Citation: Neuberger, A. & Payne, R. G. (2021). The Skewness of the Stock Market at Long Horizons. The Review of Financial Studies, 34(3), pp. 1572-1616. doi: 10.1093/rfs/hhaa048

The Skewness of the Stock Market at Long Horizons

Anthony Neuberger*

Cass Business School, City, University of London

Richard Payne

Cass Business School, City, University of London

Abstract

Higher moments of long-horizon returns are important for asset pricing but are hard to measure accurately using standard techniques. We provide theory showing that short-horizon (e.g. daily) returns can be used to construct precise estimates of long-horizon (e.g. annual) moments without making strong assumptions about the data generating process. Skewness comprises two components: skewness of short-horizon returns and a leverage effect, i.e. covariance between variance and lagged returns. We provide similar results for kurtosis. An application to US stock-index returns shows that skew is large and negative and attenuates only slowly as one moves from monthly to multi-year horizons.

* We are grateful to the editor, Stijn Van Nieuwerburgh, to two anonymous referees, and to Hendrik Bessembinder and Ian Martin for helpful comments. We also thank seminar participants at Université Paris-Dauphine, Illinois, Maynooth, Lancaster, Brunel, Norges Bank (UK) and the London Quant Group.

Corresponding author: Anthony Neuberger, Cass Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK. email: anthony.neuberger.1@city.ac.uk.

This paper makes two contributions, one methodological, the other empirical. The methodological contribution is to show that short horizon returns can be used to estimate the higher moments of long horizon returns while making only weak assumptions about the data generating process. The empirical contribution is to show that long horizon (multi-year) US equity market returns are highly negatively skewed. The skew coefficient in 1-year returns, at -1.4, is significant economically as well as statistically. We also show that skew at long horizons is entirely attributable to the leverage effect – the negative correlation between variance and lagged returns. Additionally, we use our technology to investigate the skewness at high frequencies using five-minute return data.

The existence of non-normalities in daily stock returns has been recognised for at least half a century (Mandelbrot 1963, and Fama 1965). There is good reason to believe that higher moments of returns – not just second moments – are important for asset pricing. A large theoretical literature, starting with Kraus and Litzenberger (1976), and continuing with the macroeconomic disaster research of Rietz (1988), Longstaff and Piazzesi (2004) and Barro (2006), demonstrates that heavy-tailed shocks and left-tail events in particular may play an important role in explaining asset prices. Barberis and Huang (2008) and Mitton and Vorkink (2007) argue that investors look for idiosyncratic skewness, seeking assets with lotto-type pay-offs. There is much empirical evidence suggesting that market skewness is time varying, and that it predicts future returns in both the time series (Kelly and Jiang 2014) and in the cross-section (Harvey and Siddique 2000, and Ang et al. 2006). Amaya et al. (2015) show that realized skew measures computed from intra-day return data on individual stocks can be used to sort stocks into portfolios that have significantly different excess returns, while Boyer, Mitton and Vorkink (2010) and Conrad, Dittmar and Ghysels (2013) show that high idiosyncratic skewness in individual stocks is also correlated with positive returns. Ghysels, Plazzi and Valkanov (2016) present similar results for emerging market indices.

There are two serious problems in measuring higher moments at the long horizons (e.g. years) of interest to asset pricing. First, the higher the moment, the more sensitive the estimate is to outliers. Second, the longer the horizon, the smaller the number of independent observations in any fixed span of data. We show how these problems can be mitigated by using information in short horizon returns to make better estimates of skewness and kurtosis of long horizon returns.

It is standard practice to use high frequency data to estimate the second moment of long horizon returns. Under the assumption that the price process is martingale, the annualized variance of returns is independent of the sampling frequency and the realized variance computed from high frequency returns is a good estimate of the variance of long horizon returns. But this does not hold for higher moments - there is no necessary relationship between the higher moments of long and short horizon returns. If daily returns are highly skewed but i.i.d., then annual returns will display little skew. And conversely, as Engle points out, "it is now widely recognized that asymmetric volatility models generate multi-period returns with negative skewness even if the innovations are symmetric." (Engle 2011, p455). Similar statements can be made about kurtosis.

In this paper we demonstrate how to exploit the information in short horizon returns to estimate the skewness and kurtosis of long horizon returns. The only assumptions are that the price process is martingale, and that the relevant moments exist. We prove that the skewness of long horizon returns comes from one of two sources: the skewness of short horizon returns; and the leverage effect, that is the covariance between squared returns and lagged returns. Similarly, the kurtosis of long horizon returns has just three sources: the kurtosis of short horizon returns; the covariance between cubed returns and lagged returns; and the covariance between squared returns and lagged squared returns (which we refer to as the GARCH effect). The foundation of these results is a simple expansion of long horizon returns in terms of short horizon returns and, via the martingale assumption, imposing the requirement that short horizon returns are not predictable with rolling averages of their own past levels, their own past squared values and their own past cubed values.

When we take these theoretical results to the data, we show that the skewness of the US stock market at long horizons is large and negative and due almost entirely to the leverage effect. Kurtosis in long horizon returns is driven by the GARCH effect. The negative skewness and the excess kurtosis in annual stock market returns owe virtually nothing to the shape of the daily return distribution. As a result, bootstrapping techniques, which simulate long horizon returns by repeated sampling of a population of short horizon returns (as in Fama and French 2018), do not match the leverage and GARCH effects found in the market, and so fail to reproduce the higher moments found in actual long horizon returns. Similarly, the skewness of five-minute returns estimated using a week of data, as in Amaya et al (2015), provides a poor estimate of the skewness of the weekly return.

To date, the literature has used a variety of approaches to measure the higher moments of long horizon returns. The most straightforward is to apply the standard estimators to historic returns. Kim and White (2004) show that these estimators are subject to large estimation errors.[1] They advocate the use of robust estimators such as those developed by Bowley (1920), which are based on the quantiles of the observed distribution. The attraction is that quantiles can be estimated with much greater precision than moments. This solution is used in Conrad, Dittmar and Ghysels (2013) and the methodology is further developed in Ghysels, Plazzi and Valkanov (2016). The weakness of the approach is that it assumes that the body of the distribution, which

[1] To estimate the skewness (kurtosis) of a normally distributed random variable with a standard error of 0.1 requires a sample size of 600 (2400). Even for monthly returns, this would require 50 (200) years of returns data. If returns are non-normal, the standard errors are generally substantially higher. Monthly returns on the US market over the last 50 years have a skew coefficient of -0.98; the bootstrapped standard error is 0.3.

is captured by the quantiles, is highly informative about the behaviour of the tails, which determine the higher moments.

Kelly and Jiang (2014) follow an alternative approach. They focus on the tails. They get power not by taking a very long time-series, but by exploiting the information in the cross-section of stock prices and making assumptions about the commonality in tail events. The validity of their method depends not only on the assumed decomposition of the tail component, but also on assumptions about the dependence of returns across stocks.

The options market is an attractive source of information about moments. Whereas the underlying market shows just one realization of the price process, the options market reveals the entire implied density of returns. The technology for extracting implied skewness and kurtosis from options prices is well-established (Bakshi, Kapadia and Madan 2003). However, the method can only be used on assets – such as the major market indices - that support a liquid options market. There is also a more fundamental problem in using option implied moments: they reflect risk premia as well as objective probabilities. As demonstrated by Broadie, Chernov and Johannes (2007), the wedge between the objective price process and the process as implied by option prices (the so-called risk-neutral process) can be very wide. We present evidence that option-based estimates do in fact substantially overstate the degree of skewness of the US stock market.

Our approach requires weaker assumptions than other alternatives to the standard estimators. We show by simulation that our estimators of long-horizon skewness and kurtosis are indeed substantially more powerful than standard estimators, reducing standard errors on skewness by around 60% and on kurtosis by around 30%. This is true for all the data generating processes that we use in our simulations. Focussing on skew estimation, we show that our method works pretty much equally well regardless of how skewed returns actually are and that our estimation technique is substantially more precise than a simple quantile-based skew estimator.

Previous researchers have often discussed the distribution of asset returns without making much of the distinction between simple returns and log returns. But the moments of simple and log returns differ, particularly at long horizons. To obtain our results, throughout this paper we use non-standard definitions of the moments, with our moments generally falling between the standard moments of simple returns and log returns. The definitions we use are set out and justified in section 1.2 below.

We apply our technology to the US equity market using data from the past ninety years. Our analysis suggests that the skew coefficient of monthly returns is about -1.49. This skewness attenuates only slowly with horizon. Using data from the entire 90-year span of data, we estimate the skewness of annual returns to be -1.41 and of five year returns to be -0.66. Thus, long-term investors should not think that the left-tail events that are worrisome in daily or monthly returns wash away when one aggregates to annual or longer horizons.

These estimates of skewness are far from zero. To understand their magnitude, consider a binomial model of excess stock market returns. With an annual volatility of 17% and an expected return of zero, a symmetric distribution has log returns at {+4.9%, -4.9%} each month. To get a skew of -1.49 with the same mean and volatility requires the log returns to be {+2.4%, -9.8%}.

To get a sense of the economic importance of this level of skew, consider a representative agent with CRRA preferences, a relative risk aversion coefficient γ, and a horizon of 1 year. To persuade the investor to invest 100% of their wealth in the market, the Euler condition requires

$$\mathbb{E}\left[\left(e^{r}-1\right)e^{-\gamma r}\right]=0,$$

where r is the annual log excess return. Expanding the expression to the third order, the risk premium is

$$\mathbb{E}\left[e^{r}-1\right] \approx \gamma \operatorname{var}[r]-\frac{1}{6} \gamma(3 \gamma-1) \operatorname{var}[r]^{3/2} \operatorname{skew}[r]. \tag{1}$$

With volatility of 17%, the required equity risk premium required is

Risk Aversion	Skew = 0	Skew = -1.41
$\gamma = 1$	2.9%	3.1%
$\gamma = 3$	8.7%	11.4%
$\gamma = 5$	14.5%	22.5%

With low levels of risk aversion, the skew risk premium is small. But while the variance risk premium is proportional to γ, the skew risk premium increases with γ^2 and, as the table shows, is significant at moderate levels of risk aversion.

The variance risk premium, expressed as an annual rate, is independent of horizon since variance is linear with horizon. By contrast, equation (1) shows that the component of the premium attributable to negative skew aversion increases with horizon unless the skew coefficient itself attenuates with the square root of the horizon – hence the importance of understanding the behaviour of skew with horizon. We explore this relationship in our empirical work.

1. THE THEORY

1.1 Moments of price changes

We work in a discrete time setting, $t \in \mathbb{Z}$, with a price process $P = \{P_t\}_{t\in\mathbb{Z}}$. The asset pays no dividends and the risk-free rate is normalized to zero. We are concerned with the distribution of returns from time t to $t+T$. For brevity, we refer to one time increment as a day, and T days as a month, but obviously nothing hangs on this.

The problem we are interested in is

> [**P**]: Let P be a strictly positive martingale process, whose associated daily continuously compounded returns process r, where $r_t = \ln\left(P_t/P_{t-1}\right)$, is strongly stationary. The long horizon returns process R is defined by $R_t\left(T\right) = \ln\left(P_t/P_{t-T}\right)$. How can one estimate higher moments of long horizon returns R efficiently, assuming these moments exist?

Problem **P** is difficult because it deals with returns (ratios) rather than with price changes (differences). We therefore first address a simpler problem, **P***, and use the solution as a guide to solving **P**.

The simpler problem is:

> [**P***]: Let P be a real-valued (not necessarily positive) martingale process whose associated daily *difference* process d, where $d_t = P_t - P_{t\text{-}1}$, is strongly stationary. The long horizon difference process D is defined by $D_t(T) = P_t - P_{t\text{-}T}$. How can one estimate higher moments of D efficiently, assuming these moments exist?

The particular moments that we are interested in are the variance, skewness and kurtosis of the price changes. In this section, where we are concerned with price changes, we use the standard (Pearson) definitions. Variance is the second central moment of price changes, skewness is the third central moment, scaled by the variance to the power of 3/2, and excess kurtosis (which we refer to simply as kurtosis) is the fourth central moment scaled by the squared variance, minus three. Elsewhere in the paper we use non-standard definitions.

The solution to **P*** is given by

Proposition 1

The variance, skewness and kurtosis of monthly price changes is related to the distribution of daily price changes in the following way

$$
\begin{aligned}
&\operatorname{var}\left[D(T)\right]=T\operatorname{var}\left[d\right];\\
&\operatorname{skew}\left[D(T)\right]=\left(\operatorname{skew}\left[d\right]+3\frac{\operatorname{cov}\left[y^{(1*)},d^2\right]}{\operatorname{var}\left[d\right]^{3/2}}\right)T^{-1/2};\\
&\operatorname{kurt}\left[D(T)\right]=\left(\operatorname{kurt}\left[d\right]+4\frac{\operatorname{cov}\left[y^{(1*)},d^3\right]}{\operatorname{var}\left[d\right]^{2}}+6\frac{\operatorname{cov}\left[y^{(2*)},d^2\right]}{\operatorname{var}\left[d\right]^{2}}\right)T^{-1};
\end{aligned}
\tag{2}
$$

where

$$
\begin{aligned}
y_t^{(1*)} &:= \sum_{u=1}^{T}\left(P_{t-1}-P_{t-u}\right)/T=\sum_{u=1}^{T-1}D_{t-1}\left(u\right)/T \quad \text{and}\\
y_t^{(2*)} &:= \sum_{u=1}^{T}\left(P_{t-1}-P_{t-u}\right)^2/T=\sum_{u=1}^{T-1}D_{t-1}\left(u\right)^2/T.
\end{aligned}
\tag{3}
$$

Proof: in appendix A.

Proposition 1 gives expressions for the variance, the skewness and the excess kurtosis of monthly price changes. The first result is familiar: the variance of price changes scales linearly with horizon. The second result says that skew at the monthly horizon has just two sources: daily skew and a term we call leverage. The contribution of daily skew attenuates with the square root of the aggregation horizon. The leverage term is proportional to the covariance between the current squared price change and the quantity $y^{(1*)}$, which is equal to the difference between the first lag of the price and the average price over the previous month.

The final result says that the kurtosis of monthly returns has three sources: daily kurtosis attenuating with time, the covariance between cubed price changes and $y^{(1*)}$, and the covariance between squared price changes and $y^{(2*)}$. $y^{(2*)}$ is a measure of the average squared price change over the past month.

In order to demonstrate the logic underlying Proposition 1 (and indeed the main result in this paper, Proposition 2) and also the role of the assumptions (martingale, strict stationarity), it is useful to review the proof of one part of the proposition, that concerning skewness.

Start with an algebraic decomposition of the third power of the monthly price change

$$D_t(T)^3 = \sum_{u=1}^{T} d_{t-u+1}^3 + 3\sum_{u=1}^{T-1} D_{t-u}(T-u)\, d_{t-u+1}^2 + 3\sum_{u=1}^{T-1} D_{t-u}(T-u)^2\, d_{t-u+1}. \tag{4}$$

Take conditional expectations of both sides. Here we can employ the martingale assumption with its implication that price changes are unpredictable to drop the third term on the right-hand side

$$\mathbb{E}_{t-T}\left[D_t(T)^3\right] = \sum_{u=1}^{T} \mathbb{E}_{t-T}\left[d_{t-u+1}^3\right] + 3\sum_{u=1}^{T-1} \mathbb{E}_{t-T}\left[D_{t-u}(T-u)\, d_{t-u+1}^2\right]. \tag{5}$$

Define

$$y_t^{(1*)} := \sum_{u=1}^{T-1} D_{t-1}(u)/T. \tag{6}$$

$y_t^{(1*)}$ is the difference between yesterday's price and the moving average of prices over the month ending yesterday.[2] Using strict stationarity, the conditional expectations can be replaced by unconditional expectations. Substituting $y_t^{(1*)}$ in (5) gives the unconditional third moment

[2] The asterisk superscript is used to distinguish this variable from the corresponding variable in the problem **P**; the 1 superscript refers to the first moment of price changes.

$$\mathbb{E}\left[D(T)^3\right]=T\left(\mathbb{E}\left[d^3\right]+3\mathbb{E}\left[y^{(1*)}d^2\right]\right). \tag{7}$$

As $y^{(1*)}$ is mean zero, the second expectation term on the right-hand side can be replaced by the covariance, giving

$$\mathbb{E}\left[D(T)^3\right]=T\left(\mathbb{E}\left[d^3\right]+3\operatorname{cov}\left(y^{(1*)},d^2\right)\right). \tag{8}$$

A similar argument, using the absence of serial correlation in returns, shows that

$$\mathbb{E}\left[D(T)^2\right]=T\mathbb{E}\left[d^2\right]. \tag{9}$$

The result in proposition 1 then follows immediately from the definition of the skewness coefficient.

