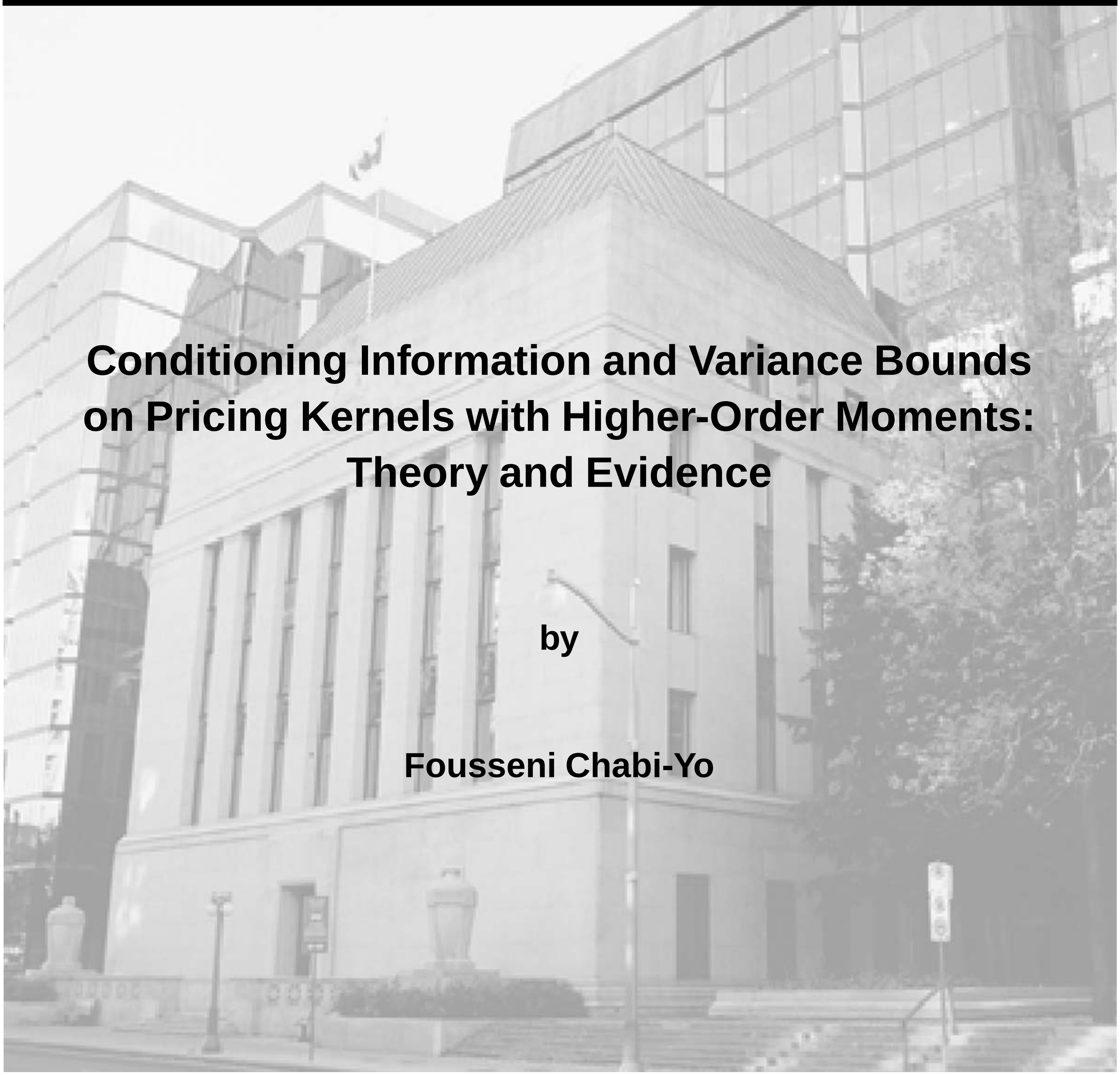
Bank of Canada Banque du Canada

Working Paper 2006-38 / Document de travail 2006-38



Conditioning Information and Variance Bounds on Pricing Kernels with Higher-Order Moments: Theory and Evidence

by

Fousseni Chabi-Yo

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2006-38

October 2006

Conditioning Information and Variance Bounds on Pricing Kernels with Higher-Order Moments: Theory and Evidence

by

Fousseni Chabi-Yo

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
fchabiyo@bankofcanada.ca

Contents

Acknowledgements

We are grateful fo René Garcia, Geert Bekaert, Raymond Kan, Jun Liu, and Eric Renault for help or suggestions. An anonymous referee provided extensive guidance, comments, and insights. This paper was presented at the 2005 Midwest Finance Association, the 2005 Econometric Society World Congress in London, England, the 2005 Northern Finance Association, and the 2005 Financial Management Association Annual Meeting in Chicago. We also thank CIRANO for making their data available.

Abstract

The author develops a strategy for utilizing higher moments and conditioning information efficiently, and hence improves on the variance bounds computed by Hansen and Jagannathan (1991, the HJ bound) and Gallant, Hansen, and Tauchen (1990, the GHT bound). The author's bound incorporates variance risk premia. It reaches the GHT bound when non-linearities in returns are not priced. The author also provides an optimally scaled bound with conditioning information, higher moments, and variance risk premia that improves on the Bekaert and Liu (2004, the BL bound) optimally scaled bound. This bound reaches the BL bound when non-linearities in returns are not priced. When the conditional first four moments are misspecified, the author's optimally scaled bound remains a lower bound to the variance on pricing kernels, whereas the BL bound does not. The author empirically illustrates the behaviour of the bounds using Bekaert and Liu's (2004) econometric models. He also uses higher moments and conditioning information to provide distance measures that improve on the Hansen and Jagannathan distance measures. The author uses these distance measures to evaluate the performance of asset-pricing models. Some existing pricing kernels are able to describe returns ignoring the impact of higher moments and variance risk premia. When accounting for the impact of higher moments and variance risk premia, these same pricing kernels have difficulty in explaining returns on the assets and are unable to price non-linearities or higher moments.

JEL classification: G12, G13, C61
Bank classification: Financial markets; Market structure and pricing

Résumé

L'auteur conçoit une stratégie pour utiliser avec efficience les moments d'ordre supérieur et l'ensemble de l'information disponible et, de la sorte, améliorer les bornes de variance calculées par Hansen et Jagannathan (1991) et par Gallant, Hansen et Tauchen (1990) (appelées ci-après « borne HJ » et « borne GHT »). La borne qu'il définit intègre les primes du risque de variance et est égale à la borne GHT lorsque les non-linéarités des rendements ne sont pas prises en considération. L'auteur calcule aussi une borne optimale qui tient compte de l'ensemble des informations, des moments d'ordre supérieur ainsi que des primes du risque de variance, et qui dépasse la borne optimale de Bekaert et Liu (2004) (« borne BL » dans la suite du résumé). En l'occurrence, la borne de l'auteur est identique à la borne BL (l'hypothèse de linéarité des rendements étant également maintenue dans ce cas). Mais la borne optimale de l'auteur demeure la limite inférieure de la variance des facteurs d'actualisation stochastiques même quand les quatre premiers moments conditionnels sont mal spécifiés. Pour illustrer de façon empirique le comportement des bornes, l'auteur met à profit les modèles économétriques de Bekaert et Liu (2004). Par ailleurs, en faisant appel aux moments d'ordre supérieur et à l'ensemble des informations, il obtient de meilleures mesures de distance que celles auxquelles parviennent Hansen et Jagannathan. Il se sert de ces mesures pour évaluer les modèles d'équilibre des actifs financiers. Certains des facteurs d'actualisation stochastiques employés arrivent à rendre compte des rendements s'il est fait abstraction de l'incidence des moments d'ordre supérieur et des primes du risque de variance. Or, une fois cette incidence prise en compte, les mêmes facteurs permettent difficilement d'expliquer les rendements obtenus et ne permettent d'évaluer ni les non-linéarités ni les moments d'ordre supérieur.

