The Skewness of the Stock Market at Long Horizons

Anthony Neuberger and Richard Payne

Cass Business School,
City, University of London

January 2018

Abstract

Moments of long-horizon returns are important for asset pricing but are hard to measure. Proxies for these moments are often used but none is wholly satisfactory. We show analytically that short-horizon (i.e. daily) returns can be used to make more much precise estimates of long-horizon (e.g. annual) moments without making strong assumptions about the data generating process. Skewness comprises two components: the skewness of short-horizon returns, and a leverage effect, which is the covariance between contemporaneous variance and lagged returns. We provide similar results for kurtosis. Applying the technology to US stock index returns, we show that skew is large and negative and does not significantly attenuate with horizon as one goes from monthly to multi-year horizons.

INTRODUCTION

This paper makes two contributions: methodological and empirical. The methodological contribution is to show how 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, at around -1.5, is economically significant. We also show that this skew at long horizons is entirely attributable to the leverage effect – the negative correlation between returns and future volatility.

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), hypothesises that heavy-tailed shocks and left-tail events in particular have an important role in explaining asset price behaviour. Barberis and Huang (2007) 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, Hodrick, Xing and Zhang, 2006). Boyer, Mitton and Vorkink (2010) and Conrad, Dittmar and Ghysels (2013) show that high idiosyncratic skewness in individual stocks too is correlated with positive returns. Ghysels, Plazzi and Valkanov (2016) show similar results for emerging market indices.

But there are two serious problems in measuring these 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 data sample. We show how these problems can be mitigated by using information in short horizon returns to make estimates of skewness and kurtosis of long horizon returns more precise.

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 volatile, then annual returns are also volatile. But if daily returns are highly skewed and i.i.d., then annual returns will show little skew. Conversely, daily return distributions can be symmetric, while annual returns are skewed (e.g. in a Heston-type model where volatility is stochastic and shocks to volatility are correlated with shocks to prices). Similar examples could be given for kurtosis.

The purpose of this paper is to demonstrate how to exploit the information in short horizon returns to estimate the skewness and kurtosis of long horizon returns. The only assumption we make about the price process is that it is martingale, and that the relevant moments exist. We prove that the skewness of long horizon returns can come from one of only two sources: the skewness of short horizon returns; and the leverage affect, that is the covariance between lagged returns and squared 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). 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. Thus, the negative skewness and the excess kurtosis in annual stock market returns owe virtually nothing to the skew and kurtosis of daily returns.

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 and advocate the use of robust estimators such as those developed by Bowley (1920), which are based on the quantiles of the observed distribution.[1] 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 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 rather by exploiting the information in the cross-section. They assume that tail risk for individual stocks is a combination of stable stock specific tail risk and time-varying market-wide tail risk. They can therefore exploit the existence of a large number of stocks to get a much more precise estimate of market-wide tail risk. The validity of the inference 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 at any point in time. The technology for extracting implied skewness and kurtosis from options prices is well-established (Bakshi, Kapadia and Madan, 2003). The method can only be used on assets – such as the major market indices - that support a liquid options market, and cannot be used for managed portfolios. But there is a more fundamental issue: implied measures 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.

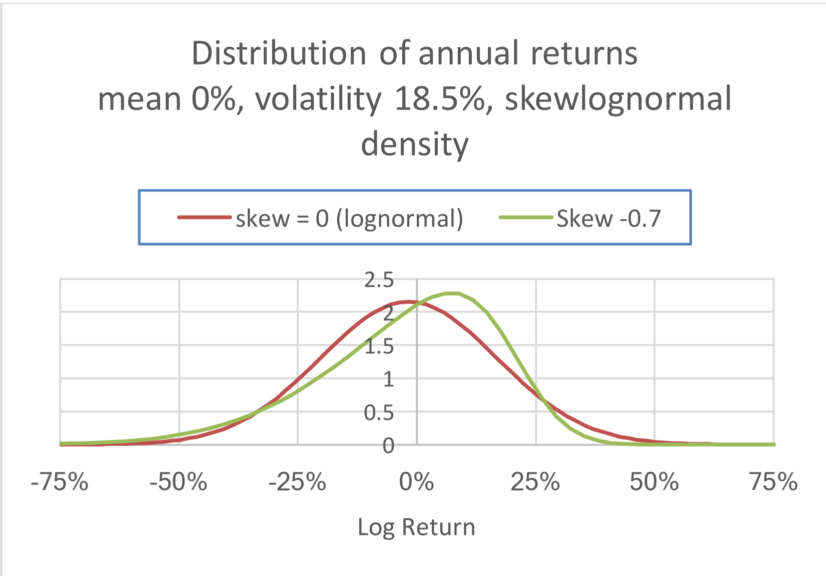
[1] To estimate the skewness (kurtosis) of a normally distributed random variable with a standard error of 0.1 requires a sample size of 600 (2400). Even for monthly returns, this would require 50 (200) years of returns data. If returns are non-normal, the standard errors are generally substantially 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.

We show by simulation that our measures of 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 of 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.

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 around -1.34. This skewness does not attenuate to any marked degree with horizon. Our central estimate is that the skewness of annual returns is -1.32 and of five year returns is -1.16. 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 an annual or longer horizon.

To illustrate visually the meaning of skew coefficients, Figure 1 shows the probability density of annual year returns on the assumption that log returns are skew normal. The skew normal is a three parameter family which includes the normal as a special case. It has been widely used in the literature to model skewed asset returns (for example by Harvey *et al*, 2010). The figure shows two distributions for returns. In both cases the mean return is zero, and the annualized volatility is 18.5%. In the one distribution, log returns are assumed to be distributed normally (which gives zero skew in the way we define skew), and in the other log returns are skew normal, with the coefficient of skewness set to -0.7.

Figure 1



It would be nice to plot the graph to match the skewness we observe in the data , but it is not possible to do so. The skew normal cannot accommodate skewness coefficients that lie outside the range (-1,+1). To illustrate annual returns with a skew of -1.32, we therefore use the binomial process. The binomial process can accommodate any level of skewness and has the added advantage that it can be readily comprehended.

The annual return takes the value $u > 1$ with probability p, and the value $d < 1$ with probability $1\text{-}p$. We fix the mean return to be 0, and the annualized volatility to be again 18.5%. In the absence of skew[2], $u = \exp(+18.5\%) = 1.203$, and $d = \exp(-18.5\%) = 0.831$, with $p = 45.4\%$. To keep the same first and second moments with a skew of -1.32, we need $u = 1.099$, and $d = 0.705$, with $p = 74.9\%$.

To get some sense of the economic importance of these levels of skew, consider the following question: how large does the equity premium have to be for the representative agent to hold all his wealth in the market portfolio? We follow the standard approach to answering this question and assume the representative agent has power utility, with constant relative risk aversion coefficient γ. We assume a horizon of 1 year.

To persuade the investor to invest 100% of their wealth in the market, the Euler condition needs to be satisfied

$$\mathrm{E}\left[\left(R-1\right)R^{-\gamma}\right]=0,$$

where R is the annual gross excess return. Expanding the expression to the third order, this can be written as

$$\mathrm{E}\left[R\right]\approx\gamma\,\mathrm{var}\left[R\right]-\frac{1}{6}\gamma\left(3\gamma-1\right)\mathrm{var}\left[R\right]^{3/2}\mathrm{skew}\left[R\right]. \tag{1}$$

With volatility of 18.5%, the equity risk premium required to get the investor to invest fully in the market depends on the coefficient of risk aversion of the investor and the skew coefficient of the market as shown below.

Risk aversion	**Skew coefficient**	
	0	**-1.32**
1	3.42%	3.70%
3	10.27%	13.61%
5	17.11%	26.86%

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 is proportional, roughly, to γ^2 and, as the table shows, is significant at fairly moderate levels of risk aversion.

The required 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 equity premium attributable to negative skew aversion actually increases with horizon unless the skew coefficient itself attenuates with the square root of the horizon – hence the

[2] Our definitions of volatility and skewness (discussed at length below) are non-standard. If returns are lognormal, our volatility is equal to the standard deviation of log returns, and our skewness is zero.

importance of understanding the behaviour of skew with horizon. We explore this relationship in our empirical work.

The rest of the paper proceeds as follows. In Section 1 we develop the theoretical relationship between low-frequency skewness and kurtosis and their high-frequency counterparts. In Section 2 we demonstrate the power of the technique through simulation. Section 3 provides an empirical application to the US stock market. Section 4 concludes.