Our main interest in Proposition 1 (and more particularly, in Proposition 2 which we are coming to) is the efficiency gain in estimating long horizon moments that comes from the use of high frequency data. It is worth therefore examining the source of those gains. The martingale assumption means that the covariance between returns and functions of lagged returns is zero, so we drop terms going from equation (4) to (5). Similar considerations apply in the case of other moments of price changes, and of the return moments we encounter in the next section. Dropping terms does not necessarily increase efficiency.[3] The existence and size of the efficiency gain from using high frequency data is an empirical issue and depends on the

[3] For an example, suppose one wants to estimate the variance of two-day price changes. One could use the two-day changes and find the sample average of $\left(d_1+d_2\right)^2$. Alternatively, one could use daily changes, throw out the $2d_1d_2$ term and estimate the sample average of $\left(d_1^2+d_2^2\right)$. If d_2 is highly skewed with a skew that is negatively correlated with d_1, so that $E\left[d_1d_2^3\right]<E\left[d_1^2d_2^2\right]$, the first is more efficient.

data generating process. We therefore explore it using both simulated and real data in sections 2 and 3 of this paper.

1.2 Moments of Returns

We now return to the problem of estimating the moments of long-horizon returns. We could simply replace the assumption that prices are martingale and price changes are stationary with an alternative assumption that log prices are martingale and log returns are stationary. Then, we could immediately recycle the technology of Proposition 1 to derive the moments of long-horizon log returns with no modification.

We choose to proceed under the alternative assumption that prices are martingale and price ratios are stationary for three main reasons. First, it is justified by the seminal work of Samuelson (1965), while the assumption that log prices are martingale has no similar theoretical foundation. Second, we want to apply the theory to the price of any financial asset. If the log prices of individual securities are martingale, the log prices of any positive portfolio of those securities, such as the market index, will in general be a supermartingale. Finally, we wish to compare the realized moments with option implied moments. This enables us to measure skewness risk premia, following the logic of papers such as Carr and Wu (2008). In a world where log prices are martingale, and there is therefore an asset risk premium, the difference between actual moments and option-implied moments is no longer a clean measure of the moment risk premium.

Thus we start from the assumption that prices are martingale and returns are stationary. The objective is to produce a result akin to Proposition 1 but applied to the moments of returns rather than to the moments of price changes. It turns out that we cannot do this with the standard definitions of variance, skewness and kurtosis, whether applied to simple returns or to log returns. In this section, we show why this is the case, and develop a set of definitions of moments that converge to the standard definitions when returns are small and which achieve our objective.

Much of the academic literature on long horizon returns concerns the moments of log returns (for example Campbell and Hentschel 1992 and, in the implied moments literature, Bakshi, Kapadia and Madan 2003). On the other hand, in the portfolio choice and investment performance literatures, moments are more often computed using simple returns (as in classic papers like Simkowitz and Beedles 1978, or Singleton and Wingender 1986, or, more recently, Bessembinder 2018).

The distinction between the distributions of simple and log returns is important, particularly at long horizons. Take the workhorse geometric Brownian model, with zero drift and annualized volatility of 17%. At the one-year horizon, the standard deviation of simple returns is 17.10% and of log returns is 17.00%. At the ten-year horizon the figures are 57.89% and 53.76% respectively. With higher moments the distinction is even more important. Simple returns are positively skewed (skew coefficients of 0.52 and 1.93 at 1 and 10 years) and leptokurtic (excess kurtosis of 0.48 and 7.28), while log returns are symmetrical and have zero excess kurtosis.

One cannot transpose the results of Proposition 1 to the world of returns using standard definitions of moments. Recall that in deriving Proposition 1, there were two crucial steps – decomposition and elimination. The decomposition step (equation (4)) uses the fact that the monthly price change is the sum of the daily price changes. This makes it possible to express the moments of monthly changes as moments and comoments of daily price changes. In the elimination step (equation (5)), the terms in the expansion that are comoments of price changes with functions of past changes are thrown out because, under the martingale assumption, the conditional expectation of price changes is zero. Decomposition by itself is not sufficient; the elimination step is necessary to provide the power for the method.

With log returns, the decomposition step works fine – the monthly log return is the sum of the daily log returns. But the elimination step does not work – the fact that the price process is martingale does not imply that the conditional expectation of the daily log return is zero. No terms can be eliminated and there is no gain in power from using daily returns. With simple returns things are no better. In this case it is the decomposition step that does not work. Simple monthly returns cannot be expressed as the sum of daily returns.

The problem is not just that the particular derivations we have used are no longer valid when applied to returns, but the results themselves do not hold. Even the simplest insight – that the variance of monthly returns is the sum of the variance of daily returns – does not hold for a general martingale process whether log returns or simple returns are used.[4]

Many alternatives to the Pearson definitions have been discussed in the statistics literature (see Joanes and Gill 1998). In the words of Balanda and McGillivray (1988) "like location, scale, and skewness, kurtosis should be viewed as a "vague concept" that can be formalized in many ways." We depart from the Pearson definitions, and use instead definitions designed to allow both the decomposition and the elimination steps to work. The definitions are derived in appendix B. Given the log return r, we define the moments ("NP moments") as

$$
\begin{aligned}
&\mathrm{var}^L[r] := \mathbb{E}\left[x^{(2L)}(r)\right] \text{ where } x^{(2L)}(r) := 2\left(e^r - 1 - r\right);\\
&\mathrm{var}^E[r] := \mathbb{E}\left[x^{(2E)}(r)\right] \text{ where } x^{(2E)}(r) := 2\left(re^r - e^r + 1\right);\\
&\mathrm{skew}[r] := \frac{\mathbb{E}\left[x^{(3)}(r)\right]}{\mathrm{var}^L[r]^{3/2}} \text{ where } x^{(3)}(r) := 6\left(\left(e^r + 1\right)r - 2\left(e^r - 1\right)\right);\\
&\mathrm{kurt}[r] := \frac{\mathbb{E}\left[x^{(4)}(r)\right]}{\mathrm{var}^L[r]^2} - 3 \text{ where } x^{(4)}(r) := 12\left(r^2 + 2\left(e^r + 2\right)r - 6\left(e^r - 1\right)\right).
\end{aligned}
\tag{10}
$$

[4] This is obviously true for simple returns since the returns over successive days combine multiplicatively. Log returns combine additively, but with the mean log return depending on the volatility, serial correlation in volatility induces serial correlation in log returns. This prevents variance being additive.

The definitions approximate the standard definitions.[5] Specifically, $x^{(n)}(r)/r^n$ tends to 1 as r tends to zero, and $x^{(n)}(r)$ lies in the range $[r^n, (e^r\text{-}1)^n]$ for $n = 2L, 2E$, 3, and 4. But before demonstrating that they do allow us to transpose Proposition 1 from changes to returns, it may be useful to say a little more about the properties of these moment definitions.

var^L is the widely used model free implied variance. It is the basis of the VIX contract – see CBOE (2018). $x^{(2L)}$ is used by Bondarenko (2014) to measure realized variance. Neuberger (2012) uses $x^{(2L)}$, $x^{(2E)}$ and $x^{(3)}$ to measure the second and third moments (the L and E notation we use comes from that paper where it distinguishes the log and entropy functions).

The NP moments have attractive properties. The geometric Brownian martingale process with constant volatility σ has variance of $\sigma^2 T$ at horizon T, and zero skewness and kurtosis. This property is not shared by arbitrary functions that approximate the square, cube and quartic functions. Furthermore, NP skewness is zero not only for the lognormal, but for any distribution where implied volatility is a symmetric function of log-moneyness, and which therefore has put-call symmetry (see Carr and Bowie 1994).

Table 1 shows the moments of daily, monthly and annual returns using the three different definitions. The underlying returns are US stock market returns 1926-2015.

[INSERT TABLE 1 ABOUT HERE]

At the daily horizon, the three approaches yield reasonably similar numbers, but at long horizons the differences become quite large. With the NP moment functions lying between simple and log power functions, NP skewness and kurtosis tend to lie between the

[5] With the martingale assumption, returns are centred on zero, so ordinary and central moments are identical. We discuss later in this section how to refine the moment definitions when we relax the martingale assumption and allow for a non-zero drift in prices.

corresponding simple and log statistics, and generally rather closer to the log moments. Figure 1 shows the times series relationship between the NP, simple and log statistics for the computation of monthly skewness. At the beginning of every year, skew is computed using overlapping 25-day returns and a rolling sample of 5 years of data. The three curves usually lie very close together and, as in Table 1, when it is possible to distinguish the three curves, the NP statistic tends to lie closer to the skewness of log returns. In the internet appendix we provide a plot similar to Figure 1 but for kurtosis; the conclusions are similar.

[INSERT FIGURE 1 ABOUT HERE]

With these definitions, we can now state the main theoretical result of this paper

Proposition 2

If P is a strongly stationary martingale process, the variance, skewness and kurtosis of monthly returns (as defined in equation (10)) are related to the distribution of daily returns as follows

$$\begin{aligned}
&\mathrm{var}^{L}\left[R(T)\right]=T\mathrm{var}^{L}\left[r\right];\\
&\mathrm{skew}\left[R(T)\right]=\left(\mathrm{skew}\left[r\right]+3\frac{\mathrm{cov}\left[y^{(1)},x^{(2E)}(r)\right]}{\mathrm{var}^{L}\left[r\right]^{3/2}}\right)T^{-1/2};\\
&\mathrm{kurt}\left[R(T)\right]=\left(\mathrm{kurt}\left[r\right]+4\frac{\mathrm{cov}\left[y^{(1)},x^{(3)}(r)\right]}{\mathrm{var}^{L}\left[r\right]^{2}}+6\frac{\mathrm{cov}\left[y^{(2L)},x^{(2L)}(r)\right]}{\mathrm{var}^{L}\left[r\right]^{2}}\right)T^{-1};
\end{aligned}\tag{11}$$

where

$$y_t^{(j)}:=\sum_{u=1}^{T-1}x^{(j)}\left(R_{t-1}(u)\right)\Big/T\quad\text{for } j=1,2L.\tag{12}$$

Proof: the proof is similar to Proposition 1; details in appendix C.

Proposition 2 implies that

- The variance of returns increases linearly with the horizon.
- The skew in daily returns generates a much smaller $\left(1/\sqrt{T}\right)$ skew in monthly returns.
- If monthly returns have significant skew, it must be through the leverage effect, the correlation between variance and lagged returns. Lagged returns are measured by $y^{(1)}$, which is the net return relative to the one month moving average over the previous month.[6]
- Kurtosis in daily returns generates a much smaller ($1/T$) kurtosis in monthly returns.
- If monthly returns are significantly leptokurtic, it is for one of two reasons:
    - because daily skew is correlated with past returns (as measured again by $y^{(1)}$);
    - or because current variance is correlated with past variance. Past variance is measured by $y^{(2L)}$, which is a function of the average realized variance over horizons of up to one month, with more recent data having more weight.

We make use of proposition 2 for estimating long horizon moments. The second, third and fourth moments of long horizon returns are expressed as the sum of variances and covariances of functions of high frequency returns. An unbiased estimate of the right-hand side variances and covariances therefore gives us an unbiased estimate of the long horizon moments.

The results are quite general; there is no presumption about any functional form for the stochastic process driving the price. For example, in a Merton (1976) jump-diffusion model, the asymmetric jump creates skewness and kurtosis in high frequency returns. The absence of any covariation between variance and lagged returns or lagged squared returns ensures that there is no leverage or GARCH effect, so skewness and kurtosis attenuate rapidly with the horizon. In a Heston (1993) model there is no skewness or conditional kurtosis in short horizon

[6] The moving average in this case is the rolling harmonic mean.

returns, but there is skewness in longer horizon returns because correlation between innovations in returns and innovations in volatility, coupled with the persistence of volatility, creates a correlation between variance and lagged returns. The persistence of volatility shocks also creates kurtosis. GARCH processes also display kurtosis in long horizon returns through the persistence of volatility shocks. To generate skewness in long horizon returns in a model from the GARCH family, one must additionally have volatility reacting asymmetrically to positive and negative return shocks (as with Heston), so creating a correlation between volatility and lagged returns.

Proposition 2 relates the skewness of long horizon returns to the leverage effect – the covariance between instantaneous realized variance and lagged returns. Neuberger (2012) relates skewness to the covariance between returns and contemporaneous changes in option implied variance. In appendix D we show that in the presence of a sufficiently rich options market, and in the absence of variance risk premia, the two estimators are both unbiased, and the option based estimator is more precise since the options market makes expected future variance observable. In section 3.7, we show that variance risk premia do in fact create a substantial wedge between the two estimates, at least in the case of the US stock market index, highlighting the importance of estimators that are not based on option prices.

The analysis assumes that prices are martingale. But Proposition 2 can readily be extended to the case where the price exhibits a constant drift. Suppose in particular that there is some constant m such that

$$\mathbb{E}_t\left[P_{t+u}\right] = P_t e^{mu} \text{ for all } t,u > 0. \tag{13}$$

With a non-zero drift m, the mean net return is non-zero, and central and ordinary moments of returns no longer coincide. This poses a problem since the moments defined in (10) are ordinary moments while it is central moments that are required to compute variance, skewness and kurtosis. The problem is readily rectified by amending the definitions of r_t and R_t to make them centered returns, so

$$\begin{aligned} r_t &:= \ln\left(P_t/P_{t-1}\right) - m; \\ R_t &:= \ln\left(P_t/P_{t-T}\right) - mT. \end{aligned} \tag{14}$$

With this reinterpretation, Proposition 2 goes through, the proof being applied to the demeaned price series $P_t^* = P_t e^{-mt}$, which is martingale.

The extension to Proposition 2 does come at a cost. The true mean m is generally not known and is estimated with error. The estimation error reduces the efficiency of the moment estimators. We explore this further in simulations.

With much evidence that the equity risk premium is not only non-zero but also time-varying, the question of the robustness of Proposition 2 to the presence of a time-varying mean return arises. For the existence of a time-varying mean to be problematic for our proposition, it would have to generate some correlation between short horizon returns and particular functions of past returns. Specifically, in deriving the variance result for some specific horizon T, the proof in appendix C requires that r is uncorrelated with $y^{(1)}(T)$, the lagged rolling return computed over the horizon T. The skewness result requires in addition that r is uncorrelated with $y^{(2E)}(T)$, lagged rolling squared returns. The kurtosis result requires that r is uncorrelated with lagged rolling returns, squared and cubed returns ($y^{(1)}$, $y^{(2L)}$, and $y^{(3)}$) calculated over a period of T. Thus time-varying expected returns are not a problem in this context per se, but they could generate bias in our estimates if the correlations mentioned above are significantly different from zero.

In our empirical work, where we apply the technology to equity market returns, we test whether these assumptions are violated in our data (section 3.6). We fail to reject the assumptions at conventional significance levels at the relevant horizons.

1.3 The term structure of moments in continuous time

So far, we have worked in a discrete time setting. Given that data is discrete, this makes it easy to implement our results in practice. The results simplify in going to continuous time and are more intuitive. We have assumed that *P* is a positive martingale. We now make the additional assumption that the process is a diffusion, in particular that *P* can be represented by a stochastic differential equation

$$dP_t / P_t = \sqrt{v_t}\, dz_t, \tag{15}$$

where v_t is predictable, and z_t is a standard Brownian process. We retain the definitions of variance, skewness and kurtosis that we use in discrete time. The counterpart to Proposition 2 in a diffusion setting is then

Proposition 3

If *P* is a strongly stationary martingale diffusion, the variance, skewness and kurtosis of *T*-period returns are related to the variance of instantaneous returns *v* as follows

$$\begin{aligned} \mathrm{var}^L(T) &= TE[v]; \\ \mathrm{skew}(T) &= 3\frac{\mathrm{cov}\left[y^{(1)}(T), v\right]}{E[v]^{3/2}} T^{-1/2}; \\ \mathrm{kurt}(T) &= 6\frac{\mathrm{cov}\left[y^{(2L)}(T), v\right]}{E[v]^{2}} T^{-1}; \end{aligned}$$

where

$$y_t^{(j)}(T) := \int_{u=0}^{T} x^{(j)}\left(R_t(u)\right) du \Big/ T \quad \text{for } j \in \{1, 2L\}.$$

Proof: in appendix E.

