REALIZED SKEWNESS

Anthony Neuberger
Warwick Business School
Warwick University
Coventry
CV4 7AL
United Kingdom

anthony.neuberger@wbs.ac.uk

April 2011

There is much evidence that the third moment of returns is important for asset pricing. But skewness is hard to measure precisely. In the case of the second moment, the measurement problem is addressed through the use of realized volatility. This paper proposes an analagous definition of realized skewness that is computed from high frequency returns, and provides an unbiased estimate of the skewness of low frequency returns. A skewness swap is characterized; just like a variance swap, it can be priced and replicated in a model free fashion. As a by-product, a revised definition of realized variance is proposed, one that enables variance swaps to be replicated exactly, even in the presence of jumps and discrete sampling. Coskewness, the marginal impact of holding an asset for a short period on the skewness of the portfolio return over the investor's horizon, is defined in a coherent fashion.

The author is grateful to comments from Mark Britten-Jones, Peter Carr, Roman Kozhan, Eberhard Mayerhofer, Paul Schneider, Neil Shephard, and seminar participants at the UBC Summer Conference, the Frankfurt MathsFinance Conference, the Oxford-Man Institute, the University of Piraeus, Morgan Stanley, and the University of Warwick.

Standard approaches to asset pricing concentrate largely on the first and second moment of returns. But there is mounting evidence that the third moment may also be important. Barro (2009) argues that the inclusion of the possibility of rare disasters (extreme negative skewness) in an otherwise reasonably standard representative-agent economy, generates equity premia and risk free rates that accord with observation. There is a substantial empirical literature going back to Kraus and Litzenberger (1976), and including more recently Harvey and Siddique (2000), Ang, Hodrick, Xing and Zhang (2006), Ang, Cheng and Xing (2006) and Xing, Zhang and Zhao (2010), that suggests that the asymmetry of the returns distribution[1] both for individual stocks and for the equity market as a whole is important for asset pricing and investment management. There is also a growing literature on skewness in bond and currency markets.

The tools at our disposal for measuring and analyzing the second moment of returns – model-free implied volatility, realized volatility - are very sophisticated. We know how to use high frequency data to compute realized volatility and so make good estimates of the variance of long period returns. We analyze option prices to compute implied variance, and use the relation between implied and realized variance to test for the existence of variance risk premia. By contrast, our tools for measuring the third moment are rather primitive. There is no established methodology for computing the skewness of long period returns from high frequency returns, and no recognized concept of realized skewness that can be traded against option implied skewness. The main aim of this paper is to fill the gap. In particular I propose definitions of realized and implied skewness that are analogous to realized and implied variance.

In the process I also refine the theory of realized variance in a way that improves on the standard definitions. Under the standard definitions, a variance swap, which pays the difference between implied and realized variance, can be replicated perfectly only in the limit of continuous trading and in the absence of jumps (Carr and Wu, 2009). Under the definitions I propose, the replication of the variance swap is exact even with jumps and discrete trading.

---

[1] It is noteworthy that the papers cited use different measures of the asymmetry of returns on stocks. I return to this in the discussion of coskewness below.

It is worth revisiting some well known properties of variance, realized and implied, to understand why realized variance has proved so useful. In derivative pricing and risk management, much use is made of stochastic models of the price process under which the instantaneous conditional volatility is a key variable. Instantaneous volatility is not directly observable. However, if returns are observed sufficiently frequently (and if market microstructure noise is sufficiently small), realized volatility is such a good estimate for actual volatility that it effectively make this latent variable observable, as Andersen, Bollerslev, Diebold and Labys (2003) show in the highly liquid foreign exchange markets.

In the area of portfolio choice too, similar arguments apply. The conditional variance of period returns (where periods in this context may be measured in months or years) plays a central role. The conditional variance is not observable. With substantial time variation in conditional variance, the unconditional distribution of period returns is a poor guide to the conditional variance. To get better estimates of the conditional variance for a particular period, high frequency returns are used to compute the realized variance over longer periods and to understand the dynamics of the conditional variance.

There is no reason to expect the third moment of returns to be any more stable than the second moment. If we are to understand the dynamics of skewness, we need to be able to exploit high frequency data to improve our estimates of the skewness of period returns. But the proper definition of realized skewness for this purpose is not obvious, and the analogy with variance may be misleading. Under the assumption that returns are serially uncorrelated, the annualized variance of high frequency and low frequency returns are similar; indeed the ratio of the two has long been used as a test for the efficiency of market prices (Lo and MacKinlay, 1988). By contrast, there is no necessary relationship between the third moments of high and low frequency returns. Under the Central Limit Theorem, if daily returns are iid then long period returns will be close to normal even if the daily return distribution is skewed. Conversely, in a Heston (1993) stochastic volatility model, instantaneous log returns are conditionally normal, while longer period returns are skewed because of the correlation between returns and volatility shocks.

By computing the skewness of daily returns we learn little if anything about the skewness of monthly returns. If we want to find a definition of realized skewness that does provide a good estimate of the skewness of period returns, we need to analyze the relationship between the variance of period returns and the realized variance of sub-period returns more deeply before trying to exploit the analogy with skewness. The critical property of realized variance is that the realized variance of sub-period returns is equal in expectation to the realized variance of period returns, at least for price processes that are martingales. Realized skewness needs to be defined in such a way that it shares this property – a property that I refer to as the Aggregation Property.

The Aggregation Property makes it possible to use the realized skewness from daily returns to estimate the conditional skewness of monthly returns with some precision. It also has another attractive feature. Just as there is a model free strategy to replicate a variance swap, a swap that pays the difference between option implied variance and realized variance, so the Aggregation Property ensures that there is a model free strategy to replicate a skew swap, one that pays the difference between option implied skew and realized skew. From the perspective of the practitioner, this allows the development of a useful risk management tool. From the academic perspective, it makes it possible to construct a pure bet on skewness, and hence explore the existence and behavior of risk premia associated with skewness, similar to the way that Carr and Wu (2009) use variance swaps to explore variance risk premia

The useful relationship between implied and realized skewness, as between implied and realized variance, comes at a price. We cannot be over-prescriptive in our definitions of either skewness or variance, but must allow the definitions to be dictated by the mathematics. We need to relax the definitions of variance and skewness to include statistics that converge to the second and third moments of the distribution. Just as the model-free implied variance is not actually the variance of period returns, so the model-free implied skewness derived in this paper only approximates the third central moment of returns.

Tampering with the standard definitions of the variance and skewness may be uncomfortable for the econometrician; the theory of statistical inference rests heavily on the standard definitions. But it should not be a worry for the asset pricing theorist. The variance of returns is important in asset pricing because of the convexity of preferences. The premium that a utility maximizing agent requires for holding a risky portfolio is proportional to the variance of returns on the portfolio multiplied by the coefficient of relative risk aversion. But the result is only approximate, and holds only for returns that are small. In the absence of a specific utility function, variance does not have priority over any other measure of dispersion that is locally proportional to squared deviations from the mean. Variance is the right measure of dispersion to use only in the case of quadratic utility. Similar remarks apply to skewness.

From the perspective of asset pricing then, stretching the definition of variance and skewness to extend to measures of dispersion and asymmetry that converge to the standard definitions for small returns is a small price to pay for having consistency in definitions between the whole period and the sub-period, and retaining expectation relationships for discontinuous and discretely sampled price processes.

If one wants to stay with the strict definition of skewness – the third central moment of log returns divided by their cubed standard deviation – one can of course do so. With a complete set of strikes, the risk neutral density of log returns can be recovered from option prices. As Bakshi, Kapadia and Madan (BKM, 2003) show, one can then recover the coefficient of skewness of the distribution. But there is no way of relating the implied skewness for a period to the skewness for the constituent sub-periods. Nor does the implied skewness then have any natural realized counterpart. If for example a trader judges that the BKM implied skewness coefficient for a specific maturity is too high, there is no trading strategy that will reliably (in the absence of strong assumptions about the underlying price process) make money for the trader if the belief is correct.

The relaxations to the standard definitions would not be necessary if we were content to work in the arithmetic world of price changes. The relaxations are needed when we want to work in the geometric world of returns or log returns. The difference between the two worlds is that under the pricing measure, prices are martingales and log prices are not. The world of price changes is altogether simpler so the first section of the paper develops a theory of realized skewness of price changes. The second section then develops and modifies the theory to apply to returns. The third section shows how the definition can be extended to forward skewness and to coskewness. The fourth section concludes.

I. Arithmetic Contracts

I.1 Notation and Terminology

S_t $\left(t \in [0,T]\right)$ is a positive adapted variable defined on a standard filtered probability space $\left(\Omega, \mathfrak{F}, \left(\mathfrak{F}_t\right)_{t\in[0,T]}, \mathbb{P}\right)$. $\mathbb{E}_t[.]$ denotes $\mathbb{E}\left[.\middle|\mathfrak{F}_t\right]$. I assume that the distribution of S_T is such that expectations of S_T and functions of S_T such as S_T^2, $\ln(S_T)$ and $S_T\ln(S_T)$ exist.