1. THE THEORY

1.1 Moments of price changes

We work in a discrete time setting, $t \in \mathbb{Z}$. The asset has discounted price P_t ("the price").We are concerned with the distribution of returns from time t to $t+T$. For brevity, we refer to the time increment as a day, and the long horizon as a month, but obviously nothing hangs on this. The term kurtosis is used specifically for excess kurtosis.

The problem we are interested in is

> [**P**]: Let $P := \{P_t \mid t = ...,0,1,...\}$ be a strictly positive martingale process, whose associated returns process r, where $r_t := P_t/P_{t-1}$, is strongly stationary. The long horizon returns process R is defined by $R_t := P_t/P_{t-T}$. How can one estimate higher moments of long horizon returns R efficiently, assuming that 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 $P := \{P_t \mid t = ...,0,1,...\}$ be a real-valued (not necessarily positive) martingale process whose associated *difference* process d, where $d_t := P_t - P_{t-1}$, is strongly stationary. The long horizon difference process D is defined by $D_t := P_t - P_{t-T}$. How can one estimate higher moments of D efficiently, assuming that these moments exist?

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

Proposition 1

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

$$
\begin{aligned}
&\operatorname{vol}[D_t] = \operatorname{vol}[d_t];\\
&\operatorname{skew}[D_t] = \left(\operatorname{skew}[d_t] + 3\frac{\operatorname{cov}\left[y_{t-1}^*, d_t^2\right]}{\operatorname{var}[d_t]^{3/2}} \right) T^{-1/2};\\
&\operatorname{kurt}[D_t] = \left(\operatorname{kurt}[d_t] + 4\frac{\operatorname{cov}\left[y_{t-1}^*, d_t^3\right]}{\operatorname{var}[d_t]^2} + 6\frac{\operatorname{cov}\left[z_{t-1}^*, d_t^2\right]}{\operatorname{var}[d_t]^2} \right) T^{-1};
\end{aligned}
\tag{2}
$$

where

$$
\begin{aligned}
y_t^* &:= \sum_{u=0}^{T-1} \left(P_t - P_{t-u}\right)/T; \quad \text{and}\\
z_t^* &:= \sum_{u=0}^{T-1} \left(P_t - P_{t-u}\right)^2/T.
\end{aligned}
\tag{3}
$$

Proof: the full proof is in the Appendix.

Proposition 1 gives expressions for the volatility (the square root of the variance rate), the skewness and the excess kurtosis of monthly price changes. The first result is familiar: the volatility of price changes is the same whether computed from monthly or daily data. The second result says that skew at the monthly horizon has just two sources: daily skew and a term we call leverage. Daily skew attenuates with horizon with the square root of time. The leverage term is proportional to the covariance between squared price changes and the quantity y^*, which is equal to the difference between the opening price on the day and the average price over the last month.

The final result says that the kurtosis of monthly returns has just three sources: daily kurtosis attenuating with time, the covariance between cubed price changes and y^*, and the covariance between squared price changes and z^*. z^* is a measure of the average squared price change over the last 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

$$
D_t^3 = \sum_{u=0}^{T-1} d_{t-u}^3 + 3\sum_{u=0}^{T-1} \left(P_{t-u-1} - P_{t-T}\right) d_{t-u}^2 + 3\sum_{u=0}^{T-1} \left(P_{t-u-1} - P_{t-T}\right)^2 d_{t-u}. \tag{4}
$$

Taking conditional expectations of both sides, the third term drops out because of the martingale assumption[3], so

[3] If the price process were not martingale, there would be an additional term in the skew, the covariance between price changes and past volatility. But there is reason to believe that any such term would be small, at least in the

$$\mathrm{E}_{t-T}\left[D_t^3\right]=\sum_{u=0}^{T-1}\mathrm{E}_{t-T}\left[d_{t-u}^3\right]+3\sum_{u=0}^{T-1}\mathrm{E}_{t-T}\left[\left(P_{t-u-1}-P_{t-T}\right)d_{t-u}^2\right]. \tag{5}$$

Define

$$y_{t-1}^* := \sum_{u=1}^{T}\left(P_{t-1}-P_{t-u}\right)/T. \tag{6}$$

y_{t-1}^* is the difference between today's opening price and the T-day moving average[4]. Using strict stationarity, the conditional expectations can be replaced by unconditional expectations. Substituting y_t^* in to (5) gives the following expression for the unconditional third moment

$$\mathrm{E}\left[D_t^3\right]=T\mathrm{E}\left[d_t^3\right]+3T\mathrm{E}\left[y_{t-1}^*d_t^2\right]. \tag{7}$$

y_{t-1}^* is mean zero, so the expectation can be replaced by the covariance, giving

$$\mathrm{E}\left[D_t^3\right]=T\left(\mathrm{E}\left[d_t^3\right]+3\operatorname{cov}\left(y_{t-1}^*,d_t^2\right)\right). \tag{8}$$

A similar argument shows that

$$\mathrm{E}\left[D_t^2\right]=T\mathrm{E}\left[d_t^2\right]. \tag{9}$$

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

1.2 Moments of Returns

The objective is to produce a result akin to Proposition 1, but one that applies to moments of returns rather than to price changes. We now work with daily returns, $r_t = P_t/P_{t-1}$, and monthly returns, $R_t\left(T\right)=P_t/P_{t-T}$; we drop the argument of *R* where it causes no confusion.

The problem is intractable if we stay with the standard definitions of moments. It is necessary to modify the definition of moments. Define

---

case of the equity market. As Bollerslev *et al* (2013, p210) say: "The most striking empirical regularities to emerge from this burgeon literature are that …returns are at best weakly positively related, and sometimes even negatively related, to past volatilities."

[4] The asterisk is used to distinguish this variable from the corresponding variable in the problem **P**.

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

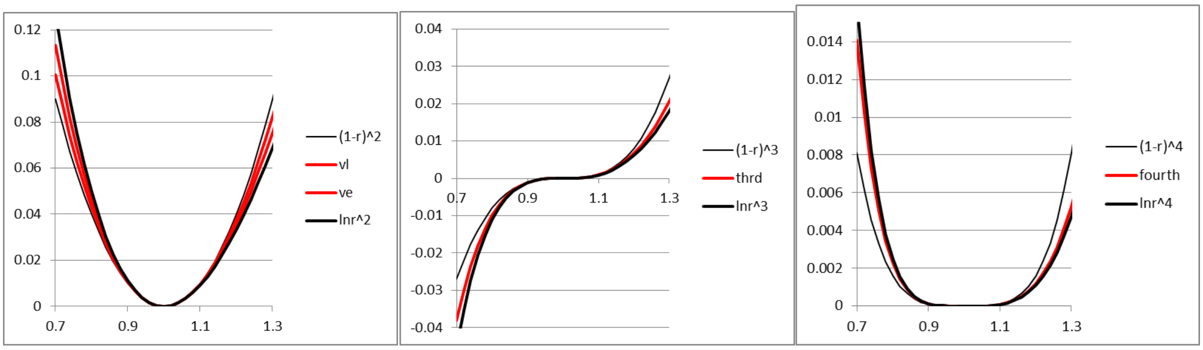
$x^{(2,L)}$ approximates the second power of log returns, as does $x^{(2,E)}$. Similarly, $x^{(3)}$ and $x^{(4)}$ approximate the third and fourth powers. This can be shown by doing a Taylor expansion, and is seen graphically in Figure 2. Modifying the definitions of moments in this way is not unprecedented. The Model Free Implied Variance is widely used by both academics and practitioners. It is defined as

$$
MFIV[R] := \mathrm{E}^{\mathrm{Q}}\left[x^{(2,L)}(R)\right], \tag{11}
$$

where the Q superscript denotes that the expectation is under the pricing measure. It also follows the definition of realized variance in Bondarenko (2014). A definition of skewness similar to the above is seen in Neuberger (2012).

We also define volatility of the return (vol[r]) as the square root of the variance rate.