The most significant difference between Propositions 2 and 3 is the dropping of the daily skewness from the long-horizon skewness, and the dropping of the daily kurtosis and the cube effect from the long-horizon kurtosis. With the diffusion assumption, the higher order moments of high frequency returns vanish. The definitions of $y^{(1)}$ (the lagged return) and $y^{(2L)}$ (the lagged realized variance) are the natural limits of their discrete time counterparts.

Proposition 3 is useful in exploring the asymptotic properties of the moments. The impact of past returns and past volatility on future volatility is unlikely to be permanent. If there is no impact beyond some time T^*, then the Central Limit Theorem then kicks in.

Corollary to Proposition 3

> If, for any $T > T^*$, $\mathrm{cov}[y^{(1)}(T), v] = \mathrm{cov}[y^{(1)}(T^*), v]$ and $\mathrm{cov}[y^{(2L)}(T), v] = \mathrm{cov}[y^{(2L)}(T^*), v]$, then skew($T$) and kurt($T$) tend to zero as T tends to infinity with the leading terms being of order $T^{-1/2}$ and T^{-1} respectively.

The effect can be seen in simulated data. If multi-year returns are generated by randomly sampling monthly returns and chaining them together, as in Fama and French (2018), $T^* = 1$ month by construction, and multi-year returns have little skewness and kurtosis.

2. SIMULATION RESULTS

2.1 Results for variance, skewness and kurtosis

We now evaluate the performance of our estimators of higher moments through a series of simulation experiments, focusing particularly on bias and efficiency. We compare our estimators both with standard methods and with a quantile-based approach.

Returns are simulated from three different models; a geometric Brownian motion (GBM), a Heston model and a Heston model with a jump component. For each model we simulate 10,000 paths for daily returns, each of length 5000 (i.e. roughly 20 years).[7] The parameters for the models are taken from Eraker (2004) who fits all of them to daily S&P-500 returns 1980-1999.

Given the parameters for a particular data generating process, the objects that we wish to measure are the standard deviation, skewness and kurtosis of 25-day returns, where these are as defined in equation (10). We construct the sample moments of overlapping 25-day returns[8] (we refer to these as 'Monthly' estimates), and compare them with the estimators from Proposition 2 (which we label 'NP'), and the true moments that we obtain by simulation. Prior to estimating moments via any method, the data from each run of the simulation are adjusted by subtracting the sample mean return for that run.

[INSERT TABLE 2 ABOUT HERE]

Results from these simulations are given in Table 2. Panel A shows the simulation results when returns are generated by a GBM, Panel B gives simulation results for the Heston model and Panel C shows the Heston model with jumps. Each panel gives statistics on the distribution of estimates from all three estimation techniques and for each of the three moments from across the 10,000 sample paths. In the discussion below, we focus on skewness and kurtosis estimates. If daily returns follow a GBM, 25-day skewness and kurtosis should both be zero.

Table 2 confirms that, on average, this is true for both estimation techniques. However, the dispersion of the estimates for the NP method are greatly reduced relative to those based on monthly returns. The standard deviations of estimates from our method are between 65% and

[7] We have also run simulations using an EGARCH model, with parameters estimated from daily value-weighted CRSP US stock returns covering the period 1980-2015. These simulation results provide no new qualitative results and are available on request.

[8] So for each simulated return path of 5,000 data points, the 'Monthly' estimator uses 4,976 overlapping 25-day returns.

80% smaller than those from the alternatives. The improvement in estimation accuracy for the NP method is most striking for skewness, but only slightly less impressive for kurtosis.

For the Heston model, we expect excess kurtosis (as the variance of daily returns is changing through time) and negative skew (as the innovations to the variance and the return are negatively correlated). Both estimation techniques pick these features up, but again the NP method results in a significant reduction in the spread of estimated moments.

For skewness, the standard deviation of estimates for the new method is around 60% smaller than the traditional approach, while for kurtosis improvements are between 40% and 50%.

Finally, Panel C shows that the addition of jumps to the Heston model does not change the conclusion that the NP estimation technique delivers a substantial improvement in accuracy. The jumps add a similar amount of variance to both estimators, thus reducing the efficiency gain from the NP method when expressed as a percentage reduction in standard deviation.

Overall, regardless of which model we choose or which moment one focusses on, use of the estimators described in Proposition 2 leads to much more precise estimates of monthly return moments. Improvements are greater for skewness estimates than they are for kurtosis and are larger for the models without jumps than they are for the model with a jump component. In the internet appendix we report results for 15000 days (60 years) and 1000 days (4 years). The conclusions remain valid. The efficiency gain in estimating skewness using the NP method is insensitive to the sample length, while in the case of kurtosis there is a significant increase in efficiency as the sample length increases.

2.2 Performance using intra-day data

Given the improvements in estimation precision we have shown from using daily data to estimate moments of monthly data, it is natural to ask how the use of intra-day data might further improve accuracy. Andersen et al. (2003) demonstrate the value of using finely sampled intra-day data to estimate volatility. We explore whether the same is true for higher moments.

We adjust our simulations from the previous section to generate data sampled at N_D equally spaced intervals across 1 day. We assume that the data generating process is the same across the day (thus ignoring issues like overnight periods) and we vary N_D between 1 (daily data) and 128. We begin by assuming that returns are generated from a Geometric Brownian Motion and then move to Heston models with and without jumps, using the same (daily) parameters as in the previous section.

[INSERT TABLE 3 ABOUT HERE]

The simulation results for the NP estimator only are given in Table 3. Panel A shows exactly what one would expect in the GBM case. The use of intra-day data increases the precision of the NP estimators of monthly moments with the ratio of the standard deviation of the daily estimator to that of the intra-day estimators equal to roughly $\sqrt{N_D}$. Thus, for example, sampling data 64 times a day rather than daily reduces the standard deviation of the distribution of moment estimates by a factor of approximately eight.

Results for the Heston model are given in Panel B. Here, the intra-day data deliver no gains in the precision with which one can estimate monthly standard deviations. For higher moments, if data are sampled 128 times per day, then the precision with which skewness and kurtosis are estimated improves by about 20%.

Panel C shows the results for the Heston specification with jumps, where sampling 128 times per day only improves the precision of the estimates of skewness and kurtosis by around 5% relative to the daily case.

These results for the Heston processes are linked to the persistence in volatility that the model displays. The (daily) mean reversion coefficient for the return variance is 0.017 and therefore volatility is close to a random walk. Sampling such a persistent process more finely does not help materially in estimating, for example, the covariance between variance and lagged returns and so our estimators derive little benefit from the use of intra-day data in this case.

2.3 Performance of the NP skew estimator across skew levels

In order to investigate how the performance of our skew estimator changes with the level of skew in returns, we take a Heston model and vary the correlation between return and variance innovations between -0.9 and +0.9 (with the former giving large negative skewness and the latter generating large positive skewness). All other parameters are set at the values from Eraker (2004). As before, for each parameter set, our simulation contains 10,000 replications of 5,000 daily returns and from these we estimate 25-day skew.

[INSERT FIGURE 2 ABOUT HERE]

The results are summarised in Figure 2. The x-axis of this figure shows the correlation parameter from the Heston model. Against each correlation parameter, we plot the average estimated skewness from our 10,000 runs, as well as the 5th and 95th percentiles of the distribution of skew estimates. Also plotted on Figure 2 is the theoretical value of the coefficient that one should obtain from the Heston model at each parameter value.

The Figure demonstrates that the NP estimator does an excellent job of tracking skewness, on average, across the range of parameters. There is a slight tendency for the estimator to be biased towards zero when the theoretical skew is large, with the largest bias around 0.05 when theoretical skewness is at a value of 0.9. The range between 5th and 95th percentiles is fairly stable at around 0.3. The bias in the estimation of the coefficient of skewness is due to the fact that it is a ratio of moments. While the estimates themselves are unbiased, the estimate of the ratio is biased.

2.4 Comparison of NP and quantile-based skew measures

In their paper on international asset allocation, Ghysels, Plazzi and Valkanov (2016) (hereafter GPV) propose a skewness estimator based on the quantiles of the return distribution. We now compare the performance of our estimator and their preferred estimator based on data simulated from a Heston model with jumps. We use exactly the same setup as in Section 2.3, except now we estimate NP skewness and GPV's quantile skewness for each simulated set of data.

The quantile skew estimator is as follows

$$6\times\frac{\int_0^{0.5}\left\{\left[q_\alpha\left(R_t\right)-q_{0.5}\left(R_t\right)\right]-\left[q_{0.5}\left(R_t\right)-q_{1-\alpha}\left(R_t\right)\right]\right\}d\alpha}{\int_0^{0.5}\left[q_\alpha\left(R_t\right)-q_{1-\alpha}\left(R_t\right)\right]d\alpha}\times\frac{\int_0^{0.5}q_\alpha\left(z\right)d\alpha}{\int_0^{0.5}q_\alpha^2\left(z\right)d\alpha}, \tag{16}$$

where R_t are returns measured at the frequency of interest (e.g. monthly), $q_\alpha(x)$ is the α'th quantile of the distribution of x and the $q_\alpha(z)$ are the quantiles of the standard Normal distribution. In their implementation, GPV approximate the integrals in the first ratio by aggregating across the following set of quantiles: [0.99, 0.975, 0.95, 0.90, 0.85, 0.80, 0.75]. The quantile estimator looks at the symmetry (or lack of it) of α and 1-α quantiles with respect to the median. This is captured by the numerator of the first ratio in the equation while the other terms are just scaling factors.

For each simulation run, we apply the quantile estimator to overlapping 25-day returns. It is worth re-iterating that the quantile estimator and the estimator proposed here are designed to target slightly different measures of skewness. The former estimates the traditional skewness coefficient whereas our estimator is of the modified skew coefficient as defined in equation (10). However, differences in these targets are minor.

The results from our comparison are displayed in Figure 3. As before, the x-axis values are the Heston correlation parameters and skewness is on the y-axis, and again we run 10000 simulations of 5000 daily returns from which we estimate 25 day skewness. The results are encouraging. The NP and the quartile mean estimates are very close together, but the precision of the NP estimator is greater. The 5^{th}-95^{th} percentile range of the quantile estimates is around 0.6 on average. The comparable number for the NP estimates is 0.5.

[INSERT FIGURE 3 ABOUT HERE]

Thus, overall, our estimation technique works well. Applied to simulated data, it is more precise than competing estimators and its precision shows little variation as the parameters of the chosen model change.

3. APPLICATION TO THE US EQUITY MARKET

In this section, we apply our technology to the US stock market. Unless stated otherwise, the returns used in this analysis run from 1935-2015 and were retrieved from Ken French's data library (mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). As in the simulation analysis from the previous section, we work with demeaned returns everywhere.

First, we document how moments of annual and monthly returns have evolved over the last ninety years, and the importance of the components of each monthly moment as described in Proposition 2. We then focus on skew and characterize the term structure of skewness and the relationship between our skew measure and those derived from options markets.

Before proceeding, it is worth addressing the question of bias. In proving Proposition 2, we showed that the NP long horizon moments can be expressed as the sum of short horizon moments and covariance terms. We need to estimate these terms without bias to ensure that the long horizon moment estimate is also unbiased.

This is not entirely straightforward. For example, the third moment requires an estimate of cov[y_{t-1} , v_t] over some period [0, S] where y_{t-1} is the lagged return and v_t is the daily variance $x^{(2E)}(r_t)$. The obvious estimator is the sample covariance

$$Q_1 := \sum_{t=1}^{S} (y_{t-1} - \overline{y})(v_t - \overline{v}) / (S-1) \quad \text{where } \overline{y} := \sum_{t=1}^{S} y_{t-1} / S.$$

But this is biased. Specifically

$$\mathbb{E}[Q_1] = c_0 - \sum_{j=1}^{S} \frac{S-j}{S(S-1)} (c_j + c_{-j}) \text{ where } c_j := \mathrm{cov}\left[y_{t+j-1}, v_t \right].$$

In our context, y is a multi-period return variable that is persistent by construction and v is an instantaneous variance which is also persistent. The cross correlations between the series are substantial, so the bias is significant. The bias of Q_1 arises from the fact that the means of y and v are estimated in-sample. We can avoid the bias by estimating the means *ex ante*. In our empirical work, we set the estimated means of y and v to their population means. Denoting these means by y_0 and v_0, the covariance estimator we use is

$$Q_2 := \sum_{t=1}^{S} (y_t - y_0)(v_t - v_0) / S.$$

We estimate the coefficients of skewness and kurtosis by scaling the third and fourth moments by the second moment in the standard way. Since the coefficients are non-linear functions of the moments they may well be biased; indeed, the simulations in the previous section provide some evidence of bias. There has been much discussion in the literature on the bias in the standard Pearson coefficient estimators which arises in the same way (see Bao 2013, for an overview); with the bias being dependent on the data generating process, correcting for bias in far from straightforward. We do not attempt to correct for bias in our coefficient estimates, leaving that for future work.

3.1 Moments of annual returns

We begin by looking at the time-variation in annual (by which we mean 250 day) return moments across our data set. The annual moments are estimated using a rolling window of 1250 days of data (and thus are autocorrelated by construction).

[INSERT FIGURE 4 ABOUT HERE]

The top left panel of Figure 4 shows the (log) market level across our sample. Its estimated volatility is shown in the top right panel (along with a 95% confidence interval). Over the 90 or so years that our data cover, US stock market volatility is initially high (around the Great Depression) and also high at the end (due to the 2008 Financial Crisis). In between volatility is smaller, punctuated by infrequent upward spikes. There was, for example, substantial market volatility around the oil price shocks of the early 1970s and the stock market crash of 1987.

The NP skew data in the bottom left panel indicates that annual stock market skew is almost always significantly negative (with a mean close to -1.25) over our 90 years of data. For a few years in the 1980s and a couple in the 1990s, skew turns marginally positive, but it is almost never significantly above zero. Times of particularly severe negative skewness include the Great Depression (with skew at -2.5) and the mid-1940s (with skew at -3) and skew has also reached a level below -2 in the most recent part of our data. Overall, there is very clear evidence of consistently large, significant and negative skew in annual US stock market returns.

The bottom right panel of Figure 4 shows estimates of annual excess return kurtosis estimated using daily data. As expected, excess kurtosis is positive on average, with a mean close to 4, and is rarely negative. Excess kurtosis is larger in the first half of the 20th century than it is in the second half. During the second half of the century there is a noticeable upwards spike in excess kurtosis around the financial crisis of 2008.

There does appear to be evidence of substantial and persistent shifts in the return distribution over time. This is inconsistent with our assumption of the return series being strictly stationary. The strict stationarity assumption was required to allow us to replace conditional with unconditional expectations and thus simplify the expressions in Propositions 1 and 2. If we are interested in conditional expectations, the absence of stationarity causes some boundary effects at the beginning and end of periods, but they are not significant.[9]

Overall, our estimates of higher moments suggest that long-term investors (i.e. those with an annual time horizon) should not assume that the negative skew and fat tails we see in daily returns wash away as the return measurement horizon is extended.

3.2 Moments of monthly returns from non-overlapping years of data

While annual moments are interesting from an investment risk perspective, previous authors have focussed on monthly measures of higher moments (e.g. tail risk and quantile-based measures of skew). Thus in this section we present the same information as in Section 3.1 but for 25-day moments. We take each year of the sample separately and use data from within that year to compute monthly volatility, skew and excess kurtosis. Time-series plots of the three

[9] When we measure monthly moments over a period say from beginning of January to end December, we use Proposition 2 and compute covariances of moments of daily returns over that year with lagged moments. The resulting estimates are not exactly estimates of the moments of monthly returns starting from 1 January to 31 December because they do not include data from the following January which are part of the December returns, and put too much weight on returns in the January of the year. These boundary effects are small, and have no net effect if stationarity is assumed.

moments estimated using the NP method and the quantile based skew measure are presented in Figure 5.[10]

[INSERT FIGURE 5 ABOUT HERE]

While Figure 5 leads to broadly the same results as Figure 4 (i.e. the US stock market return is on average very negatively skewed and displays excess kurtosis), monthly skew and monthly kurtosis are much more volatile than their annual counterparts, as one would expect. A negative correlation between monthly skew and excess kurtosis becomes clearer, however. When monthly skew is large and negative, monthly kurtosis tends to be large and positive.