$\mathbf{T} = \left\{t_0 = 0 < t_1 < \ldots < t_N = T\right\}$ is a partition of $[0,T]$. $\left\|\mathbf{T}\right\| = \max_i \left\{t_i - t_{i-1}\right\}$ is the mesh of $\mathbf{T}$. If x is a process, I will use x_i as shorthand for x_{t_i} in order to avoid excessive subscripting. δx_i denotes $x_i - x_{i-1}$. Given a function $g(.)$, $\sum^{\mathbf{T}} g\left(\delta x\right)$ denotes $\sum_{i=1}^{N} g\left(\delta x_i\right)$.

I will sometimes refer to the period $[0,T]$ as a month, and the length of the sub-period as a day, but obviously nothing hangs on this. In the present section, references to variance and skewness, relate to the distribution of price changes and not of returns. To avoid irrelevant complications with interest rates and dividends, I work throughout with forward prices, so all trades, whenever entered into, are for settlement at time T.

I.2 Variance

To examine the variance and realized variance of price changes, start from the algebraic identity

$$\left(S_T - S_0\right)^2 = \sum^{\mathbf{T}} \delta S^2 + 2\sum_{i=1}^{N-1}\left(S_i - S_0\right)\delta S_{i+1}. \tag{1}$$

Now assume S is a martingale. Take conditional expectations at time 0.

$$\mathbb{E}_0\left[\left(S_T - S_0\right)^2\right] = \mathbb{E}_0\left[\sum^{\mathbf{T}} \delta S^2\right]. \tag{2}$$

Since S is a martingale, the left hand side is the variance of S_T conditional on information at time 0. It is natural to call $\sum^{\mathbf{T}} \delta S^2$ the *realized variance* over the period on the partition **T**. Equation (2) shows that the realized variance, computed over any arbitrary partition of the period, is an unbiased estimator of the conditional variance.

So far, nothing has been said about financial markets. S_t can be interpreted as the forward price of a risky asset at time t, where the forward contract matures at time T. I assume that the market is deep and frictionless so agents can buy or sell the asset at the one price in unlimited quantities. In particular, if H is any adapted process, then there is a trading strategy on the partition **T** that yields a gain at time T of $\sum_{i=1}^{N} H_{i-1} \delta x_i$.

In the absence of risk premia, forward prices are martingales and the realized variance of S is an unbiased estimate of its conditional variance. In the presence of risk premia, S is not a martingale and the realized variance is a biased estimate of the conditional variance. The change in realized variance with the mesh of the partition gives insight into the statistical properties of mean price changes (as in Lo and MacKinlay, 1988).

But there is another financial markets interpretation of equation (1). It shows how to price *variance swap contracts*. In a variance swap, the party who receives floating is paid the realized variance over the partition **T** in return for a fixed payment K. Equation (1) shows that to avoid arbitrage, K must equal the price of a contingent claim (a "square contract") that pays $\left(S_T - S_0\right)^2$. If K were higher than this, the party who receives floating would find it cheaper to buy the square contract and then hedge dynamically by shorting $2\left(S_i - S_0\right)$ futures contracts, rebalancing each day. Indeed, the agent could make arbitrage profits by writing variance swaps at K and synthesizing them through the purchase of square contracts. Similarly, if the fixed payment were lower than its fair price, arbitrage profits would be assured by entering into a swap in the other direction, paying fixed and receiving floating, and selling square contracts.

It is natural to call the time 0 price of the square contract the *implied variance*. So the fair price of a variance swap defined on any partition **T** is the implied variance. In a world where square contracts themselves do not trade, but there are conventional puts and calls maturing at T available for all strikes, the square contract itself can be synthesized from puts and calls (Breeden and Litzenberger 1978), and the relation between implied and realized variance can be translated into a statement about the relationship between the average level of option prices and realized variance.

In particular, denote the time 0 price of the call and put with strike k by $C(k)$ and $P(k)$ respectively. The instruments themselves are denoted by the corresponding bold letters (**C(k)** and **P(k)**). To replicate the square portfolio, construct the portfolio

$$2\left\{\int_0^{S_0}\mathbf{P}(k)\,dk+\int_{S_0}^{\infty}\mathbf{C}(k)\,dk\right\}\text{ which has pay-off}$$

$$2\left\{\int_0^{S_0}\left[k-S_T\right]^+dk+\int_{S_0}^{\infty}\left[S_T-k\right]^+dk\right\}=\left(S_T-S_0\right)^2 \tag{3}$$

$$\text{where }\left[x\right]^+\equiv\max\left\{x,0\right\}.$$

The square contract can be replicated by a portfolio of out-of-the-money puts and calls.

For expositional purposes, it is easier to work with the probabilistic equation (2) rather than the algebraic equation (1). While forward prices are not generally martingales under the physical measure, all prices are martingales under the equivalent risk-adjusted or pricing measure whose existence is guaranteed by the absence of arbitrage (Harrison and Kreps, 1979). So much of the rest of paper is developed in a probabilistic setting with the probability measure $\mathbb{P}$ assumed to be a pricing measure.

The relationship between realized and implied variance stems from equation (2). To bring out its structure more clearly, write it as

For any martingale S

$$\mathbb{E}_0\left[g\left(\sum^{\mathbf{T}}\delta S\right)\right]=\mathbb{E}_0\left[\sum^{\mathbf{T}}g\left(\delta S\right)\right] \text{ where } g\left(x\right)\equiv x^2. \tag{4}$$

The g-function and the summation operator commute for any martingale. $g\left(\sum^{\mathbf{T}}\delta S\right)=g\left(S_T-S_0\right)$ is a function of S_T and can be interpreted as a contingent claim on the terminal price of the asset. With $\mathbb{P}$ a pricing measure, $\mathbb{E}_0\left[g\left(\sum^{\mathbf{T}}\delta S\right)\right]$ is the time 0 price of the contingent claim. It is the *implied* variance. $\sum^{\mathbf{T}}g\left(\delta S\right)$ is the *realized* variance. The equation says that the implied variance is the fair price for the fixed leg of a swap that pays the realized variance. This result holds exactly and is independent of the partition **T** chosen.

I.3 General Arithmetic Contracts

Now consider more general contracts that have similar properties. If g is a real-valued function and X is an adapted (scalar or vector) process, then (g, X) is said to have the *Aggregation Property* if for any times $0\le s\le t\le u\le T$

$$\mathbb{E}_s\left[g\left(X_u-X_s\right)-g\left(X_u-X_t\right)-g\left(X_t-X_s\right)\right]=0. \tag{5}$$

Applying the Law of Iterated Expectations, if (g, X) has the Aggregation Property then $\mathbb{E}_0\left[g\left(X_T-X_0\right)\right]=\mathbb{E}_0\left[\sum^{\mathbf{T}}g\left(\delta X\right)\right]$ for any partition **T**. $\sum^{\mathbf{T}}g\left(\delta X\right)$ is the *realized* characteristic; and $\mathbb{E}_0\left[g\left(X_T-X_0\right)\right]$ is the *implied* characteristic.

The obvious question is whether there are any other interesting functions of S (apart from variance) where (g, S) has the aggregation property for any martingale S. The answer (as demonstrated formally below in Proposition 1) is no.

However, we do not need to restrict ourselves to the case where $X = S$. Suppose that X is a vector-valued process $[S, V(S)]$ where $V_t(S) = Var_t[S_T]$, the conditional variance of S_T. Define **G** to be the set of analytic[2] functions g such that $(g, [S, V(S)])$ has the Aggregation Property for all martingales S.

Proposition 1: G *consists of the functions*

$$g(S,V) = h_1 S + h_2 V + h_3 S^2 + h_4 \left(S^3 + 3SV\right). \tag{6}$$

where the $\{h_i\}$ *are arbitrary constants.*

Proof: see Appendix.

Proposition 1 shows that **G** is a linear space spanned by four functions of S and V: $g = S$ and $g = V$, which are uninteresting, $g = S^2$, which is the variance and is familiar, and $g = S^3 + 3SV$, whose properties we now explore.

The implied characteristic of this function is

$$\begin{aligned} \mathbb{E}_0\left[g\left(S_T - S_0, V_T - V_0\right)\right] &= \mathbb{E}_0\left[g\left(S_T - S_0, -V_0\right)\right] \text{ (since } V_T = 0) \\ &= \mathbb{E}_0\left[\left(S_T - S_0\right)^3 - 3V_0\left(S_T - S_0\right)\right] \text{ (definition of } g) \\ &= \mathbb{E}_0\left[\left(S_T - S_0\right)^3\right] \text{ (since } S \text{ is a martingale).} \end{aligned} \tag{7}$$

The characteristic captured by g is skewness. The left hand side of equation (7) is the implied skew. The realized skew is $\sum^{\mathbf{T}} g(\delta S, \delta V) = \sum^{\mathbf{T}} \delta S^3 + 3\delta S \delta V$. The Aggregation Property means that the realized skew equals the implied skew in expectation:

$$\mathbb{E}_0\left[\sum^{\mathbf{T}} \delta S^3 + 3\delta S \delta V\right] = \mathbb{E}_0\left[\left(S_T - S_0\right)^3\right]. \tag{8}$$

---

[2] The restriction to analytic functions is for technical convenience, and could well be weakened.

