Implications of Asymmetry Risk for Portfolio Analysis and Asset Pricing*

Fousseni CHABI-YO  
Financial Markets Department, Bank of Canada

Dietmar LEISEN  
Faculty of Law and Economics, Johannes Gutenberg-Universität Mainz

Eric RENAUlT†  
Department of Economics, University of North Carolina at Chapel Hill and CIRANO, CIREQ, Montreal  

This version: June 28, 2005

---

*We thank Kenneth Judd for very helpful conservations.
†Address for correspondence: Eric Renault, Latane Distinguished Professor of Economics, Department of Economics, CB# 3305, 304 Gardner Hall, Chapel Hill, NC 27599-3305. Email: renault@email.unc.edu.
Abstract

Asymmetric shocks are common on markets; securities’ payoffs are not normally distributed and exhibit skewness. This paper studies the portfolio holdings of heterogeneous agents with preferences over mean, variance and skewness, and derives equilibrium prices. A three funds separation theorem holds, adding a skewness portfolio to the market portfolio; the pricing kernel depends linearly only on the market return and its squared value. Our analysis extends Harvey and Siddique’s (2000) conditional mean-variance-skewness asset pricing model to non-vanishing risk-neutral market variance. Empirical relevance of this extension is documented in the context of the asymmetric GARCH-in-mean model of Bekaert and Liu (2004).

Keywords: Beta pricing, portfolio choice, stochastic discount factor, (co-)skewness, skewness preferences

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 assets display a high degree of skewness. The important contribution of Harvey and Siddique (2000) renewed interest in the compensation of skewness risks and led to an active literature. 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 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 agent’s 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 show that such an “average” agent sets price. 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. Among all portfolios the skewness portfolio is the portfolio with a return “closest” to that of squared market return. Agents’ holdings of the market portfolio are proportional to individual risk-tolerance; holdings of the skewness portfolio are proportional to risk-tolerance multiplied by the

---


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.
difference between individual agent’s skew-tolerance and that of the “average” agent. 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 is 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 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 squared market return. The skewness portfolio, projection of the squared market return on primitive assets, plays the role of an additional mutual fund.

Second, we study extensively the pricing implications of SDF that are quadratic in the market

\(^3\)A work that also extends Samuelson’s analysis is Judd and Guu (2001). They present 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.
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 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 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 parameters 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. Section 5 concludes the paper. Lengthy 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 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:

\[
R_i (\sigma) = R_f + \sigma^2 a_i (\sigma) + \sigma Y_i. \tag{1}
\]

Here, \( a_i (\sigma), i = 1, ..., n, \) 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\(^4\). For future reference we denote by

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

the matrix of covariances between \( Y_k \) and cross-products \( Y_i Y_j, i, j = 1, ..., n \). Typically, asymmetry in the joint distribution of returns means that at least some matrices \( \Gamma_k, k = 1, ..., n \) are not zero.

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 \( a (\sigma) = (a_i (\sigma))_{i=1,...,n} \) 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

\[
\max_{(\omega_i)_{1 \leq i \leq n} \in \mathbb{R}^n} \mathbb{E} u \left( R_f + \sum_{i=1}^n \omega_i \cdot (R_i (\sigma) - R_f) \right). \tag{2}
\]

\(^4\)Samuelson (1970) provides a heuristic explanation of (1) that is of interest for readers accustomed to continuous-time finance models; he couches this in terms of Brownian motion and identifies \( \sigma \) with the square root of time.
Note that, for the sake of notational simplicity, we normalized the initial wealth invested 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 the following: to what extent does a Taylor approximation of \(u\) allows us to understand the local behavior of the shares \(\omega_i(\sigma)\) in the neighborhood of the zero risk \(\sigma = 0\)? Put differently, we want to characterize 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 functions \(a_i(\sigma)\) are not assumed to be constant in our analysis. For this purpose let us consider a third order Taylor expansion of \(u\) in the neighborhood of the safe return \(R_f\):

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

(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 \(u^*:\)

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

(5)

For \(i = 1, \ldots, n\) the terms \(\omega^*_i(0)\) and \(\omega''_i(0)\) are defined similar to (3) as continuity extensions. We state 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 3-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 in the limit case \(\sigma \to 0:\)

1. The optimal shares of wealth invested \(\omega_i(0), i = 1, \ldots, n\) 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 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

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

(6)

denote by the risk tolerance coefficient and the skew tolerance coefficient of the agent.

Of course the risk tolerance coefficient \( \tau \) is assumed to be positive, to capture risk aversion, while
the skew tolerance coefficient \( \rho \) is non negative, following the literature on preferences for higher order
moments (Dittmar (2002), Harvey and Siddique (2000)). This assumption may also be justified by
reference to prudence (Kimball (1990)).

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** In the limit case \( \sigma \to 0 \), the vector \( \omega (0) = (\omega_i(0))_{1 \leq i \leq n} \) of shares of wealth invested 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 trade-off
would be given by \(-Eu'(W(\sigma)) / Eu''(W(\sigma))\) with \(W(\sigma) = R_f + \sum_{i=1}^{n} \omega_i(\sigma)(R_i(\sigma) - R_f)\).
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 2.5 below).

In such a case, the formula of theorem 2.2 can be rewritten:

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

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 \( \tau \Sigma^{-1}a(0) \) in the neighborhood of \( \sigma = 0 \), we get a Sharpe ratio for optimal portfolios equivalent to:

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

Then,

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

so that

\[
P(0) = \left[\frac{a^\perp(0)\Sigma^{-1}a(0)}{a^\perp(0)\Sigma^{-1}a(0)}\right]^{1/2}.
\] (7)

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 2.5 below that these functions will not be constant, even locally in the neighborhood of \( \sigma = 0 \), as soon as the joint asset-return 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:

\[
c_k(\omega) = \frac{1}{\sigma} \frac{Cov(R_k, (\omega^\perp(R - ER))^2)}{Var(\omega^\perp(R - ER))}. \tag{8}
\]

Note that co-skewness does not depend on the scale of risk \( \sigma \). We will see below 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).

The vector \( c(\omega) = (c_k(\omega))_{1 \leq k \leq n} \) represents a multivariate notion of skewness such that investors like to get positive skewness, component-wise. This assertion is justified by the fact that the average

\[
\sum_{k=1}^{n} \omega_k c_k(\omega) = \frac{1}{\sigma} \frac{E[(\omega^\perp(R - ER))^3]}{Var(\omega^\perp(R - ER))}
\]

\[
= \frac{1}{\sigma} Skew(\omega^\perp \cdot (R - ER)) \cdot \left(Var(\omega^\perp \cdot (R - ER))\right)^{1/2}
\]
is positive if and only if the portfolio return is positively skewed. We get the following result:

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

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

where \( a' (0) = (a'_i (0))_{1 \leq i \leq n} \) is the vector of marginal risk premia and \( c = c (\omega (0)) = c (\tau \Sigma^{-1} a (0)) \).

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.

Individual preferences for positive skewness will increase, ceteris paribus, the equilibrium price of assets with positively skewed returns. This effect will appear in the equilibrium value \( a' (0) \) of risk premium slopes in the neighborhood of \( \sigma = 0 \) (see below).

