Abstract

Asymmetric shocks are common on markets; securities' payoffs are not normally distributed and exhibit skewness. Moreover, even when primary assets have symmetric payoffs, typical derivatives display a high degree of skewness. This paper studies the portfolio holdings of heterogeneous agents with preferences over mean, variance and skewness, and derives equilibrium prices. We use this to understand the structural dependencies in the variance risk and skewness risk premiums for stochastic discount factors which are quadratic in the market return. Risk-neutral variance is linked to the pricing of skewness and to derivatives. Finally we point out the importance of a conditional viewpoint to address the factor premiums on the market return and its squared counterpart.

Keywords: (Co-)Skewness

JEL: C52, D58, G11, G12
1 Introduction

Asymmetric shocks are common on markets; securities’ payoffs are not normally distributed and exhibit skewness. Moreover, even when primary assets have symmetric payoffs, typical derivatives display a high degree of skewness. The important contribution of Harvey and Siddique (2000) renewed interest in the compensation of skewness risks and lead to an active literature\(^1\). This paper revisits the pricing implications of Stochastic Discount Factors (henceforth SDF) which are quadratic in the market return, and links the price of skewness risk to derivatives and to risk-neutral variance. We particularly stress the importance of a conditional viewpoint for estimation of the skewness premium. Furthermore, while the literature is largely based on ad-hoc extensions of the CAPM where the squared market return is a priced factor (in addition to the market return) this paper provides a theoretical foundation for this practice.

Samuelson (1970) studied the limit of portfolio holdings under infinitesimal risk\(^2\) and concluded that mean-variance analysis largely characterizes the optimal portfolio problem even when the decision maker has a general concave von Neumann-Morgenstern utility function and asset returns are not normally distributed. In the presence of “small” risks it is necessary to study also the slope of portfolio holdings in the neighborhood of infinitesimal risk. This paper extends Samuelson’s analysis of financial decision making to this slope and thereby introduces skewness risk into the analysis; we derive agents’ portfolio holdings and the equilibrium allocation under mean-variance-skewness risk.

In the first part of the paper we characterize agent’s portfolio holdings using risk-tolerance and a term we call skew-tolerance which contains the third derivative of agent’s utility function. Risk-tolerance captures the mean-variance trade-off and skew-tolerance the mean-variance-skewness trade-off. Using appropriately defined “average” risk-tolerance and “average” skew-tolerance we discuss that such an “average” agent sets prices. We prove a separation theorem in which heterogeneous agents’ holdings are composed of two funds: the market portfolio and a new portfolio we call skewness portfolio. The skewness portfolio is the portfolio with a return ”closest” to that of squared market return. Holdings of the market portfolio are equal to the ratio of individual risk-tolerance to that of the “average” agent; holdings of the skewness portfolio are proportional to the difference between individual agents’ skew-tolerance and that of the “average” agent.

---


\(^2\)He studied a series of economies differing only by the amount of risk; the case of infinitesimal risk is the limit economy where all risk vanishes.
Although the return from the skewness portfolio differs from the squared market return, it remains true that any risk is compensated only through its relationship with the market, either through standard market beta or through market co-skewness which is akin to beta with respect to the squared market return. In this respect, one may say that both idiosyncratic variance and idiosyncratic skewness are not compensated in equilibrium.

In the second part of the paper we study extensively the pricing implications of an SDF which is quadratic in the market return. Although motivated by our extension of Samuelson’s small risk analysis, this part of our study is valid under very general settings and can be compared to previous literature on the pricing of skewness risks. Along the lines we revisit beta pricing under skewness as it has been considered previously by Kraus and Litzenberger (1976), Ingersoll (1987), and Harvey and Siddique (2000), among others. We also relate skewness pricing to important terms in derivatives pricing: to risk neutral variance, which has been studied extensively by Rosenberg (2000), and to the price of volatility contracts, studied by Bakshi and Madan (2000).

Our paper makes the following three contributions: First, we provide a rigorous foundation for the use of SDF which are quadratic in the market return. Most empirical studies looked at skewness extensions of the CAPM which add the squared market return as a factor. Those authors which justify this extension base their proofs on assumed separation and aggregation results or on an ad-hoc truncation of a Taylor-series expansion for the utility function at the third-order term, see, e.g., Kraus and Litzenberger (1976), Barone-Adesi (1985), Dittmar (2002). The insight of Samuelson (1970) was that the use of mean-variance analysis does not have to be based on truncated Taylor-series expansions: limits with vanishing risk justify such an analysis as an approximation\(^3\). Our extension of Samuelson’s analysis to skewness risk permits a rigorous analysis of separation and aggregation: we prove that simple market separation does not hold but that, somewhat surprisingly, the SDF depends locally on the quadratic market return.

Second, we study extensively the pricing implications of SDF that are quadratic in the market return. We shed more light on beta pricing relationships proposed by Harvey and Siddique (2000) and show that they correspond to a limit case of a zero-risk neutral variance of the market. We put forward a more general

\(^3\) A work that also extends Samuelson’s analysis is Judd and Guu (2001) where Samuelson’s analysis is also extended to an asymptotically valid theory for the trade-off between one risky asset and the riskless asset in single period setups. However, while their approach is based on bifurcation theory, our results are based directly on limits of first order conditions. Furthermore their interest is on two-agent economies with a single risky asset and potentially a derivative written on it; they do not study stochastic discount factors.
beta pricing relationship which explicitly depends on the price of the squared return on the market portfolio, or equivalently, on the market risk neutral variance. This opens the door to more extensive empirical studies of the skewness premium based on derivatives prices.

Finally, we add to the literature which aims at identifying the skewness premium. The statistical identification of a significantly positive skewness premium is generally considered a difficult task, see, e.g. Barone-Adesi, Urga, and Gagliardini (2004). We provide some empirical evidence which suggests that such premia show up in a more manifest way when they are considered with a conditional point of view, as it has been in Harvey and Siddique (2000). Our evidence is documented from simulated data on the GARCH factor model with in mean effects using the parameter estimates of Bekaert and Liu (2004). Moreover, our simulation also suggests that neglecting the market risk neutral variance — as it has been, e.g., in Harvey and Siddique (2000) — leads to a severe underestimation of skewness premium which may go so far as to invert its sign.

The remainder of the paper is organized as follows: the next section discusses portfolio choice and asset pricing in the context of infinitesimal risks. Section 3 studies quadratic pricing kernels in the conditional setup of Hansen and Richard (1987). Section 4 makes an empirical assessment of the order of magnitude of the various effects put forward in Section 3. All proofs are postponed to the appendix.

2 Static Portfolio Analysis in Terms of Mean, Variance and Skewness

Samuelson (1970) argues that, for risks that are infinitely small, optimal shares of wealth invested in each security coincide with those of a mean-variance optimizing agent. However Samuelson (1970) also derives a more general approximation theorem about higher order approximations: to further characterize the way the optimal shares vary locally in the direction of any risk, that is their first derivatives at the limit point of zero risk, one needs to push one step further the Taylor expansion of the utility function; carrying this out will lead us to a mean-variance-skewness approach.

We start here from a slight generalization of this Samuelson’s result. Following closely his exposition, let us denote by $R_i$ the (gross) return from investing $1$ in risky security $i = 1, ..., n$. The random vector $R = (R_i)_{1 \leq i \leq n}$ defines the joint probability distribution of interest, which is specified by the following
decomposition:

\[ R_i(\sigma) = R_f + \sigma^2 a_i(\sigma) + \sigma Y_i. \]  

(1)

Here, \( a_i(\sigma) \), \( i = 1, \ldots, n \), are positive functions of \( \sigma \) and \( R_f \) is the gross return on the riskless (safe) security. The \( \sigma \) parameter characterizes the scale of risk and is crucial for our analysis. In this section we are interested in approximations in the neighborhood of \( \sigma = 0 \). The small noise expansion (1) provides a convenient framework to analyze portfolio holdings and resulting equilibrium allocations for a given random vector \( Y = (Y_i)_{1 \leq i \leq n} \) with

\[ E[Y] = 0 \text{ and } Var(Y) = \Sigma, \]

where \( \Sigma \) is a given symmetric and positive definite matrix. For future reference we denote by

\[ \Gamma_k = E[YY^\perp Y_k] \]

the matrix of covariances between \( Y_k \) and cross-products \( Y_i Y_j, \ i, j = 1, \ldots, n \).

In equation (1), the term \( \sigma^2 a_i(\sigma) \) has the interpretation of the risk premium. Samuelson (1970) restricts the function \( a_i(\sigma) \) to constants; under this assumption risk premia are proportional to the squared scale of risk; we relax this restriction throughout since it would prevent us from analyzing the price of skewness in equilibrium. Throughout we refer to \( a(\sigma) = (a_i(\sigma))_{i=1,\ldots,n} \) as the vector of risk premia.

2.1 The individual investor problem

We consider an investor with Von Neumann-Morgenstern preferences, i.e. she derives utility from date 1 wealth according to the expectation over some increasing and concave function \( u \) evaluated over date 1 wealth; for given risk-level \( \sigma \) she then seeks to determine portfolio holdings \( (\omega_i)_{1 \leq i \leq n} \in \mathbb{R}^n \) that maximize her expected utility:

\[
\max_{(\omega_i)_{1 \leq i \leq n}} E\left[u\left(R_f + \sum_{i=1}^{n} \omega_i \cdot (R_i(\sigma) - R_f)\right)\right].
\]

(2)

For notational simplicity we normalized the initial invested wealth to one. The solution of this program is denoted by \( (\omega_i(\sigma))_{1 \leq i \leq n} \) and depends on the given scale of the risk \( \sigma \). The question we ask is then: to what extent does a Taylor approximation of \( u \) allow us to understand the local behavior of the shares \( \omega_i(\sigma), i = 1, \ldots, n \), in the neighborhood of the zero risk \( \sigma = 0 \)? Put differently, we want to characterize for

---

\(^4\)Samuelson (1970) provides a heuristic explanation of (1) that is of interest for readers accustomed to continuous-time finance models; he couches this terms of Brownian motion of time and identifies \( \sigma \) with the square root of time.
For $i = 1, \ldots, n$ the quantities:

$$
\omega_i(0) = \lim_{\sigma \to 0^+} \omega_i(\sigma) \quad \text{and} \quad \omega'_i(0) = \lim_{\sigma \to 0^+} \omega'_i(\sigma).
$$

(3)

Samuelson (1970) stresses that a third order Taylor expansion of $u$ is needed to do the job. We slightly extend his result by showing that it remains valid even though the function $a_i(\sigma)$ are not assumed to be constant. Let us then consider a third order Taylor expansion of $u$ in the neighborhood of the safe return $R_f$:

$$
u^*(W) = u(R_f) + u'(R_f)(W - R_f) + \frac{u''(R_f)}{2!}(W - R_f)^2 + \frac{u'''(R_f)}{3!}(W - R_f)^3.
$$

(4)

Let us denote by $(\omega^*_i(\sigma))_{1 \leq i \leq n}$ the solution of the approximated problem, i.e. $(\omega^*_i(\sigma))_{1 \leq i \leq n} \in \mathbb{R}^n$ describes the holdings of an agent with utility function $u^*$ and solves

$$
\max_{(\omega^*_i)_{1 \leq i \leq n}} E\left[u^*(R_f + \sum_{i=1}^{n} \omega^*_i \cdot (R_i - R_f))\right].
$$

(5)

For $i = 1, \ldots, n$ the terms $\omega^*_i(0)$ and $\omega''_i(0)$ are defined similar to (3) as continuity extension. We prove that Taylor expansions give tangency equivalences:

**Theorem 2.1** Under suitable smoothness and concavity assumptions, the solution to the general problem (2) is related asymptotically to that of the three-moment problem by the tangency equivalences:

$$
\omega_i(0) = \omega^*_i(0) \quad \text{and} \quad \omega'_i(0) = \omega''_i(0) \quad \text{for all} \quad i = 1, \ldots, n
$$

The intuition behind this theorem is that:

1. The optimal shares of wealth invested $\omega_i(0), i = 0, \ldots, n$, in the limit case $\sigma \to 0$ depend on its first two derivatives $u'(R_f)$ and $u''(R_f)$. Thus a second order Taylor expansion of $u$, that is a mean-variance approach provides a correct characterization of these shares.