Figure 2: approximating the moments of returns



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 volatility, skewness and kurtosis of monthly returns (as defined in equation (10)) is related to the distribution of daily returns as follows

$$
\begin{aligned}
&\operatorname{var}^L\left[R_t\right]=T\operatorname{var}^L\left[r_t\right];\\
&\operatorname{skew}\left[R_t\right]=\left(\operatorname{skew}\left[r_t\right]+3\frac{\operatorname{cov}\left[y_{t-1},x^{(2,E)}\left(r_t\right)\right]}{\operatorname{var}^L\left[r_t\right]^{3/2}}\right)T^{-1/2};\\
&\operatorname{kurt}\left[R_t\right]=\left(\operatorname{kurt}\left[r_t\right]+4\frac{\operatorname{cov}\left[y_{t-1},x^{(3)}\left(r_t\right)\right]}{\operatorname{var}^L\left[r_t\right]^{2}}+6\frac{\operatorname{cov}\left[z_{t-1},x^{(2,L)}\left(r_t\right)\right]}{\operatorname{var}^L\left[r_t\right]^{2}}\right)T^{-1};
\end{aligned}
\tag{12}
$$

where

$$
\begin{aligned}
&y_t:=\sum_{u=0}^{T-1}\left(R_t\left(u\right)-1\right)\Big/T \quad \text{and}\\
&z_t:=\sum_{u=0}^{T-1}2\left(R_t\left(u\right)-1-\ln\left(R_t\left(u\right)\right)\right)\Big/T.
\end{aligned}
\tag{13}
$$

Proof: the proof is similar to that of Proposition 1; details in the Appendix.

Proposition 2 is very similar to Proposition 1. It can be seen that

- The volatility of monthly returns is identical to the volatility of daily returns.
- 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 volatility and past returns. Past returns are measured by y, which is the net return relative to the one month moving average[5].
- Kurtosis in daily returns generates a much smaller ($1/T$) kurtosis in monthly returns.
- If annual returns are significantly leptokurtic, it is for one of two reasons:
  - because daily skew is correlated with past returns (as measured again by y);
  - or because of a GARCH effect whereby current variance is correlated with past variance. Past variance is measured by z, which is a function of the average realized variance over horizons of up to one month, again with more recent experience having more weight.

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 volatility and lagged returns and lagged squared returns (volatility is constant) 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 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 volatility and lagged returns. The persistence of

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

volatility shocks also generates kurtosis. GARCH processes also generate 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 i.e. as with Heston, a correlation between volatility and lagged returns.

1.3 Relation to Option-based realized skewness

Proposition 2 shows that the skewness of long horizon returns is related 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 this section we show how these two results are related.

The relationship can be sketched out informally. Skewness comes from leverage. Leverage at the monthly horizon is the covariance between today's realized variance and returns over the last month. By rearranging terms, it can be seen that this is the same as the covariance between today's return and realized variance over the next month. If there exists a sufficiently rich options market, we can observe the corresponding one month implied variance. In the absence of risk premia in the options market, the implied one month variance at any time is the expectation of the realized variance over the next month. We can then obtain a much more precise estimate of the leverage effect by looking not at the covariance between daily returns and realized variance over the next month, but at the covariance between daily returns and contemporaneous daily changes in monthly implied variance.

We can state the argument more precisely. Proposition 2 shows the relation between the third moment of long horizon returns and high frequency returns

$$E\left[x^{(3)}\left(R_t\right)\right]=T\left\{E\left[x^{(3)}\left(r_t\right)\right]+3\operatorname{cov}\left(y_{t-1},x^{(2,E)}\left(r_t\right)\right)\right\}. \tag{14}$$

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

$$\begin{aligned}&\operatorname{cov}\left(y_{t-1},x^{(2,E)}\left(r_t\right)\right)=\operatorname{cov}\left(r_t-1,w_{t+T-1}\right)\\&\text{where } w_t=\frac{1}{T}\sum_{u=1}^{T-1}R_{t+u-T}\left(u-1\right)x^{(2,E)}\left(r_{t+u+1-T}\right).\end{aligned} \tag{15}$$

(We take advantage of the fact that both y and r-1 are mean zero.) The variable 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 (15) or the left hand side. But suppose now that we can observe the expectation of future realized variance. Then there is potential for considerable efficiency gains. r_t-1 is known at time t; it is also mean zero. So

$$\operatorname{cov}\left(r_t-1,w_{t+T-1}\right)=\operatorname{cov}\left(r_t-1,\mathrm{E}_t\left[w_{t+T-1}\right]-h_{t-1}\right) \tag{16}$$

where h_t is any variable known at time t. Suppose we choose h_{t-1} so that it close to $\mathrm{E}_{t-1}\left[w_{t+T-1}\right]$, we can the estimate the right hand side of equation (16) much more precisely than the left hand side since the change in the expectation of future variance over the day is likely to have a far lower standard deviation than the realized variance itself.

To ensure that expectations if future variance are observable, we need to make 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 like 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 $q_{t+1} = \mathrm{E}_{t+1}\left[Tw_{t+T}\right]$. We therefore have the result that

$$\operatorname{cov}\left(r_t - 1, w_{t+T-1}\right) = \frac{1}{T}\operatorname{cov}\left(r_t - 1, q_t - q_{t-1}\right). \tag{17}$$

Assuming complete markets 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.

1.4 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. But there are advantages in going to continuous time. The results are simpler, particularly if the price process is continuous. We can also derive a simple useful test for estimating how return moments change with horizon.

We have assumed that P is a positive martingale, with a strictly stationary returns process with well-defined moments. We now make the further assumption that the process is a diffusion. We assume in particular that P can be represented by a stochastic differential equation

$$dP_t / P_t = \sqrt{v_t}\, dz_t, \tag{18}$$

where v_t is predictable, and z_t is a standard Brownian process.

We retain the definitions of variance, skewness and kurtosis that we used in the discrete time setting. The counterpart to Proposition 2 in a diffusion setting is then

Proposition 3

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

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

where

$$\begin{aligned}
y_t(T) &:= \int_{u=0}^{T}\left(R_t(u)-1\right)du\Big/T \quad \text{and}\\
z_t(T) &:= \int_{u=0}^{T}x^{(2,L)}\left(R_t(u)\right)du\Big/T.
\end{aligned}$$

Proof: the proof is in the Appendix.

The most significant difference between Propositions 2 and 3 is the dropping of the daily skewness from the period skewness, and the dropping of the daily kurtosis and the cube effect from the period kurtosis. With the diffusion assumption, the higher order moments of high frequency returns vanish. The distinction between entropy variance and log variance vanishes in the limit. The definitions of y (the lagged return) and z (the lagged realized variance) are the natural limits of their discrete time counterparts.

We show in our empirical work that, at least so far as the equity market is concerned, jumps do not play any significant role in the moments of long horizon returns. We now see that, in the absence of jumps, all skewness in period returns derives from the leverage effect, and all kurtosis comes from the GARCH effect.

The term structure of moments is a matter of considerable importance; if skewness and kurtosis tend to zero at long horizons, then these higher moments are likely to be of limited significance for longer term investors. Proposition 3 enables us to test this directly.

Corollary to Proposition 3

Given two horizons, T_1 and T_2:

$$\begin{aligned}
\operatorname{skew}(T_2) > \operatorname{skew}(T_1) &\text{ if and only if } \operatorname{cov}\left[\left(y(T_2)-\sqrt{\frac{T_2}{T_1}}y(T_1)\right),v\right] > 0;\\
\operatorname{kurt}(T_2) > \operatorname{kurt}(T_1) &\text{ if and only if } \operatorname{cov}\left[\left(z(T_2)-\frac{T_2}{T_1}z(T_1)\right),v\right] > 0.
\end{aligned} \tag{19}$$

We will use this corollary to test how the skewness of stock market returns changes with horizon.

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. 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 an EGARCH specification. For each model we simulate 10,000 paths for daily returns, each of length 5000 (i.e. roughly 20 years).

The parameters for each model are derived from fitting them to spans of daily US stock market returns. For the GBM and the Heston model, the parameters are taken from Eraker (2004). Those estimations use daily S&P-500 returns from January 2nd 1980 to December 31st 1999. The EGARCH parameters are obtained from our own fit of such a model to daily value-weighted CRSP US stock returns covering the period from January 2nd 1980 to the end of December 2015.

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 (i.e. roughly monthly) returns, where these are as defined in equation (9). We use three estimation techniques for each moment. First, we construct the sample moments of non-overlapping 25-day returns (and we refer to these subsequently as 'Monthly' estimates). Second, we measure the sample moments using overlapping 25-day returns (referred to later as 'Overlapping' estimates).[6] Finally, we implement the estimators from Proposition 2 (which we label 'NP').

Results from these simulations are given in Table 1. 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 EGARCH results. Each table 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.

Under the assumption that daily returns follow a GBM, 25-day skewness and kurtosis should both be zero. Table 1 confirms that, on average, this is true for all three estimation techniques. More importantly, the dispersion of the estimates for the NP method are greatly reduced relative to those from monthly and overlapping estimators. The standard deviations of estimates from our method are between 70% and 90% smaller than those from the alternatives. The improvement in estimation accuracy for the NP method is most striking for skewness, but only

[6] So for each simulated return path of 5,000 data points, the 'Monthly' estimator uses 200 non-overlapping 25-day returns and the 'Overlapping' estimator uses 4,976 overlapping 25-day returns.

slightly less impressive for kurtosis. Overall, for both skewness and kurtosis, the use of daily data to improve monthly moment estimates provides a substantial improvement in accuracy.