Equation (8) is significant in several respects. It shows that skewness in low frequency returns derives only in part from the skewness in high frequency returns. The second source (and indeed the only source when S is a continuously sampled continuous martingale) is the covariation between shocks to the price level and shocks to future variance.

It also shows how high frequency data can be used to get more efficient estimates of the skewness in price changes over a period than can be obtained from just the price change over the period. This improvement rests on two assumptions: the discounted price process is a martingale, and the variance of the terminal price is in the observer's information set.

Third, differences between the two sides of the equation when expectations are computed under the physical measure can be used to detect and analyze risk premia associated with skewness.

Fourth, interpreting the equation under a pricing measure, it suggests how a skew swap can be designed and replicated. A skew swap pays the difference between the implied and the realized skew. To replicate a skew swap, the variance has to be tradable as well as observable. This can be achieved through having a traded square contract, with payoff S_T^2. With $\mathbb{P}$ being a pricing measure, the price of the contract on day i is

$$P_i = \mathbb{E}_i\left[S_T^2\right] = S_i^2 + V_i. \tag{9}$$

The gain from holding one square contract for one period is

$$\delta P_{i+1} = \left(\delta S_{i+1}\right)^2 + 2S_i \delta S_{i+1} + \delta V_{i+1}. \tag{10}$$

Suppose that an agent enters into a one month skew swap at day 0, paying floating and receiving fixed, with realized skew being computed from daily prices. She uses the fixed payment to buy a contract that pays the cube of the price change over the month. She hedges by holding -3(S_i – S_0) square contracts and $3\left(\left(S_i - S_0\right)^2 - V_i\right)$ forward contracts over day i. If her initial wealth W_0 = 0, her terminal wealth is

$$
\begin{aligned}
W_T &= \left\{ \left(S_T - S_0\right)^3 \right\} - \left\{ \sum^{\mathbf{T}} \left(\delta S\right)^3 + 3\delta S \delta V \right\} \\
&-3\left\{ \sum_{i=0}^{N-1} \left(S_i - S_0\right)\left(\left(\delta S_{i+1}\right)^2 + 2S_i \delta S_{i+1} + \delta V_{i+1} \right) \right\} + 3\left\{ \sum_{i=0}^{N-1} \left(\left(S_i - S_0\right)^2 - V_i \right) \delta S_{i+1} \right\} = 0.
\end{aligned}
\tag{11}
$$

(11) shows that the agent can hedge the skew swap exactly for any finite partition **T**. So in particular it works perfectly with discrete monitoring of skewness and with jumps in the price of the underlying or in its variance. The only requirements are that the market is frictionless, and that the asset and the square contract on it can be traded at any time *t* that is used for computing the realized skew.

II. Geometric Contracts

II.1 Generalized Variance

Financial economists are interested in the behavior of returns, not price changes. It is tempting to apply the theory in the previous section directly to the log price, $s_t \equiv \ln S_t$. But the log price is not a martingale under either any pricing measure, or in general under the physical measure. In looking for a definition of implied and realized variance of returns, I start from the premise that it is important to keep the Aggregation Property. The price for this is a relaxation of the definition of variance.

Let *f* be an analytic function on the real line with the property that $\lim_{x \to 0} f\left(x\right) / x^2 = 1$. Given a process *s*, define the process v_t^f by $v_t^f\left(s\right) = E_t\left[f\left(s_T - s_t\right) \right]$. I will call $v_t^f\left(s\right)$ a *generalized variance* process for *s*[3].

The variance measures that are widely used by academics and practitioners (squared net returns and squared log returns) conform to the definition of generalized variance. Two other generalized variance measures, v^L and v^E, turn out to be important

[3] To simplify the algebra, the variance is not annualized; it tends to increase in magnitude with time to maturity.

$$
\begin{aligned}
v_t^L &= \mathbb{E}_t\left[L\left(s_T - s_t\right)\right] \text{ where } L(x) = 2\left(e^x - 1 - x\right), \\
v_t^E &= \mathbb{E}_t\left[E\left(s_T - s_t\right)\right] \text{ where } E(x) = 2\left(xe^x - e^x + 1\right).
\end{aligned} \tag{12}
$$

The two functions are plotted in Figure 1, and it can be seen that they actually lie between the two conventional measures of variance. It is easy to verify that both satisfy the definition of generalized variance.

To explain the use of the letters *L* and *E*, rewrite (12) as

$$
\begin{aligned}
v_t^L &= 2\mathbb{E}_t\left[\frac{S_T}{S_t} - 1 - \ln\left(\frac{S_T}{S_t}\right)\right] \text{ so } \mathbb{E}_t\left[\ln S_T\right] = \ln S_t - v_t^L/2; \\
v_t^E &= 2\mathbb{E}_t\left[\frac{S_T}{S_t}\ln\left(\frac{S_T}{S_t}\right) - \frac{S_T}{S_t} + 1\right] \text{ so } \mathbb{E}_t\left[S_T \ln S_T\right] = S_t \ln S_t + S_t v_t^E/2.
\end{aligned} \tag{13}
$$

In a Black-Scholes world, where the underlying asset has constant volatility σ, the price of a log contract, one that pays $\ln S_T$, is $\ln S_t - \sigma^2\left(T-t\right)/2$, so v_t^L is the implied Black-Scholes variance of the log contract. It is the same as the Model Free Implied Variance of Britten-Jones and Neuberger (2000). Similarly, v_t^E is the implied Black-Scholes variance of the contract that pays $S_T \ln S_T$. I call it entropy because of the functional similarity to entropy as used in thermodynamics and information theory.

The following properties of the log and entropy variance will be useful. From the first line of (13)

$$
\begin{aligned}
&v_t^L = 2s_t - 2\mathbb{E}_t\left[s_T\right] \text{ so} \\
&\mathbb{E}_t\left[\delta v_{t+1}^L - 2\delta s_{t+1}\right] = 0.
\end{aligned} \tag{14}
$$

Similarly, from the second line

$$
\begin{aligned}
&e^{s_t}\left(v_t^E + 2s_t\right) = 2\mathbb{E}_t\left[s_T e^{s_T}\right] \\
&\text{so } e^{s_t}\mathbb{E}_t\left[e^{\delta s_{t+1}}\left(v_t^E + \delta v_{t+1}^E + 2s_t + 2\delta s_{t+1}\right) - v_t^E - 2s_t\right] = 0 \\
&\text{which implies that } \mathbb{E}_t\left[e^{\delta s_{t+1}}\left(\delta v_{t+1}^E + 2\delta s_{t+1}\right)\right] = 0.
\end{aligned} \tag{15}
$$

II.2 Aggregation with Returns

Let **G*** be the set of analytic functions g where $\left(g,\left[s,v\left(s\right)\right]\right)$ has the Aggregation Property for any positive martingale S, where $s=\ln S$ and v is a generalized variance process for s.

Proposition 2: *the set* **G*** *comprises the following functions*

$$
\begin{aligned}
&g\left(s,v\right)=h_1 s+h_2\left(e^s-1\right)+h_3 v+h_4\left(v-2s\right)^2+h_5\left(v+2s\right)e^s \\
&\text{where } \left\{h_i\right\} \text{ are arbitrary constants, with the following constraints:} \\
&\quad -\ \text{if } h_4\neq 0, v=v^L \text{ and } h_5=0; \\
&\quad -\ \text{if } h_5\neq 0, v=v^E \text{ and } h_4=0; \\
&\quad -\ \text{if } h_4=h_5=0, v \text{ is any generalized variance measure.}
\end{aligned}
\qquad (16)
$$

Proof: see Appendix.

I now examine the properties of three particular members of **G***.

Proposition 3: *the function* $g^M\left(s\right)\equiv e^s-1$ *is a measure of expected return and has the Aggregation Property.*

Proof: The Aggregation Property follows immediately from Proposition 2 with $h_2 = 1$, and $h_1 = h_3 = h_4 = h_5 = 0$. The implied characteristic is $\mathbb{E}_0\left[S_T/S_0-1\right]$ and is the mean return. ■

With g^M the *implied return* is zero. The *realized return* is the sum of daily net returns over the month. The swap contract corresponding to g^M takes the form of a standard equity for floating swap – the receiver receives the total return on the underlying (net of the riskless interest rate) each "day" on a fixed nominal amount. The payer can hedge perfectly by going long $1/S_i$ forward contracts each day.

II.3 Variance of Returns

Proposition 2 can also be used to capture the idea of variance.

Proposition 4: *the function* $g^V(s) \equiv 2\left(e^s - 1 - s\right)$ *is a measure of variance and has the Aggregation Property.*

Proof: The Aggregation Property follows from Proposition 2 with h_1 = -2, h_2 = +2, h_3 = h_4 = h_5 = 0. The implied characteristic is $\mathbb{E}_0\left[L\left(\ln\left(S_T/S_0\right)\right)\right]$ and is a generalized variance measure. ■

The *implied variance* is v_0^L, which is the model free implied variance. The *realized variance* is $\sum^{\mathbf{T}} g^V(\delta s) = \sum^{\mathbf{T}} 2\left(e^{\delta s} - 1 - \delta s\right)$. This differs from the conventional definition of realized variance which is $\sum^{\mathbf{T}} \delta s^2$.

