0.175)	-0.010 (0.159)
$P_{2.5}(P_{97.5})$	-0.258 (-0.069)	0.009 (0.131)	0.001 (0.118)	0.069 (0.260)	0.011 (0.131)	0.002 (0.117)
$P_5(P_{95})$	-0.244 (-0.085)	0.016 (0.116)	0.007 (0.103)	0.086 (0.245)	0.017 (0.116)	0.008 (0.103)
$P_{25}(P_{75})$	-0.199 (-0.136)	0.038 (0.078)	0.029 (0.067)	0.135 (0.199)	0.038 (0.078)	0.029 (0.067)

Notes: Results are based on 5000 replications using a random sample of 1000 observations drawn from the various distributions. An allowance is made for outliers. P_x denotes the x -th percentile value of the bootstrap distributions of g , h and h^* .

in such instances. Although they do not consider the g , h and h^* statistics in their analysis, they do offer some other robust measures from the statistics literature. We follow their approach and use a mixture distribution that allows for outliers. In particular, if $\{r_t\}$ is generated from $D(\mu, \sigma)$ with probability p and from $D(\mu_2, \sigma_2)$ with probability $1 - p$, then the random numbers used for simulations are generated by $(p)D(\mu, \sigma) + (1 - p)D(\mu_2, \sigma_2)$. In the presence of outliers ($p < 1$), we need to determine the relationship of (μ_2, σ_2) with (μ, σ) . Following Kim and White, we use the daily S&P index returns to approximate $p = 0.9988$, $\mu_2 = \mu - 7$ and $\sigma_2 = 10\sigma$.

We use 5000 replications in our Monte Carlo analysis. The bootstrapping technique is used to analyse the sampling properties of the g , h and h^* estimators since their theoretical distribution is not known. These simulations shed light on the ability of the g , h and h^* statistics to capture excess kurtosis and skewness when the data are generated by a GARCH process with outliers. In particular we use the percentile method which, for a given significance level α , simply uses the $\alpha/2$ and $(1 - \alpha/2)$ percentiles of the bootstrap distribution to define the $(1 - \alpha)100\%$ confidence interval for

a given parameter. The confidence interval is then used to conduct a two-sided test for g , h and h^* . For instance, the null hypothesis, $H_0: g = 0$, is rejected at a given α if the percentile interval, $[P_{\alpha/2}, P_{1-\alpha/2}]$, for g does not include the hypothesized value of zero.

In Tables 5 and 6 we report some descriptive statistics for g , h and h^* in the GARCH models using the above-mentioned distributions for innovations. When the distribution is Gaussian, the sample mean and the sample median of g are both zero, implying that there is no evidence of skewness in the data. Further, since the 95% confidence interval for g using $P_{2.5}$ and $P_{97.5}$ is given by $[-0.057, 0.059]$, we cannot reject the claim that the data are symmetric ($g = 0$) at $\alpha = 0.05$. For the Student's t -distribution with six degrees of freedom, the symmetry is correctly captured at all significance levels. However, there is statistically significant evidence of excess kurtosis at the 10 and 5% levels. For example, at $\alpha = 0.05$, the entire ranges for h and h^* , given by $[0.011, 0.129]$ and $[0.010, 0.128]$, respectively, are positive, thus rejecting the null hypotheses $H_0: h = 0$ as well as $H_0: h^* = 0$.

Results based on the generalized Student's t -distribution for the residuals are presented in Table 6. We continue to use six degrees of freedom,

Table 7. Simulation analysis of the g , h and $(g \times h)$ distributions. GARCH models with time varying higher moments simulated with the generalized Student's t -distribution

Statistics	$\bar{\eta} \simeq 6; \bar{\lambda} \simeq -0.25$			$\bar{\eta} \simeq 6; \bar{\lambda} \simeq 0.25$		
	g	h	h^*	g	h	h^*
Median (Mean)	-0.173 (-0.172)	0.061 (0.065)	0.051 (0.055)	0.161 (0.160)	0.060 (0.064)	0.051 (0.055)
$P_{0.5}(P_{99.5})$	-0.320 (-0.020)	0.001 (0.176)	-0.007 (0.161)	0.007 (0.293)	0.002 (0.172)	-0.008 (0.164)
$P_{2.5}(P_{97.5})$	-0.281 (-0.059)	0.012 (0.147)	0.003 (0.130)	0.048 (0.263)	0.011 (0.140)	0.003 (0.129)
$P_5(P_{95})$	-0.262 (-0.078)	0.019 (0.127)	0.010 (0.114)	0.065 (0.246)	0.017 (0.124)	0.009 (0.113)
$P_{25}(P_{75})$	-0.209 (-0.136)	0.042 (0.084)	0.032 (0.072)	0.124 (0.196)	0.041 (0.083)	0.032 (0.073)

Notes: Results are based on 5000 replications using a random sample of 1000 observations drawn from the various distributions. An allowance is made for outliers. P_x denotes the x -th percentile value of the bootstrap distributions of g , h and h^* .

denoted by $\eta = 6$. One should recall that the skewness parameter falls in the interval $-1 < \lambda < 1$. When we use $\lambda = -0.25$, the 95% confidence interval for g is $[-0.258, -0.069]$. This range is entirely negative suggesting statistically significant negative skewness. Similarly, for $\lambda = 0.25$, the range $[0.069, 0.260]$ indicates positive skewness. Of course, the power of the tests will decrease (increase) as $|\lambda|$ approaches zero (one).