Table 1: simulation results for NP and standard estimators

Panel A: Geometric Brownian Motion			
Standard deviation			
	NP	Monthly	Overlap
Mean	0.0469	0.0468	0.0468
STDEV	0.0005	0.0021	0.0019
Coefficient of Skewness			
	NP	Monthly	Overlap
Mean	-0.0062	-0.0138	-0.0135
STDEV	0.0352	0.2487	0.2438
Excess Kurtosis			
	NP	Monthly	Overlap
Mean	-0.0018	-0.0264	-0.0219
STDEV	0.0702	0.3094	0.2099
Panel B: Heston model			
Standard deviation			
	NP	Monthly	Overlap
Mean	0.0468	0.0468	0.0468
STDEV	0.0023	0.0033	0.0033
Coefficient of Skewness			
	NP	Monthly	Overlap
Mean	-0.2728	-0.261	-0.2591
STDEV	0.1023	0.3555	0.3483
Excess Kurtosis			
	NP	Monthly	Overlap
Mean	1.0658	1.0046	1.0147
STDEV	0.3788	0.852	0.686
Panel C: EGARCH model			
Standard deviation			
	NP	Monthly	Overlap
Mean	0.0605	0.0603	0.0604
STDEV	0.0032	0.0048	0.0049
Coefficient of Skewness			
	NP	Monthly	Overlap
Mean	-0.6705	-0.6424	-0.6315
STDEV	0.1868	0.451	0.4366
Excess Kurtosis			
	NP	Monthly	Overlap
Mean	2.1651	1.9979	1.9925
STDEV	1.453	2.4871	2.0865

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). All three estimation techniques pick these features up, but again use of the NP method results in a significant reduction in the spread of estimation errors. For skewness, the standard deviation of estimates for the new method is around 70% smaller than those of the monthly or overlapping methods, while for kurtosis improvements are between 40% and 50%.

Finally, the estimated EGARCH model also generates negative skew and excess kurtosis and these appear in all estimation methods. Panel C shows that the NP estimation technique dominates in terms of accuracy under this model also but that the improvements it delivers are less pronounced. The standard deviation of skewness estimates is around 60% smaller for the new method but the standard deviation of the kurtosis estimates drops by only 30 to 40%.

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 GBM and Heston models than they are for the EGARCH specification. But in almost all cases, use of the NP moment estimators leads to the dispersion of estimated coefficients being reduced by 50% or more.

2.2 Simulations of the NP estimator's performance using intra-day data

Given the improvements in estimation precision that are available 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. From the results in papers such as Andersen, Bollerslev, Diebold and Labys (2003) we know that if we wish to estimate daily return variances, the use of finely sampled intra-day data is valuable. Here we explore an analogous issue but for estimation of higher moments of lower frequency returns.

Thus 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). In our simulations, we vary N_D between 1 (daily data) and 16. We start off with a benchmark case where we assume that returns are generated from a Geometric Brownian Motion and then move to a Heston model with the same (daily) parameters as in the previous section.

The simulation results for the NP estimator only and for both data generating processes are given in Table 2.

The results in Table 2 demonstrate 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 16 times a day reduces the standard deviation of the distribution of moment estimates by a factor of four relative to the daily returns case.

Table 2: Intraday simulation results

Panel A: Geometric Brownian Motion

N_D	1	2	4	8	16
			Standard deviation		
Mean	0.0469	0.0469	0.0469	0.0469	0.0469
STDEV	0.0005	0.0003	0.0002	0.0002	0.0001
			Coefficient of Skewness		
Mean	-0.006	-0.0029	-0.0014	-0.0007	-0.0004
STDEV	0.035	0.0246	0.0174	0.0122	0.0087
			Excess Kurtosis		
Mean	-0.0024	-0.0019	-0.0012	-0.0006	0
STDEV	0.0702	0.0491	0.0347	0.0245	0.0172

Panel B: Heston model

N_D	1	2	4	8	16
			Standard deviation		
Mean	0.0469	0.0469	0.0469	0.0469	0.0468
STDEV	0.0023	0.0023	0.0023	0.0023	0.0023
			Coefficient of Skewness		
Mean	-0.2709	-0.2741	-0.2765	-0.2772	-0.2781
STDEV	0.1006	0.0919	0.0865	0.0831	0.0827
			Excess Kurtosis		
Mean	1.0577	1.0753	1.077	1.0805	1.083
STDEV	0.3721	0.3372	0.3241	0.3177	0.3245

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. If data is sampled 16 times per day, then the precision with which skewness is estimated improves by about 20% and the corresponding figure for excess kurtosis is 10%.

These results are linked to the persistence in volatility that our Heston 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 than daily does not help materially in estimating, for example, the covariance between variance and lagged returns that is important in measuring skewness and so our estimators derive smaller benefit from the use of intra-day data in this case.

2.3 Performance of the NP skew estimator across skew levels

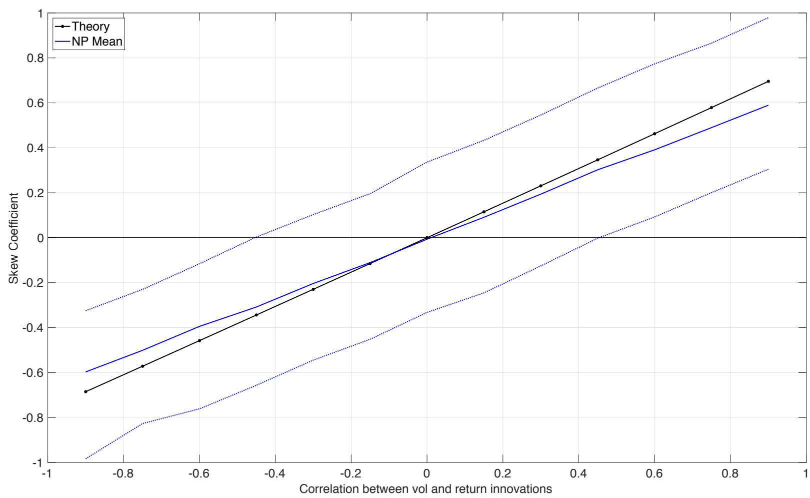
In order to investigate how the performance of our skew estimator changes with the level of skew in returns, we take a Heston model and vary the correlation between return and variance innovations between -0.9 and +0.9 (with the former giving large negative skewness and the latter generating large positive skewness). All other parameters are set at the values from Eraker

(2004). For each parameter set, our simulation contains 1,000 replications of 1,000 daily returns and from these we estimate 25-day skew.

The results are summarised in Figure 3. 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 1,000 runs, as well as the 5th and 95th percentiles of the distribution of skew estimates. Also plotted on Figure 3 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, though, with the largest bias around 0.1 when theoretical skewness is at a value of 0.9. The range between 5th and 95th percentiles is fairly stable at a value of around 0.65. The bias in the estimation of the coefficient of skewness arises due to the fact that it is a ratio of the estimated third moment to the cube of the estimated standard deviation. Estimates of both of these moment measures using the NP method are unbiased, but estimation errors in second and third moments are correlated and it is this that causes the bias in the estimated skewness.

Figure 3: Theoretical and estimated skew coefficient versus correlation parameter: mean, 5th and 95th percentiles.



Obviously, as one increases the quantity of high-frequency data points used to construct low frequency skew, estimation becomes more accurate. If one runs simulations of time-series of length 10,000, rather than 1,000, precision improves greatly, with the bias dropping to close to zero and the range between the 5th and 95th percentiles falling to around 0.2.

2.4 Comparison of NP and quantile-based skew measures

Ghysels, Plazzi and Valkanov (2016) (hereafter GPV) propose a skewness estimator based on the quantiles of the return distribution in their recent work on international asset allocation. We now compare the performance of our estimator and their preferred estimator based on data simulated from a Heston model. We use exactly the same setup as in Section 3.2.1, except now we estimate NP skewness and GPV's quantile skewness for each simulated set of data.