The conventional definition has the merit that it corresponds to the definition used in practice in the variance swap market. But, as noted by Jiang and Tian (2005; see particularly footnote 7), the replication of a standard variance swap is imperfect. The replication is only perfect in the limit case when the mesh of the partition goes to zero, and with the added assumption that the price process is continuous. By contrast, the fact that the new measure of realized variance has the Aggregation Property means that replication is perfect whatever the price path and whatever partition is chosen.

In practice, with reasonably frequent rebalancing, the two measures of realized variance are very similar. This is not surprising since they are both generalized measures of variance. Suppose one computes the monthly realized volatility of the S&P500 using daily returns. Using either measure, the mean volatility over the last ten years (2001-2010) is just over 18% (annualized). The root mean square difference between the two measures 0.06%. Since the 1950s the biggest difference between the two measures was in the month of October 1987 when the conventional measure of realized volatility 101.2%, while the alternative measure registered 98.8%.

From a risk management or portfolio allocation perspective, there does not seem a lot to choose between them. In risk management, variance is used as a handy measure of dispersion of outcomes; more tailored measures (such as Value at Risk) need to be developed for specific purposes. In the latter case, the precise measure of risk depends on the utility function. Variance is the right measure only for quadratic utility; both are serviceable approximations of risk for more general utility functions.

But it is worth exploring whether there are other reasons for preferring one measure of realized variance to the other. For the financial econometrician, the main justification for the conventional measure of realized variance is that it converges to the quadratic variation as the mesh size becomes small, and the quadratic variation is an unbiased estimate of the conditional variance of the log price process under certain conditions[4] (see Andersen, Bollerslev, Diebold and Labys, 2003, Theorem 1 and Corollary 1). The following Proposition[5] states that the new measure of realized variance also has this property, when the price process is continuous.

Proposition 5: *Let f be an analytic function on the real line such that* $\lim_{x\to 0} f(x)/x^2 = 1$. *For any continuous semimartingale s, the associated realized variance* $\sum^{\mathbf{T}} f(\delta s)$ *converges in probability to the quadratic variation of s over the period [0, T] as the mesh of the partition* **T** *goes to zero. In particular, the logarithmic and entropic variance of s equals its quadratic variation.*

Proof: see Appendix.

Thus both definitions of realized variance converge to the quadratic variation as the mesh becomes fine; the new definition has the added advantage that, with finite mesh, the realized variance is an unbiased estimate of the quadratic variation under the assumption that the price is a martingale.

---

[4] The key condition is that the predictable component of returns is predetermined. If volatility is stochastic then this condition is certainly violated under any pricing measure, and cannot be expected to hold under the physical measure, and the quadratic variation is not an unbiased estimator of the conditional variance of log returns. An implication of Propositions 4 and 5 together is that, if the price process is continuous, the quadratic variation of returns is an unbiased estimate of the logarithmic variance $\mathbb{E}_0\left[L\left(s_T - s_0\right)\right]$, under any pricing measure.

[5] I am much indebted to Eberhard Mayerhofer for the proof of this Proposition.

II.2 Skewness of Returns

Proposition 2 also shows the way to construct a definition of realized skewness of returns that closely resembles the definition already established for price changes.

Proposition 6: $g^Q\left(s, v^E\left(s\right)\right) \equiv 3v^E\left(e^s - 1\right) + K\left(s\right)$ *, where* $K\left(s\right) \equiv 6\left(se^s - 2e^s + s + 2\right)$*, is a measure of skewness and has the Aggregation Property.*

Proof: g^Q has the Aggregation Property by Proposition 2 with $h_1 = 6$, $h_2 = -12$, $h_3 = -3$, $h_4 = 0$, and $h_5 = 3$. The implied characteristic is $\mathbb{E}_0\left[K\left(\ln\left(S_T/S_0\right)\right)\right]$. $K\left(s\right) = s^3 + O\left(s^4\right)$ so this can readily be described as a measure of skewness. ■

$K(s)$ is plotted in Figure 2 and lies between two traditional measures of skewness – cubed returns and cubed log returns.

Using this measure, the *implied skewness* is

$$\begin{aligned} q_0 &= \mathbb{E}_0\left[K\left(s_T - s_0\right)\right] \\ &= 6\mathbb{E}_0\left[\frac{S_T}{S_0}\ln\frac{S_T}{S_0} + \ln\frac{S_T}{S_0} - 2\left(\frac{S_T}{S_0} - 1\right)\right] = 3\left(v_0^E - v_0^L\right). \end{aligned} \tag{17}$$

In a Black-Scholes world, all options trade on the same implied volatility, the two implied variances are equal and the implied skewness is zero. From (17), implied skewness is the price of a claim **y** that pays $y\left(S_T\right) = 6\left[\frac{S_T}{S_0}\ln\frac{S_T}{S_0} + \ln\frac{S_T}{S_0} - 2\left(\frac{S_T}{S_0} - 1\right)\right]$.

Using Bakshi and Madan (2000), the replicating portfolio of vanilla options is

$$\begin{aligned} &\int_0^{S_0} y''\left(K\right)\mathbf{P}(K)dK + \int_{S_0}^{\infty} y''\left(K\right)\mathbf{C}(K)dK + y'\left(S_0\right)\mathbf{S} - \left(y\left(S_0\right) + S_0 y'\left(S_0\right)\right)\mathbf{1} \\ &= 6\left\{\int_{S_0}^{\infty}\frac{K - S_0}{S_0 K^2}\mathbf{C}(K)dK - \int_0^{S_0}\frac{S_0 - K}{S_0 K^2}\mathbf{P}(K)dk\right\}. \end{aligned} \tag{18}$$

Implied skewness is thus the price of a portfolio that is long out of the money calls and short out of the money puts. The term skew is widely used to describe the slope of

Black-Scholes implied volatility when plotted against strike, or against the implied option delta. Equation (18) defines implied skewness precisely, and in a way that links it tightly to the skewness of the risk neutral density function.

Implied skewness is a natural measure of the asymmetry of option prices with respect to strike. If call and put prices exhibit put-call symmetry in the sense of Carr, Ellis and Gupta (1998) (so the implied Black-Scholes volatilities of two options with strike S_0/α and $S_0\alpha$ are the same for any positive α), it can readily be shown that the implied skewness is zero.

A skew swap, where the floating leg is realized skewness and the fixed leg is implied skewness, can be replicated perfectly by dynamic hedging, trading the entropy contract and the forward contract. To replicate the skew swap, it is necessary that an entropy contract with maturity *T* is traded (or equivalently, that calls or puts with maturity *T* and all possible strikes are traded, so that the entropy contract can be replicated). An agent who writes a skew swap, receiving fixed and paying floating, receives net

$$3\left(v_0^E - v_0^L\right) - \sum^{\mathbf{T}} \left\{ 3\delta v^E \left(e^{\delta s} - 1\right) + K\left(\delta s\right) \right\}. \qquad (19)$$

To hedge her position she needs to buy six log contracts, with terminal pay-off $6\ln S_T$. This costs $6\ln S_0 - 3v_0^L$ (by (13)). She also needs to hedge dynamically by holding a long position of $6/S_i$ entropy contracts on day *i* (contracts that have a pay-off of $S_T \ln S_T$) and also a short position in $3\left(2s_i + v_i^E + 4\right)/S_i$ forward contracts. The replication of the skew swap is exact.

III. Other Properties of Skewness

With a definition of skewness that has a firm theoretical basis, it is now possible to articulate and define important derived concepts such as forward skewness and coskewness.

III.1 Forward Skewness

The slope of the implied volatility surface with strike (the "smirk") is a well-attested feature of equity index markets. But it is not immediately clear how to interpret the way the smirk varies with option maturity. With our definition of skewness, there is a natural way to measure and interpret the forward implied slope.

I use a second subscript to denote maturity, so that $v^L_{t,T}$ denotes the implied log variance at time t based on options that mature at T, and I use Δ to mean an increment in maturity T (continuing to reserve the δ operator for an increment in time t). The one period forward implied log variance at time t for maturity T is

$$v^{L,F}_{t,T} = \Delta v^L_{t,T} = v^L_{t,T} - v^L_{t,T-1}. \tag{20}$$

The forward variance is the difference in price between two variance swaps: one that matures at T-1 and the other at T. The effect of going long one swap and short the other is to create a forward variance swap, where the floating leg is the realized variance $g^V\left(\delta s_T\right)$ over [T-1 , T] and the fixed leg is the time t forward variance. In the absence of term premia, the forward variance is the expectation of the future variance, and the forward variance follows a martingale. If the price process is also stationary, then the unconditional expectation of the forward variance is independent of maturity, and the unconditional expectation of the spot variance is linear in maturity. Conversely, the existence of a slope in the average term structure of forward variances can be taken as evidence of the existence of term premia in the variance market.

