BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2007-47

Implications of Asymmetry Risk for Portfolio Analysis and Asset Pricing

by Fousseni Chabi-Yo, Dietmar Leisen, and Eric Renault

Bank of Canada Working Paper 2007-47

August 2007

Implications of Asymmetry Risk for Portfolio Analysis and Asset Pricing

by

Fousseni Chabi-Yo[1], Dietmar Leisen[2], and Eric Renault[3]

[1]Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9

[2]Faculty of Law and Economics
Johannes Gutenberg-Universität Mainz

[3]Department of Economics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599-3305
and CIRANO, CIREQ, Montréal
renault@email.unc.edu

ISSN 1701-9397

Acknowledgements

We thank Scott Hendry, Kenneth Judd and Wally Speckert for very helpful conversations. We thank seminar participants at the EC^2 Conference in Rotterdam for helpful comments.

Abstract

Asymmetric shocks are common in 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. The empirical relevance of this extension is documented in the context of the asymmetric GARCH-in-mean model of Bekaert and Liu (2004).

JEL classification: C52, D58, G11, G12
Bank classification: Financial markets; Market structure and pricing

Résumé

Les chocs asymétriques sont des phénomènes courants sur les marchés. La distribution des rendements des actifs financiers ne suit pas une loi normale et est asymétrique. Les auteurs étudient la composition du portefeuille d'agents hétérogènes qui ont des préférences à l'égard de la moyenne, de la variance et de l'asymétrie de la distribution des rendements, et calculent les prix d'équilibre des actifs. Ils démontrent la validité d'un théorème de séparation à trois portefeuilles, selon lequel les agents détiennent un portefeuille dont les rendements sont répartis de façon asymétrique en plus du portefeuille standard du marché; le facteur d'actualisation stochastique ne dépend linéairement que du rendement du marché et du carré de celui-ci. Les auteurs étendent le modèle d'évaluation des actifs financiers à trois moments conditionnels (moyenne, variance et asymétrie) de Harvey et Siddique (2000) pour y inclure une variance du rendement du marché neutre à l'égard du risque et toujours supérieure à zéro. Ils testent empiriquement la pertinence de cette extension au moyen du modèle GARCH-M asymétrique de Bekaert et Liu (2004).

Classification JEL : C52, D58, G11, G12
Classification de la Banque : Marchés financiers; Structure de marché et fixation des prix

1. Introduction

Asymmetric shocks are common in 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 for skewness risks and led 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 agents' portfolio holdings using risk-tolerance and a term we call skew-tolerance which contains the third derivative of agents' utility functions. 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 the skewness portfolio. Among all portfolios, the skewness portfolio is the portfolio with a return "closest" to that of the 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 difference between the 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 coskewness 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 these 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 that 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 that justify this extension base their proofs on assumed separation and aggregation results or on ad-hoc truncation of a Taylor-series expansion for the representative agent 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.

It is important to stress that this aggregation result is not a trivial consequence of standard complete market arguments. Actually, when only considering linear portfolios in primitive asset returns, markets are not complete in terms of hedging squared market return. It turns out that, due to a preference for positive skewness, tracking the squared market return is of interest for investors. While higher order small noise expansions are beyond the scope of this paper, it can be shown (see Chabi-Yo, Ghysels and Renault (2007)) that, when taking into account investors' preferences for low kurtosis, our asset pricing model depends on the cross-sectional variance of investors' skew-tolerances. Therefore, heterogeneity of investor preferences matters for equilibrium asset pricing, precisely because nonlinear payoffs are relevant risks for investors which are not perfectly hedged.

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 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 parameter estimates of Bekaert and Liu(2004). Moreover, our simulations also suggest that neglecting the market risk neutral variance – as it has been, e.g., in Harvey and Siddique (2000) – leads to a severe underestimation of the 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 the Taylor expansion of the utility function one step further. 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 σ and R_f is the gross return on the riskless (safe) security. The σ 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 Σ is a given symmetric and positive definite matrix[4]. For future reference, we denote by

$$\Gamma_k = E\left[YY^{\perp}Y_k\right]$$

the matrix of covariances between Y_k and cross-products Y_iY_j, $i,j=1,...,n$. Typically, asymmetry in the joint distribution of returns means that at least some matrices Γ_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 σ she then seeks to determine portfolio holdings $(\omega_i)_{1\leq i\leq n}\in\mathbb{R}^n$ that maximize her expected utility

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

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 risk σ. The question we ask is then the following: to what extent does a Taylor approximation of u allow us to understand the local behaviour of the shares $\omega_i(\sigma)$ in the neighborhood of the zero risk $\sigma=0$? Put differently, we want to characterize for $i=1,...,n$ the quantities:

$$\omega_i(0)=\lim_{\sigma\longrightarrow 0^+}\omega_i(\sigma) \text{ and } \omega_i^{'}(0)=\lim_{\sigma\longrightarrow 0^+}\omega_i^{'}(\sigma). \tag{3}$$

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:

$$u^*(W)=u(R_f)+u'(R_f)(W-R_f)+\frac{u''(R_f)}{2!}(W-R_f)^2+\frac{u'''(R_f)}{3!}(W-R_f)^3. \tag{4}$$

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^*:

$$\max_{(\omega_i^*)_{1\leq i\leq n}} Eu^*\left(R_f + \sum_{i=1}^{n} \omega_i^* \cdot (R_i(\sigma) - R_f)\right) \tag{5}$$

For $i = 1, ..., n$ the terms $\omega_i^*(0)$ and $\omega_i^{*\prime}(0)$ 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:*

$$\omega_i(0) = \omega_i^*(0) \;\; \text{and} \;\; \omega_i^{'}(0) = \omega_i^{*\prime}(0) \;\; \text{for all} \; i = 1, ..., n.$$

The intuition behind this theorem is that in the limit case $\sigma \rightsquigarrow 0$:

1. The optimal shares of wealth invested $\omega_i(0)$, $i = 1, ..., n$ 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 σ, $\omega_i'(0)$, $i = 1, ..., n$ 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 denote by

$$\tau = -\frac{u^{'}(R_f)}{u''(R_f)} \text{ and } \rho = \frac{\tau^2}{2}\frac{u^{'''}(R_f)}{u^{'}(R_f)} \tag{6}$$

the risk-tolerance coefficient and the skew-tolerance coefficient of the agent.

Of course, the risk-tolerance coefficient τ is assumed to be positive, to capture risk aversion, while the skew-tolerance coefficient ρ 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 under an assumption of joint normality of returns.

Theorem 2.2 *In the limit case $\sigma \longrightarrow 0$, 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 τ in the limit case $\sigma \rightsquigarrow 0$. 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:

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

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

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

Then,

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

so that

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

This denotes, by unit of scaling risk σ, 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 behaviour 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 the risk premium potentially erase this effect. To see this, let us define the coskewness of asset k in portfolio ω as:

Definition 2.3 *The coskewness of asset k in portfolio ω is:*

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

Note that coskewness does not depend on the scale of risk σ. We will see below that this notion of coskewness 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. We can show that investors prefer positive skewness, component-wise. This assertion is justified by the fact that the average

$$\begin{aligned} \sum_{k=1}^{n}\omega_k c_k(\omega) &= \frac{1}{\sigma}\frac{E\left[\left(\omega^{\perp}(R - ER)\right)^3\right]}{Var\left(\omega^{\perp}(R - ER)\right)} \\ &= \frac{1}{\sigma}Skew\left(\omega^{\perp}\cdot(R - ER)\right)\cdot\left(Var\left(\omega^{\perp}\cdot(R - ER)\right)\right)^{\frac{1}{2}} \end{aligned}$$

is positive if and only if the portfolio return is positively skewed. We then 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:*

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

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

In other words, up to variations $a^{'}(0)$ of risk premia in equilibrium, a positive coskewness of asset k will have a positive effect on the demand for this asset with respect to common mean-variance formulas. This positive effect will be all the more pronounced when the skew-tolerance coefficient ρ 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 $\omega_s(0) = \left(\omega^{'}_{si}(0)\right)_{1 \le i \le n}$ her holdings in each of these assets; her preferences are characterized by a Von Neumann-Morgenstern utility function u_s and associated preference coefficients:

$$\tau_s = -\frac{u'_s(R_f)}{u''_s(R_f)} \text{ and } \rho_s = \frac{\tau_s^2}{2}\frac{u'''_s(R_f)}{u'_s(R_f)}. \tag{9}$$

From theorems 2.2 and 2.4 we get that:

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

Note that these formulas correspond to the case where each of the S agents would get a unit of 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, which are average risk tolerance $\overline{\tau}$ and average skew tolerance $\overline{\rho}$, such that:

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

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

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

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

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

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

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

Theorem 2.5 *In the limit case $\sigma \rightsquigarrow 0$, the equilibrium risk premium vector $a\left(\sigma\right)$ is such that the average portfolio $\overline{\omega}$ is optimal for the average investor: $\overline{\omega} = \overline{\tau}\Sigma^{-1}a\left(0\right)$, that is*

$$a\left(0\right) = \frac{1}{\overline{\tau}}\Sigma\overline{\omega}.$$

Its slope in the neighborhood of zero is given by:

$$a'_k\left(0\right) = -\overline{\rho}P^2\left(0\right)c_k\left(\overline{\omega}\right) \qquad \textit{for } k = 1, ..., K.$$

Theorem 2.5 must be interpreted as a new asset pricing model. While approximating risk premia by their limit values $a_i\left(0\right)$ would clearly lead to the Sharpe-Lintner CAPM, approximating them by higher order expansions $a_i\left(0\right) + \sigma a'_i\left(0\right)$ results in a new mean-variance-skewness asset pricing model. To see this, let us assume for notational simplicity that the total supply of the 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

$$\beta = \frac{Cov\left(R, R_M\right)}{Var\left(R_M\right)} = \frac{\Sigma\overline{\omega}}{\overline{\omega}^{\perp}\Sigma\overline{\omega}} \tag{14}$$

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 $a\left(0\right)$ of the vector of equilibrium risk premium is proportional to the vector of market betas, with a proportionality coefficient $\frac{Var(R_M)}{\overline{\tau}} = \overline{\tau}\sigma^2 P^2\left(0\right)$, which is itself increasing with market risk and market risk aversion.

The new contribution of Theorem 2.5 is encapsulated in the value

$$a_k'\left(0\right) = -\overline{\rho}P^2\left(0\right)c_k\left(\overline{\omega}\right). \tag{15}$$

It states that insofar as utility functions are not quadratic ($\overline{\rho} \neq 0$), asset k exhibits a positive skewness risk premium $a_k'\left(0\right)$ when its coskewness $c_k\left(\overline{\omega}\right)$ 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 effect. This compensation is captured through a risk premium function $a\left(\sigma\right)$ 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 $a\left(0\right)$ and $a'\left(0\right)$:

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

$$\begin{aligned} \omega_s\left(0\right) &= \frac{\tau_s}{\overline{\tau}}\overline{\omega}, \text{ and} \\ \sigma\omega_s'\left(0\right) &= \tau_s\left[\rho_s - \overline{\rho}\right]P^2\left(0\right)\Sigma^{-1}c\left(\overline{\omega}\right) = \frac{\tau_s}{\overline{\tau}^2}\left(\rho_s - \overline{\rho}\right)\Theta_{skew}. \end{aligned}$$

where

$$\Theta_{skew} = \left(Var\left(R\right)\right)^{-1} Cov\left(R, \left(R_M - ER_M\right)^2\right)$$

is called the skewness portfolio.