2. The first derivatives with respect to $\sigma$, $\omega'_i(0), i = 1, \ldots, n$ of optimal shares, in the limit case $\sigma \to 0$, depend on the utility function $u$ only through its first three derivatives $u'(R_f), u''(R_f)$ and $u'''(R_f)$. Thus a third order Taylor expansion of $u$, that is a mean-variance-skewness approach, does the job.

In the following we will analyze portfolio holdings. For future reference in this subsection we denote

$$
\tau = -\frac{u'(R_f)}{u''(R_f)} \quad \text{and} \quad \rho = \frac{\tau^2 u'''(R_f)}{2 u'(R_f)}
$$

(6)
the risk tolerance coefficient and the skew tolerance coefficient of the agent.

As far as optimal shares are concerned, the following theorem confirms that they conform to standard mean-variance formulas, that is to formulas usually obtained with an assumption of joint normality of returns:

**Theorem 2.2** The vector \( \omega(0) = (\omega_i(0))_{1 \leq i \leq n} \) of shares of wealth invested in the limit case \( \sigma \to 0 \) fulfills:

\[
\omega(0) = \tau \Sigma^{-1} a(0).
\]

The equivalence with standard formulas commonly derived under an assumption of joint normality can be understood better from the following two remarks:

1. It is known that under joint normality with a general utility function the mean-variance tradeoff would be given by

\[
- E[u'(W(\sigma))]/E[u''(W(\sigma))] \quad \text{with} \quad W(\sigma) = R_f + \sum_{i=1}^{n} \omega_i(\sigma) \cdot (R_i(\sigma) - R_f).
\]

This term plays the role of the risk-tolerance coefficient, and we directly see that this coincides with \( \tau \) in the limit case \( \sigma \to 0 \). Therefore our risk-tolerance can be interpreted as a generalization of the standard one.

2. Joint normality would imply in equilibrium, constant functions \( a_i(\sigma) \) (see theorem 2.4 below). In such a case, the formula of theorem 2.2 can be rewritten:

\[
\omega(0) = \tau \cdot \left( \text{Var}(R(\sigma)) \right)^{-1} \sigma^2 a,
\]

where \( a(\sigma) = a \) is constant. We recall that \( \sigma^2 a \) defines the vector of risk premia.

Generally speaking, following theorem 2.2, if we see optimal shares of wealth invested \( \omega(\sigma) \) as equivalent to \( \tau \Sigma^{-1} a(\sigma) \) in the neighborhood of \( \sigma = 0 \), we get a Sharpe ratio for optimal portfolios equivalent to:

\[
E \left[ \omega^\perp(\sigma) \cdot (R(\sigma) - R_f) \right] / \left( \text{Var} \left( \omega^\perp(\sigma) \cdot R(\sigma) \right) \right)^{\frac{1}{2}} = \sigma P(0).
\]

Then,

\[
\sigma^2 P^2(0) = \frac{E \left[ \omega^\perp(\sigma) \cdot (R - R_f) \right]^2}{\text{Var} \left( \omega^\perp(\sigma) \cdot R \right)} = \frac{(\tau a^\perp(0) \Sigma^{-1} \sigma^2 a(0))^2}{\tau a^\perp(0) \Sigma^{-1} (\sigma^2 \Sigma) \Sigma^{-1} \tau a(0)} = \sigma^2 \frac{(a^\perp(0) \Sigma^{-1} a(0))^2}{a^\perp(0) \Sigma^{-1} a(0)},
\]

so that

\[
P(0) = \left( \frac{a^\perp(0) \Sigma^{-1} a(0)}{a^\perp(0) \Sigma^{-1} a(0)} \right)^{\frac{1}{2}}.
\]

This denotes, by unit of scaling risk \( \sigma \), the potential performance of the set \( R \) of returns as in traditional mean variance analysis [see e.g. Jobson and Korkie (1982)]. Of course, the above analysis neglects the
variation in equilibrium of the risk premium functions \(a(\sigma)\). We are going to see in theorem 2.4 below that these functions will not be constant, even locally in the neighborhood of \(\sigma = 0\), as soon as asset return joint probability distribution features some asymmetries.

These asymmetries will actually play a double role in the local behavior of optimal shares of wealth invested. First, preferences for skewness would increase, ceteris paribus, asset demands in the direction of positive skewness. Second, market equilibrium induced variations in risk premium potentially erase this effect. To see this, let us define the co-skewness of asset \(k\) in portfolio \(\omega\) as:

**Definition 2.3** The co-skewness of asset \(k\) in portfolio \(\omega\) is:

\[
c_k(\omega) = \frac{\text{Cov}(Y_k, (\omega^\top Y)^2)}{\text{Var}(\omega^\top Y)} = \frac{\omega^\top \Gamma_k \omega}{\omega^\top \Sigma \omega}. \tag{8}
\]

We will see below (theorem 2.10 together with equation 17 and theorem 3.3 in section 3) that this notion of co-skewness is tightly related to a measure put forward by Kraus and Litzenberger (1976) (see also Ingersoll (1987), p. 100). For the optimal portfolio \(\omega(0)\), linear transformation based on equation (7) and theorem 2.2 show that

\[
c_k(\omega(0)) = \frac{1}{P^2(0)} a^\top(0) \Sigma^{-1} \Gamma_k \Sigma^{-1} a(0). \tag{9}
\]

Typically, asymmetry in the joint probability distribution of the vector \(R\) of returns means that at least some matrices \(\Gamma_k, k = 1, \ldots, n\) are not zero. We get the following result:

**Theorem 2.4** The slope \(\omega'(0)\) of the vector \(\omega(0)\) of optimal shares of wealth invested in the neighborhood of \(\sigma = 0\) is given by:

\[
\omega'(0) = \tau \Sigma^{-1} \cdot \left\{ a'(0) + \rho P^2(0) c(\omega(0)) \right\},
\]

where \(a'(0) = (a_i'(0))_{1 \leq i \leq n}\) is the vector of marginal risk premia.

In other words, up to variations \(a'(0)\) of risk premiums in equilibrium, a positive co-skewness of asset \(k\) will have a positive effect on the demand of this asset with respect to common mean-variance formulas. This positive effect will be all the more pronounced that the skew tolerance coefficient \(\rho\) is large. Of course, this interpretation is based on two implicitly maintained assumptions:

1. The skew tolerance coefficient is nonnegative, i.e. \(\frac{w''(R_f)}{w'(R_f)} > 0\). This assumption is conforms to both the literature on prudence (Kimball (1990)) and the literature on preferences for high order moments (Dittmar (2002), Harvey and Siddique (2000), Ingersoll (1987)).
2. The vector $c(\omega) = (c_k(\omega))_{1 \leq k \leq n}$ represents a multivariate notion of skewness that investors like to get positive skewness, component-wise. This assertion is justified by the fact that in average:

$$\sum_{k=1}^{n} \omega_k c_k(\omega) = \frac{E[(\omega^\top Y)^3]}{Var[\omega^\top Y]} = \frac{Skew(\omega^\top Y)}{Var[\omega^\top Y]}$$

is positive if and only if the portfolio return is positively skewed. Of course, individual preferences for positive skewness will increase, ceteris paribus, the equilibrium price of assets with positively skewed returns. This will actually appear in the equilibrium value $a'(0)$ of risk premium slopes in the neighborhood of $\sigma = 0$.

### 2.2 Equilibrium Allocations and Prices

Let us consider markets for risky assets $i = 1, 2, \ldots, n$ on which $S$ agents $s = 1, \ldots, S$ can trade. For agent $s$ we denote $\omega_s(0) = (\omega_{si}(0))_{1 \leq i \leq n}$ his holdings in each of these assets; his preferences are characterized by a Von Neumann-Morgenstern utility function $u_s$ and associated preference coefficients:

$$\tau_s = -\frac{u'_s(R_f)}{u''_s(R_f)} \quad \text{and} \quad \rho_s = \frac{\tau^2_s u'''_s(R_f)}{2 u'_s(R_f)}.$$  

From theorems 2.2 and 2.4 we get that

$$\omega_s(0) = \Sigma^{-1} \tau_s a(0) \quad \text{and} \quad \omega'_s(0) = \tau_s \Sigma^{-1} \left\{ a'(0) + \rho_s P^2(0) c(\omega(0)) \right\}$$

We consider that the net supply of each risky asset $i = 1, \ldots, n$ is exogeneous and for the sake of notational simplicity fix it to unity as a normalization. Then, in the limit case $\sigma \to 0$, the market clearing conditions can be written:

$$\sum_{s=1}^{S} \omega_s(0) = e_n, \quad \text{and} \quad \sum_{s=1}^{S} \omega'_s(0) = 0.$$  

where $\omega_s(0) = (\omega_{si}(0))_{1 \leq i \leq n}$ and $e_n$ denotes the $n$-dimensional column vector, the components of which are all equal to 1. Below, it will be convenient to consider an average investor characterized by average holdings $\bar{\omega}$, an average risk tolerance $\bar{\tau}$ and average skew tolerance $\bar{\rho}$, where

$$\bar{\omega} = \frac{1}{S} e_n, \quad \bar{\tau} = \frac{1}{S} \sum_{s=1}^{S} \tau_s, \quad \text{and} \quad \bar{\rho} = \frac{\sum_{s=1}^{S} \rho_s \tau_s}{\sum_{s=1}^{S} \tau_s}.$$  

If all individual were identical, each would buy the average portfolio $\bar{\omega}$. The role of the two average preference coefficients $\bar{\tau}$ and $\bar{\rho}$ for individual portfolios is displayed by:
Theorem 2.5 In equilibrium, in the limit case $\sigma \to 0$, the (squared) market Sharpe ratio is

$$P^2(0) = \frac{1}{\bar{\tau}^2} \bar{\omega}^1 \Sigma \bar{\omega}$$

and the optimal shares of wealth invested $\omega_s(\sigma)$ of agent $s = 1, \ldots, S$ are characterized by:

$$\omega_s(0) = \frac{\tau_s}{\bar{\tau}} \bar{\omega}, \text{ and } \omega'_s(0) = \tau_s [\rho_s - \bar{\rho}] P^2(0) \Sigma^{-1} c(\bar{\omega}).$$

The Theorem states that in the limit case $\sigma \to 0$ the vector $\omega_s(\sigma)$ of optimal shares of wealth invested is as in a standard mean-variance separation theorem. All individuals buy a share of the market portfolio $e_n$, the size of this share being determined by the comparison of individual risk tolerance $\tau_s$ with respect to average one. Preferences for skewness only play a role at the level of the slopes $\omega'_s(0)$ of the shares of wealth invested in the neighborhood of zero. A positive market co-skewness $c_k(\bar{\omega})$ will have a positive effect on the individual’s demand of asset if and only if his skew tolerance coefficient is more than the average one $\bar{\rho}$. On the contrary, if $\rho_s < \bar{\rho}$, the positive effect of asset i co-skewness on its market price makes more than a compensation of investor’s preference for skewness.

We note that heterogeneous skewness preferences give rise to a departure from standard mean-variance separation theorem only through the addition of a third mutual fund defined by a vector of shares of wealth invested proportional to $\Sigma^{-1} c(\bar{\omega})$. The role of this additional fund is actually to mimic as closely as possible in terms of mean-squared error the squared de-meaned market return.