The quantile-based skew measure, in the bottom right panel, is also negative on average, but it is less easy to see a pattern in those monthly skews than in the NP estimates. The quantile skew and NP skew measures are positively correlated, with a correlation coefficient of 0.40, but the mean levels are very different (-0.25 for the quantile measure compared with -1.54 for NP).

[INSERT FIGURE 6 ABOUT HERE]

Figure 6 compares two estimates of the skew coefficient of 25 day returns using rolling 2 year periods. The first method ("NP") is our estimate, using Proposition 2 and high frequency (daily) returns, the second method ("monthly") is computed directly from the overlapping 25 day returns using the skew definition in equation (10). The 90% confidence intervals, obtained from a stationary bootstrap with mean block length equal to 50, are shown. The NP interval is usually entirely below the x-axis, while the monthly interval almost always contains zero.[11] Thus, with the traditional approach, the skewness

---

[10] The quantile based skew measure is estimated from the set of overlapping 25 day returns that can be constructed from the selected year of daily returns.

[11] Here and elsewhere in the paper where we have employed the stationary bootstrap to compute confidence bands, we have experimented with the choice of average block length (in this particular case between 25 and 250 observations) with little influence on the qualitative nature of the results.

of monthly returns is measured with substantial error and one cannot reject the null that returns are unskewed. The NP approach delivers much more precision – the bootstrap confidence interval is on average about 60% of the width of the monthly interval – and generally the null that skew is zero can be rejected. A similar plot for kurtosis in shown in the internet appendix.

3.3 The components of 25-day skew and kurtosis

As Proposition 2 makes clear, skew in long horizon returns is driven by skew in high-frequency returns and by the leverage effect. Long-horizon kurtosis has three possible sources: kurtosis in high-frequency returns, covariation between current cubed returns and lagged returns (which we refer to as the 'Cube' component) and covariation between current and lagged squared returns (which we call the GARCH effect).

Figure 7 shows the time-series variation in the two monthly skew components for the years 1935 to 2015, with the two plots having identical vertical scales and with 95% confidence bands plotted around each estimate.

[INSERT FIGURE 7 ABOUT HERE]

The influence of skew in daily returns is negligible and almost never statistically different from zero. It is uncorrelated with the leverage effect (correlation coefficient of -0.01). Thus, both the average level of 25-day skew and its variation through time are attributable to the covariance of lagged returns and current squared returns. This covariance is almost always negative, usually significantly so and is almost never significantly greater than zero.

Similar results hold for kurtosis. Almost all of the significant positive excess kurtosis apparent in the data, as well as its time-variation in that excess kurtosis, comes from the GARCH component. The decomposition of kurtosis, analogous to Figure 7, is available in the internet appendix.

The higher moments of low frequency returns and daily returns bear little relation to one another. Low-frequency skewness and kurtosis are driven by leverage and GARCH effects respectively rather than by jumps in prices or the moments of daily data. This observation is of considerable practical importance as, for example, researchers often use moments of daily data to proxy the tail risks faced by investors who (presumably) have relatively long run investment horizons. Our results show that these proxies are largely irrelevant to the long-run investor.

3.4 Estimates of monthly moments using intra-day data

We now ask whether the use of intra-day data materially changes the estimates of monthly skew that we have obtained from daily data. To this end we have collected S&P-500 ETF returns with a 10 minute sampling frequency covering the period between the beginning of 2004 and the end of 2016. Every 25 days through this period, we estimate 25-day return standard deviations, third moments and coefficients of skewness using 250 days of daily returns, half-hourly returns and 10 minutely returns, respectively. Table 4 gives the results.

[INSERT TABLE 4 ABOUT HERE]

The key result from this table is that, in all cases, using daily data leads to monthly moment estimates that are extremely highly correlated with those from 10 minute or 30 minute data. The estimates of standard deviations and third moments from intra-day data always have a correlation with the estimates using daily data that is larger than 0.99. The coefficients of skewness measured using intra-day and daily data are somewhat less highly correlated, but the number is still around 0.9, and the average coefficient of skewness estimated from intra-day data is slightly more negative than that from daily data.

Thus, our empirical work suggests that the additional value from collecting high-frequency data to estimate the distribution of monthly or longer horizon returns is small.

3.5 The term structure of skewness and kurtosis

Our analysis thus far confirms the existence of significant negative skew and excess kurtosis in monthly and annual US stock market returns. We now examine how skewness and kurtosis vary across a range of possible horizons, from monthly to multi-year returns. Via such analysis one can ask, for example, whether investors with a 5 year investment horizon need to worry about deviations from lognormality.

[INSERT FIGURE 8 ABOUT HERE]

Figure 8 reports our estimates of the skewness and kurtosis coefficients at different horizons using the entire span of US stock market data from 1926-2015. The shaded area shows the 95% confidence intervals, obtained from bootstrapping. Our point estimates show that at horizons up to about 50 days (2 months) the distribution of returns becomes increasingly skewed reaching a level of about -1.6; the skewness coefficient then attenuates to -0.66 after 5 years. Kurtosis declines steadily from more than 10 at one week, falling below 2 after 5 years. The confidence intervals are wide, rather wider than suggested by the simulations in Table 2. The time variation in the levels of skewness and kurtosis that we observe in our data and which is not modelled in the simulations is a source of estimation error.

The very width of the confidence bands raises the question of whether the shape of the term structure is in the data or is a product of the estimation error; the confidence bands for example are consistent with a skew coefficient of -1 at all horizons. We test for this through bootstrapping the ratio skew(T)/skew(T-1) for different horizons T, and get a p-value for the hypothesis that |skew| is declining with horizon. As the corollary to Proposition 3 shows, as the leverage effect weakens, one would expect |skew| to decline with the square root of T. We test for this too. Similarly we test whether kurtosis is declining and whether it is declining faster than T. The results are shown in Table 5.

[INSERT TABLE 5 ABOUT HERE]

The results indicate that returns do become more skewed with horizon out to about two months, but beyond three months, skew attenuates as the return horizon increases. Kurtosis declines with horizon at all maturities. The decline in skewness and kurtosis is not as fast as $\sqrt{T}$ and T respectively out to at least one year, suggesting that the leverage and GARCH effects (the impact of returns and volatility on future volatility respectively) are still significant at that horizon. Beyond two years the evidence is consistent with the rate of attenuation in skewness and kurtosis predicted by the Central Limit Theorem. It should be noted that the power of the methodology to say much about the shape of the term structure of either measure is rather weak at multi-year horizons.

3.6 Return predictability and the robustness of our estimators

As discussed in Section 1.2, our estimators of long-horizon third and fourth moments rely on the assumptions that daily returns are uncorrelated with lagged rolling averages of squared returns or of cubed returns, where the length of the rolling average is equal to the horizon for which long-horizon moments are to be estimated. In applying the technology to stock index return data, it is possible that time variation in the equity risk premium could cause the assumptions to be violated. We now present empirical evidence on the size and significance of any predictability.

In the proof of Proposition 2, we show that long horizon skew can be decomposed into scaled short-horizon skewness plus two covariance terms: the first is the leverage effect and the second is the covariance between returns and average lagged squared returns. Similarly, long horizon kurtosis depends on scaled short horizon kurtosis and three covariances: the GARCH effect, the covariance between current cubed daily returns and lagged rolling returns and, finally, the covariance between current daily returns and rolling averages of lagged cubed returns. In Table 6 we present some simple regression analysis to measure each of those covariance terms for two different estimation horizons (25 and 250 days).

[INSERT TABLE 6 ABOUT HERE]

Starting with estimates relevant to skewness, the results demonstrate that for both horizons, the leverage effect (as measured by the regression coefficient of $x^{(2E)}(r)$ on the lagged value of $y^{(1)}$), is negative and significant. On the other hand, there is not a significant relationship between daily returns and the lagged variance, as measured by $y^{(2)}$, at either horizon. This suggests that the existence of time-varying mean returns does not induce a correlation between daily returns and lagged variance and does not cause a bias in our estimators. Our finding is similar to that in Bollerslev et al. (2013), who review the related literature and argue that "returns are at best weakly positively related, and sometimes even negatively related, to past volatilities".

Similarly, for the case of kurtosis, the GARCH effect is positive and significant in the specifications that relate $x^{(2L)}$ to the lagged value of $y^{(2L)}$ at both horizons, while the regressions of daily returns on $y^{(3)}$ never show a significant relationship. Thus, again, there is no evidence that daily returns are predictable with rolling averages of lagged returns. It is also evident that the cube effect, the covariance between $x^{(3)}$ and $y^{(1)}$, is never significant.

[INSERT TABLE 7 ABOUT HERE]

Table 7 contains the t-statistics for the slope coefficient in these regressions for horizons across the range that we have previously considered. The table makes clear that the conclusions drawn from Table 6 are not restricted to the measurement of 25 and 250 day skew, but hold for each individual horizon.

In sum, in the data that we study, there is no evidence to suggest that predictability of daily returns causes a distortion of our estimates of third and fourth moments.

3.7 Comparison of skewness estimates with implied and realized skew

We compare our skewness estimates with estimates that rely on data from the options market. Our option-implied measures are implied and realized skewness, as defined in Neuberger (2012) and Kozhan, Neuberger and Schneider (2013)[12].

As demonstrated in appendix D, in the absence of risk premia in the options market, the leverage term that is important for skewness can be estimated either as the covariance between current variance and lagged returns (as we do) or by the covariance between current returns and changes in option implied variance (as in Kozhan, Neuberger and Schneider (2013)).

For the purposes of this comparison our base data are monthly and run between the beginning of 1997 and the end of 2012. In each month of the sample we estimate the implied third moment at the beginning of the month; this is simply the third moment of the risk-neutral density. We also estimate the realized third moment, which is the covariance between returns and changes to implied variance. The implied and realized third moments use data only from that month.

[12] Implied and realized skewness use the NP definitions of skewness. Thus "realized skewness" here differs from the way that the term is used in Amaya et al. (2015). The former is an estimate of long horizon skewness that makes use of the covariance between returns and changes in implied variance; the latter is the skewness of high frequency returns.

We also have an end of month NP estimator of the third moment using 12 months of history (e.g. 25-day skew estimates using data from the past year). To make the NP estimator that uses a year of data and the monthly option-based estimates comparable, at the end of each month we compute simple rolling averages of realized and implied quantities over the preceding 12 months. Thus we compare annual rolling values of realized and implied skewness with NP estimation based also on the preceding year.

[INSERT TABLE 8 ABOUT HERE]

The results of our comparison are in Table 8. Looking at the third moments, there are two key observations. First, the average value obtained from the NP method is little more than half the average size of the implied third moment and around 10% smaller than the realized third moment. Second, the correlation between rolling NP and realized third moments is very strong at 0.97, while the correlation between NP and implied third moments is rather smaller at 0.92.

The difference in the levels of the third moments translates into average rolling NP skewness being considerably smaller in magnitude than estimates obtained from options data. Average NP skewness is around -0.9, while mean realized skew is more than 50% larger and implied skew more than twice as large. Implied and realized skewness are also much more volatile than NP skew and much less strongly correlated with the NP measure than third moments. The correlation between NP and realized skew is 0.45 while that between NP and implied skew is 0.13. Figure 9 confirms that, while these correlations are low, the three skewness measures display similar times series features but with clear differences in mean values.

[INSERT FIGURE 9 ABOUT HERE]

We can draw a number of conclusions from this analysis: skewness estimates based on data from the options market exhibit substantial bias because of the existence of variance risk premia. The bias seems to be more severe, unsurprisingly, for implied skewness than realized skewness. While there are variations in variance risk premia over time, there is also variation in physical skewness that is shown in the common variation of NP skewness and realized skewness. Whether these conclusions apply to returns other than those on the US stock market is an open question.

3.8 Cross-sectional sorting on skew

We close by providing a simple empirical illustration of the implications of our methods for cross-sectional portfolio sorting. We use, as a foundation, the work of Amaya et al. (2015) (hereafter ACJV) who show that portfolios of stocks with the highest mean, cubed, 5-minute returns subsequently significantly under-perform portfolios containing stocks with low mean, cubed, 5-minute returns. One particularly interesting feature of their results is the role of high frequency data. The effects they observe using five-minute returns attenuate by about fifty percent if thirty-minute returns are used instead, and become insignificant when using hourly returns (ACJV, Table 15). This raises the question whether the gain from using high frequency returns comes because investors care about the distribution of five minutely returns, or because they allow more precise estimates of skewness at the daily or longer horizon.

We take 5-minute data similar to that in ACJV and perform a similar analysis. We use the NP method to measure the skewness of lower-frequency returns from the 5-minute data and ask whether there is evidence that investors are averse to skewness at those frequencies also.[13] Thus we exploit the fact that using the NP method we can, for example, estimate daily skewness more accurately using the 5-minute data than we could by using daily returns.

[13] Our data set differs from ACJV. We use 5 minute return data from Thomson Reuters TRTH from 2009 to 2019 for the S&P-500 stocks. We filter overnight returns from the generation of the trading signal, as do ACJV, and also filter stocks which trade, on average, less than 20 times per 5 minute interval from the portfolios. ACJV use TAQ data on 3000 stocks from 1993 to 2013 and they filter stock days that have less than 80 transactions.

[INSERT TABLE 9 ABOUT HERE]

Table 9 presents return summary statistics for a portfolio that is long high skew and short low skew stocks, for skew measured from 5-minute returns, 30-minute returns and 60-minute returns (measured as in ACJV), the skew of daily returns (computed from the NP measure) and the skew of weekly returns (also using NP). In each case and on each rebalancing date, five weeks of history is used to compute skew. Rebalancing is daily and the low (high) skew portfolio is an equally weighted combination of the 1/3 of stocks at the end of the day with the lowest (highest) skew measure. As in ACJV, we omit overnight returns from the skew computations.

Like ACJV, we find that the mean return on the long-short portfolio for 5-minute skew is negative and significant. The annualized return is around -3% per annum with a Sharpe ratio of -0.66. The result is an order of magnitude less strong than in ACJV but this is likely due to our much smaller cross-section of stocks (500 versus roughly 3000) and perhaps our focus on a later period of time (2009 to 2019 versus 1993 to 2013). Also, as in ACJV, when we base our signal on the skew of 30-minute returns and 60-minute returns its power drops substantially with mean long-short portfolio returns roughly half as big as in the 5-minute case and not statistically significant.

When we move to results for daily and weekly skew, using the NP method and five minutely return data, any evidence of meaningful compensation for exposure to negative skew disappears. The mean long-short portfolio returns are never significantly different from zero and are both much less than 1 percentage point per year in magnitude. Figure 10 shows a comparison of the cumulative long-short returns for the portfolios based on the 5 minute, daily and weekly skew measurement horizons and the lack of signal in the daily and weekly skew measurements is clear.

[INSERT FIGURE 10 ABOUT HERE]

While this is just an illustrative application and not intended to be a complete cross-sectional asset pricing analysis, it helps to refine our views as to what the results of ACJV are telling us. Their results suggest that investors care about the properties of high-frequency (five minute) returns and our results provide some support for this. Investors require compensation for holding stocks that are more likely to jump down than up intra-day. But this does not imply that investors have a preference for holding stocks with greater skewness at the daily or weekly horizon. In fact, they clearly do not. Asymmetry of returns at daily and longer horizons is unrelated to mean returns in the cross-section.

4. CONCLUSIONS

Measures of the higher moments of low frequency (i.e. monthly or annual) returns on stock indices or currencies or managed portfolios are important in a variety of contexts, including risk management, portfolio selection and asset pricing. But these moments are hard to measure.

In this paper we show how short-horizon (e.g. daily) returns can be used to estimate long horizon skewness and kurtosis with impressive precision. This precision is demonstrated via a set of simulation experiments. The method is then applied to US stock market returns and estimates of long-horizon skewness and kurtosis obtained.

The analysis demonstrates that the skewness of low frequency returns has two components, the skewness in high-frequency returns and the covariance between lagged returns and current squared high-frequency returns (i.e. the leverage effect). Empirically, the latter is shown to be much more important than the former when measuring the skewness of annual or monthly US stock index returns using daily data. Similarly for kurtosis: although there are three potential sources of kurtosis at long horizons (the kurtosis of high-frequency returns, the correlation between lagged returns and current cubed high-frequency returns and the correlation between lagged and current squared high frequency returns) it is only the last of these, which we call the GARCH effect, that is significant in determining kurtosis of observed long horizon US stock index returns.