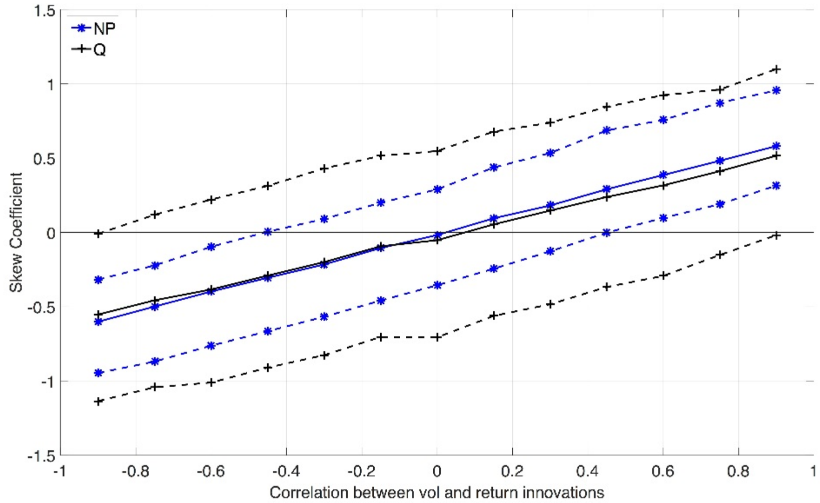
The GPV skew estimator is as follows;

$$6 \times \frac{\int_0^{0.5}\{[q_\alpha(r_t) - q_{0.5}(r_t)] - [q_{0.5}(r_t) - q_{1-\alpha}(r_t)]\}d\alpha}{\int_0^{0.5}[q_\alpha(r_t) - q_{1-\alpha}(r_t)]d\alpha} \times \frac{\int_0^{0.5} q_\alpha(z)\, d\alpha}{\int_0^{0.5} q_\alpha^2(z)\, d\alpha}$$

where r_t are returns measured at the frequency of interest (e.g. monthly), $q_\alpha(x)$ is the αth 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 GPV estimator estimates skew by looking directly 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 GPV estimator to overlapping 25-day returns. It is worth re-iterating that the GPV estimator and the estimator proposed here are designed to target slightly different measures of skewness. GPV propose an estimator of the traditional skewness coefficient whereas our estimator is of the modified skew coefficient as defined in equation (9). However, differences in these targets are minor.

Figure 4: estimated skew coefficients from the NP and quantile methods: means and 5th and 95th percentiles.



Notes: solid lines give the mean value of the estimated coefficients and the dashed lines give the 5th and 95th percentiles of the distribution of estimated coefficients. Lines marked with * symbols are for the NP estimators and those marked with + symbols are the quantile-based estimators.

The results from our comparison are displayed in Figure 4. As before, the x-axis values are the Heston correlation parameters and skewness is on the y-axis, and again we run 1000 simulations of 1000 daily returns from which we estimate 25 day skewness. The results are encouraging. The NP and the GPV mean estimates are very close together, but the precision of the NP estimator is much greater. The 5^{th}-95^{th} percentile range of the GPV estimates average around 1.3 i.e. around twice as large as that of the NP estimator.

Thus, overall, our estimation technique works well. 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 1926-2015 and were retrieved from Ken French's data library.

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 pointing out a couple of implementation issues. First, for the sake of simplicity, our theory has focussed on unconditional moments. We can, however, adapt the theory to deal with conditional moments with little difficulty. The second issue to address is the method of estimating covariances. Our skew estimator requires us to estimate terms such as $\mathrm{cov}[y_t, v_t]$ over some period $[0, S]$. The obvious estimator is the sample covariance

$$Q_1 := \sum_{t=1}^{S} \left(y_t - \bar{y}\right)\left(v_t - \bar{v}\right) / \left(S-1\right) \quad \text{where } \bar{y} := \sum_{t=1}^{S} y_t / S.$$

But this is biased. Specifically

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

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 use the martingale assumption to set the estimated mean of y to 0, while we set the estimated mean of v to its sample mean in the period prior to the sample period (empirically we use a 5-year period before the start of the sample). Denoting this mean by v_0, the covariance estimator we use is

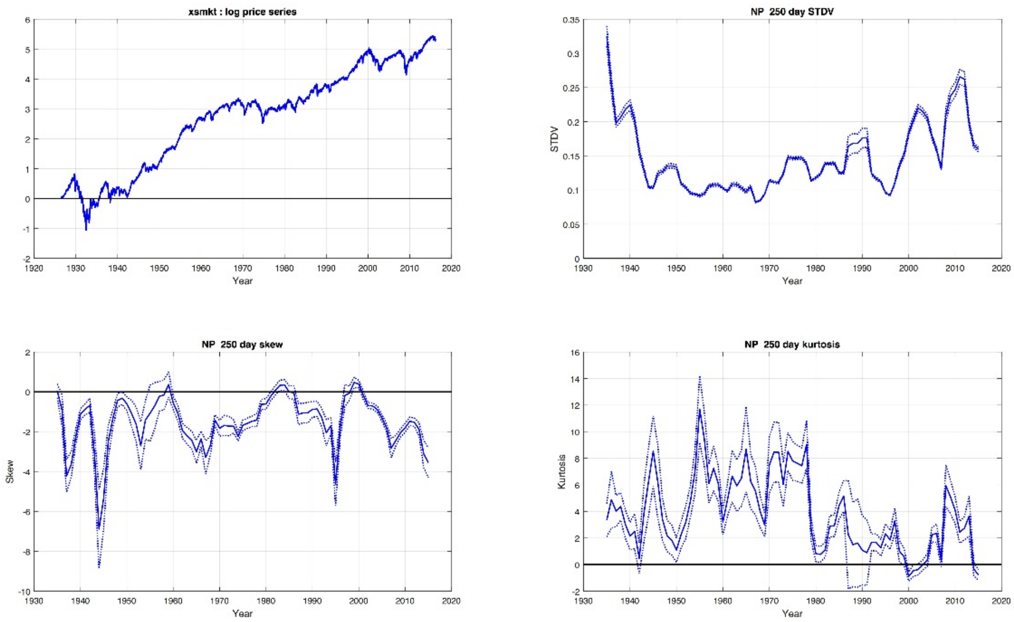
$$Q_2 := \sum_{t=1}^{S} y_t \left(v_t - v_0 \right) / S.$$

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).

The top left panel of Figure 5 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 (around 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.

Figure 5: 250-day moments of US stock market returns



The NP skew data in the bottom left panel indicates that annual stock market skew is almost always significantly negative (with a mean of -1.5) over our 90 years of data. The only exceptions to this are a couple of isolated years in the mid-1980s and late 1990s when skew is significantly positive, although very small in magnitude. Times of particularly severe negative skewness include the Great Depression (with skew below -6) and the mid-1990s (with skew around -4) and skew has also reached a level of close to -3.5 in the most recent part of our data. Interestingly, the 2008 Financial Crisis does not appear to be associated with tremendously large negative skew. 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 5 shows estimates of annual excess return kurtosis estimated from daily data. As expected, excess kurtosis is positive on average, with a mean of 2.8, and is almost never negative. Excess kurtosis is larger in the first half of the 20th century than it is in the second half, but in the second half of the century it is at its highest level in the most recent part of the data.

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 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 using data from within that year compute monthly volatility, skew and excess kurtosis. Time-series plots of the three moments estimated using the NP method plus the quantile based skew measure of GPV are presented in Figure 6.[9]

While Figure 6 leads to broadly the same results as Figure 5 (i.e. the US stock market return is on average very negatively skewed and displays excess kurtosis), as one would expect monthly skew and monthly kurtosis are much more volatile than their annual counterparts. 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 (with a mean of -0.25), but it is less easy to see a pattern in the monthly skews here than it is in the NP estimates. The quantile skew and NP skew measures are positively correlated, with a correlation coefficient of 0.40.

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 lagged returns and current cubed returns (which we refer to as the ‘Cube’ component) and covariation between current and lagged squared returns (which we will call the GARCH effect).

[9] 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.