Definition 2.6 The market portfolio is defined as the portfolio containing $\bar{\omega} = 1/S$ in risky assets. We denote the market return by

$$R_M(\sigma) = \frac{1}{n} \sum_{i=1}^{n} R_i(\sigma).$$

The skewness portfolio contains $\psi P^2(0)$ in the risky asset, where the vector $\psi$ is given as the coefficients of the affine regression of $(R_M - E[R_M])^2$ on the vector $R$ of asset returns, i.e.

$$\psi = (\text{Var}(R))^{-1} \text{Cov} \left(R, (R_M - E[R_M])^2 \right).$$

In finance we often regress individual securities’ returns on the market return. Note that in definition 2.6, however, we regress the market return on the individual securities’ return. The resulting skewness portfolio can be interpreted as giving the return “closest” to the squared market returns. Note also that $R_M(\sigma)$
describes the return on the market portfolio. We have $V\text{ar}(R) = \sigma V\text{ar}(Y) = \sigma \Sigma$ and

$$ Cov (R, (R_M - E[R_M])^2) = \sigma^3 Cov (Y, (\bar{\omega} - Y)^2) \left( \frac{S}{n} \right)^2 = \sigma^3 c(\bar{\omega}) V\text{ar}(\bar{\omega} - Y) \left( \frac{S}{n} \right)^2 = \sigma c(\bar{\omega}) \frac{S}{n} V\text{ar}(R_M), $$

so that $\psi = \sigma \frac{S}{n} V\text{ar}(R_M) \Sigma^{-1} c(\bar{\omega})$. Therefore 2.5 can be interpreted as a three-fund separation theorem:

**Theorem 2.7** In a second order approximation of the market equilibrium the portfolio $\omega_s(0) + \sigma \omega'_s(0)$ of agent $s = 1, ..., S$ consists of three mutual funds: the risk free asset, $\frac{\tau}{\bar{\tau}} \bar{\omega}$ in the market portfolio and $\tau_s(\rho_s - \bar{\rho})$ in the skewness portfolio.

Note that the composition $\psi$ of skewness portfolio depends linearly on the scaling factor $\sigma$ but that the market portfolio does not. Therefore, in the limit where $\sigma$ tends to zero the skewness portfolio will vanish and the portfolio composition will be determined by the market portfolio, as in standard mean-variance theory. Note also that the skewness portfolio will be in zero net-supply and that, according to theorem 2.5, a given investor $s$ will be short or long on it depending on the position of his skew tolerance $\rho_s$ with respect to the average one $\bar{\rho}$.

Intuitively, tracking the squared de-meaned market return allows one to benefit from possibly positive market co-skewness. Of course, this comes at a price that we can characterize through the local behavior of the risk premium in equilibrium:

**Theorem 2.8** In the limit case $\sigma \rightarrow 0$, the equilibrium risk premium vector $a(\sigma)$ is such that the average portfolio $\bar{\omega}$ is optimal for the average investor: $\bar{\omega} = \frac{\tau}{\bar{\tau}} \Sigma^{-1} a(0)$, i.e.

$$ a(0) = \frac{1}{\bar{\tau}} \Sigma \bar{\omega}, $$

and its slope in the neighborhood of zero is given by:

$$ a'(0) = -\frac{\bar{\rho}}{\bar{\tau}^2} c(\bar{\omega}) = -\frac{\bar{\rho}}{\bar{\tau}^2} \bar{\omega} \Gamma_k \bar{\omega}. $$

Note that, by comparison of theorems 2.4 and 2.8, the equilibrium slopes are precisely such that the average agent would have no motive to deviate from the market portfolio, i.e. $\omega'(0) = 0$ for the average investor.

Theorem 2.8 must be interpreted as a new asset pricing model. While approximating risk premia by their limit values $a_i(0)$ would clearly give the Sharpe-Lintner CAPM, approximating them by higher order expansions $a_i(0) + \sigma a'_i(0)$ gives a new mean-variance-skewness asset pricing model.
2.3 Stochastic Discount Factor and Beta Pricing Relationships

A convenient way to describe the implications of an asset pricing model is to characterize it through a Stochastic Discount Factor (henceforth SDF), see e.g Cochrane (2001). By definition, a SDF \( m(\sigma) \) must be able to price correctly all available securities; here we therefore need that \( E[m(\sigma)] = \frac{1}{R_f} \) and that \( E[m(\sigma) \cdot (R_f + \sigma^2 a_i(\sigma) + \sigma Y_i)] = 1 \) for \( i = 1, \ldots, n \). We define the portfolio risk \( Y_\omega \) and the risk \( Y_\overline{\omega} \) of the market portfolio by

\[
Y_\omega = \omega^\top Y \quad \text{and} \quad Y_\overline{\omega} = \overline{\omega}^\top Y = \frac{1}{\sigma S} (R_M(\sigma) - E[R_M(\sigma)]).
\]

We are then able to translate theorem 2.8 in terms of SDF:

**Theorem 2.9** The random variable

\[
m(\sigma) = \frac{1}{R_f} - \sigma \frac{1}{\tau} Y_\omega + \sigma^2 \frac{\rho}{\tau^2} Y_\overline{\omega}^2 \\
= \frac{1}{R_f} - \frac{n}{R_f S \tau} \{R_M(\sigma) - E[R_M(\sigma)]\} \\
+ \frac{\rho n^2}{R_f S^2 \tau^2} \left\{ \left( R_M(\sigma) - E[R_M(\sigma)] \right)^2 - E \left[ (R_M(\sigma) - E[R_M(\sigma)])^2 \right] \right\}
\]

is a SDF consistent with variance-skewness risk premium defined by \( a(\sigma) = a(0) + \sigma a'(0) \) where \( a(0) \) and \( a'(0) \) are given by theorem 2.8.

The conjunction of theorems 2.7 and 2.9 summarizes what we have learnt so far about portfolio choice and asset pricing from a second-order approximation of the market equilibrium with heterogeneous mean-variance-skewness preferences:

1. Due to heterogeneity in preferences, the common CAPM separation theorem is violated: different individuals may hold in equilibrium different risky portfolios. However, this difference is encapsulated in the demand for a third portfolio, defined as the skewness portfolio. Moreover, the skewness portfolio is in a new zero aggregated demand.

2. The interpretation of the skewness portfolio as the closest to squared market return implies that the pricing implications of a common two-funds separation theorem remain true in some respect. Somewhat unexpectedly, the market return alone is still able to summarize the pricing of risk. Of course, since not only market betas but also market co-skewness must be taken into account, both the actual market return and its squared value enter linearly in the pricing kernel.
Following Kraus and Litzenberger (1976) seminal paper, Harvey and Siddique (2000), and Dittmar (2002) among others have recently studied the empirical implications of a SDF which involves a quadratic function of market return. Theorem 2.9 above provides a theoretical basis for doing so. Section 3 will elaborate more on the pricing implications of such a SDF.

For each security $i = 1, \ldots, n$ we define the net return $r_i(\sigma)$ and the beta $b_i(\omega)$ with the portfolio risk by

$$r_i(\sigma) = R_i(\sigma) - R_f = \sigma Y_i + \sigma^2 a_i(\sigma), \quad r_M(\sigma) = R_M(\sigma) - R_f = \frac{1}{n} \sum_{i=1}^{n} r_i(\sigma) = \frac{\sigma S}{n} Y_\omega,$$

and

$$b_i(\omega) = \frac{\text{Cov}(Y_i, Y_\omega)}{\text{Var}(Y_\omega)}, \quad c_M(\omega) = \frac{1}{n} \sum_{i=1}^{n} c_i(\omega).$$

Note that

$$b_i(\omega) = \frac{S}{n} \cdot \frac{\text{Cov}(r_i(\sigma), r_M(\sigma))}{\text{Var}(r_M(\sigma))}, \quad c_i(\omega) = \frac{\text{Cov}(Y_i, Y^2_\omega)}{\text{Var}(Y_\omega)},$$

so that $c_M(\omega) = \frac{S \text{Cov}(Y_\omega, Y^2_\omega)}{n \text{Var}(Y_\omega)}$ cannot be expressed in simple terms of $r_M(\sigma)$. We have:

**Theorem 2.10**

$$E[r_i(\sigma)] = \sigma^2 \frac{1}{\tau} \text{Var}(Y_\omega) b_i(\omega) + \sigma^3 \frac{\rho}{\tau^2} \text{Var}(Y_\omega) c_i(\omega)$$

$$= \frac{n}{S} E[r_M(\sigma)] b_i(\omega) + \sigma^3 \frac{\rho}{\tau^2} \text{Var}(Y_\omega) \cdot (c_i(\omega) - b_i \cdot c_M(\omega))$$

We denote $r_M^{(2)}(\sigma)$ the component of $r_M^2(\sigma)$ orthogonal to $r_M(\sigma)$ and $Y^{(2)}_\omega$ the component of $Y^2_\omega$ orthogonal to $Y_\omega$. We have $r_M^{(2)}(\sigma) = \sigma^2 Y^{(2)}_\omega$, i.e.

$$\sigma^3 \cdot \text{Var}(Y_\omega) \cdot (c_i(\omega) - b_i \cdot c_M(\omega)) = \sigma^3 \cdot \text{Cov} \left( Y, Y^{(2)}_\omega \right) = \text{Cov} \left( r_i(\sigma), r_M^{(2)}(\sigma) \right).$$

The theorem tells us that

$$E[r_i(\sigma)] = \frac{n}{S} E[r_M(\sigma)] b_i(\omega) + \sigma^3 \frac{\rho}{\tau^2} \text{Cov} \left( r_i(\sigma), r_M^{(2)}(\sigma) \right).$$

We will see later that similar terms appear in $\beta$-$\gamma$ relationships in the following section. The second part of theorem relates directly to the orthogonal decomposition there.

### 3 Nonlinear Pricing Kernels

The pricing implications of a SDF formula that is quadratic with respect to the market return are studied in this section, first with a linear beta pricing point of view and second in terms of derivative pricing.
3.1 Beta Pricing

In their paper about conditional skewness in asset pricing tests, Harvey and Siddique (2000) start with the maintained assumption that the SDF is quadratic in the market return:

\[ m_{t+1} = \nu_0 t + \nu_1 t R_{Mt+1} + \nu_2 t R^2_{Mt+1}. \]  

It actually suffices to revisit our section 2 above with a conditional viewpoint to see theorem 2.9 as a theoretical justification of (13). Then, the coefficients \( \nu_0 t, \nu_1 t \) and \( \nu_2 t \) are functions of the conditioning information \( I_t \) at time \( t \).