### 2.2 Equilibrium Allocations and Prices

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

\[
\tau_s = \frac{u'_s (R_f)}{u''_s (R_f)} \quad \text{and} \quad \rho_s = \frac{\tau^2_s u'''_s (R_f)}{2 u'_s (R_f)}. \tag{9}
\]

From theorems 2.2 and 2.4 we get that:

\[
\omega_s (0) = \Sigma^{-1} \tau_s a (0), \quad \omega'_s (0) = \tau_s \Sigma^{-1} \left[ a' (0) + \rho^2_s (0) c (\omega (0)) \right]. \tag{10}
\]

Note that these formulas correspond to the case where each of the \( S \) agents would get a unit wealth to invest. Generalization to more realistic, non-uniform distributions of initial wealth would be easy to state, but this would merely complicate the notation without adding any insight to the analysis of this paper. Therefore, the only heterogeneity considered in this paper is about preferences.
An average investor will be defined by average preferences, that are average risk tolerance $\tau$ and average skew tolerance such that:

$$\tau = \frac{1}{S} \sum_{s=1}^{S} \tau_s, \text{ and } \rho = \frac{\sum_{s=1}^{S} \rho_s \tau_s}{\sum_{s=1}^{S} \tau_s}. \tag{11}$$

Note that the average skew tolerance is computed with weights proportional to risk tolerance, so that:

$$\sum_{s=1}^{S} \tau_s (\rho_s - \bar{\rho}) = 0. \tag{12}$$

We consider that the net supply of each risky asset $i = 1, \ldots, n$ is exogenous and independent of the scale of risk $\sigma$. Then, Taylor expansions of individual portfolios shares must fulfill the market clearing conditions:

$$\sum_{s=1}^{S} \omega_s (0) = S \bar{\omega}, \text{ and } \sum_{s=1}^{S} \omega'_s (0) = 0. \tag{13}$$

where $\bar{\omega}$ is the portfolio that would be selected by an average investor with characteristics $(\tau, \rho)$.

Jointly with individual asset demands (10), market clearing conditions (13) and (13) determine the Taylor expansion of the risk premium function $a(\sigma)$ in equilibrium:

**Theorem 2.5** 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: $\bar{\omega} = \tau \Sigma^{-1} a(0)$, that is

$$a(0) = \frac{1}{\tau} \Sigma \bar{\omega}. \tag{14}$$

Its slope in the neighborhood of zero is given by:

$$a'_k (0) = -\bar{\rho} p^2 (0) c_k (\bar{\omega}) = -\frac{\bar{\rho}}{\bar{\omega}^\top \Gamma_k \bar{\omega}} \text{ for } k = 1, \ldots, K. \tag{15}$$

Theorem 2.5 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. To see this, let us assume for notational simplicity that the total supply of risk-free asset is zero. Then,
the average portfolio $\overline{\omega}$ has a unit price (since we have assumed that each investor has a unit wealth) and $R_M = \overline{\omega}^\perp R$ denotes the market return. Then

$$\beta = \frac{\text{Cov} (R, R_M)}{\text{Var} (R_M)} = \frac{\Sigma \overline{\omega}}{\overline{\omega}^\perp \Sigma \overline{\omega}}$$

(14)
denotes the vector of market betas of the $n$ assets.

Thus, not surprisingly, the first part of theorem 2.5 states that the limit value $a(0)$ of the vector of equilibrium risk premium is proportional to the vector of market betas, with a proportionality coefficient $\frac{\text{Var} (R_M)}{\tau} = \tau \sigma^2 P^2 (0)$, which is itself increasing with market risk and market risk aversion.

The new contribution of theorem 2.5 is encapsulated in the value

$$a'_k (0) = -\rho P^2 (0) c_k (\overline{\omega})$$

(15)

It states that insofar as utility functions are not quadratic ($\rho \neq 0$), asset $k$ exhibits a positive skewness risk premium $a'_k (0)$ when its co-skewness $c_k (\overline{\omega})$ in the market portfolio is negative. As already explained, an asset $k$ should be preferred, ceteris paribus, when it contributes positively to the market skewness. By contrast, when it contributes negatively, investors have to be compensated for that. This effect is captured through a risk premium function $a(\sigma)$ which is not constant in the neighborhood of $\sigma = 0$, by contrast with Samuelson’s (1970) analysis.

Individual asset demands in equilibrium are then determined from the results of section 2.1, when plugging in the equilibrium values of $a(0)$ and $a'(0)$:

**Theorem 2.6** In equilibrium, in the limit case $\sigma \sim 0$, the optimal shares of wealth invested $\omega_s (\sigma)$ of agents $s = 1, \ldots, S$ are characterized by:

$$\omega_s (0) = \frac{\tau_s}{\tau} \overline{\omega}, \text{ and}$$

$$\omega'_s (0) = \tau_s [\rho_s - \overline{\rho}] P^2 (0) \Sigma^{-1} c (\overline{\omega}) = \frac{\tau_s}{\tau} (\rho_s - \overline{\rho}) \left( \text{Var} (R) \right)^{-1} \text{Cov} (R, (R_M - ER_M)^2).$$

Theorem 2.6 states that in the limit case $\sigma \sim 0$, 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 $\overline{\omega}$, the size of this share being determined by the comparison of individual risk tolerance $\tau_s$ with respect to the 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 risk. A positive market co-skewness $c_k (\overline{\omega})$ will have a positive effect on the demand of asset $k$ by agent $s$ 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 \( k \) co-skewness on its market price makes more than a compensation of investor’s preference for positive skewness.

Interestingly, the effect of individual preferences for skewness manifests itself only through one portfolio, common to all investors, and defined by the components of the vector

\[
(Var(R))^{-1} Cov(R, (R_M - ER_M)^2).
\]

In other words, heterogeneous skewness preferences give rise to a departure from the standard mean-variance theorem only through the addition of a third mutual fund, that we call the skewness portfolio and define as the portfolio with payoff:

\[
\psi_M = (Var(R))^{-1} Cov(R, (R_M - ER_M)^2) R.
\]

In order to interpret the skewness portfolio, notice that: First, \( \psi_M \) is not zero if and only if the vector

\[
c(\omega) = \frac{1}{\sigma} Cov(R, (R_M - ER_M)^2) Var(R_M)
\]

of co-skewnesses is not zero. In other words, it is not really a non-zero market skewness but a non-zero market co-skewness which give rise to a non-degenerate skewness oriented mutual fund. We recall that market skewness is just an average of market co-skewnesses:

\[
E (R_M - ER_M)^3 = \sigma (Var(R_M)) \omega'^T c(\omega).
\]

Second, when it is not zero, the skewness portfolio characterizes the variable part of the affine regression \( EL[(R_M - ER_M)^2 | R] \) of \( (R_M - ER_M)^2 \) on the vector \( R \) of asset returns:

\[
EL[(R_M - ER_M)^2 | R] = Var(R_M) + \psi_M - E\psi_M. \tag{16}
\]

We then deduce the following three-funds theorem:

**Theorem 2.7** In a second-order approximation of the market equilibrium, the risky portfolio \( \omega_s(0) + \sigma \omega'_s(0) \) of agents \( s = 1, ..., S \) gives them jointly with their investment in the riskless asset, the return

\[
\left(1 - \frac{\tau_s}{\bar{\tau}}\right) R_f + \frac{\tau_s}{\bar{\tau}} \left[ R_M + (\rho_s - \bar{\rho}) (\psi_M - R_f \pi(\psi_M)) \right],
\]

where \( \pi(\psi_M) \) is the cost of skewness portfolio \( \psi_M \), that is the sum of the components of the vector \( (Var(R))^{-1} Cov(R, (R_M - ER_M)^2) \).
To interpretation theorem 2.7 note that the standard two-funds theorem is maintained if and only if one of the following two conditions are fulfilled. Either, all market co-skewnesses are zero (and a fortiori market skewness is zero). Or, agents are homogenous in terms of preferences for skewness. In these two cases, we are back to the standard results: Agent $s$ will then choose a return which is a convex combination of the risk free return and the market return, the weighting coefficient being defined by its relative risk tolerance $\tau_s \frac{\psi_M - R_f}{\pi}$. By contrast, when none of the above two conditions is fulfilled, then agent $s$ is going to replace his share $\frac{\tau_s}{\tau} \frac{\psi_M - R_f}{\psi_M}$ of market return by the same share of an alternative portfolio which mixes the market return $R_M$ with the net present value $(\psi_M - R_f \pi (\psi_M))$ of the skewness portfolio. The weight given to the skewness portfolio is defined by the spread $(\rho_s - \bar{\rho})$ between investor’s skewness tolerance and average skewness tolerance.

An intuitive way to understand this result is the following. As will be made explicit in section 3, skewness preferences can be characterized through the price of the squared market return. However, without nonlinear derivatives, only linear combinations of primitive asset payoffs can be purchased, and therefore the skewness portfolio represents the best approximation of the variable part of $(R_M - ER_M)^2$ by a (linear) portfolio of primitive assets.

### 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 $E[m(\sigma) \cdot (R_f + \sigma^2 a_1(\sigma) + \sigma Y_i)] = 1$ for $i = 1, ..., n$. We are then able to re-express theorem 2.5 in terms of SDF:

**Theorem 2.8** The random variable:

$$m(\sigma) = \frac{1}{R_f} - \frac{1}{R_f \pi} (R_M(\sigma) - ER_M(\sigma))$$

$$+ \frac{\bar{\rho}}{R_f \pi^2} [(R_M(\sigma) - ER_M(\sigma))^2 - E(R_M(\sigma) - ER_M(\sigma))^2]$$

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 2.5.
The conjunction of theorems 2.6 and 2.8 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 skewness 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 zero aggregate demand.

2. The interpretation of the skewness portfolio as the portfolio with return closest to the 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.

This last remark allows us to compare our asset pricing model with early approaches to skewness pricing. While these approaches were formulated in terms of beta pricing, we deduce straightforwardly from theorem 2.8 that:

**Theorem 2.9** The asset pricing model associated to risk premium $a(\sigma) = a(0) + \sigma a'(0)$, with $a(0)$ and $a'(0)$ of theorem 2.5, is equivalent to the linear beta pricing relationship:

$$ER_i - R_f = \frac{1}{\tau} (Var R_M) \beta_i - \frac{\beta}{\tau^2} \left( Var \tilde{R}_M^2 \right) \gamma_i \quad \text{for } i = 1, ..., n,$$

where:

$$\tilde{R}_M = R_M - ER_M,$$

$$\beta_i = \frac{Cov (R_i, R_M)}{Var (R_M)},$$

$$\gamma_i = \frac{Cov (R_i, \tilde{R}_M^2)}{Var (\tilde{R}_M^2)}.$$

While $\beta = (\beta_i)_{1 \leq i \leq n}$, see also equation (14), is the common vector of market betas, $\gamma = (\gamma_i)_{1 \leq i \leq n}$ is the vector of betas with respect to the additional factor $\tilde{R}_M^2$. The parameters $\gamma$ are tightly related
to coskewness, since
\[ \gamma_i \text{Var}\left(\tilde{R}_M^2\right) = \text{Cov}\left(R_i, \tilde{R}_M^2\right) = c_i(\bar{\omega}) \text{Var}\left(\tilde{R}_M\right). \]

Non-zero beta coefficients show up insofar as same co-skewness coefficients are non-zero. Moreover, the price of this additional factor is proportional to the average skewness tolerance \( \bar{p} \). It has a zero price when utility functions are quadratic. Similar presentations in terms of an additional priced factor can be found in Kraus and Litzenberger (1976) as well as in Ingersoll (1987). These authors do not address the aggregation issue about investors with different preferences. However, by considering a representative investor and a third-order Taylor expansion of her utility function, they put forward a two beta-pricing relationship similar to theorem 2.9, which is also based on \( \text{Cov}\left(R_i, \tilde{R}_M^2\right) \) in addition to common betas. Note that if all agents were endowed with the same utility function \( u \), \( \bar{p} \tau^2 \) would be \( \frac{u''(R_f)}{2u'(R_f)} \) as usually derived from Taylor expansion of the representative agent utility.

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:
\[ m_{t+1} = \nu_0 t + \nu_1 t R_{Mt+1} + \nu_2 t R_{Mt+1}^2. \]  
(17)

It actually suffices to revisit our section 2 above with a conditional viewpoint to see theorem 2.8 as a theoretical justification of (17). 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 2.8, we interpret the factors coefficients as:
\[ \nu_2 t = \frac{1}{R_{ft} \bar{p}^2} > 0, \]  
(18)

and
\[ \nu_1 t = -\frac{1}{R_{ft} \bar{p}} - 2 \frac{1}{R_{ft} \bar{p}^2} E_t [R_{Mt+1}] < 0. \]  
(19)
It is worth characterizing the role of the two factors $R_{Mt+1}$ and $R^2_{Mt+1}$ in the SDF (17) in terms of beta pricing relationships. Assuming the existence of a conditionally risk-free asset (with return $R_{ft}$), we denote

$$r_{it+1} = R_{it+1} - R_{ft}$$

the net excess return of every asset $i = 1, ..., n$. 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:

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

If $\nu_1 t$ and $\nu_2 t$ are interpreted in terms of preferences of an average investor as in (18) and (19), we deduce:

$$\lambda_1 t = \frac{1}{\tau} + 2\frac{\tau}{\tau^2} (E_t [R_{Mt+1}] - R_{ft}) \text{ and } \lambda_2 t = \frac{\tau}{\tau^2}.$$

Note that $\lambda_2 t$ 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 $\lambda_2 t$ should be non-negative and all the more positive that preference for skewness is high. Similarly, $\lambda_1 t$ is expected to be positive and time varying through the market risk premium $(E_t R_{Mt+1} - R_{ft})$. To summarize:

**Theorem 3.1** Under the maintained assumption (17) of a quadratic SDF, net expected returns are given by:

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

where $\lambda_1 t$ and $\lambda_2 t$ are expected to be non negative.