Theorem 2.6 states that in the limit case $\sigma \rightsquigarrow 0$, 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 τ_s with respect to the average risk-tolerance. 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 coskewness $c_k(\overline{\omega})$ will have a positive effect on the demand for asset k by agent s if and only if his skew tolerance coefficient is more than the average $\overline{\rho}$. On the contrary, if $\rho_s < \overline{\rho}$, the positive effect of asset k coskewness on its market price results in more than a compensation of an investor's preference for positive skewness.

Interestingly, the effect of individual preferences for skewness manifests itself only through one portfolio Θ_{skew} called the skewness portfolio. Note that the payoff of the skewness mimicks, up to a constant, the affine regression $EL\left[(R_M - ER_M)^2 |R\right]$ of $(R_M - ER_M)^2$ on the vector R of asset returns:

$$EL\left[(R_M - ER_M)^2 |R\right] - Var(R_M) = \Theta_{skew}^{\perp} R - E\left(\Theta_{skew}^{\perp} R\right) \tag{16}$$

In other words, Theorem 2.6 is nothing but a three funds theorem. Due to heterogeneity of their skewness preferences, investors hold not only the risk-free asset and the market portfolio but also a position in the skewness portfolio. The standard two-funds theorem is maintained if and only if one of the following two conditions are fulfilled. Either, all market coskewnesses are zero (and a fortiori market skewness is zero) as is the case with normal returns. In this case, the skewness portfolio has just a constant payoff. 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 $\frac{\tau_s}{\overline{\tau}}$ with respect to an average investor.

By contrast, in the case of heterogeneous skewness preferences and non degenerate skewness portfolios, the weight given to the skewness portfolio is defined by the spread $(\rho_s - \overline{\rho})$ 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 $E\left[m(\sigma) \cdot \left(R_f + \sigma^2 a_i(\sigma) + \sigma Y_i\right)\right] = 1$ for $i = 1, ..., n$. We are then able to re-express theorem 2.5 in terms of SDF:

Theorem 2.7 *The random variable:*

$$m(\sigma) = \frac{1}{R_f} - \frac{1}{R_f \overline{\tau}} \left(R_M(\sigma) - ER_M(\sigma)\right) + \frac{\overline{\rho}}{R_f \overline{\tau}^2} \left(\Theta_{skew}^{\perp} R - E\left(\Theta_{skew}^{\perp} R\right)\right)$$

is a SDF consistent with the 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 2.7 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 coskewness 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 2.7 that:

Theorem 2.8 *The asset pricing model associated with 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:*

$$ER_i - R_f = \frac{1}{\overline{\tau}}(VarR_M)\beta_i - \frac{\overline{\rho}}{\overline{\tau}^2}\left(Var\left(\Theta_{skew}^{\perp}R\right)\right)\gamma_i \quad \textit{for } i = 1, ..., n,$$

where:

$$\begin{aligned}\beta_i &= \frac{Cov(R_i, R_M)}{Var(R_M)}, \\ \gamma_i &= \frac{Cov\left(R_i, \Theta_{skew}^{\perp}R\right)}{Var\left(\Theta_{skew}^{\perp}R\right)}.\end{aligned}$$

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 $\Theta_{skew}^{\perp}R$. The parameters γ are tightly related to coskewness, since

$$\gamma_i Var\left(\Theta_{skew}^{\perp}R\right) = Cov\left(R_i, \Theta_{skew}^{\perp}R\right) = c_i(\overline{\omega})\,Var\left((R_M - ER_M)^2\right).$$

Non-zero beta coefficients show up provided that the coskewness coefficients are non-zero. Moreover, the price of this additional factor is proportional to the average skewness tolerance $\overline{\rho}$. 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 regarding 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 2.8, which is also based on

$Cov\left(R_i, \Theta_{skew}^{\perp} R\right) = Cov\left(R_i, (R_M - ER_M)^2\right)$ in addition to common betas. Note that if all agents were endowed with the same utility function u, $\frac{\overline{\rho}}{\overline{\tau}^2}$ would be $\frac{u'''(R_f)}{2u'(R_f)}$ as usually derived from Taylor expansion of the representative agent utility.

To conclude, it is worth noting that additional pricing factors may be introduced by considering more accurate small noise expansions, that is expanding the risk premium function at higher orders. Straightforwardly, a second order expansion $a(\sigma) = a(0) + \sigma a'(0) + \frac{1}{2}\sigma^2 a''(0)$ would lead us to consider as pricing factors not only the squared market return, but also the cubic market return. While the former captures the market price of positive skewness, the latter concerns the market price for low kurtosis. Such an extension has been put forward by Dittmar (2002). However, as announced in the introduction, it can be shown that the representative agent paradigm used by Dittmar (2002) overlooks an additional pricing factor implied by heterogeneity of skewness preferences. More precisely (see Chabi-Yo, Ghysels and Renault (2007)), the additional factor in the pricing kernel is $R_M\left(\Theta_{skew}^{\intercal} R\right)$ with a coefficient proportional to the cross-sectional variance $\overline{\rho^2} - \overline{\rho}^2$ of skew-tolerances. In other words, investors' heterogeneity matters for asset pricing because the squared market return cannot be perfectly hedged: $\Theta_{skew}^{\intercal} R \neq R_M^2$ and in turn the pricing factor $R_M\left(\Theta_{skew}^{\intercal} R\right)$ is different from the cubic market return.

Note also that such a pricing kernel specification potentially allows one to estimate the characteristics of the cross-sectional distribution of individual preferences like mean $\overline{\rho}$ and variance $\overline{\rho^2} - \overline{\rho}^2$ even when only using aggregate price data. However, rather than a general theory of nonlinear pricing kernels implied by small noise expansions with heterogeneous investors, the focus of interest of this paper is the quadratic SDF that is observationally equivalent to the SDF of Theorem 2.7.

3. Quadratic SDF

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:

$$m_{t+1} = \nu_{0t} + \nu_{1t} R_{Mt+1} + \nu_{2t} R_{Mt+1}^2. \tag{17}$$

It actually suffices to revisit our Section 2 above with a conditional viewpoint to see Theorem 2.7 as a theoretical justification of (17). Then, the coefficients ν_{0t}, ν_{1t} and ν_{2t} are functions of the conditioning information I_t at time t. To identify (17) with Theorem 2.7, note that m_{t+1} may always be replaced by its affine regression on primitive asset returns, giving rise to the skewness portfolio.

From Theorem 2.7, we interpret the factor coefficients as:

$$\nu_{2t} = \frac{1}{R_{ft}} \frac{\overline{\rho}}{\overline{\tau}^2} > 0, \tag{18}$$

and

$$\nu_{1t} = -\frac{1}{R_{ft}} \frac{1}{\overline{\tau}} - 2 \frac{1}{R_{ft}} \frac{\overline{\rho}}{\overline{\tau}^2} E_t \left[R_{Mt+1} \right] < 0. \tag{19}$$

It is worth characterizing the role of the two factors R_{Mt+1} and R_{Mt+1}^2 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

$$\frac{1}{R_{ft}} E_t \left[r_{it+1} \right] + \nu_{1t} Cov_t \left(r_{it+1}, R_{Mt+1} \right) + \nu_{2t} Cov_t \left(r_{it+1}, R_{Mt+1}^2 \right) = E_t \left[r_{it+1} m_{t+1} \right] = 0,$$

or, using the market net excess return, we get

$$\frac{1}{R_{ft}} E_t \left[r_{it+1} \right] + \left(\nu_{1t} + 2 R_{ft} \nu_{2t} \right) Cov_t \left[r_{it+1}, r_{Mt+1} \right] + \nu_{2t} Cov_t \left[r_{it+1}, r_{Mt+1}^2 \right] = 0,$$

that is:

$$E_t \left[r_{it+1} \right] = \lambda_{1t} Cov_t \left[r_{it+1}, r_{Mt+1} \right] - \lambda_{2t} Cov_t \left[r_{it+1}, r_{Mt+1}^2 \right],$$

with:

$$\lambda_{1t} = -R_{ft}\left(\nu_{1t} + 2R_{ft}\nu_{2t}\right) \text{ and } \lambda_{2t} = R_{ft}\nu_{2t}.$$

If ν_{1t} and ν_{2t} are interpreted in terms of preferences of an average investor as in (18) and (19), we deduce:

$$\lambda_{1t} = \frac{1}{\overline{\tau}} + 2\frac{\overline{\rho}}{\overline{\tau}^2}\left(E_t\left[R_{Mt+1}\right] - R_{ft}\right) \text{ and } \lambda_{2t} = \frac{\overline{\rho}}{\overline{\tau}^2}.$$

Note that λ_{2t} is a structural by invariant in the sense that it is only time-varying through the value of preference parameters computed from the derivatives of the utility function at R_{ft}. The term λ_{2t} should be non-negative and more positive when preference for skewness is high. Similarly, λ_{1t} 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 *Let $r_{it+1} = R_{it+1} - R_f$ and $r_{Mt+1} = R_{Mt+1} - R_f$ be the net risky return and market return respectively. Under the maintained assumption (17) of a quadratic SDF, net expected returns are given by:*

$$E_t[r_{it+1}] = \lambda_{1t}Cov_t\left(r_{it+1}, r_{Mt+1}\right) - \lambda_{2t}Cov_t\left(r_{it+1}, r_{Mt+1}^2\right),$$

where λ_{1t} and λ_{2t} are expected to be non negative and increasing respectively with aggregate risk aversion and skewness tolerance. λ_{1t} and λ_{2t} define respectively the market price of market risk and of coskewness.

Note that the standard expression of the market price of market risk is modified by the fact that the price of coskewness is taken into account. λ_{1t} has actually two components which are both increasing with the average risk aversion, $1/\overline{\tau}$ and 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 λ_{1t} larger:

Corollary 3.2 *Under the assumptions of Theorem 3.1*

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

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

$$Cov_t\left(r_{Mt+1}, r_{Mt+1}^2\right) = E_t r_{Mt+1}^3 - \left(E_t r_{Mt+1}\right)\left(E_t r_{Mt+1}^2\right).$$

This notion of market coskewness has already been put forward by Harvey and Siddique (2000) and Theorem 3.1 and Corollary 3.2 correspond to their formula (7).

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

$$E_t[r_{it+1}] = \Big(\lambda_{1t} Var_t(r_{Mt+1})\Big)\beta_{iMt} - \Big(\lambda_{2t} Var_t(r_{Mt+1}^2)\Big)\delta_{iMt}, \tag{20}$$

or

$$E_t[r_{it+1}] = E_t\left[r_{Mt+1}\right]\beta_{iMt} - \lambda_{2t} Var_t(r_{Mt+1}^2)\cdot\left(\delta_{iMt} - \delta_{MMt}\beta_{iMt}\right), \tag{21}$$

where

$$\beta_{iMt} = \frac{Cov_t\left(r_{it+1}, r_{Mt+1}\right)}{Var_t(r_{Mt+1})},\ \delta_{iMt} = \frac{Cov_t\left(r_{it+1}, r_{Mt+1}^2\right)}{Var_t(r_{Mt+1}^2)}.$$

The term β_{iMt} is the standard market beta while the beta coefficient with respect to the squared market return is δ_{iMt}; up to a change in normalization, it corresponds to the measure of coskewness already introduced in Section 2. Therefore, the result of equation (21) matches exactly that of Theorem 2.8 under a conditionality.