From theorem 2.9, we interpret the factors coefficients as:

\[ \nu_2 t = \frac{1}{R_{ft}} \left( \frac{n}{S} \right)^2 \frac{\bar{\rho}}{\bar{\tau}^2} > 0, \]  

and

\[ \nu_1 t = -\frac{n}{R_{ft} S \bar{\tau}} - 2 \frac{1}{R_{ft}} \left( \frac{n}{S} \right)^2 \frac{\bar{\rho}}{\bar{\tau}^2} E_t (R_{Mt+1}) < 0. \]

It is worth characterizing the role of the two factors \( R_{Mt+1} \) and \( R^2_{Mt+1} \) in the SDF (13) in terms of beta pricing relationships. Assuming the existence of a conditionally risk-free asset (with return \( R_{ft} \)), we denote the net excess return of every asset \( i = 1, \ldots, n \). We have

\[ \frac{1}{R_{ft}} E_t [r_{it+1}] + \nu_1 t Cov_t [r_{it+1}, R_{Mt+1}] + \nu_2 t Cov_t [r_{it+1}, R^2_{Mt+1}] = E_t [r_{it+1} m_{t+1}] = 0, \]

or, using the market net excess return, we get:

\[ \frac{1}{R_{ft}} E_t [r_{it+1}] + (\nu_1 t + 2R_{ft} \nu_2 t) Cov_t [r_{it+1}, R_{Mt+1}] + \nu_2 t Cov_t [r_{it+1}, R^2_{Mt+1}] = 0, \]

that is:

\[ E_t [r_{it+1}] = \lambda_1 t Cov_t (r_{it+1}, R_{Mt+1}) - \lambda_2 t Cov_t (r_{it+1}, R^2_{Mt+1}), \]

with

\[ \lambda_1 t = -R_{ft} (\nu_1 t + 2R_{ft} \nu_2 t) \] and \( \lambda_2 t = R_{ft} \nu_2 t. \)

If \( \nu_1 t \) and \( \nu_2 t \) are interpreted in terms of preferences of an average investor as in (14) and (15), we deduce:

\[ \lambda_1 t = \frac{n}{S \bar{\tau}} + \frac{2n^2 \bar{\rho}}{S^2 \bar{\tau}^2} \left( E_t [R_{Mt+1}] - R_{ft} \right) \] and \( \lambda_2 t = \frac{n^2 \bar{\rho}}{S^2 \bar{\tau}^2}. \)
Note that $\lambda_{2t}$ is something like a structural invariant, only time varying through the value of preference parameters computed from the derivatives of the utility function at $R_{ft}$. The term $\lambda_{2t}$ should be non-negative and all the more positive that preference for skewness is high. Similarly, $\lambda_{1t}$ is expected to be positive and time varying insofar as the market risk premium $E_t[R_{Mt+1}] - R_{ft}$ is. To summarize:

**Theorem 3.1** Under the maintained assumption (13) of a quadratic SDF, net expected returns are given by:

$$E_t[r_{it+1}] = \lambda_{1t} \text{Cov}_t[r_{it+1}, r_{Mt+1}] - \lambda_{2t} \text{Cov}_t[r_{it+1}, r_{2Mt+1}].$$

If in addition, theorem 2.9 applies, $\lambda_{1t}$ and $\lambda_{2t}$ are non-negative. $\lambda_{2t} = \frac{\mu_2}{\mu_2 - \mu_1}$ is determined by average preferences for skewness while:

$$\lambda_{1t} = \frac{1}{\mu_2} + 2\lambda_{2t} E_t[r_{Mt+1}].$$

Note that $\lambda_{1t}$ has two components which are both increasing with the average risk aversion, first as $1/\mu_2$ and second as the market risk premium $E_t[r_{Mt+1}]$. When applying theorem 3.1 to the market return itself ($r_{it+1} = r_{Mt+1}$), we get even more insight on what makes $\lambda_{1t}$ large:

**Corollary 3.2** Under the assumptions of theorem 3.1

$$\lambda_{1t} = \frac{E_t[r_{Mt+1}]}{\text{Var}_t(r_{Mt+1})} + \lambda_{2t} \frac{\text{Skew}_t(r_{Mt+1})}{\text{Var}_t(r_{Mt+1})},$$

where $\text{Skew}_t(r_{Mt+1}) = \text{Cov}_t(r_{Mt+1}, r_{2Mt+1})$.

In particular, we can see that theorem 3.1 coincides with the standard Sharpe-Lintner CAPM formula when $\lambda_{2t} = 0$, that is the average preference for skewness is zero. By contrast, $\lambda_{1t}$ is augmented in the general case by an additive term which is proportional to both $\lambda_{2t}$ and

$$\text{Skew}_t(r_{Mt+1}) = \text{Cov}_t(r_{Mt+1}, r_{2Mt+1}) = E_t[r_{3Mt+1}^3] - E_t[r_{Mt+1}] \cdot E_t[r_{2Mt+1}^2].$$

This notion of market co-skewness has already been put forward by Harvey and Siddique (2000) and theorem 3.1 and corollary 3.2 correspond to their formulas (7). We define

$$\beta_{iMt} = \frac{\text{Cov}_t[r_{it+1}, r_{Mt+1}]}{\text{Var}_t(r_{Mt+1})}, \gamma_{iMt} = \frac{\text{Cov}_t[r_{it+1}, r_{2Mt+1}]}{\text{Var}_t(r_{2Mt+1})}.$$ 

It is worth rewriting the pricing relationship of theorem 3.1 and corollary 3.2 in term of betas:

$$E_t[r_{it+1}] = (\lambda_{1t} \text{Var}_t(r_{Mt+1})) \beta_{iMt} - (\lambda_{2t} \text{Var}_t(r_{2Mt+1})) \gamma_{iMt},$$

$$E_t[r_{Mt+1}] \beta_{iMt} - \lambda_{2t} \text{Var}_t(r_{2Mt+1}) (\gamma_{iMt} - \gamma_{MMt} \beta_{iMt}).$$
The term $\beta_{iMt}$ is the standard market beta while the beta coefficient with respect to the squared market return is $\gamma_{iMt}$; it is tightly related to the measure of co-skewness already introduced in section 2.

**Theorem 3.3** We denote $r_{Mt}^{(2)}$ the component of $r_{Mt}^2$ orthogonal to $r_{Mt}(\sigma)$. Then

$$\text{Var}_t(r_{Mt+1}^2) \cdot (\gamma_{iMt} - \gamma_{MMt} \beta_{iMt}) = \text{Cov}_t(r_t, r_{Mt}^{(2)}(\sigma)).$$

Therefore the result of equation (17) matches exactly that of theorem 2.10 with a conditional viewpoint.

While we had already seen in theorem 2.8 that risk premiums in equilibrium where influenced by skewness preferences in proportion of the vector $c(\bar{\omega})$ of market co-skewness coefficients, the same vector shows up in the beta pricing relationship here. Note that what Harvey and Siddique (2000) call “market co-skewness” is actually $\text{Skew}_t(r_{Mt+1}) = \gamma_{MMt} \text{Var}_t(r_{Mt+1}^2)$.

The beta pricing model (16) with a second beta coefficient interpreted in terms of co-skewness with the market is observationally equivalent to a conditional version of the three-moments CAPM first proposed by Kraus and Litzenberger (1976) (see also Ingersoll (1987), p. 100). While they put forward

$$\frac{\text{Cov}_t(R_{it+1}, (R_{it+1} - E[R_{it+1}])^2)}{\text{Cov}_t(R_{Mt+1}, (R_{Mt+1} - E[R_{Mt+1}])^2)}$$

as a measure of co-skewness, we have preferred to remain true to a genuine notion of beta coefficient as $\gamma_{iMt}$.

However, the difference between the two is just a matter of normalization and is immaterial in terms of asset pricing. In particular (17) enhances as formula (64) in Ingersoll (1987) that the beta pricing relationship differs from Sharpe-Lintner CAPM by a factor proportional to the difference between the two betas. It is however worth noticing that these authors derive this pricing relationship by using a utility function directly defined over mean, standard deviation and skewness. The small noise expansion approach of section 2 affords more theoretical underpinnings for doing so.

Normalization in terms of beta coefficient is usually convenient since it allows a direct interpretation of beta loadings in terms of factor risk premium. For instance, when $\lambda_{2t} = 0$ equation (16) applied to the market gives the usual formula: $\lambda_{1t} = P_{Mt}^{(1)}$ with

$$P_{Mt}^{(1)} = \frac{E_t[r_{Mt+1}]}{\text{Var}_t(r_{Mt+1})}.$$

However, in general $\lambda_{1t}$ and $\lambda_{2t}$ cannot be read as simple risk premium associated respectively to the two payoffs $r_{Mt+1}$ and $r_{Mt+1}^2$. Even if we assume that $r_{Mt+1}^2$ does correspond to a payoff of a portfolio available
in the market with price $\eta_t$, the risk premium on such a payoff:

$$P_{Mt}^{(2)}(\eta_t) = \frac{E_t[r_{Mt+1}^2]}{\text{Var}_t(r_{Mt+1}^2)} - R_{ft} = \frac{E_t[r_{Mt+1}^2] - R_{ft}\eta_t}{\text{Var}_t(r_{Mt+1}^2)}$$

(18)

will not coincide with $(-\lambda_{2t}\eta_t)$. The difference comes from the fact that the two factors are not orthogonal: $\lambda_{1t}$ does depend on $\lambda_{2t}$ (see corollary 3.2) and the expression of $\lambda_{2t}$ in function of the equilibrium prices is more involved:

**Theorem 3.4** If $\eta_t = E_t[m_{t+1}r_{Mt+1}^2]$ denotes the equilibrium price of a payoff $r_{Mt+1}^2$, we have:

$$\lambda_{2t} = \frac{\gamma_{MMt}F_{Mt}^{(1)} - \frac{1}{\eta_t}P_{Mt}^{(2)}(\eta_t)}{1 - \rho_{t}^2(r_{Mt+1}, r_{Mt+1}^2)},$$

where according to (18), $P_{Mt}^{(2)}(\eta_t)$ is the risk premium on the asset with payoff $r_{Mt+1}^2$ and $\rho_{t}^2(r_{Mt+1}, r_{Mt+1}^2)$ denotes the square (conditional) linear correlation coefficient between $r_{Mt+1}$ and $r_{Mt+1}^2$.

It is worth considering the limit case where $r_{Mt+1}^2$ is worthless. From (18) we derive directly:

$$\lim_{\eta_t \to 0} \frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t} = \frac{E_t[r_{Mt+1}^2]}{\text{Var}_t(r_{Mt+1}^2)}.$$  

(19)

In this limit case, one gets:

$$\lambda_{2t} = \frac{\gamma_{MMt}F_{Mt}^{(1)} - \frac{1}{\eta_t}E_t[r_{Mt+1}^2]}{1 - \rho_{t}^2(r_{Mt+1}, r_{Mt+1}^2)}$$

(20)

which actually coincides with the formula put forward by Harvey and Siddique (2000). However, this limit case appears to be at odds with a no-arbitrage condition since $\eta_t = E_t[m_{t+1}r_{Mt+1}^2]$ should be positive.

Besides, whether $\eta_t$ is significantly positive is of course an empirical question. Since from (18)

$$\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t} = \frac{E_t[r_{Mt+1}^2] - R_{ft}\eta_t}{\text{Var}_t(r_{Mt+1}^2)},$$

(21)

one may expect that considering the limit case (19), that is $\eta_t = 0$ leads to overestimate $\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t}$ and then to underestimate $\lambda_{2t}$ (see Theorem 3.4).

The fact that risk neutral variance is significantly positive is of course an empirical question. The relevant empirical issue (see section 4) is then to decide if considering only the limit case (20) leads to an economically significant underestimation of the weight $\lambda_{2t}$ of cokewness in the two factors pricing relationship (17). If it is the case, we must realize that $\lambda_{2t}$ actually depends on investors preferences for skewness as they show up either in the (market) price of squared market return or, equivalently, in the risk neutral variance of the market return.
3.2 Risk-neutral Variance and the Pricing of Asymmetry Risk

The huge expansion of derivatives markets, introducing asset payoffs which are nonlinear and often skewed functions of underlying primitive asset returns, has renewed interest in asset payoff’s skewness. For notational simplicity we consider in this subsection only options written on the market return; however, most of the results could be extended to other primitive assets.

Let us then derive the price for a contingent payoff \( h_t (R_{Mt+1}) \) of the market return. (This may depend on conditioning information.) Throughout this section we assume that a valid SDF is quadratic; this means that the price of the payoff \( h_t (R_{Mt+1}) \) coincides with the price of its (conditional) affine regression

\[
E_t [m_{t+1} h_t (R_{Mt+1})] = E_t \left[ m_{t+1} h_t^{(L2)} (R_{Mt+1}) \right].
\]

Understanding why a CAPM Sharpe-Lintner pricing does not accommodate well the pricing of derivatives is akin to show why the (conditional) affine regression \( h_t^{(L1)} (R_{Mt+1}) = EL_t [h_t (R_{Mt+1}) | R_{Mt+1}] \) of \( h_t (R_{Mt+1}) \) on \( R_{Mt+1} \) does not summarize the compensation for risk, i.e.

\[
E_t [m_{t+1} h_t (R_{Mt+1})] \neq E_t \left[ m_{t+1} h_t^{(L1)} (R_{Mt+1}) \right].
\]

One of the simplest nonlinear payoffs is \( h_t (R_{Mt+1}) = R^2_{Mt+1} \). This so-called “volatility contract” has been introduced by Bakshi, Kapadia, and Madan (2001) for derivatives pricing. We are then lead to study the difference between its price and and the price of its linear approximation \( E_t [m_{t+1} EL_t [R^2_{Mt+1} | R_{Mt+1}]] \), see below Theorem 3.6.