Note that $\lambda_1 t$ has two components which are both increasing with the average risk aversion, first as $1/\tau$ 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 $\lambda_1 t$ large:
Corollary 3.2 Under the assumptions of theorem 3.1

\[ \lambda_{1t} = \frac{E_t[r_{Mt+1}]}{Var_t(r_{Mt+1})} + \lambda_{2t} \frac{Cov_t(r_{Mt+1}, r_{Mt+1}^2)}{Var_t(r_{Mt+1})}, \]

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 market co-skewness through:

\[ Cov_t(r_{Mt+1}, r_{Mt+1}^2) = E_t r_{Mt+1}^3 - (E_t r_{Mt+1}) (E_t r_{Mt+1}^2). \]

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).

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

\[ E_t[r_{it+1}] = \left( \lambda_{1t} Var_t(r_{Mt+1}) \right) \beta_{iMt} - \left( \lambda_{2t} Var_t(r_{Mt+1}^2) \right) \delta_{iMt} \]

or

\[ E_t[r_{it+1}] = E_t [r_{Mt+1}] \beta_{iMt} - \lambda_{2t} Var_t(r_{Mt+1}^2) \cdot (\delta_{iMt} - \delta_{MMt} \beta_{iMt}) \]

where

\[ \beta_{iMt} = \frac{Cov_t(r_{it+1}, r_{Mt+1})}{Var_t(r_{Mt+1})}, \quad \delta_{iMt} = \frac{Cov_t(r_{it+1}, r_{Mt+1}^2)}{Var_t(r_{Mt+1}^2)}. \]

The term \( \beta_{iMt} \) is the standard market beta while the beta coefficient with respect to the squared market return is \( \delta_{iMt} \); up to a change in normalization, it corresponds to the measure of co-skewness already introduced in section 2. Therefore the result of equation (21) matches exactly that of theorem 2.9 with a conditional viewpoint.

As already shown in section 2, the beta pricing model (20) 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). In particular (21) shows, as does 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.

For the purpose of econometric identification (see section 4), it is convenient to interpret this difference between two betas in terms of affine regressions. In the same way we defined in section

18
2 the skewness portfolio from the affine regression of the squared market return on the vector of primitive assets, it is convenient to focus here on the part of \( r_{Mt+1}^2 \) which can be mimicked by a linear function of \( r_{Mt+1} \), conditionally on available information at time \( t \):

\[
r_{Mt+1}^2 = \frac{Cov_t \left( r_{Mt+1}^2, r_{Mt+1} \right)}{Var_t \left( r_{Mt+1} \right)} r_{Mt+1} + r_{Mt+1}^{(2)}.
\]

Then \( r_{Mt+1}^{(2)} \) is the part of \( r_{Mt+1}^2 \) which is orthogonal to \( r_{Mt+1} \). It follows straightforwardly:

**Theorem 3.3** We have:

\[
Cov_t \left( r_{it+1}, r_{Mt+1}^2 \right) = Var_t \left( r_{Mt+1}^2 \right) \cdot \left[ \delta_{iMt} - \delta_{MMt} \beta_{iMt} \right].
\]

In other words, (21) is a significant modification of a common Sharpe-Lintner CAPM pricing relationship if and only if the following two conditions are fulfilled: First, the market preference for skewness is strong enough to make the skewness price \( \lambda_{2t} \) significantly different from zero. Second, the return of asset \( i \) is significantly correlated to that part of \( r_{Mt+1}^2 \) which is orthogonal to \( r_{Mt+1} \).

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 \( \lambda_{2t} = 0 \), (20) applied to the market gives the usual formula: \( \lambda_{it} = P_{Mt}^{(1)} \) with

\[
P_{Mt}^{(1)} = \frac{E_t [r_{Mt+1}]}{Var_t (r_{Mt+1})}.
\]

However, in general \( \lambda_{it} \) and \( \lambda_{2t} \) cannot be read as simple risk premiums 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:

\[
P_{Mt}^{(2)} (\eta_t) = \frac{E_t \left[ \frac{r_{Mt+1}^2}{\eta_t} \right] - R_{ft}}{Var_t \left( \frac{r_{Mt+1}^2}{\eta_t} \right)} = \frac{E_t [r_{Mt+1}^2] - R_{ft} \eta_t}{Var_t \left( r_{Mt+1}^2 \right) \eta_t}
\]

(22)

will not coincide with \((-\lambda_{2t} \eta_t)\). The difference comes from the fact that the two factors are not orthogonal. The term \( \lambda_{it} \) 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 \( \eta_t = E_t \left[ m_{t+1} r_{Mt+1}^2 \right] \) denotes the equilibrium price of a payoff \( r_{Mt+1}^2 \), we have:

\[
\lambda_{2t} = \frac{\delta_{MMt} P_{Mt}^{(1)} - \frac{1}{\eta_t} P_{Mt}^{(2)} (\eta_t)}{1 - \rho_t^2 (r_{Mt+1}, r_{Mt+1}^2)}.
\]
where according (22), $P^{(2)}_{Mt}(\eta_t)$ is the risk premium on the asset with payoff $r^2_{Mt+1}$ and $\rho^2_t (r_{Mt+1}, r^2_{Mt+1})$ denotes the square (conditional) linear correlation coefficient between $r_{Mt+1}$ and $r^2_{Mt+1}$.

Note that, from (22) we have:

$$\lim_{\eta_t \to 0} \frac{P^{(2)}_{Mt}(\eta_t)}{\eta_t} = \frac{E_t[r^2_{Mt+1}]}{Var_t(r^2_{Mt+1})}.$$  \hspace{1cm} (23)

In this limit case, we get

$$\lambda_{2t} = \frac{\gamma_{MMt} P^{(1)}_{Mt} - E_t[r^2_{Mt+1}]}{1 - \rho^2_t (r_{Mt+1}, r^2_{Mt+1})},$$  \hspace{1cm} (24)

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 $\eta_t = E_t[m_{t+1} r^2_{Mt+1}]$ should be strictly positive. Since, from (22),

$$\frac{P^{(2)}_{Mt}(\eta_t)}{\eta_t} = \frac{E_t[r^2_{Mt+1}] - R_{ft} \eta_t}{Var_t(r^2_{Mt+1})},$$  \hspace{1cm} (25)

we expect that considering the limit case (23), i.e. considering $\eta_t = 0$, leads to overestimate $\frac{P^{(2)}_{Mt}(\eta_t)}{\eta_t}$ and therefore to underestimate $\lambda_{2t}$ (see Theorem 3.4).

Whether the shadow market price of $r^2_{Mt+1}$ is significantly positive or not is an empirical question: The relevant empirical issue (see section 4) is then to decide if considering only the limit case (24) leads to an economically significant underestimation of the weight $\lambda_{2t}$ of co-skewness in the two-factor pricing relationship (21). 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 as shown below, in the risk neutral variance of the market return.

### 3.2 Risk-Neutral Variance and the Pricing of Asymmetry Risk

The development of derivative asset markets, introducing asset payoffs which are nonlinear functions of 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 multiple primitive assets.