Classification JEL : G12, G13, C61
Classification de la Banque : Marchés financiers; Structure de marché et fixation des prix

1. Introduction

Recent years have witnessed an explosion of research that incorporates conditional skewness and conditional kurtosis in asset-pricing models (Harvey and Siddique (2000); Dittmar (2002); and others). As shown in Harvey and Siddique (2000) and Dittmar (2002), the market price of skewness risk and kurtosis risk is a key determinant in explaining the cross-section of returns. These models perform well empirically using the Hansen and Jagannathan (1991, hereafter HJ) variance bound and the Hansen and Jagannathan (1997) distance. In addition, the pricing kernels of recent models such as the non-separable utility model of Heaton (1995), incomplete-markets model of Constantinides and Duffie (1996), or polynomial pricing kernels of Bansal, Hsieh, and Viswanathan (1993) and Chapman (1997), lie inside the feasible region defined by these bounds. Although the HJ variance bound and distance are useful for asset-pricing models, they incorporate only the first two moments of asset returns. The HJ distance and variance bound use only the first two moments to evaluate the performance of non-linear pricing kernels or pricing kernels that incorporate higher-order moments. Further, studies such as by Gallant, Hansen, and Tauchen (1990, hereafter GHT), Ferson and Siegel (2001, 2003), and Bekaert and Liu (2004) suggest that the conditioning information is useful to improve the performance of asset-pricing models. Although asset-pricing models perform well empirically using the conditioning information, the GHT bound incorporates the first two conditional moments of asset returns.

In this paper, we study the use of conditioning information and derivatives to effectively increase the dimension of asset payoffs space, and hence improve the HJ distance measures and the HJ variance bound. We provide three variance bounds on pricing kernels. We first derive an efficient variance bound on pricing kernels, which we term the UCHM bound. It incorporates time-varying higher moments and variance risk premia. A large body of theory and evidence suggests that the variance risk is priced in the market (see Bakshi and Madan (2000); Bakshi, Kapadia, and Madan (2003)). Time-varying higher moments and variance risk premia are important to effectively manage risk and allocate assets, to accurately price and hedge derivative securities, and to understand the behaviour of financial asset prices. The UCHM bound is a sum of two terms. The first term is the GHT bound. The second term is a function of the first four conditional moments of asset returns and the pure variance risk premia. As shown in Bondareko (2004), the variance risk premia can be decomposed into two components. The first component is proportional to the risk premium on primitive assets. The second component is called the pure variance risk premia. It represents the part of the variance risk premia that is independent of the risk premia on primitive assets. Bondareko shows that variance risk premia are a key determinant in explaining returns that exhibit significant non-linearities (skewness). When non-linearities in returns are not priced, that is, skewness is not priced, we show that the UCHM bound reaches the GHT bound. Second, we derive a bound with unconditional higher moments and variance risk premia, which we term the HM bound. When skewness is not priced, we show that this bound reaches the HJ bound. Third, we use the scaled returns to derive the best (largest) variance bound that incorporates time-varying higher moments and variance risk premia. We term this variance the OHM bound. When non-linearities in returns are not priced, we show that the OHM bound reaches the Bekaert and Liu (2004) optimally scaled variance bound. The OHM bound has some advantageous features. First, it is efficient. Our approach optimally exploits conditioning information with higher moments, leading to a sharper bound. Second, the OHM bound is robust to the misspecification of the conditional mean, conditional variance, conditional skewness, and conditional kurtosis. The OHM bound provides a bound to the variance of the true pricing kernel even if incorrect proxies to the conditional first four moments are used. Third, we show that the OHM bound can be used to propose a diagnostic test for the first four conditional moments of asset returns if the conditional prices of derivatives are correctly specified.

Our paper also provides distance measures to evaluate asset-pricing models. We propose two distance measures. We first propose an unconditional distance measure that incorporates higher moments and variance risk premia. We term this distance the HM distance. It reaches the HJ distance when skewness is not priced in the market. We use the scaled returns to propose an optimal distance measure, which we term the OHM distance, to evaluate pricing models. We derive the best (largest) distance measure with time-varying higher moments and variance risk premia. When time-varying higher moments and variance risk premia are not important, the OHM distance reaches the distance measure obtained if we use the Bekaert and Liu (2004) scaling approach.

The remainder of this paper is organized as follows. In section 2, we derive the variance bounds that incorporate conditioning information, higher moments, and variance risk premia. In section 3, we derive the distance measures. Section 4 contains an empirical illustration of the bounds. We use Bekaert and Liu (2004) econometric models to illustrate the bounds, and explore the role of misspecification and robustness in the behaviour of the various bounds. In section 5, we use various distance measures to evaluate the performance of asset-pricing models with non-linear pricing kernels. We also investigate time-varying extensions of these pricing kernels. To do this, we use the volatility index, VIX, which is based on Standard & Poor's (S&P) 500 index option prices and different data sets. We first use hedge fund indexes. Agarwal and Naik (2004) show that a large number of equity-oriented hedge fund strategies exhibit payoffs that resemble a short position in a put option on the market index. Second, we use industry portfolios. Industry portfolios have been used in the empirical asset-pricing literature for tests of candidate asset-pricing models (Dittmar (2002)). Section 6 concludes the paper.

2. Variance Bounds on Pricing Kernels

2.1 Conditional minimum-variance pricing kernel

GHT (1990) assume that economic agents use their information set to form portfolios of risky assets and derive a variance bound on pricing kernels that incorporates conditioning information. Their bound is a function of the asset return first two moments. In this section, we assume that there is a relevant information set, I_t, available to investors and econometricians at a given point in time, and that investors use this set to form portfolios of asset payoffs and derivatives in the same assets. If this is so, investors have a larger set of assets to form their portfolios than in GHT. Intuitively, we augment the available asset space with derivatives.

We define r_{t+1} as the set of asset payoffs with finite first four conditional moments, $\vartheta_{t+1} = r_{t+1}^{(2)}$ the payoff of the "volatility contract" with components of the form $r_{it+1}r_{jt+1}$, $i \leq j$, and $h(r_{t+1})$ the payoff of derivatives. This payoff is approximated by its linear regression on asset return and the volatility contract payoff:

$$h(r_{t+1}) \simeq E_t h(r_{t+1}) + a_t [r_{t+1} - E_t r_{t+1}] + b_t [\vartheta_{t+1} - E_t \vartheta_{t+1}] + \eta_{t+1}, \tag{1}$$

with some residual risk, but the residual risk is not priced. The representation (1) states that the price of the volatility contract suffices to recover the price of derivatives. We consider the set of admissible pricing kernels that conditionally price the bond, the set of asset payoffs, and derivatives with payoff $h(r_{t+1})$. This set can be formulated as follows:

$$\mathcal{F}\left(\overline{m}_t, p_t^{\vartheta}\right) = \left\{m_{t+1} \in L^2 : E\left[m_{t+1}(1, r_{t+1}, \vartheta_{t+1})\,|I_t\right] = \left(\overline{m}_t, p_t, p_t^{\vartheta}\right)\right\}, \tag{2}$$

where $\overline{m}_t$, p_t, and p_t^{ϑ} represent the conditional price of the bond, asset returns, and the volatility contract, respectively. L^2 represents the set of random variables with a finite second moment. The payoff r_{t+1} is a return. Thus, $p_t = l$, where l is a vector column whose components are equal to 1. However, the price of the volatility contract:

$$p_t^{\vartheta} = E_t[m_{t+1}\vartheta_{t+1}] = \overline{m}_t E_t\left[\frac{m_{t+1}}{\overline{m}_t}\vartheta_{t+1}\right] = \frac{E_t^* \vartheta_{t+1}}{r_{ft}}, \tag{3}$$

is different from p_t. $E_t^*[x]$ represents the expectation of x with respect to the risk-neutral measure. For interpretation purposes, assume that there is only one risky asset. If the volatility contract is not priced, $Cov(m_{t+1}, \vartheta_{t+1}) = 0$, which indicates that $E_t\vartheta_{t+1} - E_t^*\vartheta_{t+1} = 0$.

There is a large body of theory and evidence which suggests that the volatility contract is priced in the market. Its price is easy to estimate. Bakshi and Madan (2000) show that the price of the volatility contract can be recovered from a set of OTM European calls and puts (see also Theorem 1 in Bakshi, Kapadia, and Madan (2003)). Carr and Wu (2004) theoretically and numerically show that the risk-neutral expected value of the return variance can be well approximated by a particular portfolio of options. Bondareko (2004) finds that the variance risk is priced and its risk premium is negative and economically very large. Using a regression-based analysis, he finds that the variance risk is a key determinant in explaining the performance of hedge funds. Given the evidence that the volatility contract is well priced, we consider the optimization problem:

$$\min_{m \in \mathcal{F}\left(\overline{m}_t, p_t^{\vartheta}\right)} \sigma^2 \left(m | I_t\right), \tag{4}$$

which allows us to derive the pricing kernel with minimum variance among the set of pricing kernels that correctly price returns and the volatility contract. Since this pricing kernel correctly prices the volatility contract, it should correctly price derivatives.