To enhance the understanding of skewness in this context it is first worth noting that:

\textbf{Lemma 3.5} The conditional linear regression of \( R^2_{Mt+1} \) on \( R_{Mt+1} \) is:

\[
E_t \left[ R^2_{Mt+1} | R_{Mt+1} \right] = E_t \left[ R^2_{Mt+1} \right] + 2 E_t \left[ R_{Mt+1} \right] \cdot \frac{E_t \left[ (R_{Mt+1} - E_t R_{Mt+1})^3 \right]}{\text{Var}(R_{Mt+1})} \cdot (R_{Mt+1} - E_t R_{Mt+1}) .
\]

The result of this Lemma needs to be contrasted with the Taylor expansion of \( R^2_{Mt+1} \) arround \( E_t[R_{Mt+1}] \):

\[
R^2_{Mt+1} \approx E_t[R^2_{Mt+1}] + 2 E_t[R_{Mt+1}] \cdot (R_{Mt+1} - E_t[R_{Mt+1}]) .
\]
This does not take into account the crucial role of the skewness term and casts some doubts on theories of higher moments pricing which are based on Taylor expansions. In that respect, the small noise expansion appears to be more reliable.

The price of the volatility contract can be related to risk neutral pricing popular in derivative pricing. We have:

$$\pi_t = E_t \left[ m_{t+1} R^2_{M_{t+1}} - 1 \right] R_{ft} E_t^* \left[ R^2_{M_{t+1}} \right],$$

where $E_t^*$ denotes the conditional expectation with respect to the risk neutral probability measure defined by $m$. By definition, $E_t^* \left[ R_{M_{t+1}} \right] = R_{ft}$, so that

$$\text{Var}^* (R_{M_{t+1}}) = E_t^* \left[ (R_{M_{t+1}} - R_{ft})^2 \right] = R_{ft} \cdot (\pi_t - R_{ft})$$

can be interpreted as a risk neutral variance (see Rosenberg (2000)). Since $r_{M_{t+1}} = R_{M_{t+1}} - R_{ft}$ note also that

$$\text{Var}^* (r_{M_{t+1}}) = \text{Var}^* (r^2_{M_{t+1}}) = E_t^* [r^2_{M_{t+1}}] = R_{ft} E_t [m \cdot r^2_{M_{t+1}}] = R_{ft} \cdot \eta_t.$$

Using the risk-neutral variance allows us to shed more light on the previous subsection for the limit case where $r^2_{M_{t+1}}$ was supposed to be worthless. From (18):

$$P^{(2)}_{M_{t}}(\eta_t) = \frac{E_t^* [r^2_{M_{t+1}}] - \text{Var}^* (r_{M_{t+1}})}{\text{Var}^* (r^2_{M_{t+1}})}.$$ (22)

As in the previous subsection one may expect that considering the limit case (19) with $\eta_t = 0$, i.e. with $\text{Var}^* (R_{M_{t+1}}) = 0$ leads to overestimate $P^{(2)}_{M_{t}}(\eta_t)$ and then to underestimate $\lambda_{2t}$ (see Theorem 3.4). As in the previous subsection we point out that the relevant empirical issue is to decide if considering only the limit case (20) leads to an economically significant underestimation of the weight $\lambda_{2t}$ of co-skewness in the two factors pricing relationship (17). This will be pursued in section 4.

We saw in the discussion after theorem 3.4, that the higher the price $\eta_t$ of $r^2_{M_{t+1}}$, the smaller will be the premium $P^{(2)}_{M_{t}}$ and then larger will be price $\lambda_{2t}$ of co-skewness. Therefore, one way to assess the strength of preference for skewness is to describe the factors which tend to increase ceteris paribus the risk neutral variance $\text{Var}^* (r_{M_{t+1}}) = R_{ft} \eta_t$. For doing so, we first state a useful relation between risk neutral variance $\text{Var}^* (r_{M_{t+1}})$ and the historical one $\text{Var} (r_{M_{t+1}})$:
Theorem 3.6

\[ \text{Var}^*(r_{Mt+1}) = \text{Var}(r_{Mt+1}) \cdot \left(1 - \text{Var}(r_{Mt+1}) \cdot \left(\frac{P(1)}{P_{Mt}}\right)^2\right) - P(1) E_t \left[(R_{Mt+1} - E_t R_{Mt+1})^3\right] + R_{ft} \text{Cov}_t(m_{t+1}, \varepsilon_{t+1}) , \]

where \( \varepsilon_{t+1} = R^2_{Mt+1} - E_t R^2_{Mt+1} | R_{Mt+1} \) denotes the residual of the (conditional) affine regression of \( R^2_{Mt+1} \) on \( R_{Mt+1} \).

Note that this theorem is valid under the very general assumption that a positive SDF \( m_{t+1} \) is able to price the asset of interest and in particular to define risk neutral conditional expectations as \( \frac{1}{R_{ft}} E_t^* h_t (R_{Mt+1}) = E_t [m_{t+1} h_t (R_{Mt+1})] \).

It is then worth revisiting skewness pricing by studying the factors which may potentially increase the risk neutral variance. Theorem 3.6 basically puts forward two factors. One factor is model dependent, through \( \text{Cov}_t (m_{t+1}, \varepsilon_{t+1}) \) while the other terms can be directly observed from the market return. Typically, in case of a positive return skewness (\( E_t \left[(R_{Mt+1} - E_t R_{Mt+1})^3\right] > 0 \)), risk neutral variance is inversely related to risk premium \( P(1) \). Intuitively, high risk neutral variance, that is high compensation for skewness, may compensate a low risk premium \( P(1) \). By contrast, the effect encapsulated in \( \text{Cov}_t (m_{t+1}, \varepsilon_{t+1}) \) depends in general explicitly on the SDF specification, that is on the investor preferences. There is however a case where the risk neutral variance is preference free, in the sense that it is completely determined by the observation of the risk-free interest rate and the market risk premium. This is the case of joint log normality which is an extension (see Garcia, Ghysels, and Renault (2003)) of the risk neutral valuation relationships first introduced by Brennan (1979):

**Theorem 3.7** If \((\log m_{t+1}, \log R_{Mt+1})\) is jointly normal given the conditioning information,

\[ \text{Var}^*(R_{Mt+1}) = \text{Var}(R_{Mt+1}) \cdot \left[\frac{R_{ft}}{E_t[R_{Mt+1}]}\right]^2 < \text{Var}(R_{Mt+1}). \]

This theorem confirms for a particular case the above discussion of the difference between risk neutral variance and historical one. While we expect the former to be smaller than the latter in case of positive skewness, the difference between the two is inversely related to the market risk premium.

In the general case, the role of investor preference for skewness in increasing the risk neutral variance can be characterized from the following result:
**Theorem 3.8** With a quadratic SDF,

\[ m_{t+1} = \nu_0 + \nu_1 R_{Mt+1} + \nu_2 R^2_{Mt+1} \]

the term \( \text{Cov}_t (m_{t+1}, \varepsilon_{t+1}) \) is given by:

\[ \text{Cov}_t (m_{t+1}, \varepsilon_{t+1}) = \nu_2 \text{Var}_t R^2_{Mt+1} \cdot \left( 1 - \rho_t^2 \left( R_{Mt+1}, R^2_{Mt+1} \right) \right) . \]

Therefore, we do expect that this term increases the risk neutral variance, all the more that \( R_{Mt+1} \) and \( R^2_{Mt+1} \) are weakly correlated and the average skewness tolerance \( \bar{\rho} = \nu_2 R_{ft} S^2 \bar{\tau}^2 \) is large. The main message of this subsection is that empirical assessments of risk neutral variance as recently proposed by Rosenberg (2000) from derivative asset prices may also be seen as a way to characterize preferences for skewness.

### 4 Empirical Illustration

#### 4.1 The General Issue

The empirical relevance of the asset pricing model with coskewness as developed in previous sections is encapsulated in the asset pricing equation (17):

\[ E_t [r_{Mt+1}] = E_t [r_{Mt+1}] \beta_{Mt} - \lambda_2 \text{Var}_t (r^2_{Mt+1}) (\gamma_{iMt} - \gamma_{M Mt} \beta_{Mt}) . \]

(23)

The question is: does this asset pricing equation significantly deviate from standard CAPM?, that is should we maintain a significantly positive skewness premium \( \lambda_2 t \)?

It turns out that the statistical identification of this hypothesis is difficult since, as well noticed by Barone-Adesi, Urga, and Gagliardini (2004), covariance and coskewness with market tend to be almost collinear across common portfolios, leading to hardly significant coskewness factors \( (\gamma_{iMt} - \gamma_{M Mt} \beta_{iMt}) \). To shed more light on this identification issue, let us consider the (conditional) affine regression of of asset \( i \)'s net return on market return \( r_{Mt+1} \):

\[ r_{it+1} = \alpha_{it} + \beta_{Mt} r_{Mt+1} + u_{it+1} . \]

(24)

It is clear that asset \( i \)'s coskewness can be interpreted as the covariance between the residual of this regression with squared market return:

\[ \text{Var}_t (r^2_{Mt+1}) \cdot (\gamma_{iMt} - \gamma_{M Mt} \beta_{iMt}) = \text{Cov}_t (u_{it+1}, r^2_{Mt+1}) = \text{Cov}_t (u_{it+1}, R^2_{Mt+1}) . \]

(25)
Therefore, a positive sign for $\lambda_{2t}$ can be identified only insofar as one can observe some asset returns $r_{it+1}$ with positive (negative) coskewness $\text{Cov}_t \left(u_{it+1}, r_{Mt+1}^2\right)$ and check that they command a lower (higher) expected return than explained by standard CAPM. The problem is that $\text{Cov}_t \left(u_{it+1}, r_{Mt+1}^2\right)$ will be more often than not close to zero since $u_{it+1}$ is by definition (conditionally) uncorrelated with $r_{Mt+1}$. Of course non correlation does not imply independence (except in linear market models like the Gaussian one) and one may hope that some asset $i$ exhibits a significantly positive (or negative) covariance $\text{Cov}_t \left(u_{it+1}, r_{Mt+1}^2\right)$. However, as long as $\text{Cov}_t \left(u_{it+1}, r_{Mt+1}^2\right)$ is almost zero equation (24) leads to:

$$\text{Cov}_t \left(r_{it+1}, r_{Mt+1}^2\right) \sim \beta_{iMt} \text{Cov}_t \left(r_{Mt+1}, r_{Mt+1}^2\right)$$

almost collinear with $\beta_{iMt}$ across portfolios.

To avoid such a perverse linearity effect, Barone-Adesi, Urga, and Gagliardini (2004) focus on a quadratic market model first introduced by Barone-Adesi (1985). Thanks to this specification, they estimate a slightly significantly positive coefficient $\lambda_{2t}$, at least when the risk free rate is a free parameter, not assumed to be observed by the econometrician. However, their approach is unconditional and this may explain the difficulty to identify the sign of $\lambda_{2t}$, in particular with respect to the risk free rate issue.

To remedy that, we propose here to consider instead an asymmetric GARCH in mean model recently estimated by Bekaert and Liu (2004). Since this model exhibits interesting time-varying non-linearities in the consumption process, it may allow an accurate identification of time varying conditional coskewness and in turn consumption-based preference for coskewness. The superior identification power of such a conditional approach will actually be confirmed below through a series of Monte Carlo simulations based on Bekaert and Liu (2004) parameters estimates.

4.2 The Simulation Setup

[Table 1 about here.]