Following Bakshi, Kapadia, and Madan (2003), we focus on the so called “volatility contract” defined by the payoff $R^2_{Mt+1}$. We are then led to study the difference between its price and the price of its linear approximation

$$E_t [m_{t+1} E_L [R^2_{Mt+1} | R_{Mt+1}]].$$
The role of skewness here is encapsulated in the following lemma:

**Lemma 3.5** The conditional affine regression of $R^2_{Mt+1}$ on $R_{Mt+1}$ is:

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

This lemma needs to be contrasted with Taylor expansion of $R^2_{Mt+1}$ around $E_t[R_{Mt+1}]$:

$$R^2_{Mt+1} \approx (E_tR_{Mt+1})^2 + 2(E_tR_{Mt+1})(R_{Mt+1} - E_tR_{Mt+1}).$$

Thus, a common Taylor expansion neglects the crucial role of the skewness term. In this respect, the small noise expansion approach appears to be more reliable.

The price of the volatility contract can be fruitfully interpreted in terms of risk-neutral pricing. We have:

$$\pi_t = E_t[m_{t+1}R^2_{Mt+1}] = \frac{1}{R_{ft}}E^*_t[R^2_{Mt+1}],$$

where $E^*_t$ denotes the conditional expectation with respect to a risk neutral probability measure defined by $m.$ By definition, $E^*_t[R_{Mt+1}] = R_{ft},$ so that

$$Var^*_t(R_{Mt+1}) = E^*_t[(R_{Mt+1} - R_{ft})^2] = R_{ft} \cdot (\pi_t - R_{ft})$$

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

$$Var^*_t(R_{Mt+1}) = Var^*_t(r_{Mt+1}) = E^*_t[r^2_{Mt+1}] = R_{ft} \cdot E_t[m_{t+1}r^2_{Mt+1}] = R_{ft}\eta_t.$$

Using the risk-neutral variance allows us to shed more light on the impact of a non-zero shadow price $\eta_t$ for $r^2_{Mt+1}.$ From (25) we get:

$$\frac{P^{(2)}_{Mt}(\eta_t)}{\eta_t} = \frac{E_t[r^2_{Mt+1}] - Var^*_t(R_{Mt+1})}{Var_t(r^2_{Mt+1})}.$$

As noted after theorem 3.4, we expect that an increased price $\eta_t$ of $r^2_{Mt+1}$ leads to a smaller premium $P^{(2)}_{Mt}/\eta_t$ and to a larger 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 $Var^*_t(R_{Mt+1}) = R_{ft}\eta_t.$ For doing so, we first state a useful relation between the risk-neutral variance $Var^*_t(R_{Mt+1})$ and the historical one $Var_t(R_{Mt+1})$:  

21
Theorem 3.6 The risk-neutral variance is:

\[
Var^*_t(R_{Mt+1}) = Var_t(R_{Mt+1}) \cdot \left[ 1 - \sigma^2_{mt} \left( P^{(1)}_{Mt} \right)^2 \right] - P^{(1)}_{Mt} E_t \left[ (R_{Mt+1} - E_t R_{Mt+1})^3 \right] + R_{ft} Cov_t (m_{t+1}, \varepsilon_{t+1}),
\]

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

Note that theorem 3.6 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 \( \frac{1}{R_{ft}} E_t^* h(R_{Mt+1}) = E_t \left[ m_{t+1} h(R_{Mt+1}) \right] \). It is then worth revisiting skewness pricing by studying the factor which may potentially increase the risk neutral variance. Theorem 3.6 basically puts forward two factors. One factor is model dependent, through \( Cov_t (m_{t+1}, \varepsilon_{t+1}) \) while the other terms can be directly observed from the market return. Typically, in case of a positive market return skewness (\( E_t (R_{Mt+1} - E_t R_{Mt+1})^3 > 0 \)), risk neutral variance is inversely related to risk premium \( P^{(1)}_{Mt} \). Intuitively, high risk neutral variance, that is high compensation for skewness, may compensate a low risk premium \( P^{(1)}_{Mt} \). By contrast, the effect encapsulated in \( Cov_t (m_{t+1}, \varepsilon_{t+1}) \) 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,

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

This theorem confirms in 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, i.e.

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

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

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

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 \( \bar{\rho} = \nu_2 R_{ft} \bar{\tau}^2 \) 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 co-skewness as developed in previous sections is encapsulated in the asset pricing equation (21):

\[ E_t [r_{Mt+1}] = E_t [r_{Mt+1}] \beta_{iMt} - \lambda_2 t \text{Var}_t (r^2_{Mt+1}) \cdot (\delta_{iMt} - \delta_{MMt} \beta_{iMt}). \]  \hspace{1cm} (26)

The question is: does this asset pricing equation significantly deviate from standard CAPM? That is: should we maintain a significantly positive skewness premium \( \lambda_2t \)?

It turns out that the statistical identification of this hypothesis is difficult, since, as has been noted by Barone-Adesi, Gagliardini and Urga (2004), covariance and co-skewness with the market tend to be almost collinear across common portfolios, leading to hardly significant co-skewness factors (\( \delta_{imt} - \delta_{nmt} \beta_{imt} \)). To shed more light on this identification issue, let us consider the (conditional) affine regression of asset \( i \)'s net return on market return:

\[ r_{it+1} = \alpha_{it} + \beta_{iMt} r_{Mt+1} + u_{it+1}. \]  \hspace{1cm} (27)

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

\[ \text{Var}_t (r^2_{Mt+1}) \cdot (\delta_{iMt} - \delta_{MMt} \beta_{iMt}) = \text{Cov}_t (u_{it+1}, r^2_{Mt+1}) = \text{Cov}_t (u_{it+1}, R^2_{Mt+1}). \]  \hspace{1cm} (28)
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 a linear approximation is valid, $\text{Cov}_t \left( u_{it+1}, r_{Mt+1}^2 \right)$ is almost zero, which 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, Gagliardini and Urga (2004) focus on a quadratic market model first introduced by Barone-Adesi (1985). With his specification they estimate a coefficient $\lambda_{2t}$, which is slightly significantly positive, 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 the 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’s (2004) parameter estimates.

### 4.2 The Simulation Set-up

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 stock return $\text{Log}(R_{Mt+1})$ and logarithm of bond return $\text{Log}(R_{ft+1})$:

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

\[ Y_{t+1} = c_t + AY_t + \Omega e_{t+1}, \]  

(29)

where the coefficient \( c_{it} \) of \( c_t \), \( i = 1, 2, 3 \), is an affine function of \( \text{Var}_t [Y_{it+1}] \) 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):

\[ \text{Var}_t [e_{1t+1}] = \delta_1 + \alpha (e_{1t})^2 + \beta \text{Var}_{t-1} [e_{1t}] + \xi (\text{Max} [0, -e_{1t}])^2. \]  

(30)

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.

[Table 1 about here.]

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 information \( I_t \), 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 (29) and (30) allow us to compute iteratively corresponding paths first of \( \text{Var}_t (R_{Mt+1}) = \text{Var}_t (R_{Mt+1} | R_{ft}/E_t (R_{Mt+1}))^2 \), then of \( \eta_t = \text{Var}_t (R_{Mt+1}) / R_{ft} \), of \( P_{Mt}^{(2)} (\eta_t) = \frac{E[r_{Mt+1}^2] - \text{Var}_t (R_{Mt+1})}{\text{Var}_t (r_{Mt+1}^2)} \), of \( P_{Mt}^{(2)} (\eta_t) \), and finally of \( \lambda_{2t} \) according to theorem 3.4. We recall that the limit case put forward by Harvey and Siddique (2000) corresponds to the alternative formula:

\[ \lambda_{2t}^{HS} = \frac{\gamma_{MMt} P_{Mt}^{(1)} - E[r_{Mt+1}^2]}{1 - \rho_t^2 (r_{Mt+1}, r_{Mt+1}^2)}. \]
The path of this value is also easily built from the above simulation.