Figure 6: time-variation in monthly moment estimates for US stock market

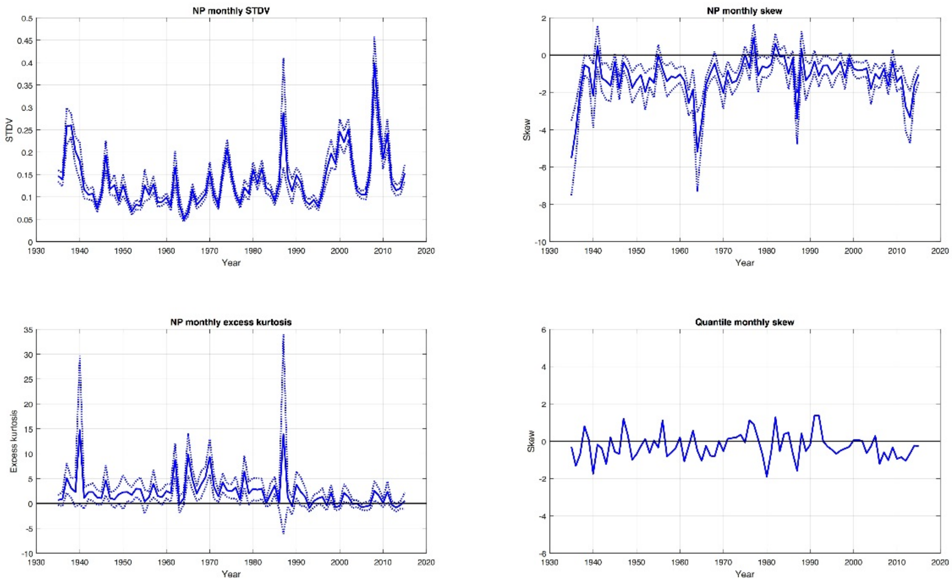


Figure 7 shows the time-series variation in the two monthly skew components for the years 1935 to 2015, with the two plots having identical vertical scales and with 95% confidence bands plotted around each estimate.

Clearly the leverage effect generates both the level and the variation in skewness. The influence of skew in daily returns is negligible and almost never statistically different from zero. Thus, both the average level of 25-day skew and its variation through time are attributable to covariation between lagged returns and current squared returns. This covariance is almost always negative, usually significantly so and is almost never significantly greater than zero.

Figure 8 shows a similar decomposition of 25-day kurtosis into its three components (i.e. daily kurtosis, the cube term and the GARCH term). As with skew, the contribution of the daily moment is close to zero and its time-series variation is small. The Cube term is also close to zero on average and so almost all the significant positive excess kurtosis apparent in the data, as well as the time-variation in that excess kurtosis, comes from the GARCH component.

Thus, we have shown here that 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 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.

Figure 7: time variation in components of monthly skew coefficient

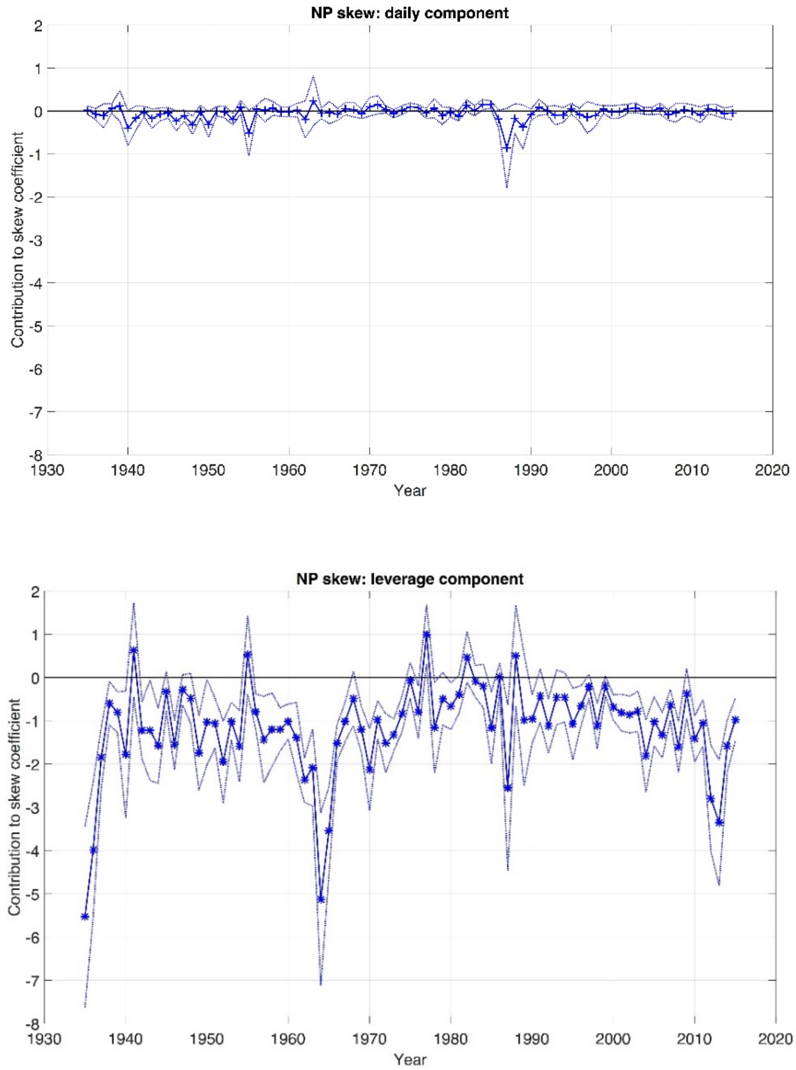
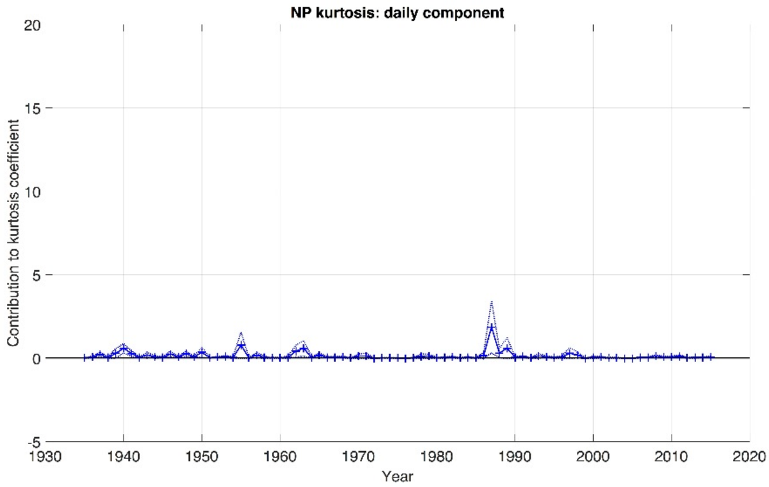
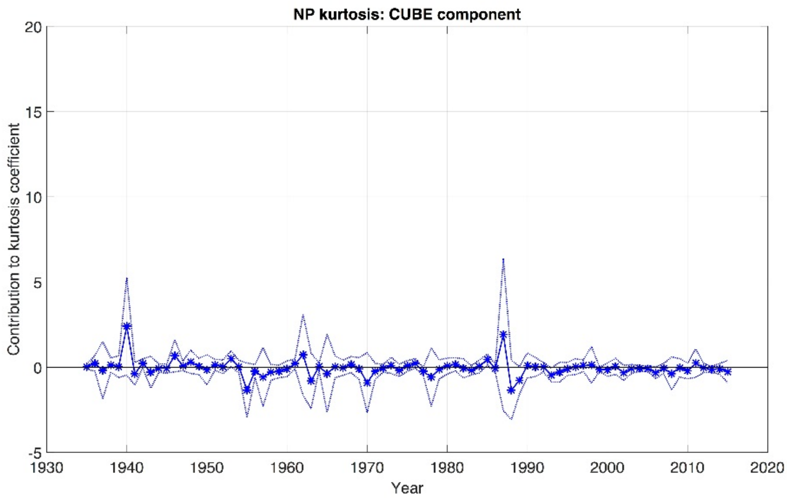
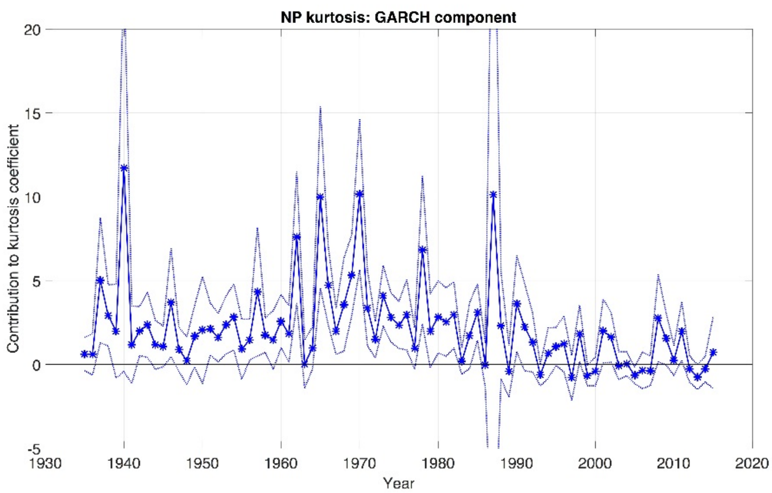


Figure 8: time variation in components of monthly excess kurtosis







3.3 Estimates of monthly skewness using intra-day data

We now focus our attention on the estimation of monthly return standard deviations and monthly skewness. We 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 3 gives the results.