Although the hypotheses for h and h^* are difficult to interpret for skewed distributions, they are broadly suggestive of excess kurtosis. For instance, for $\lambda = -0.25$, the confidence intervals for h and h^* , given by $[0.009, 0.131]$ and $[0.001, 0.118]$, respectively, are entirely positive. We would also like to point out that for all distributions the sample mean of g equals its sample median whereas for h and h^* , the mean is consistently a little higher than the median. This result may be useful in developing the asymptotic distributions for these estimators.

In order to include time-varying higher moments, we treat the parameters of the generalized Student's t -distribution as functions of the conditioning information (Hansen 1994; Jondeau and Rockinger, 2003). Consider $\tilde{\eta}_t = a_1 + b_1 \varepsilon_{t-1} + c_1 \tilde{\eta}_{t-1}$ and $\tilde{\lambda}_t = a_2 + b_2 \varepsilon_{t-1} + c_2 \tilde{\lambda}_{t-1}$ where the transformations

$$\eta_t = 2 + \frac{28}{1 + \exp(-\tilde{\eta}_t)} \quad \text{and} \quad \lambda_t = -1 + \frac{2}{1 + \exp(-\tilde{\lambda}_t)}$$

are used to ensure that $2 < \eta < \infty$ and $-1 < \lambda < 1$. For simulations we use $a_1 = -0.36$, $b_1 = 0.12$, $c_1 = 0.80$ for computing $\tilde{\eta}_t$ and $a_2 = 0.12$ (or $a_2 = -0.12$), $b_2 = 0.20$, $c_2 = 0.775$ for $\tilde{\lambda}_t$. We choose these parameter values to keep the problem tractable and comparable to the analysis with constant higher moments. The unconditional sample means of these parameters are $\bar{\eta} \simeq 6$ and $\bar{\lambda} \simeq 0.25$ (or $\bar{\lambda} \simeq -0.25$).

In Table 7, we use the exploratory analysis of the data that are generated by time-varying higher

moments and compare it with the Table 6 results that are based on constant higher moments. It is noteworthy that the time-varying η_t and λ_t parameters introduce some extra noise in the data which makes the skewness and kurtosis estimators less precise. However, the results in Tables 6 and 7 are qualitatively similar. For instance, when $\lambda = -0.25$, the 95% confidence interval for g based on time-varying higher moments, given by $[-0.281, -0.059]$, is still entirely negative. However, this interval is slightly wider than the comparable $[-0.258, -0.069]$ range implied by higher moments that do not vary with time. Similarly, although the 95% confidence intervals for h and h^* , given by $[0.009, 0.131]$ and $[0.001, 0.118]$, respectively, are slightly wider, they still infer a statistically significant excess kurtosis in the data.

In summary, we find that the exploratory data analysis of Tukey is attractive not only for its computational ease but also for its flexibility and robustness to various GARCH specifications with outliers. Monte Carlo simulations suggest that the g , h and h^* estimators are able to capture the skewness and excess kurtosis found in commonly used GARCH family of models where the higher moments may or may not vary over time. A preliminary explanatory analysis can actually be used as a tool for identifying a relevant GARCH model. We would like to point out that these estimators are likely to be robust to more complex patterns than those implied by the GARCH models that we considered in our simulations.

V. Conclusion

The vast majority of ETF return distributions point to distributions that are highly non-Gaussian when conventional measures of skewness and kurtosis are calculated. This evidence alone suggests the

inadequacy of the traditional two-parameter, i.e. mean–variance, model of portfolio diversification. Moreover, conventional measures of skewness and kurtosis might give misleading information concerning the true behaviour of financial returns. These measures are known to be sensitive to outliers and do not reveal anything about the behaviour of skewness and kurtosis across the tails of the distribution.

We found that investors should not rely on single measure of skewness and kurtosis to summarize ETF return distributions. The g , h and $(g \times h)$ distributions provide robust parameter estimates that are not always consistent with their conventional counterparts. Moreover, we find substantial variation in skewness and kurtosis for individual ETFs. The robust estimators of higher moments help us discover some trends when the funds are grouped by fund size and style of investing. We also find that ETFs with negative skewness tend to have higher Sharpe ratios. This result seems to suggest that ETF managers are perhaps ‘selling risk’ in order to maintain good Sharpe ratios and managers following a limiting downside risk strategy are incorrectly underrated. Finally, Monte Carlo simulations suggest that these exploratory techniques are able to capture patterns found in commonly used GARCH family of models.