Of course, by introducing only one risky asset, this setting does not allow us to compare co-skewness 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 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.]

One path is simulated, corresponding to 500 months. Table 2 provides summary statistics for this series. 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 (labeled “CLR”) and of $\lambda_{2t}^{HS}$ corresponding to Harvey and Siddique (2000) limit case (labeled “HS”).

The conclusions drawn 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 $\lambda_{2t}^{HS}$ 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.]

[Figure 3 about here.]

The time series of $\eta_t$ (figure 2) and of the risk premium $P_{Mt}^{(2)}(\eta_t)$ (figure 3) confirm as well that they are positive. Note also that while $\lambda_{2t}$ and $\lambda_{2t}^*$ are stationary processes — in particular first order differences ($\lambda_{2t} - \lambda_{2t-1}$) and ($\lambda_{2t}^{HS} - \lambda_{2t}^{HS-1}$) 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 ($\lambda_{2t}^{H} - \lambda_{2t-1}^{H}$).
This gives 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, Gagliardini and Urga (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 pricing must be addressed in a conditional setting. However, even an unconditional approach would not make the simplified price series $\lambda_{HS}^{2t}$ meaningful since their standard error is only 7.45, which does not compensate their negative average of (-67.82).

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 order word, 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 (17), (18), and (19) with a positive coskewness price $\lambda_{2t}$ will become increasing when the market returns exceeds its conditional expectation by more than ($\tau/2\rho$). 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 from 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 0.64%) 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 1.56%), 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 investigates the relevance of non linear pricing kernels both at the theoretical and empirical levels. We first show that considering pricing kernels that are quadratic functions of the market return is a well-founded approximation of actual expected utility behavior, at least 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 “closest” portfolio to the
squared market return. The skewness portfolio is based on all third-order cross moments; in other words, while taking heterogeneity of skewness preferences into account yields separation theorems where an additional fund show up in asset demands, it remains true that idiosyncratic risk is not priced, both in terms of variance and skewness.

While statistical identification of a positive skewness premium may be difficult since covariance and co-skewness tend to be almost collinear across common portfolios, we showed through simulated data based on an actual estimation of a factor GARCH-in-mean model 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 leads us to a generalization of the Harvey and Siddique (2000) beta pricing model for skewness; by contrast 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. Typically, while only conditional skewness of asset returns shows up in the current paper, a multiperiod setting would 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.

6 Appendix

Proof of theorems 2.2 and 2.4. The solution \( \omega (\sigma) = (\omega_i (\sigma))_{1\leq i \leq n} \) of problem (2) determines a terminal wealth

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

according to the first order conditions

\[
0 = E [u' (W (\sigma)) \cdot (R_i - R_f)] = E [h_i (\sigma)] . \tag{31}
\]

Then, setting

\[
h_i (\sigma) = u' (W (\sigma)) \cdot (\sigma a_i (\sigma) + Y_i ) ,
\]
this implies that \( E \left[ \frac{d h_i}{d \sigma} (\sigma) \right] = 0 \) and so that \( \lim_{\sigma \to 0^+} E \left[ \frac{d h_i}{d \sigma} (\sigma) \right] = 0 \). Writing out the last equality we get

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

Using the variance-covariance matrix \( \Sigma \) of the vector \( Y \) of random variables and the definition of the trisk neutral tolerance in (6) we get \( \omega (0) = \Sigma^{-1} \cdot \tau \cdot a (0) \), which ends the proof of theorem 2.2.

To prove Theorem 2.4 we take the second-order derivatives of equation (31) 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:

\[
\sum_{i=1}^{n} \omega_i' (0) \text{Cov} (Y_i, Y_k) = \frac{\rho}{\tau} \sum_{i=1}^{n} \omega_i^2 (0) E[Y_i^2 Y_k] + 2 \frac{\rho}{\tau} \sum_{i<j} \omega_i (0) \omega_j (0) E[Y_i Y_j Y_k] + \tau a_k' (0) \quad (32)
\]

Therefore equation (32) reads

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

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

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

\[\blacksquare\]

**Proof of theorems 2.5 and 2.8.** Using the definitions of \( \tau_s, \rho_s \) from equation (9), the demand equation (10) and the first market-clearing equation (13) we derive from the condition

\[
S \omega = \sum_{s=1}^{S} \omega_s (0) = \sum_{s=1}^{S} \Sigma^{-1} \tau_s a (0)
\]

that

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

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 (32) and the second market-clearing equation (13) we get from

\[
\sum_{s=1}^{S} \omega_s' (0) = \sum_{s=1}^{S} \tau_s \Sigma^{-1} \left[ c (\omega) \rho_s P^2 (0) + a' (0) \right] = 0
\]

30
that
\[ a' (0) = - \rho c (\omega) P^2 (0). \]

Thus:
\[ a'_k (0) = - \rho c_k (\omega) P^2 (0) = - \frac{\rho}{\tau} (\omega^\perp \Gamma_k \omega). \] (34)

Plugging (34) into theorem 2.4 gives:
\[ \omega'_s (0) = \tau_s \Sigma^{-1} [c (\omega) \rho_s P^2 (0) + a' (0)] = \tau_s [\rho_s - \bar{\rho}] P^2 (0) \Sigma^{-1} c (\omega). \]

\[ \textbf{Proof of theorem 2.8.} \] We look for two real numbers \( L \) and \( N \) such that, with
\[
m (\sigma) = \frac{1}{R_f} + L \cdot \left( R_M (\sigma) - E[R_M (\sigma)] \right)
+ N \cdot \left( (R_M (\sigma) - E[R_M (\sigma)])^2 - E(R_M (\sigma) - E[R_M (\sigma)])^2 \right),
\]
we have
\[ Em (\sigma) \cdot (\sigma^2 a (\sigma) + \sigma Y) = 0. \]

That is we require
\[ \frac{1}{R_f} \sigma^2 a (\sigma) + \sigma Cov (Y, m (\sigma)) = 0. \]

We want to see these equations fulfilled with \( a (\sigma) = a (0) + \sigma a' (0) \) where \( a (0) \) and \( a' (0) \) given by theorem 2.5. Then:
\[
\frac{1}{R_f} \sigma^2 a (\sigma) + \sigma Cov (Y, m (\sigma)) = 0.
\]

Let us define \( L \) and \( N \) such that
\[
L \cdot Cov (R, R_M (\sigma)) + N \cdot Cov \left( R, (R_M (\sigma) - E[R_M (\sigma)])^2 \right) = 0.
\]