Denote:

$$\begin{aligned} \mu_t &= E\left(r_{t+1} | I_t\right) \text{ and } \sigma_t^2 = E_t r_{t+1} \left(r_{t+1} - E_t r_{t+1}\right)^{'}, \\ s_t^{'} &= E_t \left(\vartheta_{t+1} - E_t \vartheta_{t+1}\right) r_{t+1}^{'} \text{ and } \kappa_t = E_t \vartheta_{t+1} \vartheta_{t+1}^{'}, \end{aligned}$$

the first four conditional moments of asset returns. We show:

Proposition 2.1 *Given the information set I_t, the pricing kernel with minimum variance for its conditional expectation, $\overline{m}_t$, is:*

$$m_{CHM} = m_{GHT} + \gamma_t^{'} \varepsilon_{t+1}, \tag{5}$$

with $m_{GHT} = \beta_t \left(r_{t+1} - \mu_t\right) + \overline{m}_t$ representing the GHT pricing kernel and

$$\varepsilon_{t+1} = \vartheta_{t+1} - E_t \vartheta_{t+1} - s_t^{'} \left(\sigma_t^2\right)^{-1} \left(r_{t+1} - \mu_t\right), \tag{6}$$

with:

$$\begin{aligned} \beta_t &= \left(p_t - \overline{m}_t \mu_t\right)^{'} \left(\sigma_t^2\right)^{-1} \text{ and } \gamma_t = \left(\sigma_{\vartheta t}^2 - s_t^{'} \left(\sigma_t^2\right)^{-1} s_t\right)^{-1} \left(p_t^{\vartheta} - \overline{p_t^{\vartheta}}\right), \\ \sigma_{\vartheta t}^2 &= \kappa_t - \left(E_t \vartheta_{t+1}\right) \left(E_t \vartheta_{t+1}\right)^{'} \text{ and } \overline{p_t^{\vartheta}} = \overline{m}_t E_t \vartheta_{t+1} + s_t^{'} \left(\sigma_t^2\right)^{-1} \left(p_t - \overline{m}_t E_t r_{t+1}\right). \end{aligned}$$

The proof of this proposition is very similar to the proof of the minimum variance pricing kernel of GHT when using the vector $\left(r_{t+1}^{'}, \varepsilon_{t+1}^{'}\right)$ in place of $r_{t+1}^{'}$. Equation (5) says that the pricing kernel with minimum variance for its conditional expectation $\overline{m}_t$ is the conditional projection of m_{t+1} onto the $\left\{z_{1t}^{'} r_{t+1},\ z_{2t}^{'} \varepsilon_{t+1} \colon \forall\ z_{1t}, z_{2t}\right\}$ space augmented with a constant payoff. The conditional variance of the pricing kernel (5) is a function of the conditional first four moments $\left(\mu_t, \sigma_t^2, s_t, \kappa_t\right)$. The matrix parameter κ_t is the fourth moment (co-kurtosis) of asset returns. The matrix s_t is related to the notion of co-skewness (see Harvey and Siddique (2000)). The quantity $\sigma_{\vartheta t}^2 - s_t^{'} \left(\sigma_t^2\right)^{-1} s_t$ denotes the variance covariance matrix of the residual ε_{t+1}, which we assume is not singular.

The parameter γ_t is determined by the correlation between the pricing kernel and the non-linear component of the volatility contract that is not spanned by primitive asset returns. This parameter is proportional to the value $p_t^{\vartheta} - \overline{p_t^{\vartheta}}$, which we interpret as a pure volatility contract risk premium. It plays an important role in the variance bound (5). The pure volatility contract risk premium is the difference between two components:

$$p_t^{\vartheta} - \overline{p_t^{\vartheta}} = \overline{m}_t \left[E_t^* \vartheta_{t+1} - E_t \vartheta_{t+1} \right] - s_t^{'} \left(\sigma_t^2 \right)^{-1} \left(p_t - \overline{m}_t E_t r_{t+1} \right). \tag{7}$$

The first component of (7) is the risk premium on the volatility contract, while the second component is proportional to the risk premium on primitive assets. When non-linearities in returns are priced, expression (7) is different from zero, and the difference between the bound derived in proposition 2.1 and the existing variance bound on pricing kernels is due to the pure variance risk premia. The parameter γ_t incorporates information about how investors deal with the uncertainty in variance. This information is important to effectively manage risk and allocate assets, to accurately price and hedge derivative securities, and to understand the behaviour of financial asset prices. The parameter γ_t can also be interpreted as the price of co-skewness. To understand this, assume that there are two assets: the risk-free and the market return. The pricing kernel specified in equation (5) is reduced to a quadratic function of the market return. The quadratic pricing kernel is used in Harvey and Siddique (2000) and, more recently, in Dittmar (2002) to investigate the role of co-skewness in asset-pricing models. When there is evidence that skewness is not important in an investment decision, the parameter γ_t is equal to zero. In that case, we say that skewness is not priced in the market and expression (5) is reduced to the pricing kernel of the capital asset-pricing model. The next proposition gives conditions under which the conditional variance of the pricing kernel specified in proposition 2.1 reaches the GHT bound.

Corollary 2.2 *Given the information set I_t, if the pure volatility contract risk premium is null, the conditional variance of m_{CHM} (see equation (5)) reaches the GHT bound.*

GHT also use conditioning information to derive an unconditional variance bound on pricing kernels. In the next section, we derive an unconditional variance bound on pricing kernels that incorporates conditioning information.

2.2 Variance bound with higher moments and conditioning information

Our goal in this section is to replicate the analysis in section 2.1 using $\mathcal{F}\left(\overline{m}, p_t^{\vartheta}\right)$ in place of $\mathcal{F}\left(\overline{m}_t, p_t^{\vartheta}\right)$ and using an unconditional projection in place of the conditional projection. We then consider the problem:

$$\min_{m \in \mathcal{F}\left(\overline{m}, p_t^{\vartheta}\right)} \sigma^2 (m). \tag{8}$$

Similarly to proposition 2.1, we show:

Proposition 2.3 *The pricing kernel, m_{UCHM}, solution to (8) is:*

$$m_{UCHM} = m^*_{GHT} + \gamma_t \varepsilon_{t+1}, \tag{9}$$

with $m^*_{GHT} = \left(p_t - \omega\mu_t\right)' \left(\sigma_t^2\right)^{-1} r_{t+1} + \omega$ *and* $\omega = \frac{\overline{m}-b_1}{1-d_1}$ *where:*

$$b_1 = Ep_t' \left(\mu_t^2 + \sigma_t^2\right)^{-1} \mu_t, \tag{10}$$

$$d_1 = E\mu_t' \left(\mu_t^2 + \sigma_t^2\right)^{-1} \mu_t. \tag{11}$$

Furthermore, the minimum variance bound with conditioning information and higher moments (hereafter, the UCHM bound) is:

$$\sigma^2_{UCHM} = \sigma^2_{GHT} + E\gamma_t' \left(\sigma^2_{\vartheta t} - s_t' \left(\sigma_t^2\right)^{-1} s_t\right) \gamma_t, \tag{12}$$

where σ^2_{GHT} is the GHT variance bound.

Proof. Let P_t be a space of payoffs at some future date on portfolios of assets and derivatives, and let P be the space of all random variables in P_t with finite unconditional second moments. Since m has a finite second moment, the unconditional least-squares projection of m onto P is the same as the conditional projection of m onto P_t. Hence, the solution to (8) is the same as (5), with $\overline{m}_t$ replaced by $\overline{m}$.[1] ■

In the case where conditional moments are replaced by unconditional moments: $\left(\mu_t, \sigma_t^2, s_t, \kappa_t\right) = \left(\mu, \sigma^2, s, \kappa\right)$, and conditional prices are replaced by unconditional prices $\left(p_t, p_t^\vartheta\right) = \left(p, p^\vartheta\right)$, the pricing kernel (9) is reduced to an unconditional minimum-variance pricing kernel. The variance of this pricing kernel is denoted the unconditional variance bound with higher moments (hereafter, HM bound). When the first four conditional moments $\left(\mu_t, \sigma_t^2, s_t, \kappa_t\right)$ and the price of the volatility contract are correctly calculated, it is easy to compute the UCHM bound. In the case where the first four conditional moments are not correctly specified, the UCHM bound is difficult to estimate. If one uses the semi-non-parametric method of Gallant, Hansen, and Tauchen (1990) to estimate conditional moments, it is possible to overestimate the true UCHM bound. In that case, the UCHM bound fails to be a lower bound for the variance of the pricing kernels.