Turning now to forward skewness, a forward skewness swap is the difference between a T- and a T-1-maturing skewness swap. Using $q^F_{t,T}$ to denote the forward skew $\Delta q_{t,T}$ the pay-off to a forward skew swap is

$$\begin{aligned}
&\sum_{u=t+1}^{T} g^Q\left(\delta s_u, \delta v^E_{u,T}\right) - \sum_{u=t+1}^{T-1} g^Q\left(\delta s_u, \delta v^E_{u,T-1}\right) - q^F_{t,T} \\
&= \left\{3\sum_{u=t+1}^{T-1} \delta v^{E,F}_{u,T}\left(e^{\delta s_u} - 1\right) + K\left(\delta s_T\right)\right\} - q^F_{t,T}.
\end{aligned} \tag{21}$$

With time measured in days, and T equal to one month, the floating leg of a forward skew swap for the last day of the month has two components. The first is the covariation between the forward variance for the last day and current returns (multiplied by 3); the second - $K\left(\delta s_T\right)$ - is the realized skew of returns on the last day.

The average term structure of the skew contains interesting information about the extent to which any leverage effect (negative correlation between returns and volatility) is persistent. If the leverage effect is short term, so that on average there is no correlation between returns today and shocks to forward variance N days ahead then, assuming the price process is stationary, the unconditional expectation of the floating leg of the forward skew swap (the curly bracketed term in equation (21)) is independent of maturity for forwards with more than N days to maturity. In the absence of term premia, the average forward skew term structure would be flat beyond N days.

III.2 Coskewness

As noted in the introduction, there is substantial evidence of the importance of the asymmetry in returns for asset pricing. The classic reference is Kraus and Litzenberger (1976) who set the problem in a one period setting with an agent who has a preference for positive skewness (and an aversion to variance). They show that the contribution of an asset to the skewness of the agent's wealth depends not on its own skewness but on its coskewness (the covariance of returns on the asset with squared returns on the market).

Single period models have proved a powerful aid to intuition, but there is reason to suppose that they may be seriously misleading when it comes to issues of asymmetry. The skewness of returns on any strategy are likely to be dependent on the time horizon; as we have seen, the link between skewness at short horizons and at long horizons is much weaker than the link in the case of variance. So the choice of horizon matters. But it is hard to explain why long-lived agents should care very much about the skewness of short horizon returns. This suggests that in single period models with skewness, the length of the period should be thought of as measured in years. But this presents two challenges: first, constraining agents to buy-and-hold strategies over many years is highly restrictive. Second, from an empirical perspective, it is hard to test predictions about long horizon returns because of data limitations.

The definition of skewness in this paper presents a simple way of addressing the difference in horizons for trading and consumption. Suppose agents care about both the realized variance and the realized skewness of returns on their total wealth (using our definitions) over the horizon *T*. Then one can reasonably conjecture that the premium they require for holding a particular asset over a particular period depends on its marginal contribution to the portfolio's terminal variance and skewness. In particular, assume that the agent has optimized the portfolio and the total return on the *t*'th day is R_t. Then \$1 at time *t*-1 becomes $\prod_{u=t}^{T} R_u$ at time *T*. Now consider the impact on the skew and variance of terminal wealth if the agent deviates from their strategy by borrowing *k* for one period and investing in a share that has a net excess return of X_t. The agent's terminal wealth will be $\left(R_t + kX_t\right)\prod_{u=t}^{T} R_u$. The expected realized variance and skewness of terminal wealth depend on the amount borrowed and can be written as $v_{t-1}^L\left(k\right)$ and $q_{t-1}\left(k\right)$ respectively.

Define the *covariance* of X_t with the optimal strategy as the marginal increase in expected realized variance of terminal wealth per unit of investment in X_t, divided by 2; and define the *coskewness* of X_t with the optimal strategy as the marginal increase in the expected realized skewness of terminal wealth per unit of investment in X_t, divided[6] by 3.

In other words

$$\text{Covariance} = \frac{1}{2}\frac{\partial v_{t-1}^L\left(k\right)}{\partial k}\Big|_{k=0}; \quad \text{coskewness} = \frac{1}{3}\frac{\partial q_{t-1}\left(k\right)}{\partial k}\Big|_{k=0} \tag{22}$$

Proposition 7: *The covariance and coskewness of X with the optimal strategy are*

[6] The division by 2 and by 3 is a normalization that ensures that in a one period world with standard definitions of variance and skewness, the definitions of covariance and coskewness coincide with the standard definitions.

$$
\begin{aligned}
&\text{Covariance} = -\mathbb{E}_{t-1}\left[\frac{X_t}{R_t}\right] = \mathbb{E}_{t-1}\left[\frac{X_t\left(R_t-1\right)}{R_t}\right];\\
&\text{Coskewness} = \mathbb{E}_{t-1}\left[2X_t\left(\ln R_t - \frac{R_t-1}{R_t}\right) + \mathbb{E}_t\delta v_{t,T}^E\right].
\end{aligned}
\tag{23}
$$

Proof: in Appendix.

The covariance shows that the impact on the variance of terminal wealth of investing in a share for one day depends only on the relation between its excess return and the contemporaneous excess return on the optimal portfolio. The consumption horizon is irrelevant. When the trading period is short and R is close to one, the covariance is close to the standard definition.

The coskewness definition shows that holding a share for a day affects the terminal realized skewness through two routes. The first depends on the relation between the excess return on the share and the contemporaneous return on the optimal portfolio. For short periods, when R is close to 1, the first term is approximately equal to the standard definition of coskewness – that is the product of the excess return on the share and the squared excess return on the portfolio. The second term depends on the relation between the return on the share and changes to expectations of the future variance of portfolio returns to the consumption horizon.

The results are interesting from the perspective of testing asset pricing models. Suppose the model being tested is an equilibrium model which admits of a representative agent who holds the market portfolio. Suppose also that agent cares about the realized variance and skewness of the distribution of terminal wealth in the sense defined above[7]. Then assets would be priced according to their covariance and coskewness with the market portfolio.

[7] If prices (in this case, the price of the risky asset and the portfolio, and the entropy contract on the portfolio) follow martingales, then true and expected realized variance and skewness are the same, and arguably investors will care more about the true quantity than the expected realized quantity. But the martingale assumption is unpalatable in an asset pricing context.

Proposition 7 provides a definition of covariance, and hence of an asset's beta with the market which is slightly different from the standard $\frac{\mathbb{E}_t\left[r_X r_M\right]}{\mathbb{E}_t\left[r_M^2\right]}$ (where little r denotes the net return, R-1, and the subscript M refers to the market portfolio) but rather should be $\frac{\mathbb{E}_t\left[r_X r_M/\left(1+r_M\right)\right]}{\mathbb{E}_t\left[r_M^2/\left(1+r_M\right)\right]}$. Whether the difference in definition would have much practical effect is an empirical question.

But the definition of coskewness is likely to be of greater practical significance. Previous practice in measuring the asymmetry of returns has varied widely. Kraus and Litzenberger (1976) use coskewness of monthly returns with the market, Harvey and Siddique (2000) use conditional coskewness of monthly returns with the market, Ang, Hodrick, Xing and Zhang (2006) use the beta of daily returns on the one month VIX index[8], Ang, Cheng and Xing (2006) use downside beta[9] based on daily return data, and Xing, Zhang and Zhao (2010) use the slope of the implied volatility of options on the stock. While the particular definition of coskewness advocated here cannot claim to be any better grounded in an acceptable asset pricing model than any of the others, it does have the particular advantage that it seeks to measure the impact of a security on the skewness of long horizon portfolio returns, rather than looking at skewness at horizons (monthly or daily) that are unlikely to be of real economic significance.

Proposition 7 shows that coskewness at long horizons is determined largely by the covariation of the asset's return with the implied variance of the market. This raises the question of how to measure coskewness. Fortunately, with the growing volume of trade in long-dated index options and in long-dated index variance swaps, the quality of data to estimate the time series for implied variance is improving rapidly.

[8] The VIX index is the implied volatility of a synthetic 1 month index option so VIX_t roughly corresponds to $\sqrt{v_{t,t+1mth}^L}$ in the notation used here.

[9] The downside beta is the covariance of returns on the asset with the market, conditional on the market return being negative, divided by the covariance of the market conditional on the same event

IV. Conclusions

In this paper, I have put forward a set of tools to improve the measurement and modeling of the skewness of asset returns. The use of high frequency data to compute realized skew facilitates the measurement of skewness and makes it easier to trace its variation over time. Exposure of the fundamental link between skewness and the covariation between returns and implied variance may help our understanding of the behavior of skewness at different horizons. With implied and realized skewness defined in such a way that a skewness swap can be replicated perfectly it is possible to explore risk premia associated with skewness without danger of contamination from the effects of model mis-specification.

The results may also be of use to practitioners. While skew swaps have been traded on a small scale, and the link between skewness on the one hand and the correlation between returns and volatility shocks on the other have been understood intuitively, the formality of the concepts presented in this paper may enable traders to speculate and hedge with much greater precision. If the definition of realized variance were modified as suggested in this paper, the risks borne by financial intermediaries writing variance swaps would be reduced, and this might be expected to improve the pricing and liquidity of these important risk management instruments.

APPENDIX

Proof of Proposition 1

It is straightforward to prove that all members of **G** have the Aggregation Property; all that is needed is to substitute (6) into (5). Proving the converse, that all analytic functions that have the Aggregation Property are in **G**, is more complicated.