Noticing that
\[
\sigma^2 \Sigma \omega = Cov (R, R_M (\sigma)), \quad \text{and}
\]
\[
\tau \sigma^3 P^2 (0) c (\omega) = \frac{\text{Var} (R_M)}{\tau} c (\omega) = \frac{1}{\tau} Cov \left( R, (R_M (\sigma) - E[R_M (\sigma)])^2 \right),
\]

31
we conclude that

\[ L = -\frac{1}{R_f \tau}, \quad \text{and} \quad N = \frac{1}{R_f \tau^2} \]

which is the announced result. ■

Proof of theorem 2.9. Note that

\[ E[R_i] - R_f = \sigma a_i(\sigma) = \sigma a_i(0) + \sigma^3 a_i'(0). \]

Then, by theorem 2.5, the vector \((ER_i - R_f)_{1 \leq i \leq n}\) can be written as:

\[ \frac{\sigma}{\tau} \sum \omega - \sigma^3 \bar{p}^2(0) c(\omega) = \frac{1}{\tau} Cov(R, R_M) - \frac{\bar{p}}{\tau^2} Cov \left( R, (R_M - E[R_M])^2 \right). \]

In others words:

\[ E[R_i] - R_f = \frac{1}{\tau} Cov(R, R_M) - \frac{\bar{p}}{\tau^2} Cov \left( R, \bar{R}_M^2 \right) \]

which corresponds to the formula of theorem 2.9. ■

Proof of Theorem 3.4. By applying (26) to the net return on the squared market return payoff

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

we get

\[ E_t[r_{Mt+1}^2] - R_{ft} \eta_t = E_t[r_{Mt+1}] \delta_{MMt} \frac{Var \left( r_{Mt+1}^2 \right)}{Var \left( r_{Mt+1} \right)} - \lambda_{2t} \cdot Var \left( r_{Mt+1}^2 \right) \cdot (1 - \rho_t^2 \left( r_{Mt+1}, r_{Mt+1}^2 \right)), \]

that is:

\[ \frac{P_{Mt}^{(2)}}{\eta_t} = \delta_{MMt} P_{Mt}^{(1)} - \lambda_{2t} \left( 1 - \rho_t^2 \left( r_{Mt+1}, r_{Mt+1}^2 \right) \right). \]

This gives the announced value for \( \lambda_{2t} \). ■

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

\[ EL_t \left[ R_{Mt+1}^2 | R_{Mt+1} \right] = E_t[R_{Mt+1}^2] + a \cdot F_{1t+1}, \quad \text{where} \quad F_{1t+1} = R_{Mt+1} - E_t[R_{Mt+1}]. \]

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,

\[ Cov_t \left( R_{Mt+1}^2 - EL_t \left[ R_{Mt+1}^2 | R_{Mt+1} \right], F_{1t+1} \right) = 0. \]

Solving this equation gives

\[ a = \frac{Cov_t \left( R_{Mt+1}^2, R_{Mt+1} \right)}{Var_t \left( R_{Mt+1} \right)}. \]
Then,

\[ EL_t \left[ R_{M+1}^2 | R_{M+1} \right] = \frac{Cov_t \left( R_{M+1}^2, R_{M+1} \right)}{Var_t \left( R_{M+1} \right)} \left( R_{M+1} - E_t[R_{M+1}] \right). \]  

(35)

But

\[ Cov_t \left( R_{M+1}^2, R_{M+1} \right) = Cov_t \left( (R_{M+1} - E_t[R_{M+1}])^2, R_{M+1} \right) \]
\[ = Cov_t \left( (R_{M+1} - E_t[R_{M+1}])^2 + 2(R_{M+1} - E_t[R_{M+1}]) E_t[R_{M+1}], R_{M+1} \right) \]
\[ = Cov_t \left( (R_{M+1} - E_t[R_{M+1}])^2, R_{M+1} \right) + 2E_t[R_{M+1}] Var_t \left( R_{M+1} \right) \]
\[ = E_t (R_{M+1} - E_t[R_{M+1}])^3 + 2E_t[R_{M+1}] Var_t \left( R_{M+1} \right). \]

Therefore,

\[ EL_t \left[ R_{M+1}^2 | R_{M+1} \right] = E_t[R_{M+1}] + \frac{(R_{M+1} - E_t[R_{M+1}]) \cdot E_t \left[ (R_{M+1} - E_t[R_{M+1}])^3 \right]}{Var_t \left( R_{M+1} \right)} + 2 \left( R_{M+1} - E_t[R_{M+1}] \right) \cdot E_t[R_{M+1}]. \]

This ends the proof. ■

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

\[ Var_t^* \left( R_{M+1} \right) = R_{ft} E_t[m_{M+1} (R_{M+1} - R_{ft})^2], \]

where

\[ (R_{M+1} - R_{ft})^2 = R_{M+1}^2 + R_{ft}^2 - 2R_{M+1}R_{ft}. \]  

(36)

But the squared market return can be rewritten as

\[ R_{M+1}^2 = EL_t \left[ R_{M+1}^2 | R_{M+1} \right] + \varepsilon_{t+1}, \text{ where } E_t[\varepsilon_{t+1}] = 0. \]

We replace this last expression into (36) and get

\[ (R_{M+1} - R_{ft})^2 = EL_t \left[ R_{M+1}^2 | R_{M+1} \right] + \varepsilon_{t+1} + R_{ft}^2 - 2R_{M+1}R_{ft} \]
\[ = E_t R_{M+1}^2 + \frac{(R_{M+1} - E_t R_{M+1}) E_t(R_{M+1} - E_t R_{M+1})^3}{Var_t \left( R_{M+1} \right)} + \]
\[ + 2 (R_{M+1} - E_t R_{M+1}) E_t R_{M+1} + \varepsilon_{t+1} + R_{ft}^2 - 2R_{M+1}R_{ft}. \]
Therefore,

\[ Var^*_t (R_{Mt+1}) = E_t R^2_{Mt+1} + \frac{R_{ft}(1 - \frac{1}{R_{ft}} E_t R_{Mt+1}) E_t (R_{Mt+1} - E_t R_{Mt+1})^3}{Var_t (R_{Mt+1})} + \]

\[ + 2R_{ft}(1 - \frac{1}{R_{ft}} E_t R_{Mt+1}) E_t R_{Mt+1} + R_{ft}Cov (m_{t+1}, \varepsilon_{t+1}) - R^2_{ft} \]

\[ = E_t R^2_{Mt+1} - P^{(1)}_{Mt}E_t (R_{Mt+1} - E_t R_{Mt+1})^3 + \]

\[ + 2R_{ft}(1 - \frac{1}{R_{ft}} E_t R_{Mt+1}) E_t R_{Mt+1} + R_{ft}Cov (m_{t+1}, \varepsilon_{t+1}) - R^2_{ft} \]

\[ = (E_t r^2_{Mt+1} - 2 (E_t r_{Mt+1})^2) - P^{(1)}_{Mt}E_t (R_{Mt+1} - E_t R_{Mt+1})^3 + R_{ft}Cov (m_{t+1}, \varepsilon_{t+1}) \]

\[ = Var_t (R_{Mt+1}) \left( 1 - \left( P^{(1)}_{Mt} \right)^2 Var_t (R_{Mt+1}) \right) \]

\[-P^{(1)}_{Mt}E_t (R_{Mt+1} - E_t R_{Mt+1})^3 + R_{ft}Cov (m_{t+1}, \varepsilon_{t+1}) \].