Thus, we show, both analytically and empirically, that information on short horizon skewness and kurtosis is close to irrelevant when it comes to measuring long horizon skew and kurtosis. This has implications for the interpretation of asset pricing results that link expected returns on stocks to skewness in returns. The skewness measures employed in these papers are often constructed from data measured over very short horizons (e.g. in Amaya et al. 2015, five minutely data are used). It is clear that they do predict returns. Our methodology allows us to distinguish between two hypotheses: whether it is the skew of five-minute returns that is important to investors and therefore commands the risk premium, or whether it is the skew of daily or weekly returns that really matters, and the use of high frequency returns makes it possible to measure this skewness with precision. At least in the specific data set we use, the evidence is clearly supportive of the first interpretation.

Our estimates of skewness of US equity market returns are broadly consistent with those obtained from the index options market in that the different methodologies deliver estimates with significant common time-series variation. But the levels of implied skewness computed from the options market are substantially higher, suggesting that they are seriously contaminated by variance and skewness risk premia.

But perhaps the most important contribution of the paper is to demonstrate the degree to which long term market returns are negatively skewed. Looking back over the last ninety years it is clear not only that monthly returns are highly negatively skewed, but also that annual returns are similarly skewed. The skew appears to attenuate significantly beyond the one year horizon, halving after 5 years. The degree of skew is economically significant in the sense that, using conventional preference assumptions, skew risk may be an important factor in risk premia for long horizon investors.

APPENDIX

A. Proof of Proposition 1

The monthly price change is the sum of daily price changes so

$$\begin{aligned}
D_t^2 &= \sum_{u=0}^{T-1} d_{t-u}^2 + 2\sum_{u=0}^{T-1}\left(P_{t-u-1} - P_{t-T}\right)d_{t-u}; \\
D_t^3 &= \sum_{u=0}^{T-1} d_{t-u}^3 + 3\sum_{u=0}^{T-1}\left(P_{t-u-1} - P_{t-T}\right)d_{t-u}^2 + 3\sum_{u=0}^{T-1}\left(P_{t-u-1} - P_{t-T}\right)^2 d_{t-u}; \\
D_t^4 &= \sum_{u=0}^{T-1} d_{t-u}^4 + 4\sum_{u=0}^{T-1}\left(P_{t-u-1} - P_{t-T}\right)d_{t-u}^3 + 6\sum_{u=0}^{T-1}\left(P_{t-u-1} - P_{t-T}\right)^2 d_{t-u}^2 \\
&\quad +4\sum_{u=0}^{T-1}\left(P_{t-u-1} - P_{t-T}\right)^3 d_{t-u}.
\end{aligned} \tag{17}$$

Taking conditional expectations at time *t-T*, the last term drops out so

$$\begin{aligned}
\mathbb{E}_{t-T}\left[D_t^2\right] &= \sum_{u=0}^{T-1} \mathbb{E}_{t-T}\left[d_{t-u}^2\right]; \\
\mathbb{E}_{t-T}\left[D_t^3\right] &= \sum_{u=0}^{T-1} \mathbb{E}_{t-T}\left[d_{t-u}^3\right] + 3\sum_{u=0}^{T-1} \mathbb{E}_{t-T}\left[\left(P_{t-u-1} - P_{t-T}\right)d_{t-u}^2\right]; \\
\mathbb{E}_{t-T}\left[D_t^4\right] &= \sum_{u=0}^{T-1} \mathbb{E}_{t-T}\left[d_{t-u}^4\right] + 4\sum_{u=0}^{T-1} \mathbb{E}_{t-T}\left[\left(P_{t-u-1} - P_{t-T}\right)d_{t-u}^3\right] \\
&\quad +6\sum_{u=0}^{T-1} \mathbb{E}_{t-T}\left[\left(P_{t-u-1} - P_{t-T}\right)^2 d_{t-u}^2\right].
\end{aligned} \tag{18}$$

Taking unconditional expectations and rearranging terms

$$\begin{aligned}
&\mathbb{E}\left[D^2\right] = T\mathbb{E}\left[d^2\right]; \\
&\mathbb{E}\left[D^3\right] = T\mathbb{E}\left[d^3\right] + 3T\mathbb{E}\left[y^{(1*)}d^2\right]; \\
&\mathbb{E}\left[D^4\right] = T\mathbb{E}\left[d^4\right] + 4T\mathbb{E}\left[y^{(1*)}d^3\right] + 6T\mathbb{E}\left[y^{(2*)}d^2\right]; \\
&\text{where } y_t^{(1*)} := \sum_{u=1}^{T}\left(P_{t-1} - P_{t-u}\right)\Big/T; \quad \text{and } y_t^{(2*)} := \sum_{u=1}^{T}\left(P_{t-1} - P_{t-u}\right)^2\Big/T.
\end{aligned} \tag{19}$$

Now

$$\mathbb{E}\left[y^{(1*)}\right]=0 \quad \text{and}$$
$$\mathbb{E}\left[y^{(2*)}\right]=\sum_{u=1}^{T}(u-1)\mathbb{E}\left[d^2\right]\big/T=\tfrac{1}{2}(T-1)\mathbb{E}\left[d^2\right]. \tag{20}$$

Replacing expectations of products in (19) by covariances and products of means we get

$$\begin{aligned}
&\mathbb{E}\left[D^2\right]=T\mathbb{E}\left[d^2\right];\\
&\mathbb{E}\left[D^3\right]=T\left\{\mathbb{E}\left[d^3\right]+3\operatorname{cov}\left(y^{(1*)},d^2\right)\right\};\\
&\mathbb{E}\left[D^4\right]-3\mathbb{E}\left[D^2\right]^2=T\left\{\mathbb{E}\left[d^4\right]-3\mathbb{E}\left[d^2\right]^2+4\operatorname{cov}\left(y^{(1*)},d^3\right)+6\operatorname{cov}\left(y^{(2*)}d^2\right)\right\}.
\end{aligned} \tag{21}$$

Using the standard definitions of skewness and excess kurtosis, the result follows.

B. The derivation of the $x^{(n)}$ functions

For n = 2, 3 and 4, we are looking for functions $x^{(m)}(r)$ of log returns r (for $m \leq n$). They must satisfy three conditions.

- They must approximate powers of r

$$\lim_{r\to 0} x^{(m)}(r)\big/r^m = 1 \text{ for all } m. \tag{22}$$

- To allow decomposition, the binomial expansion must be valid

$$x^{(n)}(r+s)=\sum_{m=0}^{n}\binom{n}{m}x^{(n-m)}(r)\,x^{(m)}(s) \quad \left(\text{where } x^{(0)}(r)=1\right). \tag{23}$$

- To allow elimination, the first moment must be mean zero under the martingale assumption. This requires that

$$x^{(1)}(r)=e^r-1. \tag{24}$$

The Solution for n = 2

We already have $x^{(1)}$. We need to find $x^{(2)}$. We seek a function $f = x^{(2)}$ where $f(r)/r^2$ tends to 1 as r tends to 0. Decomposition requires that

$$f(r+s) = f(r) + 2(e^s - 1)(e^r - 1) + f(s) \text{ for all } r, s. \tag{25}$$

Differentiate equation (25) with respect to s and set $s = 0$. Then

$$f'(r) = 2(e^r - 1). \tag{26}$$

The unique solution that also satisfies the conditions at $r = 0$ is

$$f(r) = 2(e^r - 1 - r). \tag{27}$$

It follows that

$$\begin{aligned} x^{(1)}(r) &= e^r - 1; \\ x^{(2)}(r) &= 2(e^r - 1 - r); \end{aligned} \tag{28}$$

is the unique solution for $n = 2$.

The Solution for n = 3

We now seek functions f and g where $f = x^{(2)}$, $g = x^{(3)}$, and $f(r)/r^2$ and $g(r)/r^3$ tend to 1 as r tends to 0. Again we have $x^{(1)} = e^r$-1. The decomposition requirement is

$$g(r+s) = g(r) + 3(e^s - 1) f(r) + 3 f(s)(e^r - 1) + g(s) \text{ for all } r, s. \tag{29}$$

Differentiate equation (29) with respect to s first once, and then again, and in each case set s = 0. Then using the limit properties of f and g at zero we get

$$\begin{aligned} g'(r) &= 3f(r); \\ g''(r) &= 3f(r) + 6(e^r - 1). \end{aligned} \tag{30}$$

The unique solution to this equation set that also satisfies the limit conditions at $r = 0$ is

$$\begin{aligned} f(r) &= 2(re^r + 1 - e^r); \\ g(r) &= 6((e^r + 1)r + 2(e^r - 1)). \end{aligned} \tag{31}$$

This gives the unique solution in the case of n=3 as

$$
\begin{aligned}
&x^{(1)}(r) = e^r - 1; \\
&x^{(2)}(r) = 2\left(re^r + 1 - e^r\right); \\
&x^{(3)}(r) = 6\left(\left(e^r + 1\right)r - 2\left(e^r - 1\right)\right);
\end{aligned} \tag{32}
$$

Note that the definition of $x^{(2)}$ is different from the case where $n = 2$. The fact that both solutions are unique shows that there is no single definition of $x^{(2)}$ that works for all n.

A Solution for n = 4

We seek functions f, g and h where $f(r)/r^2$, $g(r)/r^3$ and $h(r)/r^4$ all tend to 1 as r tends to 0. The decomposition requirement is

$$
h(r+s) = h(r) + 4\left(e^s - 1\right)g(r) + 6f(s)f(r) + 4g(s)\left(e^r - 1\right) + h(s). \tag{33}
$$

Differentiate equation (33) with respect to s three times, and set $s = 0$. Then

$$
\begin{aligned}
&h'(r) = 4g(r); \\
&h''(r) = 4g(r) + 12f(r); \\
&h'''(r) = 4g(r) + 6f'''(0)f(r) + 24\left(e^r - 1\right).
\end{aligned} \tag{34}
$$

Eliminating g and h from this equation set, we get a differential equation for f

$$
f' + kf = 2\left(e^r - 1\right) \text{ where } k := 1 - f'''(0)/2. \tag{35}
$$

There is now a family of solutions, indexed by k

$$
f = \frac{2}{1+k}\left\{e^r - 1 + \frac{e^{-kr} - 1}{k}\right\}, \tag{36}
$$

together with corresponding solutions for g and h. We choose to set $k = 0$ to equate $x^{(2)}$ in this case to its value when $n = 2$. We then have

$$
\begin{aligned}
&x^{(1)}(r) = e^r - 1; \\
&x^{(2)}(r) = 2\left(e^r - 1 - r\right); \\
&x^{(3)}(r) = 6\left(\left(e^r + 1\right)r - 2\left(e^r - 1\right)\right); \\
&x^{(4)}(r) = 12\left(r^2 + 2\left(e^r + 2\right)r - 6\left(e^r - 1\right)\right).
\end{aligned} \tag{37}
$$

C. Proof of Proposition 2

Start from the binomial expansion (23). Using the strong stationarity assumption (but not the martingale assumption) we get

$$\begin{aligned}&\mathbb{E}\left[x^{(n)}\left(R(k)\right)-x^{(n)}\left(\left(R(k-1)\right)\right)\right]=\\&\mathbb{E}\left[x^{(n)}(r)+\sum_{m=1}^{n-1}\binom{n}{m}x^{(n-m)}\left(R_{t-1}(k-1)\right)x^{(m)}(r_t)\right].\end{aligned} \tag{38}$$

Summing these equations from k =2 to T gives (for n – 2, 3 or 4)

$$\mathbb{E}\left[x^{(n)}\left(R(T)\right)\right]=T\mathbb{E}\left[x^{(n)}(r)+\sum_{m=1}^{n-1}\binom{n}{m}y^{(n-m)}x^{(m)}(r)\right]. \tag{39}$$

For n =1, it is easy to show that

$$\mathbb{E}\left[x^{(1)}\left(R(T)\right)\right]=T\mathbb{E}\left[x^{(1)}(r)+y^{(1)}x^{(1)}(r)\right]. \tag{40}$$

If we assume that the daily net return $x^{(1)}(r)$ is mean zero, and also that it is uncorrelated with past returns, as measured by $y^{(1)}$ then equation (40) shows that the multi-period net return is also mean zero, as is the lagged return $y^{(1)}$.

Now consider the monthly variance. Take equation (39) with n = 2, recalling that $x^{(2)} = x^{(2L)}$ when n =2, and retaining the assumption that returns are mean zero and uncorrelated with past returns, gives

$$\mathbb{E}\left[x^{(2L)}\left(R(T)\right)\right]=T\mathbb{E}\left[x^{(2L)}(r)\right]. \tag{41}$$

This is the first line of Proposition 2.

If we add the further assumption that returns are uncorrelated with squared lagged returns as measured by $x^{(2E)}$, then with n =3 we derive

$$\mathbb{E}\left[x^{(3)}\left(R(T)\right)\right]=T\mathbb{E}\left[x^{(3)}(r)+3y^{(1)}x^{(2)}(r)\right]. \tag{42}$$

Since $y^{(1)}$ has mean zero we can replace the expectation of the product by the covariance, normalize by the variance and get the second line of the Proposition.

Finally, take $n = 4$, and assume that $y^{(3)}$ is also uncorrelated with returns, we get

$$\mathbb{E}\left[x^{(4)}\left(R(T)\right)\right]=T\mathbb{E}\left[x^{(4)}(r)+4y^{(1)}x^{(3)}(r)+6y^{(2L)}x^{(2L)}(r)\right]. \tag{43}$$

Scaling by $\mathbb{E}\left[x^{(2L)}\left(R(T)\right)\right]^2$ and making use of the fact that $\mathbb{E}\left[x^{(2L)}\left(R(T)\right)\right]=T\mathbb{E}\left[x^{(2L)}(r)\right]$ and that $\mathbb{E}\left[y^{(2L)}\right]=T(T-1)\mathbb{E}\left[x^{(2L)}(r)\right]/2$, and subtracting 3 from both sides, we get the third line of the Proposition.

D. Relation with an option price-based skewness estimator

Proposition 2 shows that

$$\mathbb{E}\left[x^{(3)}(R)\right]=T\left\{\mathbb{E}\left[x^{(3)}(r)\right]+3\operatorname{cov}\left(y^{(1)},x^{(2E)}(r)\right)\right\}. \tag{44}$$

By reordering the terms, the leverage term can be written as

$$\begin{aligned}&\operatorname{cov}\left(y_t x^{(2E)}(r_t)\right)=\operatorname{cov}\left(\mathrm{e}^{r_t}-1,w_{t+T-1}\right)\\&\text{where } w_t=\frac{1}{T}\sum_{u=1}^{T-1}\mathrm{e}^{R_{t+u-T}(u-1)}x^{(2E)}\left(r_{t+u+1-T}\right).\end{aligned} \tag{45}$$

(We take advantage of the fact that both $y^{(1)}$ and e^r-1 are mean zero.) w is a measure of average future realized variance. In estimating the skewness of long horizon returns, it makes little difference whether one estimates the right-hand side of equation (45) or the left hand side. But if the expectation of future realized variance is observable, we can greatly increase efficiency by estimating $\operatorname{cov}\left(\mathrm{e}^{r_t}-1, \mathbb{E}_t\left[w_{t+T-1}\right]-\mathbb{E}_{t-1}\left[w_{t+T-1}\right]\right)$ rather than $\operatorname{cov}\left(\mathrm{e}^{r_t}-1, w_{t+T-1}\right)$.

To ensure that expectations of future variance are observable, we need two further assumptions

1. that the options market is complete, so that in particular we can replicate (and hence price) the so-called "entropy contract" that pays $x^{(2\mathrm{E})}(P_{t+T}/P_{t+1})$;
2. that the price of options, as well as of the underlying, are martingale (ie there is no volatility or jump risk premium).

The significance of the entropy contract is two-fold: the price of the entropy contract (like the log contract) at inception is equal to its Black-Scholes implied variance. Second, the contract, when delta-hedged, generates the cash flow Tw_{t+T}. Denote the price of the entropy contract at time t+1 by q_{t+1} then the absence of risk premia means that $q_{t+1}=\mathbb{E}_{t+1}\left[Tw_{t+T}\right]$. We therefore have the result that

$$\operatorname{cov}\left(\mathrm{e}^{r_t}-1, w_{t+T-1}\right)=\frac{1}{T}\operatorname{cov}\left(\mathrm{e}^{r_t}-1, q_t-q_{t-1}\right). \tag{46}$$

Assuming a sufficiently rich options market and the absence of variance risk premia, the leverage effect can be estimated from the covariance between changes in the implied variance of the entropy contract and contemporaneous returns.