Bekaert and Liu (2004) estimate a GARCH factor model with in mean effects for the trivariate process of logarithm $X_{t+1}$ of consumption growth, logarithm of market return $\log (R_{Mt+1})$ and logarithm of bond return $\log (R_{ft+1})$. We denote

$$Y_{t+1} = [Y_{1t+1}, Y_{2t+1}, Y_{3t+1}]' = [X_{t+1}, \log (R_{Mt+1}), \log (R_{ft+1})]' .$$
The model assumes the dynamics

$$Y_{t+1} = c_t + AY_t + \Omega e_{t+1}, \quad (26)$$

where the coefficient $c_i$ of $c_t$, $i = 1, 2, 3$ is an affine function of $Var_t [Y_{t+1}]$ and all the time variation in volatility is driven by time varying uncertainty in consumption growth: the conditional probability distribution of $e_{t+1}$ given information $I_t$ is normal with zero mean and a diagonal covariance matrix, the coefficients of which are constant except the first one which follows an asymmetric GARCH(1,1):

$$Var_t [e_{1t+1}] = \delta_1 + \alpha (e_{1t})^2 + \beta \Var_t [e_{1t}] + \xi (\Max [0, -e_{1t}])^2. \quad (27)$$

To limit parameter proliferation, they assume that all the off-diagonal coefficients of the matrix $\Omega$ are zero except in the first column; in other words the consumption shock is the only factor. For sake of normalization, the diagonal coefficients of $\Omega$ are fixed to the value 1. Table 1 gives the parameters estimates provided by Bekaert and Liu (2004) on monthly US data. These estimates will be considered below as true population values for simulating a sample path.

A convenient feature of the above model for our purpose is that, since it maintains a conditional joint normality assumption for log-consumption and log-market return, it allows us to apply theorem 3.7 to assess the risk neutral variance without need of a preference specification. More precisely, insofar as the log-pricing kernel is, given $I_t$, a linear combination of the first two components of $Y_{t+1}$, as it is not only in the Lucas (1978) consumption-based CAPM with isoelastic preferences but also more generally in the Epstein and Zin (1991) recursive utility model, we are sure that theorem 3.7 applies.

Then, our simulation set-up is as follows: For a given simulated path of the process $(Y_{t+1})$, specifications (26) and (27) allow us to compute iteratively corresponding paths of $\Var^* (r_{Mt+1}) = \Var (r_{Mt+1}) \cdot (R_{ft}/E_t [R_{Mt+1}])^2$, $\eta_t = \Var^* (r_{Mt+1}) / R_{ft}$, $P_{Mt}^{(2)}(\eta_t) = \frac{E_{t} r_{Mt+1}^2 - \Var^* (r_{Mt+1})}{\Var(r_{Mt+1})}$, $P_{Mt}^{(2)}$, and finally $\lambda_{2t}$ according to theorem 3.4. We recall that the limit case put forward by Harvey and Siddique (2000) corresponds to the alternative formula:

$$\lambda_{2t}^{HS} = \gamma_{MM} P_{Mt}^{(1)} - \frac{E_{t} [r_{Mt+1}]}{\Var(r_{Mt+1})},$$

the path of which is also easy to build from the above simulation.

Of course, by introducing only one risky asset, this setting does not allow us to compare coskewness across portfolios. However, the focus of our interest here is to get time series of $\lambda_{2t}$ and $\lambda_{2t}^{HS}$ in order to assess their
sign and their differences both date by date and in average. Note moreover that return skewness in this market is not as trivial as log-normality may lead you to think. Over two periods, temporally aggregated asset returns will feature some sophisticated skewness, first due to the asymmetric effect in the variance dynamics and second due to time varying risk premium. A detailed characterization of induced dynamic skewness pricing is beyond the scope of this paper.

4.3 Monte Carlo Results

[Table 2 about here.]

[Figure 1 about here.]

All the simulated paths considered correspond to 500 months. The main message conveyed by these simulated series is well summarized by figure 1 where we plot on the same graph both the path of $\lambda_{2t}$ corresponding to our formula for the price of coskewness and of $\lambda^{HS}_{2t}$ corresponding to Harvey and Siddique (2000) limit case.

The conclusions draw from this graph are twofold: First, while the series of $\lambda_{2t}$ does show a positive price for coskewness as expected (4.25 in average), the series $\lambda^{HS}_{2t}$ displays some implausible huge negative price of coskewness ($-67.82$ in average). This tends to prove that neglecting the price $\eta_t$ of squared net returns (or equivalently the risk neutral variance) leads to a severe underestimation of coskewness price, so severe that it may reverse the direction of the effect of coskewness in asset prices.

[Figure 2 about here.]

The time series of $\eta_t$ (figure 2) and risk premium $P^{(2)}_{Mt}(\eta_t)$ (figure 3) as well confirm that they are positive. Note also that while $\lambda_{2t}$ and $\lambda^{HS}_{2t}$ are stationary processes — in particular first order differences ($\lambda_{2t} - \lambda_{2t-1}$) and ($\lambda^{HS}_{2t} - \lambda^{HS}_{2t-1}$) have a zero time average — the former is more stable than the latter: the standard error of the series ($\lambda_{2t} - \lambda_{2t-1}$) is only 4.93 while it is 8.75 for ($\lambda^{HS}_{2t} - \lambda^{HS}_{2t-1}$). This give some support to our interpretation of $\lambda_{2t}$ as a kind of preference-based structural invariant, which is time varying only through the value of utility derivatives at point $R_{ft}$.

Second, our simulations confirm that the positive sign of the price for coskewness should be hardly identifiable in an unconditional setting. While the series $\lambda_{2t}$ does show a positive average price of 4.25 for coskewness, it comes with a standard error of 4.06. This may explain why Barone-Adesi, Urga, and
Gagliardini (2004) were so much in trouble to identify a positive price in an unconditional setting. They actually get a t-statistic of 1.01, which has the same order of magnitude as our informal assessment. Of course, a rigorous unconditional study should not be simply based on time averages. By contrast, figure 1 shows that spot values of the process series $\lambda_{2t}$ may cover the full interval between 0 and 20, making them likely significant for a number of dates. This enhances the important contribution of Harvey and Siddique (2000) who stress that coskewness spring must be addressed in a conditional setting. However, even an unconditional approach would not make the simplified price series $\lambda_{2t}^{HS}$ meaningful since their standard error is only 7.45, which does not compensate their negative average of $-67.82$.

[Figure 3 about here.]

Overall, we conclude that there should be a positive price for coskewness, but not so high and hardly identifiable in an unconditional setting. One way to interpret the limited level of this price is to realize that buying the squared net market return commands a positive risk premium (see figure 3) which, by theorem 3.4 leads to lower the price $\lambda_{2t}$. This does not mean that skewness is worthless but only that, by lemma 3.5, a part of its value is already captured by the linear pricing of squared return. In other words, a positive skewness implies a positive correlation between market return and squared market return, so that the two components of asset prices cannot be interpreted separately.

Finally, one ought to realize that quadratic pricing kernels cannot be more than local approximations of a true pricing kernel, for instance in the neighborhood of small risk as in section 2. In particular, while a representative agent with a convex utility function would imply that the pricing kernel is decreasing with respect to the market return, this cannot be the case on the full range of returns with a quadratic function. More precisely, a quadratic pricing kernel as characterized by (13), (14), and (15) with a positive coskewness price $\lambda_{2t}$ will become increasing when the market returns exceeds its conditional expectation by more than $(S/\rho)$. This kind of paradoxical increasing shape of pricing kernels for large levels of market return already showed up in the empirical evidence documented by Dittmar (2002). Of course, a negative $\lambda_{2t}$ as in the case of the zero-price $\eta_t$ approximation would produce an even weirder behavior with increasing pricing kernel for any value of the market return below its expectation.

[Figure 4 about here.]

[Figure 5 about here.]
As far as Dittmar’s paradox is concerned, it does not mean that one should give up nonlinear polynomial pricing kernels because their decreasing shape cannot be enforced on the whole range of possible market returns. One must only remember that polynomial approximations are local and ought to be used cautiously. For instance, it is clear that market information about risk neutral variance or equivalently about the price $\eta_t$ of squared net market return may be helpful for a better control of a quadratic pricing kernel on the range of interest. Since this information may be in practice backed out of observed derivative asset prices, it is worth checking how it works on simulated paths. Figure 4 displays the pricing kernel surface as well as its time average as a function of the net market return. This figure is obtained with our value of $\eta_t$ (time average of $6.4 \cdot 10^{-3}$) which determines the coefficients $\lambda_{1t}$ and $\lambda_{2t}$ of the pricing kernel by application of corollary 3.2 and theorem 3.4. No paradoxical behavior of the pricing kernel is observed in this figure: on the range of interest for the net market return, the pricing kernel is always decreasing. If now one increases the value of $\eta_t$, by fixing somewhat arbitrarily the price of the squared market return at the level 1.02, which in turns implies a time-varying $\eta_t$ (with a time average of $15.6 \cdot 10^{-3}$), one gets figure 5. Then, one may observe that, by contrast with figure 4, on the same range of values of the market return, the aforementioned increasing shape of the pricing kernel for large returns may show up.

5 Conclusion

This paper investigated the relevance of non linear pricing kernels both at the theoretical and empirical levels. We first showed that pricing kernels that are quadratic functions of the market return is a well-founded approximation of actual expected utility behavior that aims to characterize locally the demand for risky asset in the neighborhood of zero risk. Such quadratic pricing kernels disclose some pricing for skewness, but only through co-skewness with the market. Heterogeneous agents hold the market portfolio and the skewness portfolio; the latter being the “closests” thing to the squared market return is based on all third-order cross-moments. In other words, while taking heterogeneity of skewness preferences into account yields separation theorems where idiosyncratic skewness risk plays a role for asset demands, it remains true that idiosyncratic risk is not priced, both in terms of variance and skewness.

While statistical identification of positive skewness premium may be difficult since covariance and co-skewness tend to be almost collinear across common portfolios, we showed through simulated data calibrated
on actual estimation of a factor GARCH model of returns with in mean effect that a conditional set-up is much more informative to capture relevant nonlinearities in pricing kernels. Such non-linearities imply some level of risk neutral variance for the market which cannot be neglected. This observation lead us to a generalization of the Harvey and Siddique (2000) beta pricing model for skewness; by contrasts with theirs, our model considers the price of the squared market return as a free parameter whose actual value might be backed out from observed derivative asset prices.

Although conditional, our study is purely static in the sense that investors only maximize a one-period utility function. An intertemporal extension of this study is still work in progress; it will point out the role of various kinds of asymmetries in a dynamic setting. Typically, while only conditional skewness of asset returns shows up in the current paper, a multiperiod setting will also enhance the role of dynamic asymmetry, that is some instantaneous correlation between asset returns and their volatility process. Such an effect has been dubbed leverage effect by Black (1976) and specific leverage-based dynamic risk premia should be the result of non-myopic intertemporal optimization behavior of investors with preferences for skewness.