2.3 Optimally scaled variance bound under higher moments

The conditional higher moments are not easy to compute. In this section, we derive a variance bound that remains a lower bound to the variance of pricing kernels even if conditional higher

[1] This argument is similar to the proof of Theorem A.2 in Hansen and Richard (1987).

moments are misspecified. To do this, we scale the risky asset returns with the conditioning random variable, $z_1 \in I_t$, that is believed to capture time variation in expected returns. Thus, the scaled return is $z_{1t}^{'} r_{t+1}$. In addition, we scale the non-linear component of the volatility contract that is not spanned by primitive assets with the conditioning variable $z_{2t} \in I_t$. Thus, the scaled payoff is $z_{2t}^{'} \varepsilon_{t+1}$. We then consider the payoff $z_t^{'} g_{t+1}$ with $z_t^{'} = \left(z_{1t}^{'}, z_{2t}^{'}\right)$ and $g_{t+1}^{'} = \left(r_{t+1}^{'}, \varepsilon_{t+1}^{'}\right)$, where ε_{t+1} is defined in (6). There exists an HJ bound based on the scaled payoff $z_t^{'} g_{t+1}$:

$$\sigma^2\left(\overline{m}, z_t^{'} g_{t+1}\right) = \frac{\left(E\left(z_t^{'}\pi_t\right) - \overline{m}E\left(z_t^{'} g_{t+1}\right)\right)^2}{Var\left(z_t^{'} g_{t+1}\right)}, \tag{13}$$

where $\pi_t^{'} = \left(p_t^{'}, p_t^{\vartheta\prime} - \overline{p_t^{\vartheta}}^{\prime}\right)$. We call expression (13) the scaled variance bound with higher-order moments. The relevant question we ask is: what conditioning variable z_t yields the best (largest) scaled variance bound with higher-order moments? This is a problem of variational calculus. We call this bound the "Optimally scaled bound under Higher Moments" (hereafter, the OHM bound). The OHM bound is:

$$\sigma_{OHM}^2 = \sup_{z_t \in I_t} \sigma^2\left(\overline{m}, z_t^{'} g_{t+1}\right). \tag{14}$$

This bound is the highest variance bound that incorporates higher moments when the conditioning information is used. To derive the solution to (14), we consider the following notation:

$$a_1 = E\left(p_t^{'}\left(\mu_t \mu_t^{'} + \sigma_t^2\right)^{-1} p_t\right), \tag{15}$$

$$a_2 = E\left(p_t^{\vartheta} - s_t^{'}\left(\sigma_t^2\right)^{-1} p_t\right)^{'}\left(\sigma_{\vartheta t}^2 - s_t^{'}\left(\sigma_t^2\right)^{-1} s_t\right)^{-1}\left(p_t^{\vartheta} - s_t^{'}\left(\sigma_t^2\right)^{-1} p_t\right), \tag{16}$$

$$b_2 = E\left(E_t\vartheta_{t+1} - s_t^{'}\left(\sigma_t^2\right)^{-1} E_t r_{t+1}\right)^{'}\left(\sigma_{\vartheta t}^2 - s_t^{'}\left(\sigma_t^2\right)^{-1} s_t\right)^{-1}\left(p_t^{\vartheta} - s_t^{'}\left(\sigma_t^2\right)^{-1} p_t\right), \tag{17}$$

$$d_2 = E\left(E_t\vartheta_{t+1} - s_t^{'}\left(\sigma_t^2\right)^{-1} E_t r_{t+1}\right)^{'}\left(\sigma_{\vartheta t}^2 - s_t^{'}\left(\sigma_t^2\right)^{-1} s_t\right)^{-1}\left(E_t\vartheta_{t+1} - s_t^{'}\left(\sigma_t^2\right)^{-1} E_t r_{t+1}\right), \tag{18}$$

and show:

Proposition 2.4 *The solution, z_t^*, to the maximization problem*

$$\sigma_{OHM}^2 = \sup_{z_t \in I_t} \sigma^2\left(\overline{m}, z_t^{'} g_{t+1}\right)$$

is given by:

$$z_t^{*'} = \left(z_{1t}^{*'}, z_{2t}^{*'}\right),$$

with

$$z_{1t}^{*'} = \left(\mu_t \mu_t^{'} + \sigma_t^2\right)^{-1}\left(p_t - \omega\mu_t\right), \tag{19}$$

and

$$z_{2t}^{*'} = \left(\sigma_{\vartheta t}^2 - s_t^{'}\left(\sigma_t^2\right)^{-1} s_t\right)^{-1}\left(p_t^{\vartheta} - \overline{p_t^{\vartheta}}\right). \tag{20}$$

So the optimally scaled payoff is $z_t^{*'} g_{t+1} = z_{1t}^{*'} r_{t+1} + z_{2t}^{*'} \varepsilon_{t+1}$. *Furthermore, the maximum bound with higher moments has two components:*

$$\sigma_{OHM}^2 = \left[\frac{a_1(1-d_1) + \overline{m}^2 d_1 - 2\overline{m} b_1 + b_1^2}{1-d_1}\right] + \left[\overline{m}^2 d_2 - 2\overline{m} b_2 + a_2\right], \tag{21}$$

where a_1, b_1, *and* d_1 *are defined in (10), (11), (15) and* a_2, b_2, d_2 *are defined in (16), (17), and (18). Each component of the maximum bound is positive.*

Proof. Bekaert and Liu (2004) give the solution to $\sup_{z_t \in I_t} \sigma^2\left(\overline{m}, z_{1t}^{'} r_{t+1}\right)$. Using $g_{t+1}^{'} = \left(r_{t+1}^{'}, \varepsilon_{t+1}^{'}\right)$ in place of $r_{t+1}^{'}$ in the proof provided by Bekaert and Liu (2004), we obtain:

$$z_t^* = \begin{pmatrix} \mu_t \mu_t^{'} + \sigma_t^2 & 0 \\ 0 & \sigma_{\vartheta t}^2 - s_t^{'}\left(\sigma_t^2\right)^{-1} s_t \end{pmatrix}^{-1} \left(\begin{pmatrix} p_t \\ p_t^{\vartheta} - \overline{p_t^{\vartheta}} \end{pmatrix} - \omega \begin{pmatrix} \mu_t \\ 0 \end{pmatrix}\right).$$

Substituting the optimally scaled payoff $z_t^{*'} g_{t+1}$ in $\sigma^2\left(\overline{m}, z_t^{'} g_{t+1}\right)$, we obtain the maximum bound with higher moments. ■

The optimal scaling factor $z_t^{*'} = \left(z_{1t}^{*'}, z_{2t}^{*'}\right)$ depends on the conditional distribution function through the first four conditional moments $\left(\mu_t, \sigma_t^2, s_t, \kappa_t\right)$. When these moments are known to econometricians or researchers, and if p_t^{ϑ} and p_t are correctly specified, we show the relation between the OHM and the UCHM bound.

Proposition 2.5 *Consider the payoffs* r_{t+1} *and* ϑ_{t+1} *with conditional prices* p_t *and* p_t^{ϑ}. *Assume that the first four conditional moments of asset payoffs are* $\left(\mu_t, \sigma_t^2, s_t, \kappa_t\right)$; *then the OHM bound is:*

$$\sigma_{OHM}^2 = \sigma_{UCHM}^2. \tag{22}$$

Proof. The UCHM bound represents the efficient way of using conditional information. Thus, it follows that:

$$\sigma^2\left(\overline{m}, z_t^{'} g_{t+1}\right) \leq \sup_{z_t} \sigma^2\left(\overline{m}, z_t^{'} g_{t+1}\right) \leq \sigma_{UCHM}^2.$$

From proposition 2.4, we know that σ_{OHM}^2 has the form described in (21). The variance of $z_t^{*'} g_{t+1}$ is:

$$Var\left(z_t^{*'} g_{t+1}\right) = Var\left(z_{1t}^{*'} r_{t+1}\right) + Var\left(z_{2t}^{*'} \varepsilon_{t+1}\right) = \sigma_{UCHM}^2.$$

We substitute $z_t^{*'}$ in this variance and use the definition of a_1, b_1, d_1 and a_2, b_2, d_2 to obtain $\sigma_{OHM}^2 = \sigma_{UCHM}^2$. ■

2.4 Relation to Bekaert and Liu (2004) and Ferson and Siegel (2001, 2003)

This paper is related to the Bekaert and Liu (2004; hereafter the BL bound) article. BL find the scaling factor that yields the largest HJ bound. Their variance bound is a function of the first two moments of asset returns. The BL bound uses only asset payoffs, whereas in this paper we use asset payoffs and derivatives. The BL optimally scaled bound is:

$$\sigma_{OSB}^2 = \frac{a_1 (1 - d_1) + \overline{m}^2 d_1 - 2\overline{m} b_1 + b_1^2}{1 - d_1}, \tag{23}$$

where a_1, b_1, and d_1 are defined in (15), (10), and (11). If the conditional skewness is not priced, $p_t^{\vartheta} = \overline{p_t^{\vartheta}}$ and the optimally scaled bound with higher moments collapses to the Bekaert and Liu (2004) optimally scaled bound:

$$\sigma_{OHM}^2 = \sigma_{OSB}^2.$$

Ferson and Siegel (2001) use conditioning information efficiently to solve for unconditionally minimum variance portfolios. Since there is a duality between HJ frontiers and the mean standard deviation frontiers, there exists a variance bound that is observationally equivalent to the Ferson and Siegel mean standard deviation frontiers. As mentioned in Bekaert and Liu (2004), this bound is not as sharp as the Bekaert and Liu bound because it restricts the portfolio weight to have a sum of one. Ferson and Siegel (2003) assume correct specification of the conditional moments and empirically illustrate the variance bound on the pricing kernel. Their bound is often close to, but lower than, the Gallant, Hansen, and Tauchen (1990) bound. The bounds derived in this paper are sharper than the Gallant, Hansen, and Tauchen and the Bekaert and Liu (2004) bounds. Consequently, they are sharper than the Ferson and Siegel bounds.