E. Proof of Proposition 3

For any function $x^{(j)}(r)$, where $j\in\{2L,2E,3,4\}$, defined in (10), we can write the corresponding moment as

$$\begin{aligned} M^{(j)}(T) &:= \mathbb{E}\left[x^{(j)}\left(R_t(T)\right)\right], \\ &= \int_{u=0}^{T}\mathbb{E}\left[dx^{(j)}_{t+u-T,u}\right] \text{ where } dx^{(j)}_{v,u} := x^{(j)}\left(R_{v+du}(u+du)\right)-x^{(j)}\left(R_v(u)\right). \end{aligned} \tag{47}$$

Using the stationarity assumption

$$M^{(j)}(T)=\int_{u=0}^{T}\mathbb{E}\left[dx_{t,u}^{(j)}\right]. \tag{48}$$

By the diffusion assumption

$$dx_{t,u}^{(j)}=\frac{dx^{(j)}}{dr}\frac{dP_t}{P_t}+\tfrac{1}{2}\left(\frac{d^2x^{(j)}}{dr^2}-\frac{dx^{(j)}}{dr}\right)v_t du. \tag{49}$$

The martingale assumption ensures that the first term is mean zero, so

$$M^{(j)}(T)=\tfrac{1}{2}\mathbb{E}\left[v_t\int_{u=0}^{T}\left(\frac{d^2x^{(j)}}{dr^2}-\frac{dx^{(j)}}{dr}\right)du\right]. \tag{50}$$

Substituting for the appropriate function x we derive expressions for the second, third and fourth moments

$$\begin{aligned}
M^{(2L)}(T)&=T\mathbb{E}\left[v_t\right];\\
M^{(3)}(T)&=3\mathbb{E}\left[v_t\int_{u=0}^{T}\left(\mathrm{e}^{R_t(u)}-1\right)du\right];\\
M^{(4)}(T)&=6\mathbb{E}\left[v_t\int_{u=0}^{T}x^{(2L)}\left(R_t(u)\right)du\right].
\end{aligned} \tag{51}$$

Replace the expectation of the products by the covariance plus the product of the means

$$\begin{aligned}
M^{(2L)}(T)&=T\mathbb{E}\left[v\right];\\
M^{(3)}(T)&=3\operatorname{cov}\left(\int_{u=0}^{T}\left(\mathrm{e}^{R(u)}-1\right)du,v\right);\\
M^{(4)}(T)&=6\operatorname{cov}\left(\int_{u=0}^{T}x^{(2L)}\left(R(u)\right)du,v\right)+3M^{(2L)}(T)^2.
\end{aligned} \tag{52}$$

By using the definitions of skewness and kurtosis, proposition 3 follows immediately.

REFERENCES

Amaya, D., P. Christoffersen, K. Jacobs, and A. Vasquez. 2015. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118:135-167.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579-625.

Ang, A., R. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259-299.

Bakshi, G., N. Kapadia, and D. Madan. 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies* 16:101-143.

Balanda, K., and H. MacGillivray. 1988. Kurtosis: a critical review. *American Statistician*, 42:111-119.

Bao, Y. 2013. On sample skewness and kurtosis. *Econometric Reviews* 32:415-448.

Barberis, N., and M. Huang. 2008. Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review* 98:2066-2100.

Barro, R. 2006. Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics* 121:823-866.

Bessembinder, H. 2018. Do stocks outperform treasury bills? *Journal of Financial Economics* 129:440-457.

Bollerslev, T., D. Osterrieder, N. Sizova, and G. Tauchen. 2013. Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics* 108: 409-424.

Bondarenko, O. 2014. Variance trading and market price of variance risk. *Journal of Econometrics* 180:81-97.

Bowley, A. 1920. *Elements of statistics*, New York: Charles Scribner Sons.

Boyer, B., T. Mitton, and K. Vorkink. 2010. Expected idiosyncratic skewness. *Review of Financial Studies* 23:169-202.

Broadie, M., M. Chernov, and M. Johannes. 2007. Model specification and risk premia: Evidence from futures options. *Journal of Finance* 62:1453-1490.

Campbell, J., and L. Hentschel. 1992. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31:281-318.

Carr, P., and J. Bowie. 1994. Static simplicity. *Risk* 7:45-50.

Carr, P. and L. Wu. 2008. Variance risk premiums. *Review of Financial Studies* 22:1311-1341.

CBOE (Chicago Board Options Exchange), 2018, VIX White Paper – Cboe Volatility Index. http://www.cboe.com/micro/vix/vixwhite.pdf.

Conrad, J., R. Dittmar, and E. Ghysels. 2013. Ex ante skewness and expected stock returns. *Journal of Finance* 68:85-124.

Engle, R. 2011. Long-term skewness and systemic risk. *Journal of Financial Econometrics* 9:437-468.

Eraker, B. 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59:1367-1403.

Fama, E. 1965. The Behavior of Stock-Market Prices. *Journal of Business* 38:34-105.

Fama, E., and K. French. 2018. Long-Horizon Returns. *Review of Asset Pricing Studies* 8:232–252.

Ghysels, E., A. Plazzi, and R. Valkanov. 2016. Why invest in emerging markets? The role of conditional return asymmetry. *Journal of Finance* 71:2145-2192.

Harvey, C., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263-1295.

Heston, S. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6:327-343.

Joanes, D., and C. Gill. 1998. Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)* 47:183-189.

Kelly, B., and H. Jiang. 2014. Tail risk and asset prices. *Review of Financial Studies* 27:2841-2871.

Kim, T., and H. White. 2004. On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1:56-73.

Kozhan, R., A. Neuberger, and P. Schneider. 2013. The skew risk premium in the equity index market. *Review of Financial Studies* 26:2174-2203.

Kraus, A., and R. Litzenberger. 1976. Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085-1100.

Longstaff, F., and M. Piazzesi. 2004. Corporate earnings and the equity premium. *Journal of Financial Economics* 74:401-421.

Mandelbrot, B. 1963. The Variation of Certain Speculative Prices. *Journal of Business* 36:394-419.

Merton, R. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3:125-144.

Mitton, T., and K. Vorkink. 2007. Equilibrium under-diversification and the preference for skewness. *Review of Financial Studies* 20:1255-1288.

Neuberger, A. 2012. Realized skewness. *Review of Financial Studies* 25:3423-3455.

Rietz, T. 1988. The equity risk premium a solution. *Journal of Monetary Economics* 22:117-131.

Samuelson, P. 1965. Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review* 6:41-49.

Simkowitz, M., and W. Beedles. 1978. Diversification in a three-moment world. *Journal of Financial and Quantitative Analysis* 13:927-941.

Singleton, J., and J. Wingender. 1986. Skewness persistence in common stock returns. *Journal of Financial and Quantitative Analysis* 21:335-341

Table 1: moments of US equity market returns using different methodologies

	Volatility			*Skewness*			*Kurtosis*		
	Simple	***NP***	***Log***	***Simple***	***NP***	***Log***	***Simple***	***NP***	***Log***
Daily	16.91%	16.93%	16.94%	-0.11	-0.32	-0.42	16.71	17.33	17.55
Monthly	18.32%	18.42%	18.52%	0.21	-0.44	-0.65	9.47	7.72	7.54
Annual	19.81%	20.83%	21.59%	-0.07	-0.93	-1.11	1.24	2.75	2.92

The table shows the moments of returns on the CRSP universe of US stocks 1926-2015 (data from Ken French's website). "Daily" is trading day return, "monthly" is 25 trading day return, and "annual" is 250 trading day return. "Simple" and "Log" use the Pearson definitions of moments as applied respectively to simple and to log returns. "NP" uses the moment definitions in (10). "Volatility" is the annualized standard deviation, using a 250 day year.

Table 2: simulation results for NP and standard estimators

Panel A: Geometric Brownian Motion

	NP	Monthly	True
Standard deviation			
Mean	0.047	0.047	0.047
Stdev	0.001	0.002	
Coefficient of Skewness			
Mean	-0.006	-0.005	0.000
Stdev	0.035	0.150	
Excess Kurtosis			
Mean	-0.002	-0.019	0.000
Stdev	0.070	0.217	

Panel B: Heston model

	NP	Monthly	True
Standard deviation			
Mean	0.047	0.047	0.047
Stdev	0.002	0.003	
Coefficient of Skewness			
Mean	-0.277	-0.252	-0.280
Stdev	0.099	0.248	
Excess Kurtosis			
Mean	1.064	0.999	1.128
Stdev	0.373	0.669	

Panel C: Heston model with jumps

	NP	Monthly	True
Standard deviation			
Mean	0.049	0.049	0.049
Stdev	0.003	0.004	
Coefficient of Skewness			
Mean	-0.281	-0.274	-0.300
Stdev	0.191	0.297	
Excess Kurtosis			
Mean	1.372	1.315	1.467
Stdev	0.699	0.919	

The table reports the outcome of a simulation of 10,000 sample price paths, each of length 5000 days. The three panels give estimation results for cases where prices are created from Geometric Brownian motion, a Heston stochastic volatility model and a Heston stochastic volatility model with jumps, respectively. The objects of interest are the standard deviation, skewness and kurtosis of 25 day returns. These are estimated using two methods: NP uses the methodology of Proposition 2, while Monthly uses standard estimators based on overlapping 25 day observations. "True" is the observed skewness, computed using the Pearson definition and overlapping monthly observations from 50,000 daily data points.

Table 3: intraday simulation results

Panel A: Geometric Brownian Motion

N_D	1	2	8	64	128
			Standard deviation		
Mean	0.0469	0.0469	0.0469	0.0469	0.0469
Stdev	0.0005	0.0003	0.0002	0.0001	0.0001
			Coefficient of Skewness		
Mean	-0.0059	-0.0029	-0.0007	-0.00003	-0.00003
Stdev	0.0345	0.0246	0.0122	0.0043	0.0031
			Excess Kurtosis		
Mean	-0.0024	-0.0019	-0.0006	0.00004	-0.00003
Stdev	0.0702	0.0491	0.0245	0.0085	0.0061

Panel B: Heston model

N_D	1	2	8	64	128
			Standard deviation		
Mean	0.0469	0.0469	0.0469	0.0469	0.0469
Stdev	0.0023	0.0023	0.0023	0.0023	0.0023
			Coefficient of Skewness		
Mean	-0.2709	-0.2741	-0.2772	-0.2771	-0.2766
Stdev	0.1006	0.0919	0.0831	0.0812	0.0803
			Excess Kurtosis		
Mean	1.0577	1.0753	1.081	1.0740	1.0797
Stdev	0.3721	0.3372	0.3177	0.3082	0.3041

Panel C: Heston model plus jumps

N_D	1	2	8	64	128
			Standard deviation		
Mean	0.0492	0.0491	0.0492	0.0492	0.0492
Stdev	0.0026	0.0025	0.0025	0.0025	0.0025
			Coefficient of Skewness		
Mean	-0.2701	-0.2813	-0.2836	-0.2857	-0.2809
Stdev	0.1935	0.1894	0.1803	0.1810	0.1813
			Excess Kurtosis		
Mean	1.3795	1.3802	1.3702	1.3988	1.3874
Stdev	0.6866	0.6871	0.6538	0.6452	0.6698

The table reports results from 10,000 simulations, each of length 5000 days where prices are sampled N_D times per day. The three panels correspond to Geometric Brownian motion, and a Heston stochastic volatility model with and without jumps. The objects of interest are the standard deviation, skewness and kurtosis of 25 day returns. They are estimated using the methodology of Proposition 2.

Table 4: estimation of US stock market skewness with intra-day data

	10min	30min	Daily
Standard Deviation			
Mean	0.053	0.053	0.056
Stdev	0.026	0.025	0.029
Corr(Daily)	0.996	0.996	1.000
Coefficient of skewness			
Mean	-1.179	-1.155	-1.112
Stdev	0.477	0.468	0.466
Corr(Daily)	0.893	0.888	1.000
Coefficient of kurtosis			
Mean	1.426	1.421	0.630
Stdev	1.8485	1.935	1.500
Corr(Daily)	0.926	0.943	1.000

The table uses high frequency return data on the SPY S&P-500 ETF 2004-2016. The table summarises the distribution of the estimates of standard deviation, skewness and kurtosis of 25 day returns using overlapping 250 day spans of data, using data at different sampling frequencies. The rows headed Corr(Daily) give the correlation between the time series of a particular moment estimated using the sampling frequency in the column header with a time series of the same moment estimated using daily data. The estimates use the technology of Proposition 2.

Table 5: shape of the term structure of skewness and kurtosis

Horizon (days)	skew	pr{|skew| declining}	pr{|skew| declining faster than √T}	excess kurtosis	pr{kurtosis declining}	pr{kurtosis declining faster than T}
10	-1.22	0.00	0.00	11.31	0.27	0.00
25	-1.49	0.05	0.00	9.69	0.94	0.00
50	-1.61	0.50	0.00	9.12	0.95	0.00
75	-1.60	0.96	0.00	8.26	1.00	0.00
125	-1.53	0.98	0.01	7.03	0.99	0.02
250	-1.41	0.99	0.08	5.62	1.00	0.09
500	-1.21	1.00	0.31	4.22	0.99	0.33
1000	-0.77	0.95	0.47	2.12	0.91	0.52
1250	-0.66	0.94	0.44	1.61	0.84	0.49

The table is based on daily returns on the US stock market 1935-2015. It shows the estimated skewness and kurtosis of the market at different horizons. A stationary bootstrap with mean block length equal to 250 days is used to test the hypotheses that $|\text{skew}(T)| < |\text{skew}(T\text{-}1)|$, and that $\sqrt{T}|\text{skew}(T)| < \sqrt{(T\text{-}1)}|\text{skew}(T\text{-}1)|$; the table shows the *p*-stats. The kurtosis hypotheses are that $\text{kurt}(T) < \text{kurt}(T\text{-}1)$, and that $T\,\text{kurt}(T) < (T\text{-}1)\text{kurt}(T\text{-}1)$.

Table 6: regression estimates of covariance terms in long horizon skewness and kurtosis

Panel A: h = 25					
Dependent Var	**Constant (x10^{-3})**	**$y^{(1)}$ (x10^{-3})**	**$y^{(2)}$**	**$y^{(3)}$**	**R^2**
$x^{(2E)}$	0.090	-2.806			0.042
	[33.41]	[-9.00]			
r	-0.172		0.147		0.002
	[-1.72]		[1.77]		
$x^{(3)}$	-0.001	0.012			0.000
	[-1.77]	[0.33]			
$x^{(2L)}$	0.040		0.042		0.093
	[7.04]		[7.85]		
r	-0.028			-0.726	0.002
	[-0.43]			[-1.78]	

Panel B: h = 250					
Dependent Var	**Constant (x10^{-3})**	**$y^{(1)}$ (x10^{-3})**	**$y^{(2)}$**	**$y^{(3)}$**	**R^2**
$x^{(2E)}$	0.091	-0.749			0.034
	[32.77]	[-12.93]			
r	-0.119		0.008		0.000
	[-1.13]		[0.98]		
$x^{(3)}$	-0.001	-0.001			0.000
	[-1.77]	[-0.31]			
$x^{(2L)}$	0.027		0.005		0.061
	[5.15]		[11.16]		
r	-0.043			-0.026	0.001
	[-0.68]			[-1.21]	

The table is based on daily returns on the US stock market 1935-2015. It reports results of regressions of returns on lagged rolling averages of squared and cubed returns, respectively, plus results of regressions of squared returns on lagged variance and lagged average rolling returns and, finally, results of regressions of cubed returns on lagged rolling returns. Each regression is run for two horizons, i.e. 25 and 250 days. Robust *t*-statistics are in parentheses.

Table 7: regression estimates of covariance terms across all horizons

Dep Var	RHS Var	25	125	250	375	500	750	1000	1250
$x^{(2E)}$	$y^{(1)}$	-9.00	-11.94	-12.93	-12.95	-13.03	-12.20	-10.49	-9.63
r	$y^{(2E)}$	1.77	1.38	0.98	0.83	0.97	1.04	1.15	0.97
$x^{(3)}$	$y^{(1)}$	0.33	0.22	-0.31	-0.69	-0.95	-1.44	-1.45	-1.42
$x^{(2L)}$	$y^{(2L)}$	7.85	10.80	11.16	11.06	11.08	11.41	10.59	10.27
r	$y^{(3)}$	-1.78	-1.50	-1.21	-1.22	-1.22	-1.48	-1.60	-1.46

This table is based on daily returns on the US stock market 1935-2015. It shows the robust *t*-statistics for the slope coefficient from the regression of the variable named in the first column on the variable named in the second column.