Appendix

Proof of theorems 2.2 and 2.4. We denote by $h_i$ a function with

$$h_i(\sigma) = u'(W(\sigma)) \cdot (\sigma a_i(\sigma) + \sigma Y_i) .$$

The solution $\omega(\sigma) = (\omega_i(\sigma))_{1 \leq i \leq n}$ of problem (2) determines a terminal wealth

$$W(\sigma) = R_f + \sum_{i=1}^{n} \omega_i(\sigma)(R_i(\sigma) - R_f) ,$$

according to the first-order conditions

$$0 = E \left[u'(W(\sigma))(R_i(\sigma) - R_f)\right] = E \left[h_i(\sigma)\right] . \quad (A.1)$$

This implies in particular that $E \left[\frac{dh_i}{d\sigma}(\sigma)\right] = 0$ and so $\lim_{\sigma \to 0^+} E \left[\frac{dh_i}{d\sigma}(\sigma)\right] = 0$. Writing out the last equality we get

$$\sum_{i=1}^{n} \omega_i(0) Cov(Y_i, Y_k) = -\frac{u'(R_f)}{u''(R_f)} a_k(0) .$$

Using the variance-covariance matrix $\Sigma$ of the vector $Y$ of random variables and the definition of the risk-tolerance in (6) we then get $\omega(0) = \Sigma^{-1} \tau \cdot a \cdot (0)$, which ends the proof of theorem 2.2.
To prove Theorem 2.4 we take second-order derivatives of equation (A.1) and get

$$\lim_{\sigma \to 0^+} E \left[ \frac{d^2 h_i}{d^2 \sigma} (\sigma) \right] = 0.$$ 

Writing this out and using definition (6) we get

$$\sum_{i=1}^n \omega_i'(0) \text{Cov}(Y_i, Y_k) = \frac{\rho}{\tau} \sum_{i=1}^n \omega_i^2(0) E[Y_i^2 Y_k] + 2\rho \sum_{i<j}^n \omega_i(0) \omega_j(0) E[Y_i Y_j Y_k] + \tau a_k'(0) \quad (A.2)$$

Therefore equation (A.2, A.3) reads

$$\omega'(0) = \tau \Sigma^{-1} \left[ c(\omega(0)) \rho \frac{1}{\tau^2} \text{Var}[\omega^+ R] + a'(0) \right]$$

Proof of theorems 2.5 and 2.8. Using the definitions of \(\tau_s, \rho_s\) of equation (2.2), demand equations (10) and the first market-clearing equation (11) we derive from the condition \(e_n = \sum_{s=1}^S \omega_s(0) = \sum_{s=1}^S \Sigma^{-1} \tau_s a(0)\) that

$$a(0) = \frac{1}{\tau} \Sigma \omega. \quad (A.4)$$

Using the results of Theorem 2.2 that \(\omega_s(0) = \Sigma^{-1} \tau_s a(0)\) this implies

$$\omega_s(0) = \frac{\tau_s}{\tau} \omega.$$ 

Looking at equation (8) we then check that \(c_k(\omega_s(0)) = c_k(\omega)\). Using equation (A.4) and the second market-clearing equation (11) we get from \(0 = \sum_{s=1}^S \omega_s'(0) = \sum_{s=1}^S \tau_s \Sigma^{-1} \left[ c(\omega) \rho S^2 (0) + a'(0) \right]\) that

$$a'(0) = -\bar{p} c(\omega) P^2(0).$$

Using the description of \(c(\omega)\) in equation (9) and using equation (A.4) we get

$$a'(0) = -\bar{p} a(0) \Sigma^{-1} \Gamma_k \Sigma^{-1} a(0) = -\bar{p} \frac{1}{\tau^2} (\omega^+ \Gamma_k \omega) \quad (A.5)$$

Using the term describing \(\omega_s'(0)\) in theorem 2.4 and equation (A.5) we find

$$\omega_s'(0) = \tau_s \Sigma^{-1} \left[ c(\omega_s(0)) \rho_s P^2(0) + a'(0) \right] = \frac{\tau_s}{\tau^2} [\rho_s - \bar{p}] P^2(0) \Sigma^{-1} c(\omega).$$

28
Proof of theorem 2.9. We are looking for a variable $m(\sigma)$ with the property that

$$0 = E[m(\sigma) \cdot (R_k(\sigma) - R_f)] = Cov(m(\sigma), R_k(\sigma)) + E[m(\sigma)] \cdot E[R_k(\sigma) - R_f],$$

i.e. that the zero- and first order risk premium of each asset $k$ fulfill

$$Cov(m(\sigma), R_k(\sigma)) = \frac{1}{R_f} \sigma^2 (a_k(0) + \sigma a_k'(0)).$$ \hspace{1cm} (A.6)

From equations (A.4, A.5) follows:

$$\sigma^2 a_k(0) = \frac{1}{\tau S} Cov \left( \sum_{i=1}^{n} \sigma Y_i, R_k(\sigma) \right) = Cov \left( \frac{1}{\tau} \sigma \omega, R_k(\sigma) \right),$$ \hspace{1cm} (A.7)

$$\sigma^2 a_k'(0) = -Cov \left( \frac{\bar{\rho}}{(\tau S)^2} \left( \sum_{i=1}^{n} \sigma Y_i \right)^2, R_k(\sigma) \right) = -Cov \left( \frac{\bar{\rho}}{\tau^2} (\sigma Y_\omega)^2, R_k(\sigma) \right).$$ \hspace{1cm} (A.8)

Therefore,

$$-\frac{1}{R_f} \sigma^2 (a_k(0) + \sigma a_k'(0)) = Cov \left( -\frac{1}{R_f \bar{\tau}} \sigma \omega + \frac{\bar{\rho}}{R_f \bar{\tau}^2} (\sigma Y_\omega)^2, R_k(\sigma) \right),$$

and to identify this with $Cov(m(\sigma), R_k(\sigma))$ (equation (A.6)) it suffices to choose:

$$m(\sigma) = \xi - \frac{1}{R_f \bar{\tau}} \sigma Y_\omega + \frac{\bar{\rho}}{R_f \bar{\tau}^2} (\sigma Y_\omega)^2$$

for a suitable constant $\xi$. From $E[m] = \frac{1}{R_f}$ we get

$$\xi = \frac{1}{R_f} - \frac{\bar{\rho}}{R_f \bar{\tau}^2} E \left[ (\sigma Y_\omega)^2 \right],$$

i.e.

$$m = \frac{1}{R_f} - \frac{1}{R_f \bar{\tau}} \sigma Y_\omega + \frac{\bar{\rho}}{R_f \bar{\tau}^2} \left( (\sigma Y_\omega)^2 - E[(\sigma Y_\omega)^2] \right)$$

Using the relationship $\sigma Y_\omega = \frac{n}{2}(R_M(\sigma) - E[R_M(\sigma)])$ between $Y_\omega$ and the market return $R_M(\sigma)$ we can write

$$m = \frac{1}{R_f} - \frac{n}{R_f S \bar{\tau}} (R_M(\sigma) - E[R_M(\sigma)]) + \frac{\bar{\rho} m^2}{R_f S^2 \bar{\tau}^2} \left( (R_M(\sigma) - E[R_M(\sigma)])^2 - E[(R_M(\sigma) - E[R_M(\sigma)])]^2 \right).$$

Proof of theorem 2.10. Based on equations (A.7, A.8) we get

$$E[r_i(\sigma)] = E[R_i(\sigma)] - R_f = \sigma^2 a_k(0) + \sigma a_k'(0) = \sigma^2 \frac{1}{\tau} \text{Cov}(Y_\omega, R_k(\sigma)) + \sigma^2 \frac{\bar{\rho}}{\tau^2} \text{Cov}(Y_\omega^2, R_k(\sigma))$$

$$= \sigma^2 \frac{1}{\tau} \text{Cov}(Y_\omega, Y_k) + \sigma^3 \frac{\bar{\rho}}{\tau^2} \text{Cov}(Y_\omega^2, Y_k)$$

$$= \sigma^2 \frac{1}{\tau} \text{Var}(Y_\omega) b_i(\bar{\omega}) + \sigma^3 \frac{\bar{\rho}}{\tau^2} \text{Var}(Y_\omega)c_i(\bar{\omega})$$
Applying this to the market net return \( r_M(\sigma) = \frac{1}{n} \sum_{i=1}^{n} r_i(\sigma) \) we get, since \( \frac{1}{n} \sum_{i=1}^{n} b_i(\bar{\omega}) = \frac{S}{n} \) that

\[
E[r_M(\sigma)] = \sigma^2 \frac{1}{\tau} \text{Var}(Y_{\bar{\omega}}) \frac{S}{n} + \sigma^3 \frac{\bar{\xi}}{\tau} \text{Var}(Y_{\bar{\omega}}) c_M(\bar{\omega}) \frac{S}{n}
\]

and from this

\[
E[r_i(\sigma)] = \frac{n}{S} E[r_M(\sigma)] b_i + \sigma^3 \frac{\bar{\xi}}{\tau} \text{Var}(Y_{\bar{\omega}}) (c_i(\bar{\omega}) - b_i \cdot c_M(\bar{\omega}))
\]

Let us denote \( Y_{\bar{\omega}}^{(2)} = Y_{\bar{\omega}}^2 - EL[Y_{\bar{\omega}}^2 | Y_{\bar{\omega}}] \) the component of \( Y_{\bar{\omega}}^2 \) that is orthogonal to \( Y_{\bar{\omega}} \) and \( r_M^{(2)}(\sigma) = r_M^2(\sigma) - EL[r_M^2(\sigma) | r_M(\sigma)] \) the component of \( r_M^2(\sigma) \) that is orthogonal to \( r_M(\sigma) \). We have

\[
Y_{\bar{\omega}}^{(2)} = Y_{\bar{\omega}}^2 - \frac{\text{Cov}(Y_{\bar{\omega}}, Y_{\bar{\omega}}^2)}{\text{Var}(Y_{\bar{\omega}})} Y_{\bar{\omega}} \text{ and so } c_i(\bar{\omega}) - b_i \cdot c_M(\bar{\omega}) = \text{Cov} \left( Y_{\bar{\omega}}, Y_{\bar{\omega}}^{(2)} \right)
\]

Since \( r_M(\sigma) = 1 + \sigma Y_M + \sigma^2 a_M(\sigma) \) we get for a suitable constant \( \psi \) that \( r_M^2(\sigma) = \psi + (2\sigma^2 a_M(\sigma) + 2\sigma) Y_M + \sigma^2 Y_M^2 \) we see that the orthogonal component of \( r_M^2(\sigma) \) and \( \sigma^2 Y_M^2 \) coincide. This ends the proof. ■

**Proof of Theorem 3.4.** We apply (17) to the asset net return:

\[
r_{it+1} = \frac{r_{Mt+1}^2}{\eta_t} - R_{jt}.
\]

We get:

\[
E_t[r_{Mt+1}^2] - R_{jt} \eta_t = E_t [r_{Mt+1}^2] \frac{\text{Var} (r_{Mt+1}^2)}{\text{Var} (r_{Mt+1})} - \lambda_2 \frac{\text{Var} (r_{Mt+1}^2)}{\text{Var} (r_{Mt+1})} \left( 1 - \rho_t^2 (r_{Mt+1}, r_{Mt+1}^2) \right),
\]

that is:

\[
\frac{p_{Mt}^{(2)}}{\eta_t} = \gamma_{MM} \frac{F_{Mt}^{(1)}}{\eta_t} - \lambda_2 \left( 1 - \rho_t^2 (r_{Mt+1}, r_{Mt+1}^2) \right).
\]

This gives the announced value for \( \lambda_2 \). ■

**Proof of Lemma 3.5.** The conditional linear regression of \( R_{Mt+1}^2 \) on \( R_{Mt+1} \) is of the form

\[
EL_t \left[ R_{Mt+1}^2 | R_{Mt+1} \right] = E_t R_{Mt+1}^2 + a F_{1t+1},
\]

where

\[
F_{1t+1} = R_{Mt+1} - E_t R_{Mt+1}.
\]

The residual of the conditional linear regression of \( R_{Mt+1}^2 \) on \( R_{Mt+1} \), that is \( R_{Mt+1}^2 - EL_t \left[ R_{Mt+1}^2 | R_{Mt+1} \right] \), is orthogonal to \( F_{1t+1} \). Consequently,

\[
\text{Cov}_t \left( R_{Mt+1}^2 - EL_t \left[ R_{Mt+1}^2 | R_{Mt+1} \right], F_{1t+1} \right) = 0.
\]
Solving this equation gives

\[ a = \frac{\text{Cov}_t (R_{Mt+1}^2, R_{Mt+1})}{\text{Var}_t (R_{Mt+1})}. \]

Then,

\[ EL_t [R_{Mt+1}^2 | R_{Mt+1}] = \frac{\text{Cov}_t (R_{Mt+1}^2, R_{Mt+1})}{\text{Var}_t (R_{Mt+1})} (R_{Mt+1} - E_t R_{Mt+1}). \quad (A.9) \]