2.5 Relation to Snow (1991)

The present paper is also related to Snow (1991). Snow assumes that the pricing kernel must be a positive random variable, and it should correctly price the set of asset returns r_{t+1} and the call option $\left(\omega' r_{t+1}\right)^+$ with $\omega \in \mathbb{R}^n$. He then uses Holder's inequality to derive a lower bound on the δ^{th} moments of the pricing kernel m[2]:

$$\left(E\left[m^{\delta}\right]\right)^{\frac{1}{\delta}} \geq \lambda(\delta) = \sup_{p \in P} \frac{E\pi\left(p^+\right)}{E\left[p^{+\rho}\right]^{\frac{1}{\rho}}}, \tag{24}$$

where $\frac{1}{\rho} + \frac{1}{\delta} = 1$ and $\pi(x)$ represents the price of the portfolio x, and P represents the set $\left\{p = \omega' r_{t+1} : \omega \in \mathbb{R}^n\right\}$ of asset returns under consideration. From expression (24), it can be seen

[2] We would like to thank the referee for suggesting that we investigate the relationship between the unconditional variance bound with higher-order moments and Snow's (1991) bound.

that Snow provides a direct link between the δ^{th} moments of the pricing kernel and the ρ^{th} moments of asset returns. Snow's bound has some similarities to our unconditional bound with higher moments. Snow's variance (2^{th} moments) bound depends on the variance of the option payoff $\left(\omega' r_{t+1}\right)^{+}$. Therefore, it depends on the higher moments of the asset returns. This paper provides an unconditional variance bound on pricing kernels that depends on the skewness and kurtosis of asset returns. However, there are also many differences between our bound and Snow's bound, so that our respective papers should be viewed as complements rather than substitutes. First, our unconditional variance bound has a structural interpretation in terms of asset returns mean, variance, skewness, and kurtosis, while Snow's variance bound does not. We relate our bound to the Hansen and Jagannathan variance bound and show that if skewness is not priced, our bound reaches the Hansen and Jagannathan variance bound. There is no such interpretation for Snow's bound. Second, the computation of Snow's bound requires knowledge of the option price $\pi\left(p^{+}\right)$, which is not known. In his empirical implementation, Snow assumes that $\pi\left(p^{+}\right) = \pi\left(p\right)$ and computes the lower bound on the δ^{th} moments of a pricing kernel m using three data sets: small firms, large firms, and small and large firms. He then shows that the moments of the returns of small firms contain information about the pricing kernel that is not contained in the moments of the returns of large firms. Even though the results found in Snow (1991) are interesting, it is useful to point out that the assumption $\pi\left(p^{+}\right) = \pi\left(p\right)$ allows Snow's bound to depend only on asset-return moments. This assumption ignores the price of the call option. This price is an interesting component that can be used to capture the risk premium on the volatility contract p^2. As shown in section 2.1, the price of the volatility contract is closely related to the market price of skewness. Our unconditional variance bound depends not only on higher-order moments (co-skewness, co-kurtosis), but also on the volatility contract risk premium. Third, the lower bounds obtained in this paper are derived without a positivity requirement on pricing kernels, whereas Snow considers positive pricing kernels.

3. Implied Distance Measure

3.1 Distance measures

Consider the set, $\mathcal{F}\left(\overline{m}, p^{\vartheta}\right)$, of admissible pricing kernels that price the bond, the set of assets payoff, and the volatility contract. Let h_{t+1} be the payoff of risky assets or derivatives and let y_{t+1} be the pricing kernel of a pre-specified asset-pricing model. The price assigned by this pricing kernel should belong to $\mathcal{F}\left(\overline{m}, p^{\vartheta}\right)$. When the pre-specified asset-pricing model is false, $y_{t+1} \notin \mathcal{F}\left(\overline{m}, p^{\vartheta}\right)$ and there is a strictly positive distance between y_{t+1} and the set $\mathcal{F}\left(\overline{m}, p^{\vartheta}\right)$. This implies a positive pricing error of model y_{t+1} on payoff h_{t+1}; that is, $\left|E\left(y_{t+1}h_{t+1}\right) - E\left(m_{t+1}h_{t+1}\right)\right| > 0$ for all $m_{t+1} \in \mathcal{F}\left(\overline{m}, p^{\vartheta}\right)$. Similarly to Hansen and Jagannathan (1997), we define the distance measure with higher moments, which we call the HM distance:

$$\delta_{HM} = \min_{m \in \mathcal{F}\left(\overline{m}, p^{\vartheta}\right)} \|y - m\|, \tag{25}$$

where $\|x\| = \sqrt{E\left(x^2\right)}$ is the usual norm. Following Hansen and Jagannathan (1997, hereafter HJ), we obtain:

$$\delta_{HM} = \left[E\left(y_{t+1}\widetilde{r}_{t+1} - \widetilde{\pi}\right)' \left(E\widetilde{r}_{t+1}\widetilde{r}'_{t+1}\right)^{-1} E\left(y_{t+1}\widetilde{r}_{t+1} - \widetilde{\pi}\right)\right]^{\frac{1}{2}}, \tag{26}$$

where $\widetilde{r}_{t+1} = \left(r_{t+1}, \vartheta_{t+1}\right)$. The value $\widetilde{\pi} = \left(p, p^{\vartheta}\right)$ is the price of $\widetilde{r}_{t+1}$. The value δ_{HM} is the maximum pricing error for the set of portfolios based on asset returns and derivatives with the norm of the portfolio return equal to one. To see the relationship between the distance (26) and the HJ distance, we rewrite (26) as:

$$\delta^2_{HM} = \delta^2_{HJ} + \widetilde{\delta}^2, \tag{27}$$

where:

$$\delta^2_{HJ} = E\left(y_{t+1}r_{t+1} - p\right)' \left(Er_{t+1}r'_{t+1}\right)^{-1} E\left(y_{t+1}r_{t+1} - p\right), \tag{28}$$

and $\widetilde{\delta}^2 = E\left(y_{t+1}\varepsilon_{t+1} - \left(p^{\vartheta\prime} - \overline{p^{\vartheta}}'\right)\right)' \left(Var\left(\varepsilon_{t+1}\right)\right)^{-1} E\left(y_{t+1}\varepsilon_{t+1} - \left(p^{\vartheta\prime} - \overline{p^{\vartheta}}'\right)\right)$.

The value δ^2_{HJ} is the HJ distance. The value $\widetilde{\delta}^2 = \delta^2_{HM} - \delta^2_{HJ}$ is the deviation of the HM distance from the HJ distance. This value is a function of the asset return first four moments and the pure volatility contract risk premium. If non-linearities in volatility returns are not priced, $\widetilde{\delta}^2 = 0$ and the HM distance reaches the HJ distance.

The distance measure δ_{HM} is still unconditional. To incorporate conditioning information in this measure, we use the scaling argument of the previous section. We scale the returns and the residual ε_{t+1} with conditioning variables and derive the distance measure based on the scaled payoffs:

$$\delta^2\left(y_{t+1}, z'_t g_{t+1}\right) = \frac{\left(E\left(y_{t+1}z'_t g_{t+1} - z'_t \pi_t\right)\right)^2}{E\left(z'_t g_{t+1}\right)^2}, \tag{29}$$

with $z'_t = \left(z'_{1t}, z'_{2t}\right)$. We then ask the following question: what conditioning variable z_t yields the best (largest) scaled distance measure with higher moments?

$$\delta^2 = \sup_{z_t \in I_t} \delta^2\left(y_{t+1}, z'_t g_{t+1}\right). \tag{30}$$

The next theorem gives the solution to (30).