Table 8: comparing NP skewness with option-based skew measures

Third moment			
	NP	**KNS Realized**	**Implied**
Mean (x100)	-0.037	-0.041	-0.066
Stdev (x100)	0.069	0.063	0.074
Corr with NP	1.000	0.974	0.916

Skewness			
	NP	**KNS Realized**	**Implied**
Mean	-0.876	-1.387	-1.879
Stdev	0.414	0.612	0.529
Corr with NP	1.000	0.451	0.126

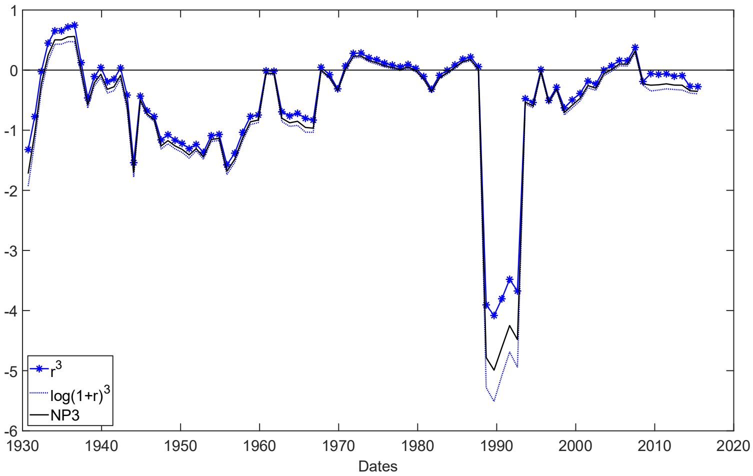
The table is based on daily prices of options and futures on the S&P500 index 1997-2012. It shows estimates of the third moment and skewness of monthly returns. Three methods are used: the NP method of Proposition 2 (computed using a one year rolling window), the realized skewness method of Kozhan, Neuberger and Schneider (2013), and estimation from the option implied distribution.

Table 9: skew-sorted long-short portfolio returns: summary statistics

	Skew measurement horizon				
	5-min	30-min	60-min	Daily	Weekly
Mean	-0.031	-0.017	-0.019	-0.003	0.006
Stdev	0.046	0.051	0.055	0.056	0.060
t-statistic	-2.03	-0.975	-1.047	-0.153	0.306
Sharpe	-0.665	-0.320	-0.344	-0.050	0.100

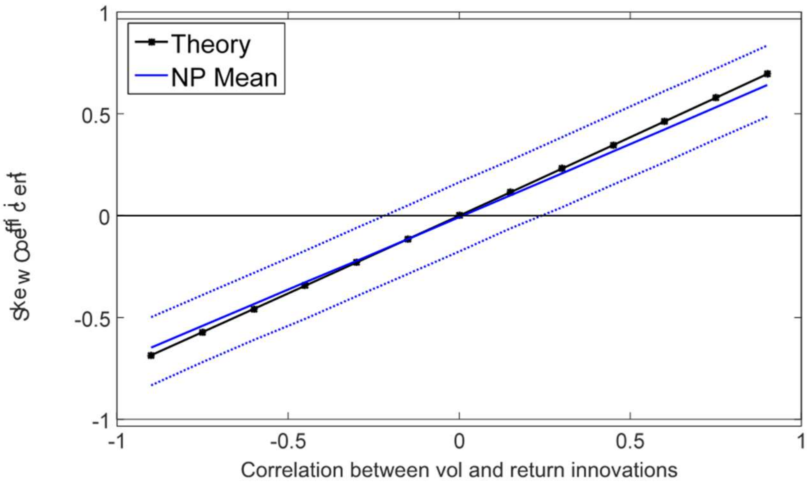
The table is based on 5-minute return data from Thomson Reuters TRTH 2009-19 for S&P-500 stocks. It contains summary statistics for daily returns on zero-investment portfolios where long positions are taken in stocks with positive skew and short positions are taken in stocks with negative skew. Rebalancing is daily and 5 weeks of data are used to calculate skew. *n*-minute skew is computed as the actual skew of n-minute returns while daily and weekly skew are computed using the NP method from 5-minute returns. Mean returns, standard deviations and Sharpe ratios are all annualized. The *t*-statistic tests the null that the mean return is zero.

Figure 1: rolling monthly skewness of returns using the simple, log and NP definitions.



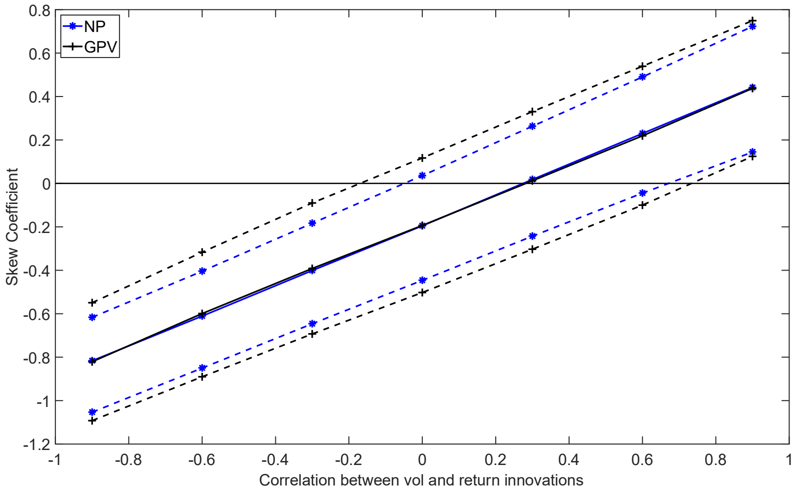
This figure is based on daily returns on the US stock market 1926-2015. It plots measures of 25 day skewness, computed at the beginning of each year using overlapping 25 day returns and 1250 days of data. The three lines are for computations based on cubed simple returns, cubed log returns and the NP measure, $x^{(3)}$.

Figure 2: theoretical and estimated skew coefficient versus correlation parameter: mean, 5th and 95th percentiles.



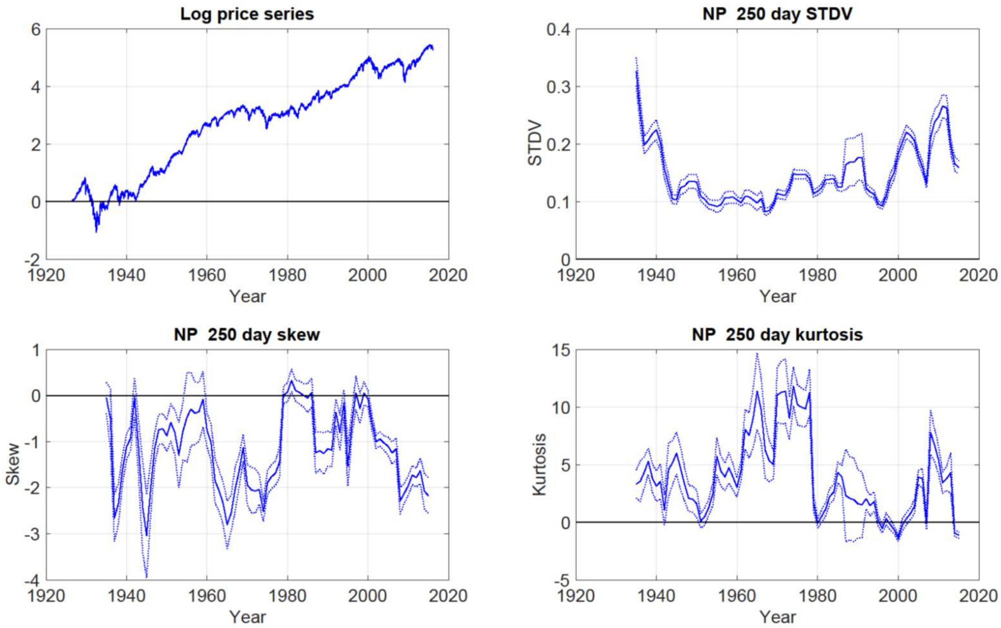
The graph summarizes results from simulating 10,000 sample price paths, each of length 5000 days, for each set of parameters. The underlying model is a Heston model with correlation coefficient given by the value shown on the x-axis. The object of interest is the skewness of 25 day returns, estimated using the methodology of Proposition 2. The mean estimate is shown together with the 5^{th} and 95^{th} percentiles. The theoretical skew coefficient is also plotted.

Figure 3: estimated skew coefficients from the NP and quantile methods: means and 5th and 95th percentiles.



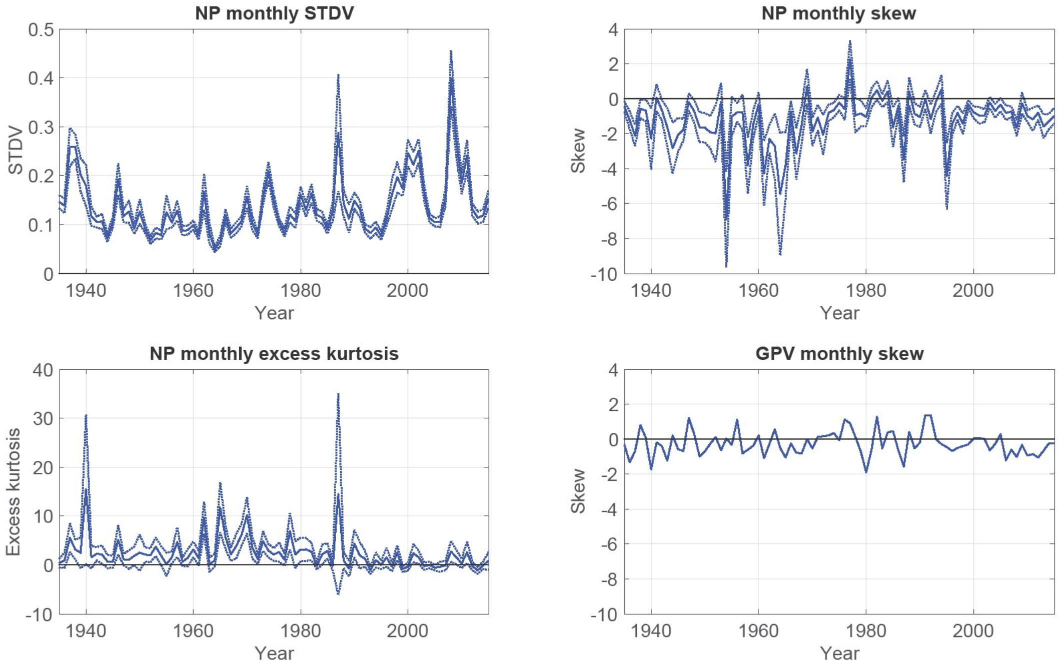
The graph reports the outcome of a simulation of 10,000 sample price paths, each of length 5,000 days. The paths are generated by a Heston stochastic volatility model with jumps and with the correlation coefficient between volatility and return innovations plotted on the x-axis. The object of interest is the skewness of 25-day returns, estimated using the methodology of Proposition 2 (marked by *) and the quantile based approach of Ghysels, Plazzi and Valkanov (2016) (marked by +). The mean estimates are shown together with the 5th and 95th percentiles.

Figure 4: 250-day moments of US stock market returns



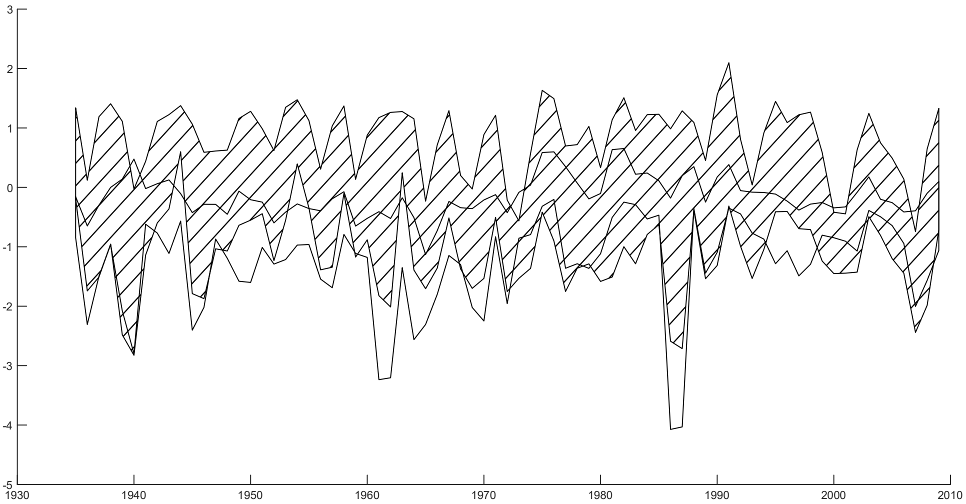
The figures are based on daily returns on the US stock market 1926-2015. The top left hand graph shows the cumulative log return. The other three graphs show rolling estimates computed from 1250 days of data of the volatility, skewness and excess kurtosis of 250 day returns. The dashed lines show the 90% confidence band.

Figure 5: time-variation in monthly moment estimates for US stock market



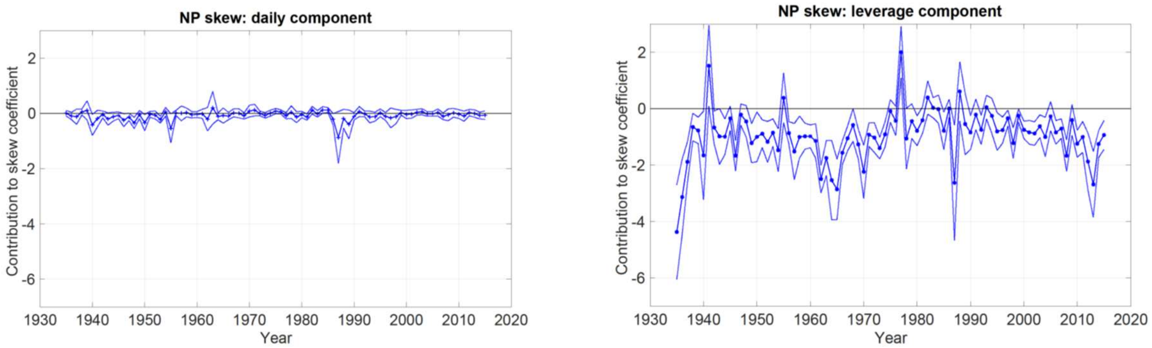
The figures are based on daily returns on the US stock market 1926-2015. The first three graphs show estimates of the volatility, skewness and excess kurtosis of 25-day returns computed from non-overlapping 250 days of data using the technology of Proposition 2. The dashed lines show the 90% confidence band. The lower right hand graph shows estimates of skewness from quantile estimation.

Figure 6: NP and monthly skew, 90% bootstrap confidence intervals



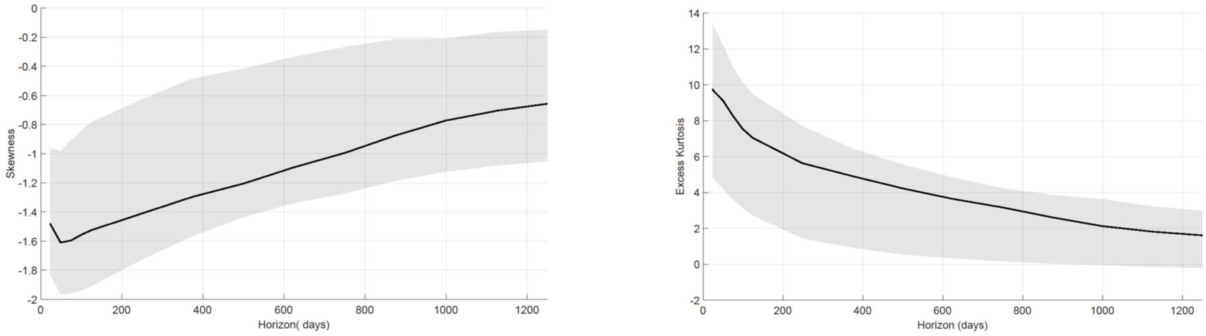
The figure shows the 90% confidence intervals for two estimates of the skewness of 25-day returns on the US stock market index 1926-2015, using 500 days of data. NP (plain) uses the technology of proposition 2 on daily data. Monthly (hatched) computes the moments using the overlapping 25-day returns. The confidence interval is obtained using a stationary bootstrap in which the block length is set equal to 50 days.