Let $\tilde{\eta}$ be a random variable with $\mathbb{E}[\tilde{\eta}]=0$ and $\mathbb{E}\left[\tilde{\eta}^2\right]=\alpha$, and g an analytic function that has the Aggregation Property. We consider two processes with $t\in\{0,1,2\}$. The first is given by $S_0 = S_1 = 0$, $S_2 = \tilde{\eta}$. S is clearly martingale. The process (S, V) is (0, α), (0, α), ($\tilde{\eta}$, 0). For the Aggregation Property to hold

$$\mathbb{E}\left[g(\tilde{\eta},-\alpha)\right]=\mathbb{E}\left[g(0,0)+g(\tilde{\eta},-\alpha)\right]. \tag{A-1}$$

It follows that $g(0, 0) = 0$.

The second process for S is shown below

$$0\rightarrow\begin{cases} u & \rightarrow \quad u+\tilde{\eta} \qquad \Pr=\pi \\ d & \rightarrow \quad d \qquad\quad \Pr=1-\pi \end{cases}. \tag{A-2}$$

The process S is martingale provided that ud < 0 and $\pi u+(1-\pi)d=0$. The process (S, V) is

$$(0,V_0)\rightarrow\begin{cases} (u,\alpha) & \rightarrow \quad (u+\tilde{\eta},0) \qquad \Pr=\pi \\ (d,0) & \rightarrow \quad (d,0) \qquad \Pr=1-\pi \end{cases}$$
$$\text{where } V_0=\pi\left(u^2+\alpha^2\right)+(1-\pi)d^2. \tag{A-3}$$

For the Aggregation Property to hold

$$\begin{gathered}\mathbb{E}\left[\pi g(u+\tilde{\eta},-V_0)+(1-\pi)g(d,-V_0)\right]= \\ \mathbb{E}\left[\pi g(u,\alpha-V_0)+\pi g(\tilde{\eta},-\alpha)+(1-\pi)g(d,-V_0)+(1-\pi)g(0,0)\right].\end{gathered} \tag{A-4}$$

Simplifying, and using the fact that $g(0, 0) = 0$, gives

$$\mathbb{E}\left[g\left(u+\tilde{\eta},-V_0\right)\right]=\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]+g\left(u,\alpha-V_0\right), \tag{A-5}$$

for arbitrary u and V_0[10]. Take the limit of (A-5) as $u \to 0$

$$\mathbb{E}\left[g\left(\tilde{\eta},-V_0\right)\right]=\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]+g\left(0,\alpha-V_0\right). \tag{A-6}$$

Take the derivative of (A-6) with respect to V_0

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-V_0\right)\right]=g_2\left(0,\alpha-V_0\right), \tag{A-7}$$

where the subscript denotes the partial derivative. Now take limits as $V_0 \to \alpha$

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-\alpha\right)\right]=g_2\left(0,0\right), \tag{A-8}$$

Since (A-8) holds for any random variable $\tilde{\eta}$ with $\mathbb{E}\left[\tilde{\eta}\right]=0$ and $\mathbb{E}\left[\tilde{\eta}^2\right]=\alpha$, $g_2(S, V)$ must take the form

$$g_2\left(S,V\right)=a+B\left(V\right)S+C\left(V\right)\left(S^2+V\right), \tag{A-9}$$

for some constant a and functions B and C. But substituting (A-9) back into (A-7) shows that $C(v)$ is a constant, which we denote by $2c$. Integrating (A-9) gives

$$g\left(S,V\right)=aV+S\int_0^V B\left(W\right)dW+c\left(2S^2+V\right)+D\left(S\right), \tag{A-10}$$

where D again is an arbitrary function. It is easy to verify that (A-10) does indeed satisfy(A-6). Substituting it into the more general (A-5) shows that the following equation must be satisfied if g is to have the Aggregation Property

[10] We require that u and V_0 can take any values within some neighborhood; we do not require that they can take any value. We derive g's properties within that neighbourhood and then use the assumption that g is analytic to extend the function to the real plane.

$$u\int_{\alpha-V_0}^{-V_0} B(W)dW + \mathbb{E}\left[D(u+\tilde{\eta}) - D(\tilde{\eta}) - D(u)\right] = 0, \qquad \text{(A-11)}$$

for arbitrary u, V_0 and random variable $\tilde{\eta}$ with zero mean. For this to hold, differentiating (A-11) with respect to V_0 gives $B(\alpha - V_0) = B(-V_0)$, so $B(V)$ must be a constant, which we denote by $3b$.

Let

$$\tilde{\eta} = \tilde{\eta}^*(\kappa) \equiv \begin{cases} 1+\sqrt{\kappa} & \Pr = 1/2 \\ 1-\sqrt{\kappa} & \Pr = 1/2 \end{cases} \qquad \text{(A-12)}$$

for some $|\kappa| < 1$. Substitute into (A-11), divide by $\kappa/2$ and take limits as $\kappa \to 0$

$$\begin{aligned} &D''(u) - D''(0) = 6bu, \text{ so} \\ &D(u) = d_0 + d_1 u + d_2 u^2 + bu^3, \end{aligned} \qquad \text{(A-13)}$$

for any d_0, d_1, and d_2. So g must take the form

$$\begin{aligned} &g(S,V) = h_1 S + h_2 V + h_3 S^2 + h_4\left(S^3 + 3SV\right) \\ &\text{where } h_1 = d_1, h_2 = a + c, h_3 = 2c + d_2 \text{ and } h_4 = b. \end{aligned} \qquad \text{(A-14)}$$

$d_0 = 0$ since $g(0,0) = 0$. ■

Proof of Proposition 2

The proof is similar to the proof of Proposition 1, but with the added problem that the form of the variance function is not known. The proof that all members of **G*** have the Aggregation Property is straightforward. We focus on the converse.

Let $\tilde{\eta}$ be a random variable with $\mathbb{E}\left[e^{\tilde{\eta}}\right] = 1$ and $\mathbb{E}\left[f(\tilde{\eta})\right] = \alpha$, and g a function that has the Aggregation Property. We consider two processes with $t \in \{0,1,2\}$. The first is given by $s_0 = s_1 = 0$, $s_2 = \eta$. $S = e^s$ is clearly martingale. The process (s, v) is $(0, \alpha)$, $(0, \alpha)$, $(\eta, 0)$. For the Aggregation Property to hold

$$\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]=\mathbb{E}\left[g\left(0,0\right)+g\left(\tilde{\eta},-\alpha\right)\right]. \tag{A-15}$$

It follows that $g(0, 0) = 0$.

The second process for s is

$$0 \rightarrow \begin{cases} u & \rightarrow \quad u+\tilde{\eta} \qquad \Pr = \pi \\ d & \rightarrow \qquad d \qquad \Pr = 1-\pi \end{cases}. \tag{A-16}$$

The process S is martingale provided that $ud < 0$ and $\pi e^u + \left(1-\pi\right)e^d = 1$. The process (s, v) is

$$\left(0,v_0\right) \rightarrow \begin{cases} \left(u,\alpha\right) & \rightarrow \quad \left(u+\tilde{\eta},0\right) \qquad \Pr = \pi \\ \left(d,0\right) & \rightarrow \quad \left(d,0\right) \qquad \Pr = 1-\pi \end{cases} \tag{A-17}$$
$$\text{where } v_0 = \pi\mathbb{E}\left[f\left(u+\tilde{\eta}\right)\right]+\left(1-\pi\right)f\left(d\right).$$

For the Aggregation Property to hold

$$\begin{gathered}\mathbb{E}\left[\pi g\left(u+\tilde{\eta},-v_0\right)+\left(1-\pi\right)g\left(d,-v_0\right)\right]= \\ \mathbb{E}\left[\pi g\left(u,\alpha-v_0\right)+\pi g\left(\tilde{\eta},-\alpha\right)+\left(1-\pi\right)g\left(d,-v_0\right)+\left(1-\pi\right)g\left(0,0\right)\right].\end{gathered} \tag{A-18}$$

Simplify, and use the fact that $g(0, 0) = 0$, to give

$$\mathbb{E}\left[g\left(u+\tilde{\eta},-v_0\right)\right]=\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]+g\left(u,\alpha-v_0\right), \tag{A-19}$$

for arbitrary u and v_0. Take the limit of (A-19) as $u \rightarrow 0$

$$\mathbb{E}\left[g\left(\tilde{\eta},-v_0\right)\right]=\mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right]+g\left(0,\alpha-v_0\right), \tag{A-20}$$

Take the derivative of (A-20) with respect to v_0

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-v_0\right)\right]=g_2\left(0,\alpha-v_0\right), \tag{A-21}$$

where the subscript denotes the partial derivative. Now take limits as $v_0 \to \alpha$

$$\mathbb{E}\left[g_2\left(\tilde{\eta},-\alpha\right)\right]=g_2\left(0,0\right), \tag{A-22}$$