As already shown in Section 2, the beta pricing model (20) with a second beta coefficient interpreted in terms of coskewness 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

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^2_{Mt+1} which can be mimicked by a linear function of r_{Mt+1}, conditionally on available information at time t:

$$r^2_{Mt+1} = \frac{Cov_t\left(r^2_{Mt+1}, r_{Mt+1}\right)}{Var_t\left(r_{Mt+1}\right)} r_{Mt+1} + r^{(2)}_{Mt+1}.$$

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

Theorem 3.3 *We have:*

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

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 λ_{2t} significantly different from zero; second, the return of asset i is significantly correlated to that part of r^2_{Mt+1} which is orthogonal to r_{Mt+1}.

Normalization in terms of the beta coefficients 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: $\lambda_{1t} = P^{(1)}_{Mt}$ with

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

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

$$P^{(2)}_{Mt}\left(\eta_t\right) = \frac{E_t\left[\frac{r^2_{Mt+1}}{\eta_t}\right] - R_{ft}}{Var_t\left(\frac{r^2_{Mt+1}}{\eta_t}\right)} = \frac{E_t[r^2_{Mt+1}] - R_{ft}\eta_t}{Var_t\left(r^2_{Mt+1}\right)}\eta_t \tag{22}$$

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

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

$$\lambda_{2t} = \frac{\delta_{MMt} P^{(1)}_{Mt} - \frac{1}{\eta_t} P^{(2)}_{Mt}(\eta_t)}{1 - \rho^2_t\left(r_{Mt+1}, r^2_{Mt+1}\right)},$$

where according to (22), $P^{(2)}_{Mt}(\eta_t)$ *is the risk premium on the asset with payoff* r^2_{Mt+1} *and* $\rho^2_t\left(r_{Mt+1}, r^2_{Mt+1}\right)$ *denotes the square (conditional) linear correlation coefficient between* r_{Mt+1} *and* r^2_{Mt+1}.

Note that, from (22) we have:

$$\lim_{\eta_t \rightsquigarrow 0} \frac{P^{(2)}_{Mt}(\eta_t)}{\eta_t} = \frac{E_t[r^2_{Mt+1}]}{Var_t\left(r^2_{Mt+1}\right)}. \tag{23}$$

In this limit case, we get

$$\lambda_{2t} = \frac{\delta_{MMt} P^{(1)}_{Mt} - \frac{E_t[r^2_{Mt+1}]}{Var_t\left(r^2_{Mt+1}\right)}}{1 - \rho^2_t\left(r_{Mt+1}, r^2_{Mt+1}\right)}, \tag{24}$$

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

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

we expect that considering the limit case (23), i.e. considering $\eta_t = 0$, leads one to overestimate $\frac{P^{(2)}_{Mt}(\eta_t)}{\eta_t}$ and therefore to underestimate λ_{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 λ_{2t} of coskewness in the two-factor pricing relationship (21). If it is the case, we must realize that λ_{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

In order to shed more light on the difference between the general skewness-pricing formulas of Theorem 3.4 (jointly with Theorem 3.1 and Corollary 3.2) and the limit case (24) put forward by

Harvey and Siddique (2000), it is worth interpreting their implied assumption "$\eta_t = 0$" in terms of market risk neutral variance.

More precisely, the conditional risk neutral variance $Var_t^*(R_{Mt+1})$ of the market return is defined from the probability density function $R_{ft}m_{t+1}$:

$$\begin{aligned} Var_t^*(R_{Mt+1}) &= E_t\left(R_{ft}m_{t+1}R_{Mt+1}^2\right) - \left(E_t\left(R_{ft}m_{t+1}R_{Mt+1}\right)\right)^2 \\ &= R_{ft}\left(E_t\left(m_{t+1}R_{Mt+1}^2\right) - R_{ft}\right) \\ &= R_{ft}E_t m_{t+1}\left(R_{Mt+1} - R_{ft}\right)^2 \\ &= R_{ft}\eta_t. \end{aligned}$$

In other words, quantitative assessment of η_t is akin to the pricing of the "volatility contract" R_{Mt+1}^2 (see Bakshi, Kapadia and Madan (2003)) or the direct evaluation of the market risk neutral variance.

Typically, the risk neutral variance can be inferred from observed option prices (see Rosenberg (2000)). For instance, in a standard conditional log-normal setting, a simple extension of Brennan (1979) risk neutral valuation relationships (see Garcia, Ghysels and Renault (2003)) will give:

Theorem 3.5 *If* $(\log m_{t+1}, \log R_{Mt+1})$ *is jointly normal given the conditioning information,*

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

Put differently, in the case of conditional log-normality, the risk neutral variance will be smaller than the objective variance, even more so when the market risk premium is large. We want however to argue here that in the general case, pushing the risk neutral variance to zero as in Harvey and Siddique (2000) implicitly amounts to neglecting the possibly positive price of the component of the volatility contract which cannot be hedged by primitive asset returns.

More precisely, we can show:

Theorem 3.6

(i) The risk neutral variance $Var_t^*(R_{Mt+1})$ *of the market return is:*

$$Var_t(R_{Mt+1}) + R_{ft}\left(E_t\left(m_{t+1}\psi_{Mt+1}\right) - \frac{E_t\left(\psi_{Mt+1}\right)}{R_{ft}}\right) + R_{ft}E_t\left(m_{t+1}\varepsilon_{t+1}\right)$$

where $\psi_{Mt+1} = \Theta_{skew}^{\perp} R_{t+1}$ *is the payoff of the skewness portfolio defined in Section 2 and:*

$$\varepsilon_{t+1} = \left(R_{Mt+1} - ER_{Mt+1}\right)^2 - \left(\psi_{Mt+1} - E\psi_{Mt+1}\right).$$

(ii) With a quadratic SDF

$$m_{t+1} = \nu_{0t} + \nu_{1t} R_{Mt+1} + \nu_{2t} R_{Mt+1}^2$$

we have:

$$E_t\left[m_{t+1}\varepsilon_{t+1}\right] = \nu_{2t} Var_t\left(\varepsilon_{t+1}\right)$$

Theorem 3.6 confirms that the difference $Var_t^*\left(R_{Mt+1}\right) - Var_t\left(R_{Mt+1}\right)$ has two components:

The first component is determined by the risk premium on the skewness portfolio. As seen in Section 2, investors with a strong skewness tolerance will want to hold this portfolio and we may then expect a negative risk premium to be associated with it. To analyze further the sign of the risk premium on the skewness portfolio, let us apply Theorem 2.8 to $R_i = \Theta_{skew}^{\perp} R$. Straightforward computation gives

$$E\left(\Theta_{skew}^{\perp} R\right) - R_f = \frac{1}{\overline{\tau}} E\left(R_M - ER_M\right)^3 - \frac{\overline{\rho}}{\overline{\tau}^2} Var\left(\Theta_{skew}^{\perp} R\right).$$

In other words, only a rather unlikely strongly positive market skewness could prevent the risk premium on the skewness portfolio from being negative. Such a negative risk premium would explain why the risk neutral market variance may be much smaller than the objective one and even possibly zero as in Harvey and Siddique (2000).

However, the second component of the risk neutral variance, namely $E_t\left(m_{t+1}\varepsilon_{t+1}\right)$, should be positive with a quadratic SDF. Typically, $\nu_{2t} = \frac{1}{R_{ft}} \frac{\overline{\rho}}{\overline{\tau}^2}$ will be significantly positive in the case of high skewness tolerance, and appears to be multiplied by the part $Var_t\left(\varepsilon_{t+1}\right)$ of the total variance $Var_t\left(\left(R_{Mt+1} - E_t R_{Mt+1}\right)^2\right)$ which is not hedged by primitive assets.

This is the reason why we conclude that it is safer to consider a strictly positive risk neutral market variance and in turn a strictly positive coefficient η_t in Theorem 3.4.

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 (21):

$$E_t\left[r_{it+1}\right] = E_t\left[r_{Mt+1}\right]\beta_{iMt} - \lambda_{2t} Var_t(r^2_{Mt+1}) \cdot \left(\delta_{iMt} - \delta_{MMt}\beta_{iMt}\right). \tag{26}$$

The question is: does this asset pricing equation significantly deviate from standard CAPM? That is: should we maintain a significantly positive skewness premium λ_{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 coskewness with the market tend to be almost collinear across common portfolios, leading to marginally significant coskewness factors $(\delta_{imt} - \delta_{mmt}\beta_{imt})$. To shed more light on this identification issue, let us consider the (conditional) affine regression of asset i's net return on market return:

$$r_{it+1} = \alpha_{it} + \beta_{iMt} r_{Mt+1} + u_{it+1}. \tag{27}$$

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

$$Var_t\left(r^2_{Mt+1}\right) \cdot \left(\delta_{iMt} - \delta_{MMt}\beta_{iMt}\right) = Cov_t\left(u_{it+1}, r^2_{Mt+1}\right) = Cov_t\left(u_{it+1}, R^2_{Mt+1}\right). \tag{28}$$

Therefore, a positive sign for λ_{2t} can be identified only insofar as one can observe some asset returns r_{it+1} with positive (negative) coskewness $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$ and check that they command a lower (higher) expected return than explained by standard CAPM. The problem is that $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$ 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 $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$. However, as long as a linear approximation is valid, $Cov_t\left(u_{it+1}, r^2_{Mt+1}\right)$ is almost zero, which leads to:

$$Cov_t\left(r_{it+1}, r^2_{Mt+1}\right) \sim \beta_{iMt} Cov_t\left(r_{Mt+1}, r^2_{Mt+1}\right)$$

almost collinear with β_{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 λ_{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 in identifying the sign of λ_{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 nonlinearities 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 $Log\left(R_{Mt+1}\right)$ and logarithm of bond return $Log\left(R_{ft+1}\right)$:

$$Y_{t+1} = [Y_{1t+1}, Y_{2t+1}, Y_{3t+1}]' = [X_{t+1}, Log\left(R_{Mt+1}\right), Log\left(R_{ft+1}\right)]'.$$

The model assumes the dynamics

$$Y_{t+1} = c_t + AY_t + \Omega e_{t+1}, \tag{29}$$

where the coefficient c_{it} of c_t, $i = 1, 2, 3$, is an affine function of $Var_t\left[Y_{it+1}\right]$ 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 for the first one which follows an asymmetric GARCH(1,1):

$$Var_t\left[e_{1t+1}\right] = \delta_1 + \alpha\left(e_{1t}\right)^2 + \beta Var_{t-1}\left[e_{1t}\right] + \xi\left(Max\left[0, -e_{1t}\right]\right)^2. \tag{30}$$

To limit parameter proliferation, they assume that all the off-diagonal coefficients of the matrix Ω are zero except in the first column; in other words the consumption shock is the only factor. For the sake of normalization, the diagonal coefficients of Ω are fixed to the value 1. Table 2 gives the estimated parameters 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 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 $Var_t^*\left(R_{Mt+1}\right) = Var_t\left(R_{Mt+1}\right)\left[R_{ft}/E_t\left(R_{Mt+1}\right)\right]^2$, then of $\eta_t = Var_t^*\left(R_{Mt+1}\right)/R_{ft}$, of $\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t} = \frac{E_t[r_{Mt+1}^2] - Var_t^*(R_{Mt+1})}{Var_t\left(r_{Mt+1}^2\right)}$, of $P_{Mt}^{(2)}\left(\eta_t\right)$, and finally of λ_{2t} according to Theorem 3.4. We recall that the limit case put forward by Harvey and Siddique (2000) corresponds to the alternative formula:

$$\lambda_{2t}^{HS} = \frac{\gamma_{MMt} P_{Mt}^{(1)} - \frac{E_t r_{Mt+1}^2}{Var_t\left(r_{Mt+1}^2\right)}}{1 - \rho_t^2\left(r_{Mt+1}, r_{Mt+1}^2\right)}.$$

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 coskewness across portfolios. However, we vary exogenously both the asset i's beta and coskewness and compare the relative expected return error of Harvey and Siddique (2000) and the CAPM model with respect to the expected return decomposition given in our model (see equation 20).