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 above 0.9, and the average coefficient of skewness estimated from intra-day data is more negative than that from daily data.

Table 3: estimation US stock market skewness with intra-day data

Standard Deviation			
	10min	30min	Daily
Mean	0.053	0.053	0.056
STDEV	0.026	0.025	0.029
Corr(Daily)	0.996	0.996	1.000
Third moment			
	10min	30min	Daily
Mean	-0.00042	-0.00040	-0.00048
STDEV	0.00083	0.00077	0.00097
Corr(Daily)	0.999	0.999	1.000
Coefficient of skewness			
	10min	30min	Daily
Mean	-1.37	-1.34	-1.29
STDEV	0.63	0.59	0.71
Corr(Daily)	0.944	0.936	1.000

Thus, while the simulations show that the use of intra-day data can lead to improvements in estimation precision, in our empirical work it seems that the additional value from collecting high-frequency data is small. Of course these results are based on an examination of monthly return moments. It is possible that intra-day data could be much more informative for a researcher/investor interested in, for example, weekly moments.

3.4 The term structure of skewness

Our analysis thus far confirms the existence of significant negative skew in monthly and annual US stock market returns. It is reasonable to ask how measured skewness varies across a range of possible horizons, from monthly to multi-year return skew. Via such analysis one can ask, for example, whether investors with a 5 year investment horizon need to worry about skewness and begin to comment upon the compensation they might demand for holding portfolios with skewed return series.

As Proposition 3 and its corollary make clear, the variation in skewness with horizon is driven by $cov(y(T), v)$ where T is horizon. In Table 4 we present estimates of skewness for T ranging from 25 days (roughly 1 month) to 1250 days (roughly 5 years). We present these numbers for the full sample of data and then separately for data up to the end of 1970 and data after 1970. Alongside each estimate of skewness we give a t-test of the null hypothesis that the skew at that horizon is significantly different from 250-day (i.e. annual) skewness. This t-test is built using the results in the corollary to Proposition 3. Figure 9 plots the skew term structures (for the full sample and the two subsamples respectively).

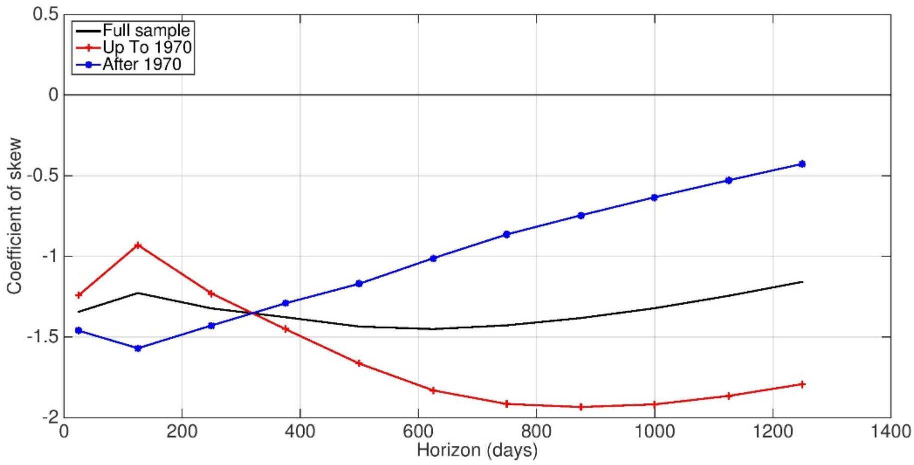
Table 4

	Full sample		**Up to 1970**		**After 1970**	
Horizon	**Skew**	**t-test**	**Skew**	**t-test**	**Skew**	**t-test**
25	-1.344	-0.181	-1.241	-0.164	-1.462	-0.094
125	-1.228	1.097	-0.930	2.218	-1.570	-1.686
250	-1.323	-	-1.123	-	-1.429	-
375	-1.378	-1.152	-1.452	-3.015	-1.291	3.764
500	-1.436	-1.450	-1.665	-3.600	-1.170	3.882
625	-1.451	-1.197	-1.831	-3.708	-1.012	4.326
750	-1.428	-0.764	-1.916	-3.486	-0.864	4.822
875	-1.383	-0.308	-1.935	-3.124	-0.746	5.091
1000	-1.322	0.168	-1.918	-2.725	-0.634	5.565
1125	-1.245	0.710	-1.866	-2.261	-0.529	6.123
1250	-1.156	1.191	-1.793	-1.846	-0.427	6.554

Looking first at the term structure for the full sample, it is remarkably flat across horizons. Monthly skew is roughly -1.3 and 5-year skew is just above -1.2. Statistically, no estimate of skew at any horizon is different from annual skew. Thus, worryingly for long term investors and risk managers, there is no sign that negative skewness disappears at long horizons.

The index returns from the early part of the sample show a different pattern. Skew tends to become more negative with horizon, and very significantly so. In this subsample, 1-year skew is close to the figure from the full sample (-1.12 versus -1.24) but 5 year skew is more than 50% larger than its full sample counterpart (-1.8 versus -1.2).

Figure 9: the term structure of skew coefficients



The data from the more recent subsample of data (from 1970 to today) show the opposite pattern. Skew is negative and relatively large in magnitude for short horizons, but it then attenuates with horizon, starting from -1.46 at a monthly horizon and reaching -0.43 at a 5-year horizon. This reduction in the magnitude of the skew in returns is strongly significant. Thus in recent data, asymmetry in returns is more pronounced at shorter sampling frequencies. Having said this, 5 year returns are still far from normal as a skew of -0.43 represents very strong asymmetry.

3.5 Comparison of skewness estimates with implied and realized skew

Finally, we compare the skewness figures obtained from our method with those that rely on data from the options market. Our option-implied measures are the implied and realized skewness presented and employed in Neuberger (2012) and Kozhan, Neuberger and Schneider (2013).

As Section 1.3 demonstrates, 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 volatility and lagged returns (as we do) or by the covariance between current returns and changes in option implied volatilities (as Kozhan, Neuberger and Schneider (2013) do).

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 have an estimate of the realized and implied third moment using data only from that month. We also have an end of month NP estimator of the same two quantities 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.

Table 5: comparing NP skewness with option-based skew measures

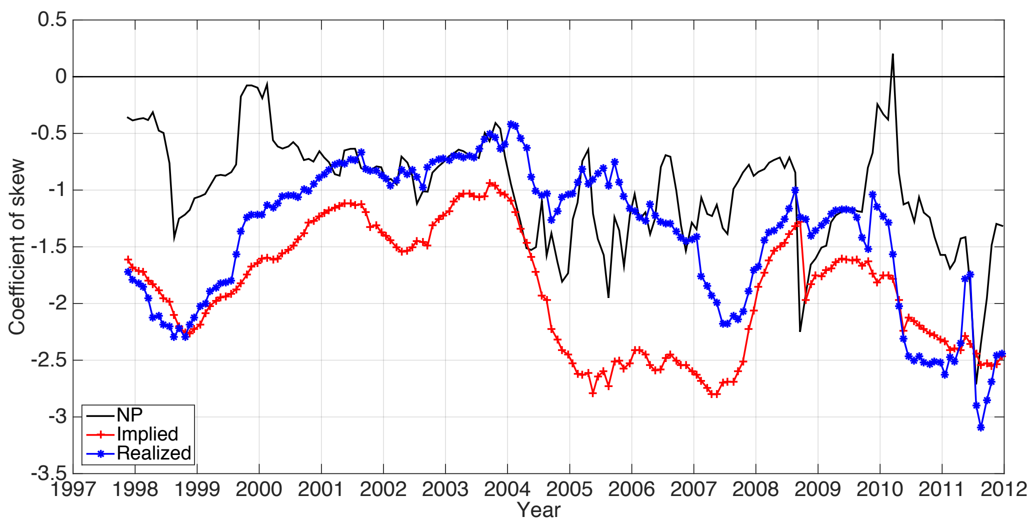
Third moment			
	NP	**Implied**	**Realized**
Mean	-0.00036	-0.00066	-0.00041
STDEV	0.00065	0.00074	0.00063
Corr with NP	1.000	0.912	0.972

Skewness			
	NP	**Implied**	**Realized**
Mean	-0.998	-1.879	-1.387
STDEV	0.461	0.529	0.612
Corr with NP	1.000	0.500	0.321

The results of our comparison are in Table 5. Looking first at the third moments, there are two key observations. First, the average value obtained from the NP method is on average about 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 somewhat smaller at 0.91.