Acknowledgements

This article has benefited immensely from insightful comments of an anonymous referee. We also thank Carlos Morales for useful comments. Any errors are ours alone.

References

- Akgiray, V. and Booth, G. G. (1988) The stable-law model of stock returns, *Journal of Business and Economic Statistics*, **6**, 51–7.
- Badrinath, S. G. and Chatterjee, S. (1988) On measuring skewness and elongation in common stock return distributions: the case of the market index, *Journal of Business*, **61**, 451–72.
- Badrinath, S. G. and Chatterjee, S. (1991) A data-analytic look at skewness and elongation in common stock return distributions, *Journal of Business and Economic Statistics*, **9**, 223–33.
- Christoffersen, P., Heston, S. and Jacobs, K. (2006) Option valuation with conditional skewness, *Journal of Econometrics*, **131**, 253–84.
- Dittmar, R. (2002) Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *The Journal of Finance*, **57**, 369–403.
- Dutta, K. K. and Babbel, D. F. (2005) Extracting probabilistic information from the prices of interest rate options, *Journal of Business*, **78**, 841–70.
- Elton, E., Gruber, M., Brown, S. and Goetzmann, W. (2003) *Modern Portfolio Theory and Investment Analysis*, 6th edn, Wiley, New York.
- Hansen, B. E. (1994) Autoregressive conditional density estimation, *International Economic Review*, **35**, 705–30.
- Harvey, C. and Siddique, A. (2000a) Conditional skewness in asset pricing tests, *The Journal of Finance*, **55**, 1263–95.
- Harvey, C. and Siddique, A. (2000b) Time-varying conditional skewness and the market risk premium, *Research in Banking and Finance*, **1**, 27–60.
- Hoaglin, D. (1985) Using quantiles to study shape, in *Exploring Data Tables, Trends, and Shapes* (Eds) D. Hoaglin, F. Mosteller and J. Tukey, Wiley, New York, pp. 417–60.
- Jondeau, E. and Rockinger, M. (2003) Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements, *Journal of Economics Dynamic and Control*, **27**, 1699–737.
- Kim, T.-H. and White, H. (2004) On more robust estimation of skewness and kurtosis, *Finance Research Letters*, **1**, 56–73.
- Kon, S. (1984) Models of stock returns – a comparison, *Journal of Finance*, **39**, 149–65.
- Leland, H. (1999) Beyond mean-variance: performance measurement in a nonsymmetrical world, *Financial Analysts Journal*, **55**, 27–36.
- Mills, T. (1995) Modelling skewness and kurtosis in the London stock exchange FT-SE index return distributions, *The Statistician*, **44**, 323–32.
- Reilly, F. and Brown, K. (2003) *Investment Analysis and Portfolio Management*, 7th edn, South-Western, Mason, OH.
- Tukey, J. (1977) *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

Appendix: Morningstar Style Box

The Morningstar style box is a tool that represents the characteristics of a fund in a graphical format. There are two pieces of data that determine where the fund falls within the style box. The first piece of data is market capitalization; that is, the size of a fund. Large funds, medium-sized funds and small funds are placed in the top row, the middle row and the bottom row, respectively, of the style box. Morningstar calculates the market capitalization of each stock in the fund and accounts for its weighting in the fund in order to arrive at a number that best represents how the fund is positioned in the style box.

Further, Table A1 shows that the Morningstar style box incorporates the size of the fund relative to the region where the fund invests. That is, an ETF that tracks the United States with a market capitalization of \$300 (in millions of dollars) is categorized as a small fund; whereas, an ETF tracking Canada with the same market capitalization is categorized as a medium-sized fund.

For the sample in this article, Morningstar defined 89 large ETFs, 14 medium-sized ETFs and 9 small ETFs. In order to conduct a detailed analysis given the style of investing and ensure a large enough sample with respect to subgroups, it was necessary to combine medium-sized funds with small-sized funds. This sum, 23 funds, is referred to as small funds in this article.

Table A1. Market capitalization breakpoints in millions of US dollars

Region	Large	Medium	Small
United States	9225.41	1585.19	449.14
Canada	3896.93	754.35	168.71
Latin America	3769.03	1058.55	291.29
Europe	8037.51	1501.40	353.77
Japan	3793.53	757.13	210.01
Australia/New Zealand	3626.43	703.62	125.00
Asia ex-Japan	1981.89	297.78	87.20

The other factor that determines a fund's placement in the style box is its investment style. Morningstar uses a number of statistics for each stock in a fund (long-term projected earnings growth, historical earnings growth, sales growth, price/projected earnings, price-to-book, etc. relative to other stocks in its market-cap range) and calculates a growth score and a value score – each score will range from 0 to 100. Morningstar arrives at a stock's investment style by subtracting its value score from its growth score. A stock with a strongly negative score is assigned to value, and one with a strongly positive score is assigned to growth. Those in between are categorized as blend. The fund's overall style is based on the weighted average of the style scores for all of its stocks.