The focus of our interest here is to get time-series of λ_{2t} and λ_{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 one 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

When drawing time series from the Bekaert and Liu (2004) estimated model described above, several experiments are performed.

First, a long time series for say a 500 month-long path, is informative in several respects. Due to the stationarity of the stochastic processes of interest, time-averages over 500 months allow us to compare conditional and unconditional quantities. Moreover, the dynamic features of the observed simulated path are of interest.

Of course, one could argue that our conclusions are only valid for one particular simulated path. This is the reason why we also perform an extensive Monte Carlo experiment by simulating 1000 sample paths, each one with a length of 500 months. This will allow us to show that cross-simulation variability is sufficiently small to ensure the robutness of conclusions drawn from one specific simulated path.

Note that, since conditional pricing is our focus of interest, it would be meaningless to want to assess it through averages over a large number of paths. For instance, volatility clustering must be assessed on a given sample path while averaging across paths would push volatility towards its constant unconditional mean. For the same reason, we are going to see that the market price of coskewness is significant only with a conditional point of view. This is the reason why we study in particular a single 500 month-long simulated path throughout this section.

Figure 1 displays the associated sample paths for both the market price of coskewness λ_{2t} (as characterized by Theorem 3.4) and of its limit value λ_{2t}^{HS} defined by (24). See Table 2 for same summary statistics.

A first striking observation is that while the λ_{2t} path confirms that the market price of coskewness is positive (4.25 on average), the λ_{2t}^{HS} path diaplays some implausibly hugely negative prices of coskewness (-67.82 on average). In other words, neglecting the price η_t of squared net returns leads to a severe underestimation of the price of coskewness, so severe that it may reverse the direction of the effect of coskewness in asset prices.

This conclusion is statistically significant and not related to a specific sample path. According to Table 3, the cross-paths standard deviation of the time-average of λ_{2t} over 1000 simulated paths is only 0.2758 (for a general average of 4.2917) while it is only 0.5388 for λ_{2t}^{HS} (for a general average of -67.7317).

As explained in Section 3.1, the reason why λ_{2t}^{HS} is too small is that it overestimates the risk premium $\frac{P_{Mt}^{(2)}(\eta_t)}{\eta_t}$ on the squared market return. It is worth noticing that the correct formula (22) always leads to a nonnegative risk premium on the squared net market return as displayed on Figure 2. However, by contrast with the limit case considered by Harvey and Siddique (2000), the shadow price η_t of the squared net market return is significantly positive: 0.64% on average with a standard deviation (of time-averages over 1000 simulated paths) of only 0.00052%.

As already pointed out, the advantage of considering a specific simulated path is to enhance the differences between conditional and unconditional quantities. While the time series of λ_{2t} does show a positive average price of 4.25 for coskewness (4.2917 among 1000 paths), it comes with a standard deviation error of 0.2758 (average value of standard deviations over 1000 paths is equal to 4.06).

This result may explain why Barone-Adesi, Gagliardini and Urga (2004) had such difficulty in identifying 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 λ_{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 λ_{2t}^{HS} meaningful, since their standard error is only 7.45, which does not compensate for their negative average of -67.82.

Overall, we conclude that there should be a positive price for coskewness, but that it is not high and difficult to identify 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 2) which, by Theorem 3.4 results in a decrease in the price λ_{2t}. This does not mean that skewness is worthless, but only that, by (26), 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.

Of course, one ought to realize that quadratic pricing kernels cannot be more than local approximations of a true pricing kernel, for instance, as 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 λ_{2t} will become increasing when the market returns exceed 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 surfaced in the empirical evidence documented by Dittmar (2002). Of course, a negative λ_{2t}, as in the case of the zero-price η_t approximation, would produce an even more bizarre behaviour, with an increasing pricing kernel for any value of the market return below its expectation.

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 η_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 from derivative asset prices, it is worth checking how it works on simulated paths. Figure 3 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 η_t (time average of 0.64%) which determines the coefficients λ_{1t} and λ_{2t} of the pricing kernel by application of Corollary 3.2 and Theorem 3.4. No paradoxical behaviour 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 one now increases the value of η_t, by fixing somewhat arbitrarily the price of the squared market return at the level 1.02, which in turn implies a time-varying η_t (with a time average of 1.56%), one gets the results shown in Figure 4. Then, one may observe that, by contrast with Figure 3, on the same range of values of the market return, the aforementioned increasing shape of the pricing kernel for large returns may show up. While Figures 3 and 4 are about time averages over one particular sample path, the general average and the standard deviation of these time averages are computed over 1000 sample paths. This leads to the pricing kernel plots with 5% confidence bounds provided in figures 5 and 6.

Finally, to assess the economic significance of the pricing improvements brought by our model, we compare it with both the limit case of Harvey and Siddique (2000) (HS hereafter) and with a standard CAPM as well. In order to do this, we plot as functions of characteristics a hypothetical asset i (coefficient beta-i and coskewness delta-i), the relative errors on expected returns between our model and CAPM (Figure 7) and between our model and HS (Figure 8). Not surprisingly, the CAPM risk premium may be too high in the case of a joint occurrence of positive betas and positive coskewness. Typically, our model predicts that a positive coskewness, since it meets investors' concern for positive skewness, may slightly erase the role of market beta as a measure of risk to be compensated. Note however that the relative pricing errors of CAPM do not exceed 1 or 2 per-cent. Pricing errors of the HS specification may be much more significant. They may be between 10 and 20 per-cent when a large positive coskewness is not well taken into account due to the negative sign λ_{2t}^{HS}.

5. Conclusion

This paper investigates the relevance of non-linear pricing kernels both at the theoretical and empirical level. We first show that considering pricing kernels that are quadratic functions of the market return is a well-founded approximation of actual expected utility behaviour, at least to characterize locally the demand for risky assets in the neighborhood of zero risk. Such quadratic pricing kernels disclose some pricing for skewness, but only through coskewness 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 emerges 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 coskewness 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 nonlinearities 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; in contrast with theirs, our model considers the price of the squared market return as a free parameter whose actual value might be backed out of observed derivative asset prices.

Although conditional, our study is purely statistical in the sense that investors only maximize a one-period utility function. Typically, while only conditional skewness of asset returns show 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 the leverage effect by Black (1976) and specific leverage-based dynamic risk premia should be the result of non-myopic intertemporal optimization behaviour of investors with preferences for skewness.

6. Appendix

PROOF OF THEOREMS 2.2 AND 2.4. The solution $\omega(\sigma) = (\omega_i(\sigma))_{1 \le i \le n}$ 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

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

Then, setting

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

this implies that $E\left[\frac{dh_i}{d\sigma}(\sigma)\right] = 0$ and so that $\lim_{\sigma \rightsquigarrow 0^+} E\left[\frac{dh_i}{d\sigma}(\sigma)\right] = 0$. 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 Σ of the vector Y of random variables and the definition of the risk 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

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

Writing this out and using definition (6) we get:

$$\begin{aligned} \sum_{i=1}^{n} \omega_i'(0)\, Cov(Y_i, Y_k) &= \frac{\rho}{\tau} \sum_{i=1}^{n} \omega_i^2(0)\, E[Y_i^2 Y_k] + 2\frac{\rho}{\tau} \sum_{i<j}^{n} \omega_i(0)\, \omega_j(0)\, E[Y_i Y_j Y_k] + \tau a_k'(0) \quad (32) \\ &= \frac{\rho}{\tau} \frac{1}{\sigma^2} Cov\left[\left(\omega^{\perp}(0) R\right)^2, Y_k\right] + \tau a_k'(0). \end{aligned}$$

Therefore equation (32) reads

$$\begin{aligned} \omega'(0) &= \tau \Sigma^{-1} \left[c(\omega(0)) \frac{\rho}{\tau^2} \frac{1}{\sigma^2} Var\left[\omega^{\perp} R\right] + a'(0)\right] \\ &= \tau \Sigma^{-1} \left[c(\omega(0)) \frac{\rho}{\tau^2} \frac{1}{\sigma^2} \left[\tau a^{\perp}(0)\, \Sigma^{-1} \left(\sigma^2 \Sigma\right) \Sigma^{-1} \tau a(0)\right] + a'(0)\right] \\ &= \tau \Sigma^{-1} \left[c(\omega(0))\, \rho P^2(0) + a'(0)\right]. \end{aligned}$$

■

PROOF OF THEOREMS 2.5 AND 2.6. Using the definitions of τ_s, ρ_s from equation (9), the demand equation (10) and the first market-clearing equation in (13) we derive from the condition

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

that

$$a(0) = \frac{1}{\overline{\tau}}\Sigma\overline{\omega}. \tag{33}$$

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

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

Looking at equation (8) we then check that $c_k(\omega_s(0)) = c_k(\overline{\omega})$. Using (32) and the second market-clearing equation in (13) we get from

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

that

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

Thus:

$$a_k'(0) = -\overline{\rho}c_k(\overline{\omega})P^2(0). \tag{34}$$

Plugging (34) into Theorem 2.4 gives:

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

We find:

$$\begin{aligned} P^2(0)\Sigma^{-1}c(\overline{\omega})c(\overline{\omega}) &= \frac{1}{\sigma}P^2(0)\Sigma^{-1}\frac{Cov\left(R,\left(\overline{\omega}^{\perp}(R-ER)\right)^2\right)}{Var\left(\overline{\omega}^{\perp}(R-ER)\right)} \\ &= \frac{1}{\sigma}\sigma^2P^2(0)(Var(R))^{-1}\frac{Cov\left(R,\left(\overline{\omega}^{\perp}(R-ER)\right)^2\right)}{Var\left(\overline{\omega}^{\perp}(R-ER)\right)} \\ &= \frac{1}{\sigma}(Var(R))^{-1}Cov\left(R,\left(\overline{\omega}^{\perp}(R-ER)\right)^2\right) \end{aligned}$$

Given that $R_M = \overline{\omega}^{\perp}R$,

$$P^2(0)\Sigma^{-1}c(\overline{\omega})c(\overline{\omega}) = \frac{1}{\sigma}(Var(R))^{-1}Cov\left(R,(R_M - ER_M)^2\right)$$