Since (A-22) holds for any random variable $\tilde{\eta}$ with $\mathbb{E}\left[e^{\tilde{\eta}}\right]=1$ and $\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]=\alpha$, $g_2(s, v)$ must take the form

$$g_2\left(s,v\right)=a+B\left(v\right)\left(e^s-1\right)+C\left(v\right)\left(f\left(s\right)+v\right), \tag{A-23}$$

for some constant *a* and functions *B* and *C*. Substituting (A-23) back into (A-20) shows that $C(v)$ is a constant, which we denote by $2c$. Integrating (A-23) gives

$$g\left(s,v\right)=av+\left(e^s-1\right)\int_0^v B\left(w\right)dw+cv\left(2f\left(s\right)+v\right)+D\left(s\right), \tag{A-24}$$

where *D* again is an arbitrary function. It is easy to verify that (A-24) does indeed satisfy (A-20). Substituting it into the more general (A-19) shows that the following equation must be satisfied if *g* is to have the Aggregation Property

$$\begin{aligned}\left(e^u-1\right)\int_{\alpha-v_0}^{-v_0} B\left(w\right)dw-2c\left\{v_0\mathbb{E}\left[f\left(u+\tilde{\eta}\right)-f\left(u\right)\right]+\left(f\left(u\right)-v_0\right)\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]\right\}\\ +\mathbb{E}\left[D\left(u+\tilde{\eta}\right)-D\left(\tilde{\eta}\right)-D\left(u\right)\right]=0\end{aligned} \tag{A-25}$$

For a random variable $\tilde{\eta}$ and $p\in\left[0,1\right]$ define

$$\tilde{\eta}_p \equiv \begin{cases}\tilde{\eta} & \Pr = p\\ 0 & \Pr = 1-p\end{cases}. \tag{A-26}$$

If $\mathbb{E}\left[e^{\tilde{\eta}}\right]=1$ and $\mathbb{E}\left[f\left(\tilde{\eta}\right)\right]=\alpha$ then $\mathbb{E}\left[e^{\tilde{\eta}_p}\right]=1$ and $\mathbb{E}\left[f\left(\tilde{\eta}_p\right)\right]=\alpha_p$. Putting $\tilde{\eta}_p$ into (A-25) gives

$$\left(e^u - 1\right)\int_{p\alpha - v_0}^{-v_0} B(w)\,dw - 2cp\left\{v_0\mathbb{E}\left[f(u+\tilde{\eta}) - f(u)\right] + \left(f(u) - v_0\right)\mathbb{E}\left[f(\tilde{\eta})\right]\right\} \\ + p\mathbb{E}\left[D(u+\tilde{\eta}) - D(u) - D(\tilde{\eta})\right] - (1-p)D(0) = 0. \tag{A-27}$$

By setting $p = 0$, we can see that $D(0) = 0$. Since the other terms in (A-27) are linear in the arbitrary scalar p so must the first term be, which implies that B is constant, and can be denoted by b. So (A-25) can be simplified to

$$b\left(e^u - 1\right)\mathbb{E}\left[f(\tilde{\eta})\right] - 2c\left\{v_0\mathbb{E}\left[f(u+\tilde{\eta}) - f(u)\right] + \left(f(u) - v_0\right)\mathbb{E}\left[f(\tilde{\eta})\right]\right\} \\ + \mathbb{E}\left[D(u+\tilde{\eta}) - D(\tilde{\eta}) - D(u)\right] = 0 \tag{A-28}$$

Let

$$\tilde{\eta} = \tilde{\eta}^*(\kappa) \equiv \begin{cases} \ln\left(1+\sqrt{\kappa}\right) & \Pr = 1/2 \\ \ln\left(1-\sqrt{\kappa}\right) & \Pr = 1/2 \end{cases} \tag{A-29}$$

for some $|\kappa| < 1$. Substitute into (A-28), divide by $\kappa/2$ and take limits as $\kappa \to 0$

$$2b\left(e^u - 1\right) - 2c\left\{v_0\left(f''(u) - f'(u) - 2\right) + 2f(u)\right\} \\ + \left(D''(u) - D'(u) - D''(0) - D'(0)\right) = 0 \tag{A-30}$$

Since (A-30) holds for arbitrary v_0

$$c = 0, \text{ or } f''(u) - f'(u) - 2 = 0. \tag{A-31}$$

If $c \neq 0$, we can solve for f using its limit properties at 0 to give

$$f(u) = 2\left(e^u - 1 - u\right) = L(u). \tag{A-32}$$

The general solution for D from (A-30) is

$$D(u) = d_1 u + d_2\left(e^u - 1\right) + (8c - 2b)ue^u + 4cu^2 \tag{A-33}$$

where d_1 and d_2 are arbitrary scalars. Putting (A-33) into (A-24) gives

$$\begin{aligned} &g(s,v) = h_1 s + h_2\left(e^s - 1\right) + h_3 v + h_4 (v-2s)^2 + h_5 e^s (v+2s), \text{ where} \\ &h_1 = d_1; \quad h_2 = d_2; \quad h_3 = a+b-4c; \quad h_4 = c; \quad h_5 = b+4. \end{aligned} \tag{A-34}$$

Finally, substituting for g into (A-19) gives

$$\begin{aligned} &\mathbb{E}\left[g\left(u+\tilde{\eta},-v_0\right)\right] - \mathbb{E}\left[g\left(\tilde{\eta},-\alpha\right)\right] - g\left(u,\alpha-v_0\right) = \\ &h_4\left(4u+2v_0-2\alpha\right)\left(\mathbb{E}[2\tilde{\eta}]+\alpha\right) + h_5\left(e^u-1\right)\left(\mathbb{E}\left[2\tilde{\eta}e^{\tilde{\eta}}\right]-\alpha\right). \end{aligned} \tag{A-35}$$

For this to be zero, as required for Aggregation, one of three conditions is necessary

$$\begin{aligned} &1) \quad h_4 = h_5 = 0; \\ &2) \quad h_4 = 0, \text{ and } \mathbb{E}\left[f\left(\tilde{\eta}\right)\right] = \mathbb{E}\left[2\tilde{\eta}e^{\tilde{\eta}}\right], \text{so } f = E; \\ &3) \quad h_5 = 0, \text{ and } \mathbb{E}\left[f\left(\tilde{\eta}\right)\right] = \mathbb{E}\left[-2\tilde{\eta}\right], \text{so } f = L. \end{aligned} \tag{A-36}$$

■

Proof of Proposition 5

By assumption, f is an analytic function which can be written as

$$\begin{aligned} &f(x) = x^2 + h(x), \\ &\text{where } h(x) = \sum_{k\geq 3} a_k x^k. \end{aligned} \tag{A-37}$$

and the convergence radius of f is infinity, so

$$\limsup_{k\to\infty} \sqrt[k]{\left|a_k\right|} = 0. \tag{A-38}$$

Let X be a continuous semimartingale. Assume, without loss of generality, that $X(0) = 0$. For a function g, define the *g-variation* of X as

$$V(X,g)_t \equiv \lim_{n\to\infty} V^{(n)}(g)_t \text{ (if it exists), where}$$
$$V^{(n)}(X,g)_t \equiv \sum_{i=1}^{[nt]} g\left(X_{i/n} - X_{(i-1)/n}\right), \tag{A-39}$$

where [x] denotes the integer part of x. Let T_N be the following increasing sequence of stopping times

$$T_N \equiv \inf\left\{t > 0 \,\middle|\, |X_t| \geq N\right\}, \tag{A-40}$$

and define the stopped process

$$X_t^N \equiv X_{t\wedge T_N}. \tag{A-41}$$

For $k \geq 3$,

$$\begin{aligned}
\sum_{i=1}^{[nt]} \left|X_{i/n}^N - X_{(i-1)/n}^N\right|^k &= \sum_{i=1}^{[nt]} (2N)^k \left|\frac{X_{i/n}^N - X_{(i-1)/n}^N}{2N}\right|^k \\
&\leq (2N)^k \sum_{i=1}^{[nt]} \left|\frac{X_{i/n}^N - X_{(i-1)/n}^N}{2N}\right|^3 \\
&= (2N)^{k-3} \sum_{i=1}^{[nt]} \left|X_{i/n}^N - X_{(i-1)/n}^N\right|^3.
\end{aligned} \tag{A-42}$$

Hence

$$\left|\sum_{k=1}^{M} \left(a_k \sum_{i=1}^{[tn]} \left(X_{i/n}^N - X_{(i-1)/n}^N\right)^k\right)\right| \leq \left\{\sum_{k=1}^{M} |a_k| (2N)^{k-3}\right\} \left\{\sum_{i=1}^{[tn]} \left|X_{i/n}^N - X_{(i-1)/n}^N\right|^3\right\}. \tag{A-43}$$

(A-38) implies that the first term on the right hand side tends to a finite limit as M tends to infinity. Denoting the limit by c_N,

$$\left|V^{(n)}\left(X^N, h\right)_t\right| \leq c_N \left\{\sum_{i=1}^{[nt]} \left|X_{i/n}^N - X_{(i-1)/n}^N\right|^3\right\} \text{ for all } N. \tag{A-44}$$