Proposition 3.1 *The solution $z^{*\prime}_t = \left(z^{*\prime}_{1t}, z^{*\prime}_{2t}\right)$ to the maximization problem (30) is given by*

$$\begin{aligned}
z^*_{1t} &= \left(\mu_t\mu'_t + \sigma^2_t\right)^{-1}\left(p_t - E_t y_{t+1} r_{t+1}\right), \\
z^*_{2t} &= \left(\sigma^2_{\vartheta t} - s'_t\left(\sigma^2_t\right)^{-1} s_t\right)^{-1}\left(p^{\vartheta\prime}_t - \overline{p^{\vartheta}_t}' - E_t y_{t+1}\varepsilon_{t+1}\right).
\end{aligned}$$

So the optimal distance measure with higher-order moments (hereafter, the OHM distance) is

$$\delta^2_{OHM} = \delta^2_{BL} + \widetilde{\delta}^2, \tag{31}$$

with:

$$\delta^2_{BL} = E\left[(E_t y_{t+1} r_{t+1} - p_t)' \left(\mu_t \mu_t' + \sigma_t^2\right)^{-1} (E_t y_{t+1} r_{t+1} - p_t)\right], \tag{32}$$

and

$$\widetilde{\delta}^2 = E\left[\left(E_t y_{t+1} \varepsilon_{t+1} - \left(p_t^{\vartheta\prime} - \overline{p_t^{\vartheta\prime}}\right)\right)' \left(\sigma^2_{\vartheta t} - s_t' \left(\sigma_t^2\right)^{-1} s_t\right)^{-1} \left(E_t y_{t+1} \varepsilon_{t+1} - \left(p_t^{\vartheta\prime} - \overline{p_t^{\vartheta\prime}}\right)\right)\right],$$

where y_{t+1} is the pre-specified pricing kernel. δ^2_{BL} represents the optimal distance measure using the Bekaert and Liu (2004) scaling approach. We call this distance the BL distance.

Proof. The proof is similar to the proof of proposition 2.4. Specifically, if y_{t+1} is constant, propositions 2.4 and 3.1 are identical.[3] ■

It is useful to point out that Hansen and Jagannathan (1997) also provide a distance measure for positive pricing kernels. In their empirical analysis, Hansen and Jagannathan find that the requirement that the pricing kernel must be positive does not make a big difference. Following their approach, the theoretical set-up provided in this section can be used to derive a distance measure, for positive pricing kernels, that incorporates higher moments. We intend to address this issue in future research.

3.2 Estimation of parameters

Assume we have an asset-pricing model with a proxy pricing kernel y_{t+1}. We will examine asset-pricing models in which the proxy pricing kernel is a linear function of a constant and a vector of variable factors, f_{t+1}. Let us define $F_{t+1}' = \left[1, f_{t+1}'\right]$, and let the vector of parameters be $b' = \left[b_0', b_1'\right]$. Thus, the pricing kernel proxy is

$$y_{t+1} = b' F_{t+1},$$

where F_{t+1} is the $k \times 1$ factor vector, and b is the $k \times 1$ coefficient vector. A big advantage of linear factor models is that they can be solved analytically. Non-zero elements of b indicate the relevance of a factor as a determinant of the pricing kernel. Define also the vector of returns $\mathbf{R}_{t+1} = \left(r_{t+1}, \mathbf{r}^{\upsilon}_{t+1}\right)$ with $\mathbf{r}^{\upsilon}_{t+1} = (\vartheta_{it+1}/p^{\upsilon}_{it})_{i=1,\dots,n}$ where $\mathbf{r}^{\upsilon}_{t+1}$ represents the vector of returns on the volatility contract. Similarly to the Hansen and Jagannathan (1997) framework, the estimate

[3] The proof is available from the author on request.

$\widehat{b}$ of b can be chosen to minimize δ_{HM} using the standard generalized method of moments (GMM) approach.

To estimate b, define the pricing error vector $g = E\left(y_{t+1}\mathbf{R}_{t+1} - \mathbf{1_N}\right)$, and its sample counterpart

$$g_T\left(b\right) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t y_t - \mathbf{1_N},$$

where T represents the number of time-series observations and $\mathbf{N}$ the number of assets and volatility contracts under consideration. Let W_T be a sample estimate of $W = E\left(\mathbf{R}_{t+1}\mathbf{R}_{t+1}^{'}\right)^{-1}$. By squaring (25), $\widehat{b}$ can be chosen as

$$\widehat{b} = \arg\min\delta_{HM}^2 = \arg\min g_T^{'}\left(b\right)W_T g_T\left(b\right). \tag{33}$$

Equation (33) is a standard GMM problem, but it is not the optimal GMM of Hansen (1982). The optimal GMM uses the weighting matrix $W_T = S_T^{-1}$, where S_T is a consistent estimator of $\left[TVar\left(g_T\right)\right]$. The weighting matrix, $W = E\left(\mathbf{R}_{t+1}\mathbf{R}_{t+1}^{'}\right)^{-1}$, proposed by Hansen and Jagannathan (1997), is invariant across asset-pricing models. We prefer the Hansen and Jagannathan (1997) weighting matrix because it allows us to compare different asset-pricing models. In this case, our weighting matrix is:

$$W = \begin{bmatrix} Er_{t+1}r_{t+1}^{'} & Er_{t+1}\mathbf{r}_{t+1}^{v'} \\ Er_{t+1}^{'}\mathbf{r}_{t+1}^{v} & E\mathbf{r}_{t+1}^{v}\mathbf{r}_{t+1}^{v'} \end{bmatrix}^{-1}. \tag{34}$$

As shown in expression (34), the matrix W_T is a function of the asset-returns covariance, skewness, and kurtosis matrix. The Hansen and Jagannathan weighting matrix depends only on the asset returns covariance. Using the first-order conditions of (33), it can be shown that the analytical solution for $\widehat{b}$ is

$$\widehat{b} = \left(D_T^{'}W_T D_T\right)^{-1}\left(D_T^{'}W_T p\right) \text{ with } D_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t F_t^{'}.$$

Following Hansen (1982), the asymptotic variance of $\widehat{b}$ is given by

$$var\left(\widehat{b}\right) = \frac{1}{T}\left(D_T^{'}W_T D_T\right)^{-1}D_T^{'}W_T S_T W_T D_T\left(D_T^{'}W_T D_T\right)^{-1}.$$

For the optimal GMM, the J-test is obtained with

$$J = g_T^{'}\left(\widehat{b}\right)var\left[g_T\left(\widehat{b}\right)\right]^{-1}g_T\left(\widehat{b}\right)\xrightarrow{d}\chi^2\left(\mathbf{N}-k\right).$$

Note that the distribution of δ_{HM} is not standard under the assumption that the true δ_{HM} equals zero. Jagannathan and Wang (1996) show that the distribution of $T\delta_{HJ}^2$ involves a weighted sum of $n-k\ \chi^2\left(1\right)$ statistics, where n is the number of assets and k is the number of estimated parameters.

Similarly to Jagannathan and Wang (1996), it can be shown that the distribution of $T\delta^2_{HM}$ involves a weighted sum of $\mathbf{N}-k$ $\chi^2(1)$ statistics. The weights are the $\mathbf{N}-k$ non-zero eigenvalues of

$$X = S_T^{1/2} W_T^{1/2} \left[I_{\mathbf{N}} - W_T^{1/2} D_T \left(D_T' W_T D_T \right)^{-1} D_T' W_T^{1/2'} \right] W_T^{1/2} S_T^{1/2'},$$

where $I_{\mathbf{N}}$ is the $\mathbf{N}$ dimensional identity matrix. $S_T^{1/2}$ and $W_T^{1/2}$ are the upper-triangular matrices obtained from the Cholesky decompositions of S_T and W_T. It can be shown that the matrix X has exactly $\mathbf{N}-k$ non-zero and positive eigenvalues. If we denote $\lambda_1, ..., \lambda_{\mathbf{N}-k}$ the eigenvalues of X, then the asymptotic sampling distribution of the HM distance is

$$T\delta^2_{HM} \xrightarrow{d} t_\chi = \sum_{j=1}^{\mathbf{N}-k} \lambda_j \chi_i,$$

where χ_1,...,$\chi_{\mathbf{N}-k}$ are independent $\chi^2(1)$ random variables. To determine the p-values, $p(\delta_{HM}=0)$, of the test $\delta_{HM}=0$ under the null hypothesis that the true distance δ_{HM} is zero, one needs to simulate the statistic t_χ. The standard errors for the estimates of the HM and HJ distance are calculated under the alternative hypothesis that the true distance is not equal to zero as in equation (45) of Hansen and Jagannathan (1997). The approach described in this section can be used to estimate b and compute the p-values when the conditioning information is used. In this case, scaled returns will be used, instead of returns.