Therefore

$$\sigma\omega_s'(0) = \tau_s [\rho_s - \overline{\rho}] (Var(R))^{-1} Cov\Big(R, (R_M - ER_M)^2\Big)$$

is ends the proof. ■

PROOF OF THEOREM 2.7. We first find two real numbers L and N such that, with

$$\begin{aligned} m(\sigma) &= \frac{1}{R_f} + L \cdot \Big(R_M(\sigma) - E[R_M(\sigma)]\Big) \\ &\quad + N \cdot \Big((R_M(\sigma) - E[R_M(\sigma)])^2 - E(R_M(\sigma) - E[R_M(\sigma)])^2 \Big), \end{aligned}$$

we have:

$$Em(\sigma) \cdot \big(\sigma^2 a(\sigma) + \sigma Y\big) = 0.$$

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:

$$\begin{aligned} & \frac{1}{R_f}\frac{\sigma^2}{\overline{\tau}}\Sigma\overline{\omega} - \frac{1}{R_f}\sigma^3\overline{\rho}P^2(0)\, c(\overline{\omega}) \\ + \;& L \cdot Cov(R, R_M(\sigma)) + N \cdot Cov\Big(R, (R_M(\sigma) - E[R_M(\sigma)])^2\Big) = 0. \end{aligned}$$

Let us define L and N such that

$$\begin{aligned} L \cdot Cov(R, R_M(\sigma)) &= -\frac{1}{R_f}\frac{\sigma^2}{\overline{\tau}}\Sigma\overline{\omega}, \\ N \cdot Cov\Big(R, (R_M(\sigma) - E[R_M(\sigma)])^2\Big) &= \frac{1}{R_f}\sigma^3\overline{\rho}P^2(0)\, c(\overline{\omega}). \end{aligned}$$

Noticing that

$$\begin{aligned} \sigma^2\Sigma\overline{\omega} &= Cov(R, R_M(\sigma)), \quad \text{and} \\ \overline{\tau}\sigma^3 P^2(0)\, c(\overline{\omega}) &= \sigma\frac{Var(R_M)}{\overline{\tau}} c(\overline{\omega}) = \frac{1}{\overline{\tau}} Cov\Big(R, (R_M(\sigma) - E[R_M(\sigma)])^2\Big), \end{aligned}$$

we conclude that

$$L = -\frac{1}{R_f\overline{\tau}}, \text{ and } N = \frac{1}{R_f}\frac{\overline{\rho}}{\overline{\tau}^2}$$

Second, since the pricing kernel correctly prices the primitive assets:

$$Em(\sigma)R = 1$$

it can be replaced by its projection on the set of primitive assets and the market return:

$$\begin{aligned} m(\sigma) &= E(m(\sigma)|R_M, R) \\ &= \frac{1}{R_f} - \frac{1}{R_f\overline{\tau}}(R_M(\sigma) - ER_M(\sigma)) + \frac{\overline{\rho}}{R_f\overline{\tau}^2}\Theta_{skew}^{\perp}(R - E(R)) \end{aligned}$$

where

$$\Theta_{skew} = (Var(R))^{-1} Cov\left(R, (R_M - ER_M)^2\right)$$

which is the announced result. ■

Proof of theorem 2.8. 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:

$$\frac{\sigma}{\overline{\tau}}\Sigma\overline{\omega} - \sigma^3\overline{\rho}P^2(0)c(\overline{\omega}) = \frac{1}{\overline{\tau}}Cov(R, R_M) - \frac{\overline{\rho}}{\overline{\tau}^2}Cov\left(R, (R_M - E[R_M])^2\right).$$

From theorem 2.6, we have

$$\Theta_{skew} = (Var(R))^{-1} Cov\left(R, (R_M - ER_M)^2\right)$$

Thus

$$Cov\left(R, \Theta_{skew}^{\perp}R\right) = Cov\left(R, (R_M - ER_M)^2\right)$$

Hence,

$$\frac{\sigma}{\overline{\tau}}\Sigma\overline{\omega} - \sigma^3\overline{\rho}P^2(0)c(\overline{\omega}) = \frac{1}{\overline{\tau}}Cov(R, R_M) - \frac{\overline{\rho}}{\overline{\tau}^2}Cov\left(R, \Theta_{skew}^{\perp}R\right)$$

Defining

$$\gamma_i = \frac{Cov\left(R_i, \Theta_{skew}^{\perp}R\right)}{Var\left(\Theta_{skew}^{\perp}R\right)}$$

we deduce for each $i = 1, ..., K$:

$$E[R_i] - R_f = \frac{1}{\overline{\tau}}Cov(R, R_M) - \frac{\overline{\rho}}{\overline{\tau}^2}Var\left(\Theta_{skew}^{\perp}R\right)\gamma_i$$

which corresponds to the formula of Theorem 2.8. ■

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

$$r_{it+1} = \frac{r^2_{Mt+1}}{\eta_t} - R_{ft}$$

we get

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

that is:

$$\frac{P^{(2)}_{Mt}}{\eta_t} = \delta_{MMt}P^{(1)}_{Mt} - \lambda_{2t}\left(1 - \rho_t^2\left(r_{Mt+1}, r^2_{Mt+1}\right)\right).$$

This gives the announced value for λ_{2t}. ■

Proof of Theorem 3.5. Assume that the joint process $\left(m_{t+1}, R_{t+1}^{;}\right)$ is conditionally lognormal. Then,

$$\begin{bmatrix} Log\left(m_{t+1}\right) \\ LogR_{Mt+1} \end{bmatrix} / I_t \rightsquigarrow N\left[\begin{bmatrix} \mu_{mt} \\ \mu_{Mt} \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & \sigma_{mrt} \\ \sigma_{mrt} & \sigma^2_{Mt} \end{bmatrix}\right].$$

Let us denote

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

The market return risk neutral variance $Var_t^*\left(R_{Mt+1}\right)$ is

$$Var_t^*\left(R_{Mt+1}\right) = E_t^* R^2_{Mt+1} - R^2_{ft}, \quad \text{with } E_t^* R^2_{Mt+1} = R_{ft}E_t m_{t+1} R^2_{Mt+1}.$$

We know that

$$Log\left(m_{t+1}R^2_{Mt+1}\right) = Log\left(m_{t+1}\right) + 2Log\left(R_{Mt+1}\right).$$

Therefore,

$$\begin{aligned} E_t[m_{t+1}R^2_{Mt+1}] &= \exp\left(\mu_{mt} + 2\mu_{Mt} + 0.5\sigma_t^2 + 2\sigma^2_{Mt} + 2\sigma_{mrt}\right) \\ &= \exp\left(-\mu_{mt} - 0.5\sigma_t^2\right)\exp\left(2\mu_{Mt} + 2\sigma^2_{Mt}\right)\exp\left(-2\mu_{Mt} - \sigma^2_{Mt}\right) \cdot \\ &\quad \cdot \left[\exp\left(\mu_{mt} + \mu_{Mt} + 0.5\sigma_t^2 + 0.5\sigma^2_{Mt} + \sigma_{mrt}\right)\right]^2. \end{aligned}$$

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

$$\exp\left(\mu_{mt} + \mu_{Mt} + 0.5\sigma_t^2 + 0.5\sigma^2_{Mt} + \sigma_{mrt}\right) = 1,$$

and therefore

$$
\begin{aligned}
E_t[m_{t+1}R_{Mt+1}^2] &= \exp\left(-\mu_{mt} - 0.5\sigma_t^2\right)\exp\left(2\mu_{Mt} + 2\sigma_{Mt}^2\right)\exp\left(-2\mu_{Mt} - \sigma_{Mt}^2\right) \\
&= R_{ft}\frac{ER_{Mt+1}^2}{\left(E_tR_{Mt+1}\right)^2}.
\end{aligned}
$$

Consequently,

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

■

Proof of Theorem 3.6. Assume that

$$
m_{t+1} = \nu_{0t} + \nu_{1t}R_{Mt+1} + \nu_{2t}R_{Mt+1}^2.
$$

We denote

$$
\varepsilon_{t+1} = \left(R_{Mt+1} - ER_{Mt+1}\right)^2 - E\left[\left(R_{Mt+1} - ER_{Mt+1}\right)^2 |R_{t+1}\right]
$$

the residual of the linear regression of $\left(R_{Mt+1} - ER_{Mt+1}\right)^2$ on R_{Mt+1}. This residual can be written as:

$$
\varepsilon_{t+1} = \left(R_{Mt+1} - ER_{Mt+1}\right)^2 - \left(\psi_{Mt+1} - E\psi_{Mt+1}\right)
$$

with

$$
\psi_{Mt+1} = E\left[\left(R_{Mt+1} - ER_{Mt+1}\right)^2 |R_{t+1}\right]
$$

Note that:

$$
\psi_{Mt+1} = \Theta_{skew}^{\perp}R_{t+1}
$$

Therefore,

$$
Cov_t\left(m_{t+1}, \varepsilon_{t+1}\right) = \nu_{2t}Var_t\left(\varepsilon_{t+1}\right).
$$

■

References

[1] Bakshi, G., N. Kapadia, and D. Madan. 2003. "Stock Return Characteristics, Skew Laws, and Differential Pricing of Individual Equity Options." *Review of Financial Studies* 16: 101-143.

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

[3] Barone-Adesi, G. 1985. "Arbitrage Equilibrium with Skewed Asset Returns." *Journal of Financial and Quantitative Analysis* 20: 299-313.

[4] Barone-Adesi, G., G. Urga, and P. Gagliardini. 2004. "Testing Asset Pricing Models with Coskewness." *Journal of Business and Economic Statistics* 22: 474-485.

[5] Bekaert, G., and J. Liu. 2004. "Conditioning Information and Variance Bounds on Pricing Kernels." *The Review of Financial Studies* 17: 339-378.

[6] Black, F. 1976. "The Pricing of Commodity Contracts." *Journal of Financial Economics* 3: 167-179.

[7] Brennan, M. J. 1979. "The Pricing of Contingent Claims in Discrete Time Models." *Journal of Finance* 24: 53-68.

[8] Chabi-Yo, F., E. Ghysels, and E. Renault. 2007. "Disentangling the Effects of Heterogeneous Beliefs and Preferences on Asset Prices." Discussion paper, UNC.

[9] Cochrane, J.H. 2001. *Asset Pricing*. Princeton University Press.

[10] Dittmar, R. F. 2002. "Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from Cross section of Equity Returns." *Journal of Finance* 57: 368-403.

[11] Epstein, L., and S. Zin. 1991. "Substitution, Risk Aversion and the Temporal Behaviour of Consumption and Asset Returns: An Empirical Analysis." *Journal of Political Economy* 99: 263-286.

[12] Garcia, R., E. Ghysels, and E. Renault. 2003. *The Econometrics of Option Pricing*, in Y. Ait Sahalia and L. P. Hansen, ed.: The Handbook of Financial Econometrics (North Holland, *forthcoming*)

[13] Hansen, L., and S. Richard. 1987. "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models." *Econometrica* 55: 587-613.

[14] Harvey, C.R., and A. Siddique. 2000. "Conditional Skewness in Asset Pricing Tests." *Journal of Finance* 55: 1263-1295.

[15] Ingersoll, J. 1987. *Theory of Financial Decision Making*. New Jersey: Rowman & Littlefield.

[16] Jobson, J. D., and B. M. Korkie. 1982. "Potential Performance and Tests of Portfolio Efficiency." *Journal of Financial Economics* 10: 433-466.

[17] Juud, G., and S. Guu. 2001. "Asymptotic Methods for Asset Market Equilibrium Analysis." *Economic Theory* 18: 127-157.

[18] Kimball, M. 1990. "Precautionary Saving in the Small and in the Large." *Econometrica* 58: 53-73.

[19] Kraus, A., and R. H. Litzenberger. 1976. "Skewness Preference and the Valuation of Risk Assets." *Journal of Finance* 31: 1085-1100.