So for each N and for all $t < T_N$ the h-variation of X is bounded. Since X is continuous by assumption, its cubic variation converges to zero in probability as n tends to infinity (see Lepingle, 1976, and Jacod, 2008). Hence, by (A-44), the h-variation of X equals 0 on [0, T_N], and the f-variation of X equals the quadratic variation on the same interval. Since T_N is an increasing sequence of stopping times tending to infinity, $T_N \wedge t$ increases to t a.s.. Furthermore, for $N' > N$, we have $V(X^N, f)_t = V(X^{N'}, f)_t$ on $\left[0, T_N \wedge t\right]$, hence the f-variation of X is well defined by the sequence $V(X^N, f)_t$ and equals the quadratic variation of X. The result holds true for any function f that satisfies (A-37) and (A-38). In particular, it holds for $f = L$ and $f = E$. ■

Proof of Proposition 7

For covariance

$$\begin{aligned} v_t^L(k) &= -2E_t\left[\ln\frac{W_T}{W_t}\right] = -2E_t\left[\ln\left(R_{t+1}+kX\right)\prod_{u=t+2}^{T}R_u\right] \\ \text{Covariance} &= \frac{1}{2}\frac{\partial v_t^L(k)}{\partial k}\Big|_{k=0} = -E_t\left[\frac{X}{R_{t+1}}\right]. \end{aligned} \tag{A-45}$$

For coskewness

$$\begin{aligned} q_t(k) &= 6E_t\left[\left(1+\frac{W_T}{W_t}\right)\ln\frac{W_T}{W_t}\right] \\ &= 6E_t\left[\left\{1+\left(R_{t+1}+kX\right)\prod_{u=t+2}^{T}R_u\right\}\ln\left(R_{t+1}+kX\right)\prod_{u=t+2}^{T}R_u\right]. \end{aligned} \tag{A-46}$$

So

$$\frac{\partial q_t(k)}{\partial k}\Big|_{k=0} = 6E_t\left[X\prod_{u=t+2}^{T}R_u\ln\left(R_{t+1}\prod_{u=t+2}^{T}R_u\right)+\left\{1+R_{t+1}\prod_{u=t+2}^{T}R_u\right\}\frac{X}{R_{t+1}}\right]. \tag{A-47}$$

Plugging in the definition of coskewness, of entropy variance and using the martingale property of returns, coskewness is given by

$$\frac{1}{3}\frac{\partial v_t^L\left(k\right)}{\partial k}\Big|_{k=0} = E_t\left[2X\left(\ln R_{t+1} + 1/R_{t+1}\right) + Xv_{t+1}^E\right]. \qquad \text{(A-48)}$$

■

References

Andersen, T., T. Bollerslev, F. Diebold and P. Labys, 2003, "Modeling and Forecasting Realized Volatility", *Econometrica*, 71(2), 579-625.

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

Ang, A., R. Hodrick, Y. Xing and X. Zhang, 2006, "The Cross-Section of Volatility and Expected Returns", *Journal of Finance*, 61(1), 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(1), 101-143.

Bakshi, G., and D. Madan, 2000, "Spanning and Derivative Security Valuation", *Journal of Financial Economics*, 55, 205-238.

Barro, R., 2009, "Rare Disasters, Asset Prices, and Welfare Costs", *American Economic Review*, 99(1), 243-264.

Breeden, D.T., and R.H.Litzenberger, 1978, "Prices of State Contingent Claims Implicit in Option Prices", *Journal of Business*, 51(4), 621-651.

Britten-Jones, M., and A. Neuberger, 2000, "Option Prices, Implied Price Processes, and Stochastic Volatility", *Journal of Finance*, 55(2), 839-866.

Carr, P., K. Ellis and V. Gupta, 1998, "Static Hedging of Exotic Options", *Journal of Finance*, 53(3), 1165-1190.

Carr, P., and L. Wu, 2009, "Variance Risk Premiums", *Review of Financial* Studies, 22, 1311-1341.

Harrison, J. M., and D. M. Kreps, 1979. "Martingales and arbitrage in multiperiod securities markets," *Journal of Economic Theory*, 20(3), 381-408.

Harvey, C.R., and A. Siddique, 2000, “Conditional Skewness in Asset Pricing Tests”, *Journal of Finance*, 55(3), 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), 327-343.

Jacod, J., 2008, “Asymptotic Properties of Power Variations and Associated Functionals of Semimartingales”, *Stochastic Processes and their Applications*, 118, 517-559.

Jiang, G. J., and Y.S. Tian, 2005, “The Model-Free Implied Volatility and Its Information Content”, *Review of Financial Studies*, 18(4), 1305-1342.

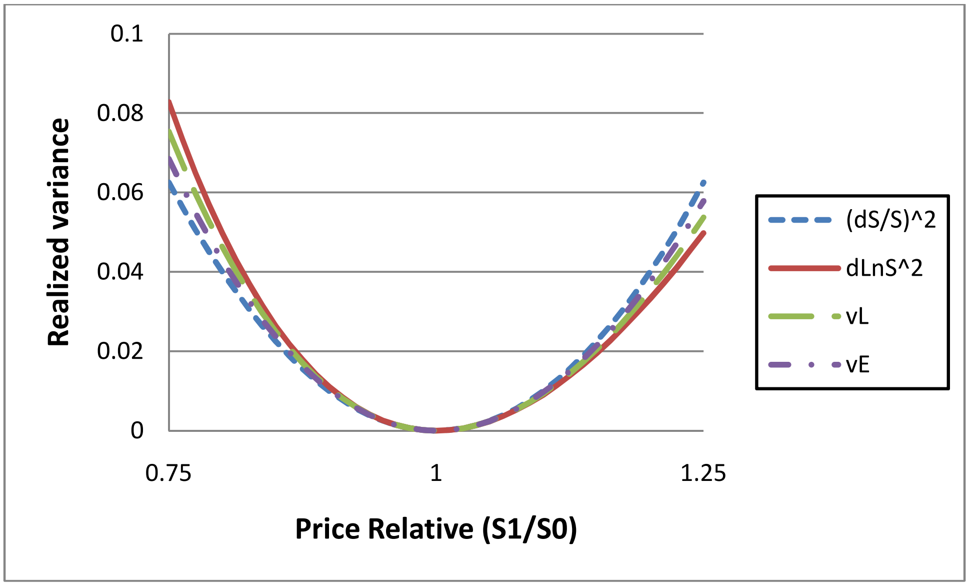
Kraus, A., and R. H. Litzenberger, 1976, “Skewness Preference and the Valuation of Risk Assets”, *Journal of Finance*, 31(4), 1085-1100.

Lepingle, D., 1976, “ La variation d'ordre p des semi-martingales”, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36, 295-316.

Lo, A., and A. C. MacKinlay, 1988, “Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test”, *Review of Financial Studies*, 1(1), 41-66.

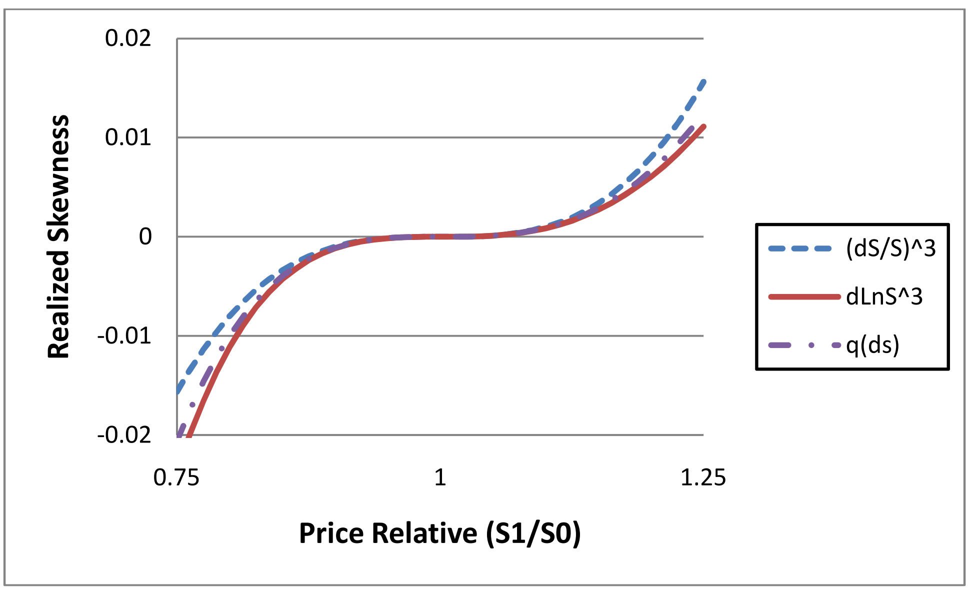
Xing, Y., X. Zhang, and R. Zhao, 2010, “What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns?”, *Journal of Financial and Quantitative Analysis*, 45(3), 641-662.

Figure 1: Alternative Measures of Variance



The chart shows four alternative measures of realized variance: squared net return $(R-1)^2$, squared log return $(\ln R)^2$, the log measure $2(R-1-\ln R)$ and the entropy measure $2(R\ln R-R+1)$, where R is the price relative.

Figure 2: Alternative Measures of Skewness



The chart shows three alternative measures of realized skewness: cubed net return $(R-1)^3$, cubed log return $(\ln R)^3$, and $6(1+R)\ln R-12(R-1)$ which is the one used in the definition of realized skew. R is the price relative.