Figure 7: time variation in components of monthly skew coefficient



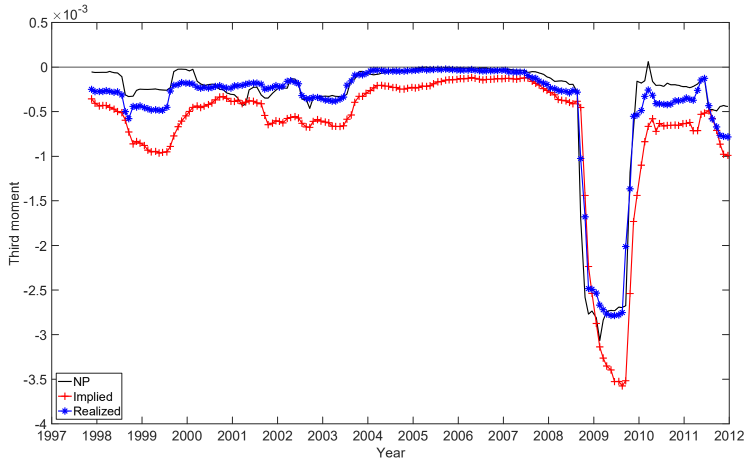
The two figures are based on daily returns on the US stock market 1926-2015. They show the components of the skewness of 25 day returns computed from non-overlapping 250 days of data using the technology of Proposition 2. The two components are daily skewness and leverage. The dashed lines show the 90% confidence interval.

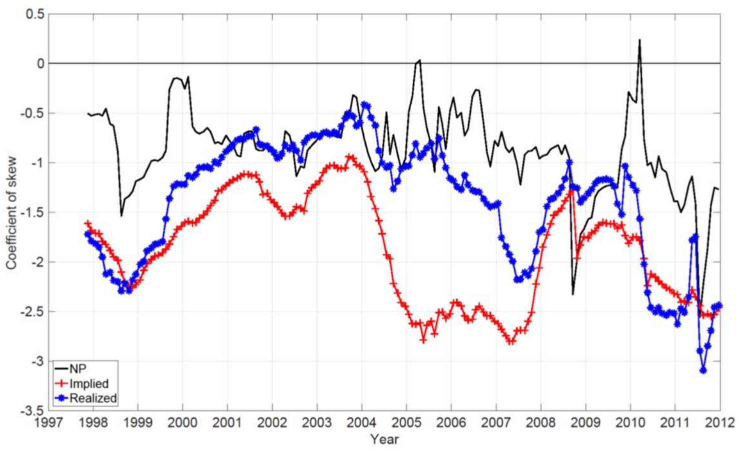
Figure 8: term structure of skewness and kurtosis 1926-2015



The figures are based on daily returns on the US stock market 1926-2015. They show the skewness and kurtosis of returns at different horizons using the technology of Proposition 2. The shaded area in each figure shows the 95% confidence intervals derived from a stationary bootstrap with average block length equal to 250 days.

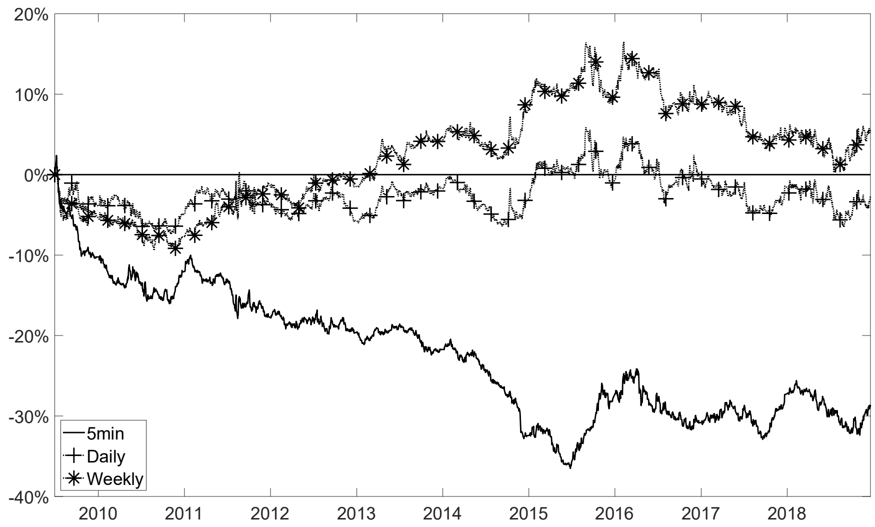
Figure 9: estimated third moment and skew coefficients using three methods





The graph is derived from daily options and futures prices on the S&P500 index 1997-2012. It shows estimates of the third moments and skewness of monthly returns. The estimates are reported as annual rolling averages. Three methods are used: the NP method of Proposition 2, estimation from the option implied distribution, and the realized skewness method of Kozhan, Neuberger and Schneider (2013).

Figure 10: cumulative long-short returns from sorting stocks on skewness



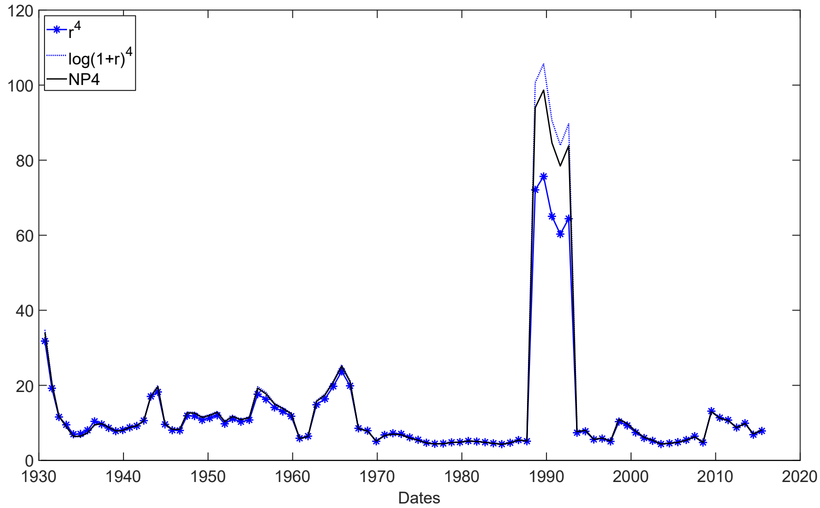
The figure is based on 5-minute return data from Thomson Reuters TRTH 2009-19 for S&P-500 stocks It shows the cumulative return of a long-short portfolio constructed from sorting stocks on skewness. Rebalancing is daily, and skewness is measured using 5 weeks of data. Sorting is done on 5-minute skew, daily skew and weekly skew respectively.

INTERNET APPENDIX

The following tables and figures present results from additional simulation experiments that support the discussion in the text. Figure A.1 is analogous to Figure 1, but for kurtosis rather than skewness. Figures A.2 and A.3 are analogous to Figures 6 and 7, but for kurtosis rather than skewness. Figure A.4 is similar to Figure 8 in showing the term structure of skewness, but shows the term structure for the two halves of the period separately.

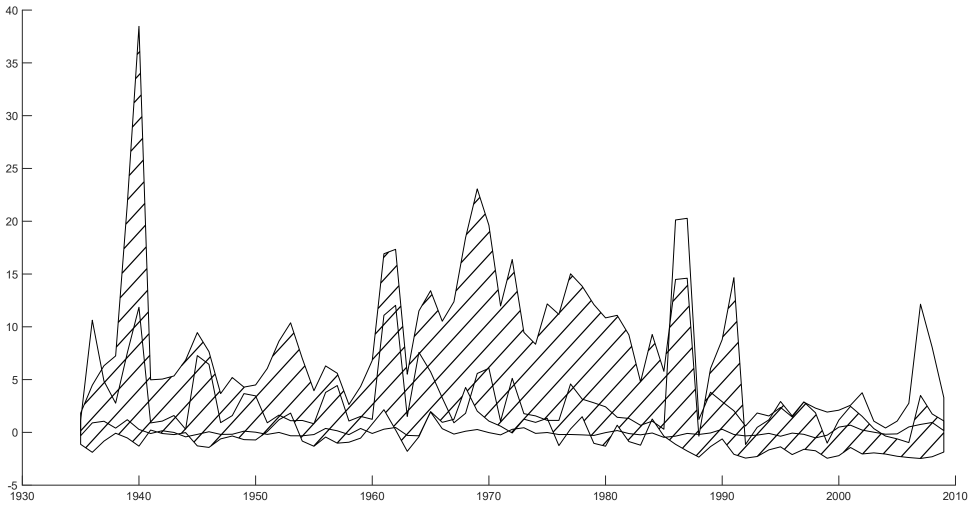
Tables A.1 and A.2 present simulation results similar to Table 2, but using simulations of samples containing 1,000 and 15,000 days rather than 5,000 days.

Figure A.1: rolling monthly kurtosis of returns using the simple, log and NP definitions



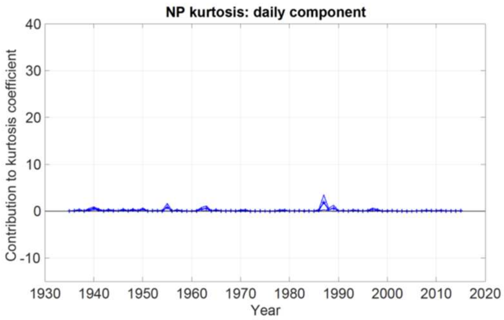
This figure is based on daily returns on the US stock market 1926-2015. It plots measures of 25 day kurtosis, computed at the beginning of each year using overlapping 25 day returns and 1250 days of data. The three lines are for computations based on the fourth power of simple returns, the fourth power of log returns and the NP measure, $x^{(4)}$.

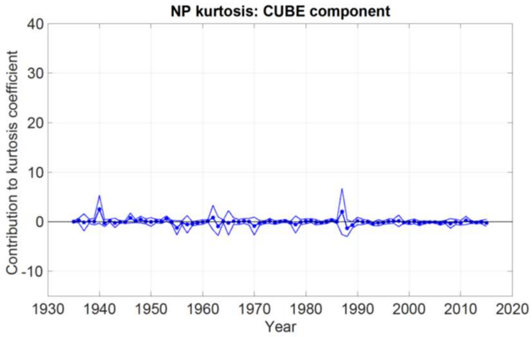
Figure A.2: NP and monthly kurtosis, 90% confidence intervals

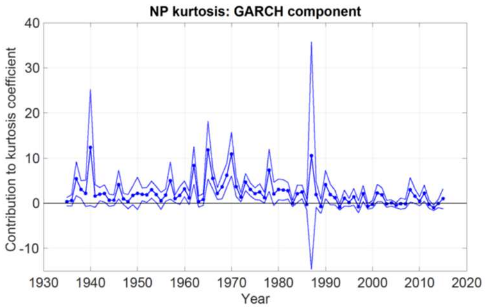


The figure shows the 90% confidence intervals for two estimates of the kurtosis of 25-day returns on the US stock market index 1926-2015, using 500 days of data. NP (plain) uses the technology of proposition 2 on daily data. Monthly (hatched) computes the fourth moment using overlapping 25-day returns. Both are then scaled by the same squared variance estimate. The confidence interval is obtained using a stationary bootstrap where the mean block length is set to 50 days.

Figure A.3: time variation in components of monthly excess kurtosis

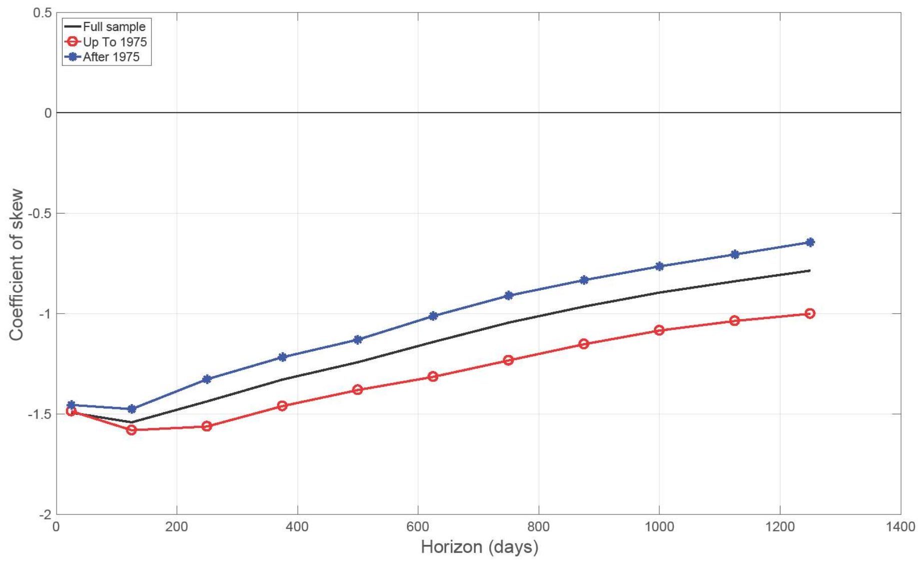






The figures are based on daily returns on the US stock market 1926-2015. They show the components of the kurtosis of 25 day returns computed from non-overlapping 250 days of data using the technology of Proposition 2. The three components are daily kurtosis, cube and GARCH. The dashed lines show the 90% confidence intervals.

Figure A.4: the term structure of skewness by sub-period



The graph is based on daily returns on the US stock market 1926-2015. It shows the skewness of returns market at different horizons using the technology of Proposition 2. Estimates are reported for the full sample, and also for the first and second half of the sample separately.

Table A.1: simulation results for NP and standard estimators (1000 days)

Panel A: Geometric Brownian Motion			Panel B: Heston model			Panel C: Heston model with jumps		
Standard deviation			Standard deviation			Standard deviation		
	NP	Monthly		NP	Monthly		NP	Monthly
Mean	0.047	0.047	Mean	0.047	0.046	Mean	0.049	0.049
Stdev	0.001	0.0043	Stdev	0.0051	0.007	Stdev	0.0056	0.0075
Coefficient of Skewness			Coefficient of Skewness			Coefficient of Skewness		
	NP	Monthly		NP	Monthly		NP	Monthly
Mean	-0.007	0.001	Mean	-0.251	-0.174	Mean	-0.256	-0.199
Stdev	0.078	0.511	Stdev	0.203	0.662	Stdev	0.344	0.72
Excess Kurtosis			Excess Kurtosis			Excess Kurtosis		
	NP	Monthly		NP	Monthly		NP	Monthly
Mean	-0.014	-0.096	Mean	0.862	0.625	Mean	1.073	0.835
Stdev	0.157	0.423	Stdev	0.616	0.913	Stdev	0.996	1.136

The table reports the outcome of a simulation of 10,000 sample price paths, each of length 1000 days. The three panels give estimation results for cases where prices are created from Geometric Brownian motion, a Heston stochastic volatility model and an EGARCH model, respectively. The objects of interest are the standard deviation, skewness and kurtosis of 25 day returns. These are estimated using two methods: NP uses the methodology of Proposition 2, while Monthly uses standard estimators based on overlapping 25 day observations.

Table A.2: simulation results for NP and standard estimators (15000 days)

Panel A: Geometric Brownian Motion		
Standard deviation		
	NP	Monthly
Mean	0.047	0.047
Stdev	0.0003	0.0011
Coefficient of Skewness		
	NP	Monthly
Mean	-0.006	-0.007
Stdev	0.02	0.142
Excess Kurtosis		
	NP	Monthly
Mean	-0.001	-0.006
Stdev	0.041	0.126

Panel B: Heston model		
Standard deviation		
	NP	Monthly
Mean	0.047	0.047
Stdev	0.0013	0.0019
Coefficient of Skewness		
	NP	Monthly
Mean	-0.275	-0.273
Stdev	0.06	0.207
Excess Kurtosis		
	NP	Monthly
Mean	1.107	1.101
Stdev	0.233	0.457

Panel C: Heston model with jumps		
Standard deviation		
	NP	Monthly
Mean	0.049	0.049
Stdev	0.0015	0.002
Coefficient of Skewness		
	NP	Monthly
Mean	-0.283	-0.289
Stdev	0.115	0.235
Excess Kurtosis		
	NP	Monthly
Mean	1.428	1.416
Stdev	0.432	0.591

The table reports the outcome of a simulation of 10,000 sample price paths, each of length 15000 days. The three panels give estimation results for cases where prices are created from Geometric Brownian motion, a Heston stochastic volatility model and an EGARCH model, respectively. The objects of interest are the standard deviation, skewness and kurtosis of 25 day returns. These are estimated using two methods: NP uses the methodology of Proposition 2, while Monthly uses standard estimators based on overlapping 25 day observations.