[20] Lucas, R. Jr. 1978. "Asset Prices in an Exchange Economy." *Econometrica* 46: 1429-1445.

[21] Markowitz, H. 1952. "Portfolio Selection." *Journal of Finance* 7: 77-91.

[22] Rosenberg J. 2000. "Asset Pricing Puzzles: Evidence from options markets." Stern School of Business Working Paper.

[23] Samuelson, P. A. 1970. "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Mean, Variance and Higher Moments." Review of Economic Studies 37: 537-42.

Notes

[1]Dittmar (2002) and Barone-Adesi, Urga, and Gagliardini (2004) study if skewness risk is priced and its price. Chang, Johnson, and Schill (2004) test whether Fama-French factors proxy for skewness and higher moments.

[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.

[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. Futhermore 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.

[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 σ with the square root of time.

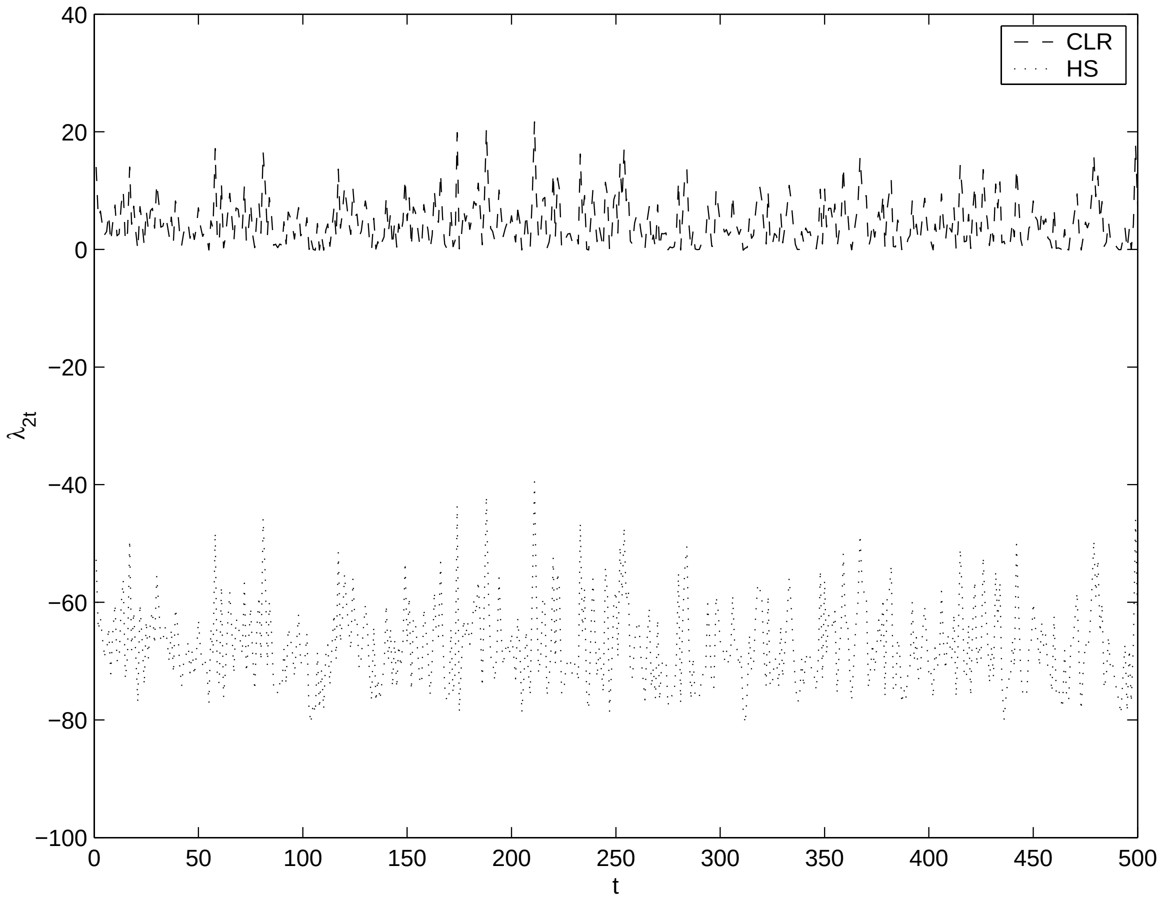

Figure 1: Price of coskewness inferred from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004). HS indicates the price of coskewness corresponding to Harvey and Siddique (2000); CLR indicates the price of coskewness corresponding to our formula.

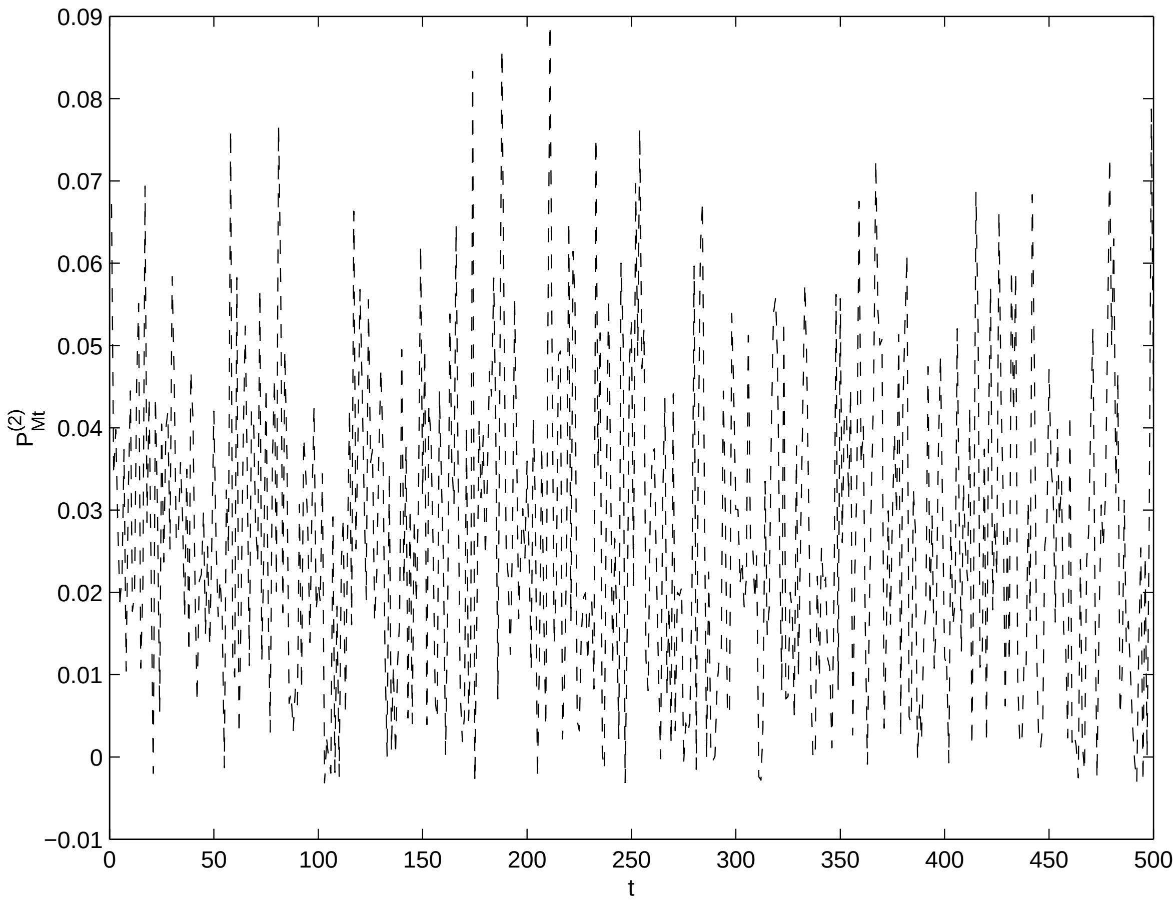

Figure 2: Risk premium on the squared net return inferred from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004)

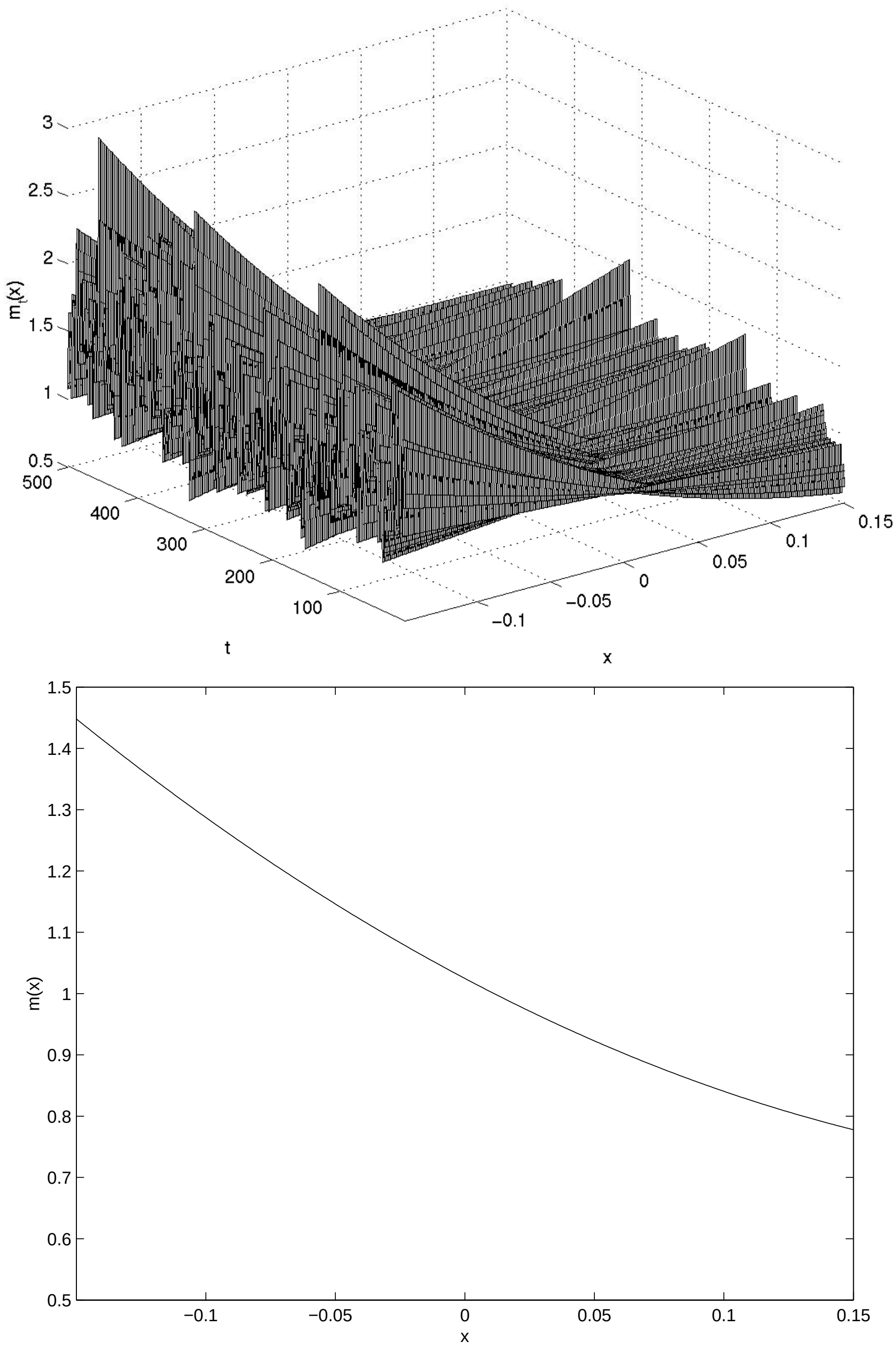

Figure 3: Quadratic pricing kernel inferred from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004). In the upper graph, we plot the pricing kernel m_{t+1} as a function of time and r_{Mt+1}. In the lower graph we plot the average pricing kernel $\Sigma_{t=1}^{T} \frac{1}{T} m_{t+1}$.