But,

\[
\text{Cov}_t (R_{Mt+1}^2, R_{Mt+1}) = \text{Cov}_t (R_{Mt+1} - E_t R_{Mt+1} + E_t R_{Mt+1})^2, R_{Mt+1} \]

\[
Cov_t ((R_{Mt+1} - E_t R_{Mt+1})^2 + 2 (R_{Mt+1} - E_t R_{Mt+1}) E_t R_{Mt+1}, R_{Mt+1}) \]

\[
Cov_t ((R_{Mt+1} - E_t R_{Mt+1})^2, R_{Mt+1}) + 2 E_t R_{Mt+1} \text{Var}_t (R_{Mt+1}) \]

\[
= E_t (R_{Mt+1} - E_t R_{Mt+1})^3 + 2 E_t R_{Mt+1} \text{Var}_t (R_{Mt+1}).
\]

Therefore,

\[
EL_t [R_{Mt+1}^2 | R_{Mt+1}] = E_t R_{Mt+1}^2 + \frac{(R_{Mt+1} - E_t R_{Mt+1}) E_t (R_{Mt+1} - E_t R_{Mt+1})^3}{\text{Var}_t (R_{Mt+1})} +
\]

\[
2 (R_{Mt+1} - E_t R_{Mt+1}) E_t R_{Mt+1}.
\]

This ends the proof. ■

**Proof of Theorem 3.6.** Let us first note that

\[ \text{Var}^* (R_{Mt+1}) = R_{ft} E_t \left[ m_{t+1} (R_{Mt+1} - R_{ft})^2 \right], \]

where

\[ (R_{Mt+1} - R_{ft})^2 = R_{Mt+1}^2 + R_{ft}^2 - 2 R_{Mt+1} R_{ft}. \quad (A.10) \]

But the squared market return can be rewritten as

\[
R_{Mt+1}^2 = EL_t [R_{Mt+1}^2 | R_{Mt+1}] + \varepsilon_{t+1}
\]

where,

\[ E_t \varepsilon_{t+1} = 0. \]

We replace this last expression into (A.10) and get

\[
(R_{Mt+1} - R_{ft})^2 = EL_t [R_{Mt+1}^2 | R_{Mt+1}] + \varepsilon_{t+1} + R_{ft}^2 - 2 R_{Mt+1} R_{ft}
\]

\[
= E_t R_{Mt+1}^2 + \frac{(R_{Mt+1} - E_t R_{Mt+1}) E_t (R_{Mt+1} - E_t R_{Mt+1})^3}{\text{Var}_t (R_{Mt+1})} +
\]

\[
2 (R_{Mt+1} - E_t R_{Mt+1}) E_t R_{Mt+1} + \varepsilon_{t+1} + R_{ft}^2 - 2 R_{Mt+1} R_{ft}.
\]
Therefore,

\[
\text{Var}^* (R_{Mt+1}) = E_t R_{Mt+1}^2 + \frac{R_{ft} \left( 1 - \frac{1}{R_{ft}} E_t R_{Mt+1} \right) E_t (R_{Mt+1} - E_t R_{Mt+1})^3}{\text{Var}_t (R_{Mt+1})} + 2R_{ft} \left( 1 - \frac{1}{R_{ft}} E_t R_{Mt+1} \right) E_t R_{Mt+1} + R_{ft} \text{Cov} (m_{t+1}, \varepsilon_{t+1}) - R_{ft}^2
\]

\[= E_t R_{Mt+1}^2 - P_{Mt}^{(1)} E_t (R_{Mt+1} - E_t R_{Mt+1})^3 + 2R_{ft} \left( 1 - \frac{1}{R_{ft}} E_t R_{Mt+1} \right) E_t R_{Mt+1} + \text{Cov} (m_{t+1}, \varepsilon_{t+1}) - R_{ft}^2
\]

\[= \text{Var}(r_{Mt+1}) \left( 1 - \left( P_{Mt}^{(1)} \right)^2 \text{Var}(r_{Mt+1}) \right) - P_{Mt}^{(1)} E_t (R_{Mt+1} - E_t R_{Mt+1})^3 + R_{ft} \text{Cov} (m_{t+1}, \varepsilon_{t+1})
\]

This ends the proof. □

**Proof of Theorem 3.7.** Assume that the joint process \((m_{t+1}, R_{t+1})\) is conditionally lognormal; let us denote here

\[
\left[ \begin{array}{c} \log (m_{t+1}) \\ \log R_{Mt+1} \end{array} \right] / I_t \sim N \left( \left[ \begin{array}{c} \mu_{mt} \\ \mu_{Mt} \end{array} \right], \left[ \begin{array}{cc} \sigma_t^2 & \sigma_m^{mt} \\ \sigma_m^{mt} & \sigma_{Mt}^2 \end{array} \right] \right).
\]

The market return risk neutral variance \(\text{Var}^* (R_{Mt+1}) = E_t[R_{Mt+1}^2] - R_{ft}^2 = R_{ft} E_t[m_{t+1}R_{Mt+1}^2] - R_{ft}^2\). We know that:

\[
\log (m_{t+1} R_{Mt+1}^2) = \log (m_{t+1}) + 2 \log (R_{Mt+1})
\]

Therefore,

\[
E_t m_{t+1} R_{Mt+1}^2 = \exp (\mu_{mt} + 2\mu_{Mt} + 0.5 \sigma_t^2 + 2 \sigma_{Mt}^2 + 2 \sigma_m^{mt})
\]

\[= \exp (-\mu_{mt} - 0.5 \sigma_t^2) \exp (2\mu_{Mt} + 2 \sigma_{Mt}^2) \exp (-2\mu_{Mt} - \sigma_{Mt}^2) \times \]

\[\left[ \exp (\mu_{mt} + \mu_{Mt} + 0.5 \sigma_t^2 + 0.5 \sigma_{Mt}^2 + \sigma_{mrt}) \right]^2.
\]

But

\[1 = E_t m_{Mt+1} = \exp (\mu_{mt} + \mu_{Mt} + 0.5 \sigma_t^2 + 0.5 \sigma_{Mt}^2 + \sigma_{mrt}) \]

and therefore,

\[E_t [m_{t+1} R_{Mt+1}^2] = \exp (-\mu_{mt} - 0.5 \sigma_t^2) \exp (2\mu_{Mt} + 2 \sigma_{Mt}^2) \exp (-2\mu_{Mt} - \sigma_{Mt}^2) = R_{ft} \frac{E_t[R_{Mt+1}^2]}{(E_t[R_{Mt+1}])^2}.
\]

Consequently,

\[
\text{Var}^* (R_{Mt+1}) = R_{ft}^2 \left( \frac{E_t R_{Mt+1}^2}{E_t[R_{Mt+1}]} \right)^2 - R_{ft}^2 = \text{Var}(R_{Mt+1}) \left( \frac{R_{ft}}{E_t[R_{Mt+1}]} \right)^2 < \text{Var}(R_{Mt+1}).
\]
Proof of Theorem 3.8. Assume that

\[ m_{t+1} = \nu_0 t + \nu_1 R_{Mt+1} + \nu_2 R^2_{Mt+1}. \]

Then,

\[ \text{Cov}_t (m_{t+1}, \varepsilon_{t+1}) = \nu_2 \text{Cov}_t (R^2_{Mt+1}, \varepsilon_{t+1}). \]

But

\[
\begin{align*}
\text{Cov}_t (R^2_{Mt+1}, \varepsilon_{t+1}) &= \text{Cov}_t \left( R^2_{Mt+1}, R^2_{Mt+1} - \frac{(R_{Mt+1} - E_t R_{Mt+1}) \text{Cov}_t (R^2_{Mt+1}, R_{Mt+1})}{\text{Var}_t (R_{Mt+1})} \right) \\
&= \text{Var}_t (R^2_{Mt+1}) - \frac{\text{Cov}_t^2 (R^2_{Mt+1}, R_{Mt+1})}{\text{Var}_t (R_{Mt+1})} \\
&= \text{Var}_t (R^2_{Mt+1}) \left[ 1 - \frac{\text{Cov}_t^2 (R^2_{Mt+1}, R_{Mt+1})}{\text{Var}_t (R^2_{Mt+1}) \text{Var}_t (R_{Mt+1})} \right] \\
&= \text{Var}_t (R^2_{Mt+1}) \left[ 1 - \rho_t^2 (R^2_{Mt+1}, R_{Mt+1}) \right].
\end{align*}
\]
References


Figure 1: Price of coskewness inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004). HS indicates the price of coskewness corresponding to Harvey and Siddique (2000); CLR indicates the price of coskewness corresponding to our formula.
Figure 2: Price of squared net return inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004).
Figure 3: Risk premium on the squared net return inferred from simulated data according to the Factor GARCH in mean estimated in Bekaert and Liu (2004)
Figure 4: Quadratic pricing kernel inferred from simulated data according to the Factor GARCH in mean of Bekaert and Liu (2004). Above we plot the pricing kernel $m_{t+1}$ as a function of $t+1$ and $r_{Mt+1}$. Below we plot the average pricing kernel $\sum_{t=1}^{T} \frac{1}{T} m_{t+1}$. 
Figure 5: Fixing the price of the squared market return at the level 1.02, which in turns implies a time varying $\eta_t$, we infer the quadratic pricing kernel for simulated data according to the Factor GARCH in mean of Bekaert and Liu (2004). Above we plot the pricing kernel $m_{t+1}$ as a function of $t + 1$ and $r_{Mt+1} = x$. Below we plot the average pricing kernel $\Sigma_{t=1}^{T} \frac{1}{T} m_{t+1}$.
Table 1: Parameter estimates for the GARCH-in-mean of Bekaert and Liu (2004), taken from their table 1. This table should be read similar to a matrix-vector multiplication: the values in first column are obtained by multiplying the first row, respectively the fifth row, with the entries in the corresponding row.

<table>
<thead>
<tr>
<th></th>
<th>$Y_{1t}$</th>
<th>$Y_{2t}$</th>
<th>$Y_{3t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1t+1}$</td>
<td>0.0030</td>
<td>0.361</td>
<td>-0.029</td>
</tr>
<tr>
<td>$Y_{2t+1}$</td>
<td>0.0056-162.65 · $\text{Var} (\varepsilon_{1t})^2$</td>
<td>-0.198</td>
<td>0.738</td>
</tr>
<tr>
<td>$Y_{3t+1}$</td>
<td>0.0188 - 58.02 · $\text{Var} (\varepsilon_{1t})^2$</td>
<td>-1.734</td>
<td>1.029</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Var}<em>t (\varepsilon</em>{1t+1})$</td>
<td>0.000019</td>
<td>-0.0265</td>
<td>0.0008</td>
</tr>
<tr>
<td>$\text{Var}<em>t (\varepsilon</em>{2t+1})$</td>
<td>0.000014</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\text{Var}<em>t (\varepsilon</em>{3t+1})$</td>
<td>0.006134</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma_{13}$</td>
<td>-0.0564</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>3.182</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
averages over time

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_t$</td>
<td>0.64%</td>
</tr>
<tr>
<td>$\lambda_{2t}$</td>
<td>4.25</td>
</tr>
<tr>
<td>$\lambda_{2t} - \lambda_{2t-1}$</td>
<td>0</td>
</tr>
<tr>
<td>$\lambda^{HS}_{2t}$</td>
<td>-67.82</td>
</tr>
<tr>
<td>$\lambda^{HS}<em>{2t} - \lambda^{HS}</em>{2t-1}$</td>
<td>0</td>
</tr>
</tbody>
</table>

standard deviations over time

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{2t}$</td>
<td>4.06</td>
</tr>
<tr>
<td>$\lambda^{HS}_{2t}$</td>
<td>7.45</td>
</tr>
<tr>
<td>$\lambda_{2t} - \lambda_{2t-1}$</td>
<td>4.93</td>
</tr>
<tr>
<td>$\lambda^{HS}<em>{2t} - \lambda^{HS}</em>{2t-1}$</td>
<td>8.75</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics for our Monte-Carlo simulation