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 -1.0, while mean realized skew is around 40% larger and implied skew almost twice as large. Implied and realized skewness are also much more volatile than NP skew and less strongly correlated with the NP measure than third moments. Having said this, the correlation between NP and implied skew is still approximately 0.5, showing that the two measures have much in common. Figure 10 confirms that the three skewness measures display very similar times series features, despite a clear difference in average value.

Figure 10: time variation in estimated skew coefficients from three methods



4. Conclusions

Measures of the higher moments of low-frequency (i.e. monthly or quarterly) 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 low frequency skewness and kurtosis with impressive precision. This precision is demonstrated via a set of simulation experiments in which returns are generated from a few popular data generating processes (e.g. a Heston model and a GARCH model). 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 practice in US stock index returns. Thus, we show, both analytically and empirically, that information on high-frequency skewness and kurtosis is close to irrelevant when it comes to measuring low frequency skew and kurtosis.

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 evidence at longer horizons is more equivocal. The degree of skew is economically significant in the sense that, using conventional preference assumptions, skew risk premia may not be that much smaller than variance risk premia.

REFERENCES

Ang, A., Chen, J. and Xing, Y., 2006. Downside risk. *Review of Financial Studies*, 19(4), pp.1191-1239.

Ang, A., Hodrick, R.J., Xing, Y. and Zhang, X., 2006. The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), pp.259-299.

Bakshi, G., Kapadia, N. and Madan, D., 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, *16*(1), pp.101-143.

Barberis, N. and Huang, M., 2007. Stocks as lotteries: The implications of probability weighting for security prices (No. w12936). *National Bureau of Economic Research*.

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

Bollerslev, T., Osterrieder, D., Sizova, N. and Tauchen, G., 2013. Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2), pp.409-424.

Bondarenko, O., 2014. Variance trading and market price of variance risk. *Journal of Econometrics*, *180*(1), pp.81-97.

Bowley, A.L., 1920. *Elements of statistics* (Vol. 2). PS King.

Boyer, B., Mitton, T. and Vorkink, K., 2010. Expected idiosyncratic skewness. *Review of Financial Studies*, *23*(1), pp.169-202.

Broadie, M., Chernov, M. and Johannes, M., 2007. Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, *62*(3), pp.1453-1490.

Conrad, J., Dittmar, R.F. and Ghysels, E., 2013. Ex ante skewness and expected stock returns. *The Journal of Finance*, *68*(1), pp.85-124.

Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance*, *59*(3), pp.1367-1403.

Ghysels, E., Plazzi, A. and Valkanov, R.I., 2016. Why invest in emerging markets? The role of conditional return asymmetry. *The Journal of Finance, 71(5),* 2145-2192.

Harvey, C. R., Liechty, J. C., Liechty, M. W., & Müller, P. 2010. Portfolio selection with higher moments. *Quantitative Finance*, 10(5), 469-485.

Harvey, C.R. and Siddique, A., 2000. Conditional skewness in asset pricing tests. *The Journal of Finance*, *55*(3), pp.1263-1295.

Heston, S.L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, *6*(2), pp.327-343.

Kelly, B. and Jiang, H., 2014. Tail risk and asset prices. *Review of Financial Studies*, *27*(10), pp.2841-2871.

Kraus, A. and Litzenberger, R.H., 1976. Skewness preference and the valuation of risk assets. *The Journal of Finance*, *31*(4), pp.1085-1100.

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

Mitton, T. and Vorkink, K., 2007. Equilibrium under-diversification and the preference for skewness. *Review of Financial Studies*, *20*(4), pp.1255-1288.

Neuberger, A., 2012. Realized skewness. *Review of Financial Studies*, *25*(11), pp.3423-3455.

Longstaff, F.A. and Piazzesi, M., 2004. Corporate earnings and the equity premium. *Journal of financial Economics*, *74*(3), pp.401-421.

Rietz, T.A., 1988. The equity risk premium a solution. *Journal of Monetary Economics*, *22*(1), pp.117-131.

APPENDIX

Proof of Proposition 1

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

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

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

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

Taking unconditional expectations and rearranging terms

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

Now

$$\begin{aligned}
&\mathrm{E}\left[y^*\right] = 0 \quad \text{and}\\
&\mathrm{E}\left[z^*\right] = \sum_{u=1}^{T}(u-1)\mathrm{E}\left[d^2\right]\Big/T = \tfrac{1}{2}(T-1)\mathrm{E}\left[d^2\right].
\end{aligned} \tag{4}$$

Replacing expectations of products in (3) by covariances and products of means we get

$$
\begin{aligned}
&\mathrm{E}\left[D^2\right]=T\mathrm{E}\left[d^2\right];\\
&\mathrm{E}\left[D^3\right]=T\left\{\mathrm{E}\left[d^3\right]+3\operatorname{cov}\left(y^*,d^2\right)\right\};\\
&\mathrm{E}\left[D^4\right]-3\mathrm{E}\left[D^2\right]^2=T\left\{\mathrm{E}\left[d^4\right]-3\mathrm{E}\left[d^2\right]^2+4\operatorname{cov}\left(y^*,d^3\right)+6\operatorname{cov}\left(z^*d^2\right)\right\}.
\end{aligned}
\tag{5}
$$

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

Proof of Proposition 2

Applying similar arguments to those used in the previous proof (algebraic decomposition, take conditional expectations, drop terms using the martingale property, replace conditional by unconditional expectations) we get

$$
\begin{aligned}
&\mathrm{E}\left[\ln R\right]=T\mathrm{E}\left[\ln r\right];\\
&\mathrm{E}\left[R\ln R\right]=T\mathrm{E}\left[r\ln r\right]+T\mathrm{E}\left[yr\ln r\right];\\
&\mathrm{E}\left[\left(\ln R\right)^2\right]=T\mathrm{E}\left[\left(\ln r\right)^2\right]+2T\mathrm{E}\left[w\ln r\right];\\
&\text{where } y_t:=\sum_{u=1}^{T}\left(\frac{P_{t-1}}{P_{t-u}}-1\right)\Big/T \text{ and } w_t:=\sum_{u=1}^{T}\ln\left(P_{t-1}/P_{t-u}\right)\Big/T.
\end{aligned}
\tag{6}
$$

Substituting these into the definitions of v^L, v^E, s and k gives

$$
\begin{aligned}
&\mathrm{E}\left[x^{(2L)}\left(R\right)\right]=T\mathrm{E}\left[x^{(2L)}\left(r\right)\right];\\
&\mathrm{E}\left[x^{(2E)}\left(R\right)\right]=T\left\{\mathrm{E}\left[x^{(2E)}\left(r\right)\right]+2\mathrm{E}\left[yr\ln r\right]\right\};\\
&\mathrm{E}\left[x^{(3)}\left(R\right)\right]=T\left\{\mathrm{E}\left[x^{(3)}\left(r\right)\right]+6\mathrm{E}\left[yr\ln r\right]\right\};\\
&\mathrm{E}\left[x^{(4)}\left(R\right)\right]=T\left\{\mathrm{E}\left[x^{(4)}\left(r\right)\right]+24\mathrm{E}\left[w\ln r\right]+24\mathrm{E}\left[yr\ln r\right]\right\}.
\end{aligned}
\tag{7}
$$

With the definition of var^L the first line gives the first part of Proposition 2.

Note that r and y are independent, and both are mean zero so

$$
6\mathrm{E}\left[yr\ln r\right]=3\operatorname{cov}\left(y,x^{(2E)}\left(r\right)\right).
\tag{8}
$$

This together with the definition of skewness and the third line of equation (7) gives the second part of the Proposition.

Finally, the fourth line of (7) gives

$$
\begin{aligned}
&\mathrm{E}\left[x^{(4)}\left(R\right)\right]=T\left\{\mathrm{E}\left[x^{(4)}\left(r\right)\right]+6\mathrm{E}\left[zx^{(2L)}\left(r\right)\right]+4\mathrm{E}\left[ys\left(r\right)\right]\right\}\\
&\text{where } z:=2\left(y-w\right).
\end{aligned}
\tag{9}
$$

Now

$$\begin{aligned} &\mathrm{E}\left[y\right]=0 \quad \text{and} \\ &\mathrm{E}\left[z\right]=\sum_{u=1}^{T}\left(u-1\right)\mathrm{var}^{L}\left[r\right]/T=\tfrac{1}{2}\left(T-1\right)\mathrm{var}^{L}\left[r\right]. \end{aligned} \tag{10}$$

Replace expectations of products by covariances, and products of expectations, substitute into the definition of kurtosis, and the final part of the proposition follows.