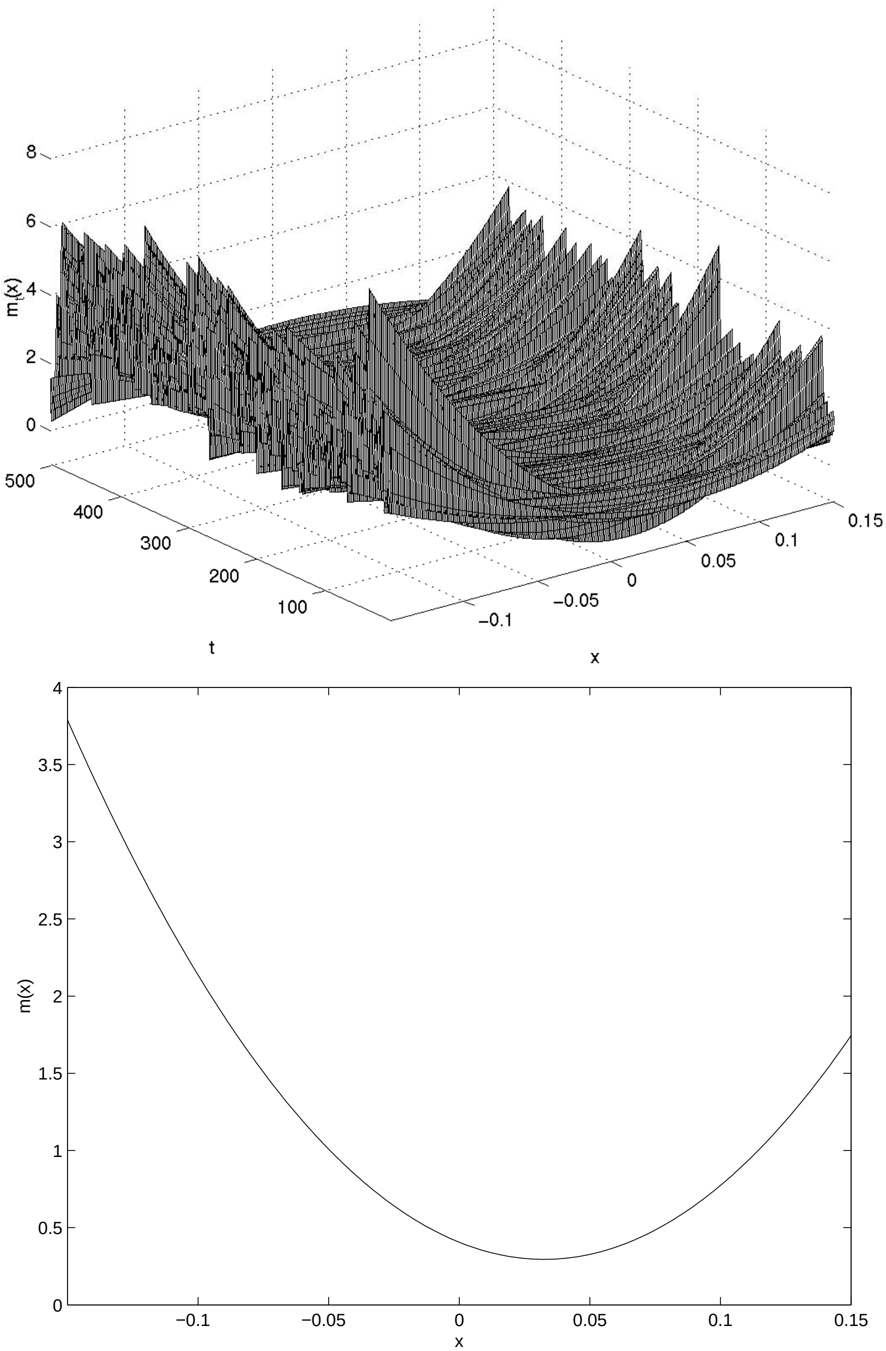

Figure 4: Fixing the price of the squared market return at the level 1.02, which in turns implies a time varying η_t, we infer the quadratic pricing kernel from a simulated time-series of the Factor GARCH-in-mean used in Bekaert and Liu (2004). In the upper graph, we plot the pricing kernel m_{t+1} as a function of time and $r_{Mt+1} = x$. In the lower graph, we plot the average pricing kernel $\Sigma_{t=1}^{T} \frac{1}{T} m_{t+1}$.

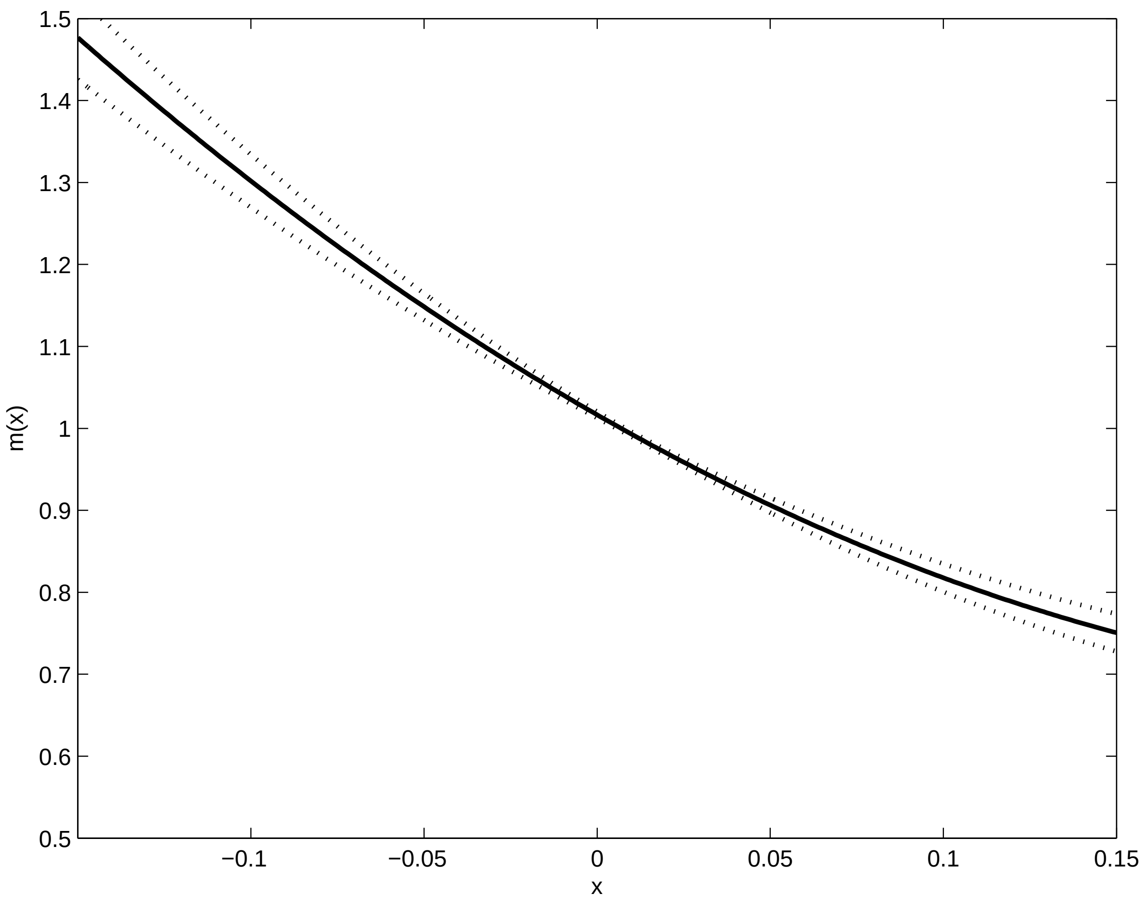

Figure 5: We plot the average of pricing kernels inferred from a simulated time-series of the Factor GARCH-in mean model: $m(x) = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{T}\sum_{t=1}^{T} m_t^{(k)}(x)\right)$ where T represents the number of simulated time series of return and $K = 1000$ the number of replications. For each replication, we simulated a time-series of $T = 500$ observations. $\frac{1}{T}\sum_{t=1}^{T} m_t^{(k)}(x)$ is the average pricing kernel for the k^{th} replication. The number of replication is $K = 1000$. We also plot the average pricing kernel 5% confidence interval.

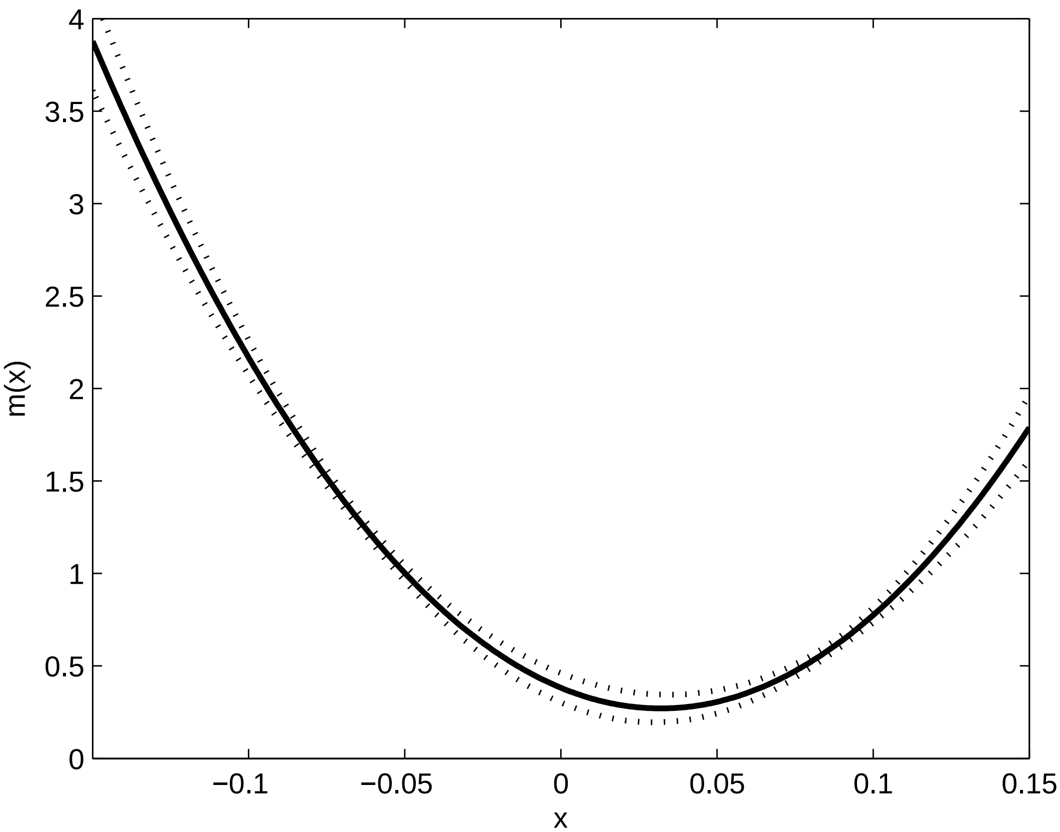

Figure 6: We plot the average of pricing kernels inferred from a simulated time-series of the Factor GARCH-in mean model when the price of the squared market return is fixed at the level of 1.02 which in turns implies a time-varying η_t: $m(x) = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{T}\sum_{t=1}^{T} m_t^{(k)}(x)\right)$ where T represents the number of simulated time series of return and $K = 1000$ the number of replications. For each replication, we simulated a time-series of $T = 500$ observations. The number of replication is $K = 1000$. $\frac{1}{T}\sum_{t=1}^{T} m_t^{(k)}(x)$ is the average pricing kernel for the k^{th} replication. We also plot the average pricing kernel 5% confidence interval.