This ends the proof. ■

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

Then,

\[
\begin{bmatrix}
\text{Log} (m_{t+1}) \\
\text{Log} R_{Mt+1}
\end{bmatrix}
\sim N
\begin{bmatrix}
\mu_{mt} \\
\mu_{Mt}
\end{bmatrix}
, 
\begin{bmatrix}
\sigma^2_{mt} & \sigma_mrt \\
\sigma_mrt & \sigma^2_{Mt}
\end{bmatrix}.
\]

Let us denote

\[ c_{mt} = E_t m_{t+1} R^2_{Mt+1}. \]

The market return risk neutral variance \(Var^*_t (R_{Mt+1})\) is

\[ Var^*_t (R_{Mt+1}) = E^*_t R^2_{Mt+1} - R^2_{ft}, \quad \text{with } E^*_t R^2_{Mt+1} = R_{ft} E_t m_{t+1} R^2_{Mt+1}. \]

We know that

\[ \text{Log} (m_{t+1} R^2_{Mt+1}) = \text{Log} (m_{t+1}) + 2 \text{Log} (R_{Mt+1}). \]

Therefore,

\[ E_t[m_{t+1} R^2_{Mt+1}] = \exp (\mu_{mt} + 2 \mu_{Mt} + 0.5 \sigma^2_{Mt} + 2 \sigma^2_{mrt}) \]

\[ = \exp (-\mu_{mt} - 0.5 \sigma^2_{t}) \exp (2 \mu_{Mt} + 2 \sigma^2_{Mt}) \exp (-2 \mu_{Mt} - \sigma^2_{Mt}) \cdot \]

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

But \(E_t[m_{t+1} \cdot R_{Mt+1}] = 1\) is equivalent to

\[ \exp (\mu_{mt} + \mu_{Mt} + 0.5 \sigma^2_{t} + 0.5 \sigma^2_{Mt} + \sigma_{mrt}) = 1, \]

34
and therefore
\[
E_t[m_{t+1}R_{Mt+1}^2] = \exp \left( -\mu_{mt} - 0.5\sigma_t^2 \right) \exp \left( 2\mu_{Mt} + 2\sigma_{Mt}^2 \right) \exp \left( -2\mu_{Mt} - \sigma_{Mt}^2 \right)
= R_{ft} \frac{ER_{Mt+1}^2}{(E_tR_{Mt+1})^2}.
\]

Consequently,
\[
Var_t^* (R_{Mt+1}) = R_{ft}^2 \frac{E_tR_{Mt+1}^2}{(E_tR_{Mt+1})^2} - R_{ft}^2 = Var_t (R_{Mt+1}) \left( \frac{R_{ft}}{E_tR_{Mt+1}} \right)^2 < Var_t (R_{Mt+1}).
\]

\section*{Proof of Theorem 3.8.}
Assume that
\[
m_{t+1} = \nu_{0t} + \nu_{1t}R_{Mt+1} + \nu_{2t}R_{Mt+1}^2.
\]
Then,
\[
Cov_t (m_{t+1}, \varepsilon_{t+1}) = \nu_{2t}Cov_t \left( R_{Mt+1}^2, \varepsilon_{t+1} \right).
\]
But
\[
Cov_t \left( R_{Mt+1}^2, \varepsilon_{t+1} \right) = Cov_t \left( R_{Mt+1}^2 - (R_{Mt+1} - E_tR_{Mt+1}) Cov_t \left( R_{Mt+1}^2, R_{Mt+1} \right) \right)
= Var_t \left( R_{Mt+1}^2 \right) - \frac{Cov_t^2 \left( R_{Mt+1}^2, R_{Mt+1} \right)}{Var_t \left( R_{Mt+1} \right)}
= Var_t \left( R_{Mt+1}^2 \right) \left[ 1 - \frac{Cov_t^2 \left( R_{Mt+1}^2, R_{Mt+1} \right)}{Var_t \left( R_{Mt+1} \right) Var_t \left( R_{Mt+1} \right)} \right]
= Var_t \left( R_{Mt+1}^2 \right) \left[ 1 - \rho_t^2 \left( R_{Mt+1}^2, R_{Mt+1} \right) \right].
\]
References


Figure 1: Price of coskewness inferred from a simulated time-series of the Factor GARCH-in-mean used 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 a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004).
Figure 3: Risk premium on the squared net return inferred from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004)
Figure 4: Quadratic pricing kernel inferred from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004). In the upper graph, we plot the pricing kernel $m_{t+1}$ as a function of $t+1$ and $r_{Mt+1}$. In the lower graph 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 from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004). In the upper graph, we plot the pricing kernel $m_{t+1}$ as a function of $t+1$ and $r_{Mt+1} = x$. In the lower graph, we plot the average pricing kernel $\Sigma_{t=1}^{T} \frac{1}{T} m_{t+1}$. 
<table>
<thead>
<tr>
<th>Equations</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1t+1}$</td>
<td>$c_t$</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
</tr>
<tr>
<td>$Y_{2t+1}$</td>
<td>$0.0056 - 162.65 (e_{1t})^2$</td>
</tr>
<tr>
<td></td>
<td>(0.0006) (0.0001)</td>
</tr>
<tr>
<td>$Y_{3t+1}$</td>
<td>$0.0188 - 58.02 (e_{1t})^2$</td>
</tr>
<tr>
<td></td>
<td>(0.0083) (0.0003)</td>
</tr>
<tr>
<td>$Var_t(e_{1t+1})$</td>
<td>constant</td>
</tr>
<tr>
<td></td>
<td>0.000019</td>
</tr>
<tr>
<td></td>
<td>0.000018</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>0.000014</td>
</tr>
<tr>
<td></td>
<td>(0.000002)</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.006134</td>
</tr>
<tr>
<td></td>
<td>(0.00103)</td>
</tr>
<tr>
<td>$\sigma_{13} = -0.0564$</td>
<td>$\sigma_{12} = 3.182$</td>
</tr>
<tr>
<td></td>
<td>(0.1425)</td>
</tr>
</tbody>
</table>

Table 1: This table reproduces the results of the Factor GARCH in mean estimated by Bekaert and Liu (2004). Here $\delta_2 = Var_t(e_{2t+1})$ and $\delta_3 = Var_t(e_{3t+1})$. 
averages over time
\[ \eta_t \quad 0.64\% \]
\[ \lambda_{2t} \quad 4.25 \]
\[ \lambda_{2t} - \lambda_{2t-1} \quad 0 \]
\[ \lambda_{2t}^{HS} \quad -67.82 \]
\[ \lambda_{2t}^{HS} - \lambda_{2t-1}^{HS} \quad 0 \]

standard deviations over time
\[ \lambda_{2t} \quad 4.06 \]
\[ \lambda_{2t}^{HS} \quad 7.45 \]
\[ \lambda_{2t} - \lambda_{2t-1} \quad 4.93 \]
\[ \lambda_{2t}^{HS} - \lambda_{2t-1}^{HS} \quad 8.75 \]

Table 2: Summary statistics for our Monte-Carlo simulation