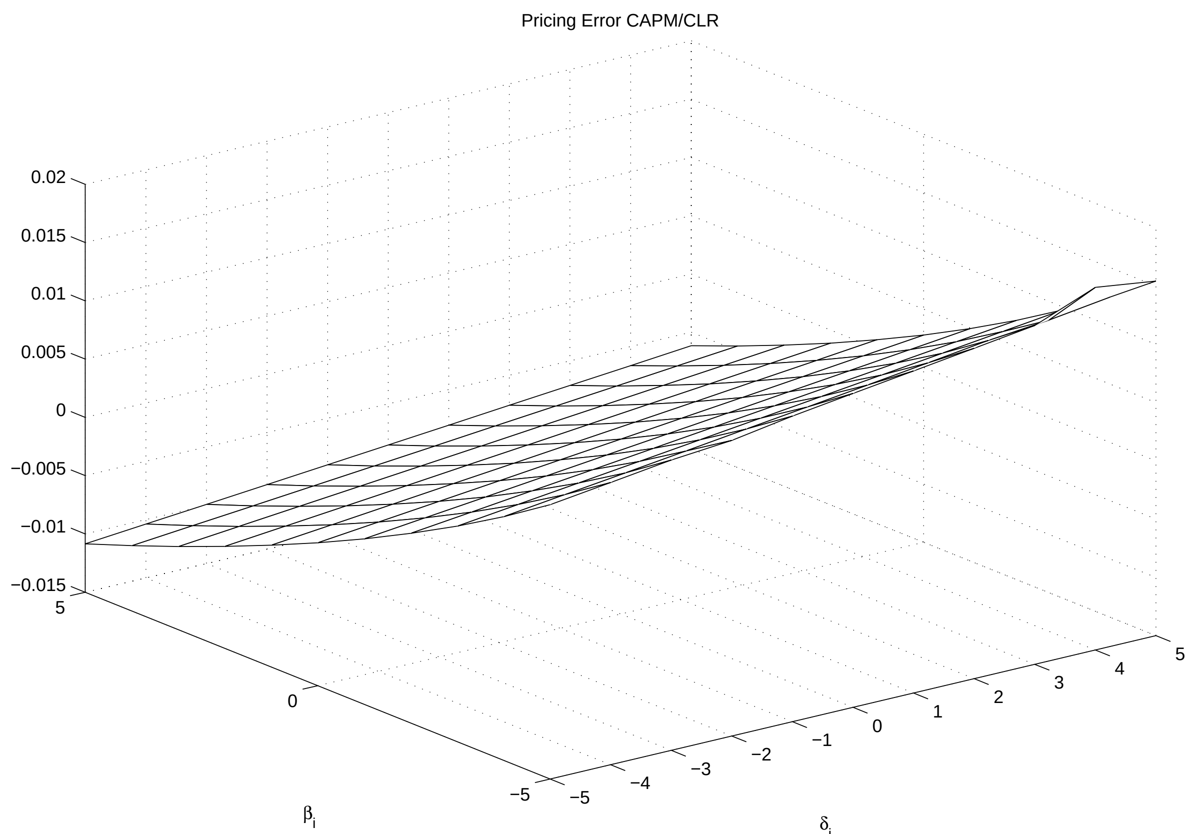

Figure 7: For each replication, we compute the relative expected return error of the CAPM versus the expected return given by the CLR model (see equation (20)) as a function of β_{iMt} and δ_{iMt}:

$$((E_t R_{it+1})_{CAPM} - (E_t R_{it+1})_{CLR}) \,/\, (E_t R_{it+1})_{CLR} \,.$$

We then compute the average of the relative expected return error over K replications and plot the average of the relative expected return error as a function of the asset i's beta and coskewness.

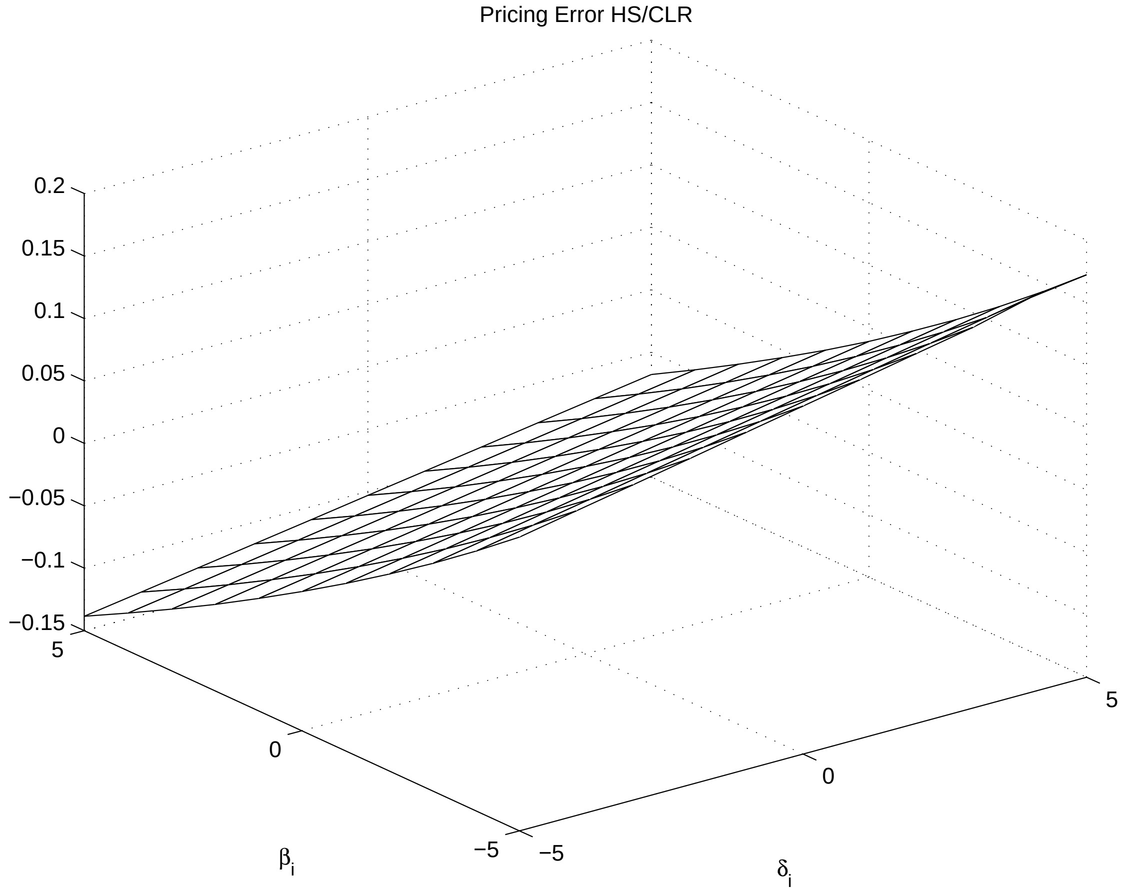

Figure 8: For each replication, we compute the relative expected return error of the HS expected return decomposition versus the expected return given by the CLR model (see equation (20)) as a function of β_{iMt} and δ_{iMt}:

$$((E_t R_{it+1})_{HS} - (E_t R_{it+1})_{CLR}) \,/\, (E_t R_{it+1})_{CLR}\,.$$

We then compute the average of the relative expected return error over K replications and plot the average of the relative expected return error as a function of the asset i's beta and coskewness.

Equations		Coefficients		
	c_t	Y_{1t}	Y_{2t}	Y_{3t}
Y_{1t+1}	0.0030	0.361	-0.029	0.008
	(0.0005)	(0.033)	(0.022)	(0.005)
Y_{2t+1}	0.0056-162.65$Var_t\left[e_{1t+1}\right]$	-0.198	0.738	-0.0002
	(0.0006) (0.0001)	0.031	(0.037)	(0.0043)
Y_{3t+1}	$0.0188 - 58.02Var_t\left[e_{1t+1}\right]$	-1.734	1.029	0.077
	(0.0083) (0.0003)	0.005	(0.014)	(0.034)
	constant	α	β	ξ
$Var_t\left(e_{1t+1}\right)$	0.000019	-0.0265	0.0008	0.2705
	0.000018	(0.0807)	(0.7898)	(0.0426)
δ_2	0.000014	0	0	0
	(0.000002)			
δ_3	0.006134	0	0	0
	(0.00103)			
	$\sigma_{13} = -0.0564$		$\sigma_{12} = 3.182$	
	(0.1425)		(0.003)	

Table 1: This table reproduces the results of the Factor GARCH in mean estimated by Bekaert and Liu (2004). Here $\delta_2 = Var_t\left(e_{2t+1}\right)$ and $\delta_3 = Var_t\left(e_{3t+1}\right)$.

$K = 1$ replication		
Parameters	Mean	standard deviation
η_t	0.64%	0.0076413%
λ_{2t}	4.25	4.06
$\lambda_{2t} - \lambda_{2t-1}$	-67.82	7.45
λ_{2t}^{HS}	0	4.93
$\lambda_{2t}^{HS} - \lambda_{2t-1}^{HS}$	0	8.75

Table 2: Summary statistics for one particular Monte-Carlo simulation

$K = 1000$ replications

Parameters	Mean	standard deviation
η_t	0.64%	0.00051634%
λ_{2t}	4.2917	0.2758
$\lambda_{2t} - \lambda_{2t-1}$	0.00014460	0.0129
λ_{2t}^{HS}	−67.7317	0.5388
$\lambda_{2t}^{HS} - \lambda_{2t-1}^{HS}$	0.00038303	0.0234

Table 3: Summary statistics for $K = 1000$ replications in our Monte-Carlo simulation. For each replication (sample path), we compute the average of parameters and then take the average over replications. We also compute the standard deviation of the "average parameters" over replications.

$K = 1000$ replications

Parameters	averages of "standard deviation"	standard deviations of "standard deviation"
η_t	0.0072216%	0.0010233%
λ_{2t}	4.0694	0.2876
$\lambda_{2t} - \lambda_{2t-1}$	4.8847	0.3094
λ_{2t}^{HS}	7.4418	0.4408
$\lambda_{2t}^{HS} - \lambda_{2t-1}^{HS}$	8.6955	0.4726

Table 4: Summary statistics for $K = 1000$ replications in our Monte-Carlo simulation. For each replication, we compute standard deviations of parameters and then take the averages of the "average standard deviation" over replications. We also find the standard deviation of the "average standard deviation" over replications.