3.3 Economic significance of the distance measures

Hansen and Jagannathan (1997) and Campbell and Cochrane (2000) provide economic interpretation of the Hansen and Jagannathan distance measure, δ_{HJ}. We follow these authors and give two interpretations of the distance measure with higher moments.

The first interpretation is related to the expected return error for a portfolio of basis assets and derivatives. Consider a portfolio of assets and derivatives, and assume that the payoffs of these derivatives can be spanned by the basis asset returns and the volatility contract (see equation (1)). The return on this portfolio is $\theta'\mathbf{R}_{t+1}$. The true expected return for this portfolio, when priced with the true pricing kernel m_{t+1}, is

$$E\theta'\mathbf{R}_{t+1} = r_f\theta'\mathbf{1_N} - r_f cov\left(m_{t+1}, \theta'\mathbf{R}_{t+1}\right),$$

with $Em_{t+1} = r_f^{-1}$. Assume that the proxy pricing kernel prices correctly asset returns and the return on the volatility contracts. The expected return computed with the proxy pricing kernel y_{t+1} when $Ey_{t+1} = Em_{t+1}$ is

$$E^y\theta'\mathbf{R}_{t+1} = r_f\theta'\mathbf{1_N} - r_f cov\left(y_{t+1}, \theta'\mathbf{R}_{t+1}\right).$$

Hence, the expected return error is

$$E\theta^{'}\mathbf{R}_{t+1} - E^{y}\theta^{'}\mathbf{R}_{t+1} = r_{f}cov\left(y_{t+1} - m_{t+1}, \theta^{'}\mathbf{R}_{t+1}\right).$$

Using the Cauchy-Schwartz inequality, it can be shown that

$$\left|E\theta^{'}\mathbf{R}_{t+1} - E^{y}\theta^{'}\mathbf{R}_{t+1}\right| = r_{f}\left|cov\left(y_{t+1} - m_{t+1}, \theta^{'}\mathbf{R}_{t+1}\right)\right| \leq r_{f}\left\|y_{t+1} - m_{t+1}\right\|.\left\|\theta^{'}\mathbf{R}_{t+1}\right\|. \quad (35)$$

The inequality (35) holds as an equality when the portfolio return, $\theta^{'}\mathbf{R}_{t+1}$, is perfectly correlated with $y_{t+1} - m_{t+1}$. From the first-order conditions of (25), it can be shown that $y_{t+1} - m_{t+1} = \varphi^{'}\mathbf{R}_{t+1}$ with $\varphi = Wg$. Thus, the portfolio with shares $\theta = \varphi/\delta_{HM}$ is the maximally mispriced portfolio with norm equal to one. Substituting back θ into (35) and recognizing that $E\varphi^{'}\mathbf{R}_{t+1} = 0$ gives:

$$\frac{\left|E^{y}\varphi^{'}\mathbf{R}_{t+1}\right|}{\sigma\left(\varphi^{'}\mathbf{R}_{t+1}\right)} = r_{f}\delta_{HM} = r_{f}\delta_{HJ}\sqrt{1 + \frac{\widetilde{\delta}^{2}}{\delta_{HJ}^{2}}}. \quad (36)$$

The left-hand side of (36) is the maximum possible expected return error for a portfolio of basis assets and derivatives per unit of standard deviation under the assumption that the true pricing kernel and the proxy pricing kernel have the same mean. The intuition behind (36) is the following. Assume that two pricing kernels are estimated using the distance measure (25). Among these two pricing kernels, the one with the lowest value $r_{f}\delta_{HM}$ is the best, in the sense that it gives the lowest maximum expected return error for a portfolio of basis assets and derivatives. It is useful to point out that $r_{f}\delta_{HJ}$ is the Hansen and Jagannathan (1997) maximum expected return error for a portfolio of basis assets (only) per unit of standard deviation. If non-linearities contained in derivatives are not priced, $\widetilde{\delta} = 0$, and (36) coincides with the Hansen and Jagannathan maximum expected return error.

The second interpretation is related to the expected return error for a portfolio of basis assets only. Assume that, although non-linearities matter, investors are interested in the expected return error of a portfolio of basis assets only. The expected return error for this portfolio is

$$E\theta_{1}^{'}r_{t+1} - E^{y}\theta_{1}^{'}r_{t+1} = r_{f}cov\left(y_{t+1} - m_{t+1}, \theta_{1}^{'}r_{t+1}\right).$$

The first-order conditions of (25) imply that $y_{t+1} - m_{t+1} = \varphi^{'}\mathbf{R}_{t+1}$. Partitioning φ as $(\varphi_{1}, \varphi_{2})$ and substituting back this equality into the expected return error gives:

$$E\theta_{1}^{'}r_{t+1} - E^{y}\theta_{1}^{'}r_{t+1} = r_{f}cov\left(\varphi_{1}^{'}r_{t+1}, \theta_{1}^{'}r_{t+1}\right) + r_{f}cov\left(\varphi_{2}^{'}\mathbf{r}_{t+1}^{v}, \theta_{1}^{'}r_{t+1}\right). \quad (37)$$

When non-linearities matter, the second component in the right-hand side of equation (37) is a function of the higher moments of asset returns and the volatility contract risk premium. To compare the expected return error, (37), to the Hansen and Jagannathan (1997) maximum expected return error, we consider the Hansen and Jagannathan (1997) portfolio shares $\theta_1 = \gamma/\delta_{HJ}$ with $\gamma = \left(Er_{t+1}r_{t+1}^{'}\right)^{-1}(Ey_{t+1}r_{t+1} - 1_n)$. Using the share θ_1, Hansen and Jagannathan (1997) show that the maximum expected return error for a portfolio of basis assets (only) is:

$$\delta_{err} = \left|E^y\gamma^{'}r_{t+1}\right|_{HJ} = r_f\sigma_{HJ}\sigma\left(\gamma^{'}r_{t+1}\right). \tag{38}$$

Using the share θ_1 and substituting back this share into equation (37) gives the expected return error for a portfolio of basis assets (only) when accounting for non-linearities or higher moments:

$$\left|E^y\gamma^{'}r_{t+1}\right| = \left|r_f cov\left(\varphi_1^{'}r_{t+1}, \gamma^{'}r_{t+1}\right) + r_f cov\left(\varphi_2^{'}\mathbf{r}_{t+1}^{v}, \gamma^{'}r_{t+1}\right)\right|.$$

Thus, the maximum expected return error for a portfolio of basis assets (only) when accounting for non-linearities or higher moments is:

$$\delta_{err} = \left|E^y\gamma^{'}r_{t+1}\right|_{HM} = \left|r_f cov\left(\varphi_1^{'}r_{t+1}, \gamma^{'}r_{t+1}\right)\right| + \left|r_f cov\left(\varphi_2^{'}\mathbf{r}_{t+1}^{v}, \gamma^{'}r_{t+1}\right)\right|. \tag{39}$$

Equation (39) is the maximum expected return error for a portfolio of basis assets (only) when the distance measure with higher moments is used. It will be useful in the empirical illustration (see section 5) to compare $\left|E^y\gamma^{'}r_{t+1}\right|_{HM}$ and $\left|E^y\gamma^{'}r_{t+1}\right|_{HJ}$ and investigate whether higher-order moments help to have an accurate measure of the expected excess return for a portfolio of basis assets (only).

4. Illustration of the Variance Bounds: A Simulation Exercise

Do the variance bounds with higher moments contain information about the distribution of pricing kernels that is not contained in the HJ, GHT, and BL bounds? To shed light on this question, we use a simulation exercise. The BL econometric models are considered as a benchmark for comparison purposes. Implementation of these bounds requires knowledge of the conditional price of the volatility contract and conditional moments. To compute conditional moments, we consider econometric models estimated by BL. To compute the conditional price of the volatility contract, we assume that we live in a world with the pricing kernel of the form:

$$m_{t+1} = \phi_t\left(\frac{C_{t+1}}{C_t}\right)^{\theta_1}(R_{Mt+1})^{\theta_2}, \tag{40}$$

where $\frac{C_{t+1}}{C_t}$ is the gross consumption growth, R_{Mt+1} is the return on the market portfolio, θ_1, θ_2 are constant parameters, and ϕ_t may be constant or a time t parameter. Most consumption-based asset-pricing models produce a pricing kernel of the form (40). Under the assumption that the joint-process asset return and the pricing kernel are conditionally lognormally distributed, it can be shown that the price of the volatility contract is:

$$p_t^{\vartheta} = r_{ft} \frac{E_t \vartheta_{t+1}}{E_t \vartheta_{t+1} - \sigma_t^2}, \tag{41}$$

where r_{ft} is the conditional risk-free return.[4] We use the same data set and the econometric models proposed in Bekaert and Liu (2004).[5] The results obtained with the several BL econometric models are similar. We report the results only for the regime-switching model with time-varying transition probability (hereafter the TP RS model). With a likelihood-ratio test, BL cannot reject the TP RS at the 5 per cent level. The TP RS model exhibits interesting time-varying non-linearities in the asset return and consumption process. We use the estimated TP RS parameters as the true population values for the simulation. The conditional moments derived from the TP RS will be considered as the true conditional moments. To compute the misspecified conditional moments, we use the constrained vector autoregression (VAR) model (hereafter CO VAR) estimated in Bekaert and Liu (2004). With a likelihood-ratio test, BL reject the CO VAR model at the 5 per cent level with a p-value of 0.0000. To illustrate the variance bounds, we simulate asset returns based on the econometric model described above. Simulations use 15,500 observations where the first 500 observations are discarded. The OHM bound has three interesting properties:

Efficiency and predictability with higher moments We explore the efficiency and the predictability with higher moments. We empirically investigate whether higher-order moments may account for predictability in asset returns. In Figure 1, Graph A presents the variance bounds when data are simulated from the TP RS model. Four important results stand out in this graph. First, the difference between the HJ and the HM bounds reveals little predictability, although the difference between these bounds is sharper for small $\overline{m}$'s. Second, the difference between the OHM and the BL bound reveals considerable predictability. In addition, the difference between the UCHM and the GHT bound is considerable. When the pricing kernel mean is in the neighbourhood of 0.995, the OHM bound is 40 per cent higher than the BL bound, while the UCHM bound is 25 per cent higher than the GHT bound. The difference between the bounds that incorporate higher moments and the HJ bound reveals considerable predictability: the OHM bound is 75 per cent higher than the HJ bound, while the BL bound is 20 per cent higher than the HJ bound. Additionally, the UCHM bound is 40 per cent higher than the HJ bound, while the GHT bound is 20 per cent higher than the HJ bound. This predictability is due to (i) the market return's conditional higher moments, and (ii) the market return's pure volatility contract risk premium. This result shows that conditioning

[4] The proof of this formula is available from the author on request.

[5] We would like to thank Bekaert and Liu for providing us with their data set and parameter estimates.

variables that contain information about higher moments and the volatility contract risk premium help to better predict future returns. Surprisingly, the difference between the UCHM bound and the larger OHM bound is huge, with the OHM bound being larger. There are two potential explanations for this. First, the difference may be due to parameter uncertainty risk. Second, the lognormality assumption used to compute the conditional price of the volatility contract (41) may account for this difference. To examine this issue more closely, Graph B in Figure 1 presents the OHM bound with conditional moments calculated from the TP RS model and the conditional price of the volatility contract calculated from the CO VAR model. When the conditional price of the volatility contract is misspecified, Graph B reveals that the OHM bounds are below the bound calculated with the true conditional price (i.e., the price calculated from the TP RS model using (40)). The UCHM bound underestimates the true lower bound on the variance of pricing kernels. The difference between the UCHM bound calculated with the misspecified conditional price and the true conditional price is quite small. This leads us to conclude that the difference between the OHM and UCHM bound may be due to uncertainty risk.

Diagnostic We investigate whether the OHM bound can be used as a diagnostic tool for the specification of the first four conditional moments. Results are displayed in Figure 2. Graphs A and B present the bounds with data simulated according to the TP RS model and conditional moments calculated from the CO VAR model. Two results stand out. First, as shown in Graph A, the OHM bound highlights the misspecification of the first four conditional moments, while the BL bound does not. Second, as shown in Graph B, the GHT and UCHM bounds fail to highlight the misspecification of the first four conditional moments.

Robustness Figure 2 presents the bounds with data simulated according to the TP RS model and conditional moments calculated from the CO VAR model. When the first four conditional moments are misspecified, Graph A shows that the OHM bound underestimates the bound calculated with the true conditional moments. Graph B shows that the UCHM and GHT bounds calculated with misspecified conditional moments (moments calculated with the CO VAR model) quite overestimate the bound calculated with the true conditional moments (moments calculated with the TP RS model).

5. Performance of Asset-Pricing Models

We first present asset-pricing models of interest. Second, we provide a simple model-free approach to compute the price of the volatility contract, since the variance bounds and the distance measures depend on the price of the volatility contract. Third, we discuss the performance of these asset-pricing models and their implications in using two independent data sets. We use hedge fund returns and industry portfolio returns.[6]

5.1 Asset-pricing models

We evaluate asset-pricing models with linear and non-linear pricing kernels. We also investigate time-varying extensions of these models. The linear and non-linear pricing kernels include the capital asset-pricing model (CAPM), the Fama and French (1993) (hereafter FF) pricing kernel, and the quadratic pricing kernel of Harvey and Siddique (2000, hereafter HS). A big advantage of linear factor models is that they can be solved analytically. In the following, we briefly describe these models. We first consider the pricing kernel implied by the CAPM and its time-varying extensions:

$$\begin{aligned} m_{t+1}^{CP(1)} &= b_0 + b_1 r_{Mt+1}, \\ m_{t+1}^{CP(2)} &= b_0 + b_1 r_{Mt+1} + c_0 z_t, \\ m_{t+1}^{CP(3)} &= (b_0 + c_0 z_t) + (b_1 + c_1 z_t)\, r_{Mt+1}, \end{aligned} \tag{42}$$

where r_{Mt+1} is the excess return on the market portfolio, and $b_i's$ and $c_i's$ are constant parameters in the model. Second, we consider the Harvey and Siddique (2000) model and its time-varying extensions[7]:

$$\begin{aligned} m_{t+1}^{HS(1)} &= b_0 + b_1 r_{Mt+1} + b_2 r_{Mt+1}^2, \\ m_{t+1}^{HS(2)} &= (b_0 + c_0 z_t) + b_1 r_{Mt+1} + b_2 r_{Mt+1}^2, \\ m_{t+1}^{HS(3)} &= (b_0 + c_0 z_t) + (b_1 + c_1 z_t)\, r_{Mt+1} + (b_2 + c_2 z_t)\, r_{Mt+1}^2. \end{aligned} \tag{43}$$

The third linear model is the Fama and French three-factors model and its time-varying extensions. We choose this model for its successful performance in cross-sectional stock pricing and mutual fund pricing:

$$\begin{aligned} m_{t+1}^{FF(1)} &= b_0 + b_1 r_{Mt+1} + b_2 r_{SMBt+1} + b_3 r_{HMLt+1}, \\ m_{t+1}^{FF(2)} &= (b_0 + c_0 z_t) + b_1 r_{Mt+1} + b_2 r_{SMBt+1} + b_3 r_{HMLt+1}, \\ m_{t+1}^{FF(3)} &= (b_0 + c_0 z_t) + (b_1 + c_1 z_t)\, r_{Mt+1} + (b_2 + c_2 z_t)\, r_{SMBt+1} + (b_3 + c_3 z_t)\, r_{HMLt+1}, \end{aligned} \tag{44}$$

[6]We also repeat the analysis with the 25 Fama and French portfolio returns. The results are not tabulated and are available on request. Conclusions are similar.

[7]We do not investigate the cubic pricing kernel for the following reason. Dittmar (2002) shows that the cubic market return does not improve the performance of the pricing kernel when the market return is measured without human capital. The market return used in this paper is measured without human capital.

where r_{SMBt+1} (small minus big) is constructed as the difference in returns on small and big stocks. This factor captures risk related to size; r_{HMLt+1} (high minus low) is constructed as the difference in returns between high and low book-to-market stocks. This factor captures the book-to-market ratio.

5.2 A model-free approach to estimate the price of the volatility contract

It is well known that the volatility contract tends to change unpredictably over time. However, it is less understood whether investors require compensation for the volatility contract risk and, if so, to what extent. This issue has a number of important asset-pricing implications. Because the volatility contract is not a tradable asset and its market price is not observable, it is difficult to estimate its price. Previous researchers (Bakshi, Kapadia, and Madan (2003), Bondareko (2004), and Carr and Wu (2004)) relied on different assumptions in order to infer the volatility contract risk premium from prices of traded options. We propose a model-free approach to estimate the price of the volatility contract ϑ_{t+1}. This price is calculated in the following manner. Equation (3) states that the price of the volatility contract, ϑ_{t+1}, is:

$$p_{it}^{\vartheta} = \frac{\sigma_{it}^{*2}}{r_{ft}} + r_{ft}, \tag{45}$$

with:

$$\sigma_{it}^{*2} = E_t \frac{m_{t+1}}{E_t m_{t+1}} \left(r_{it+1} - r_{ft}\right)^2 \text{ with } E_t m_{t+1} = \frac{1}{r_{ft}}.$$

As articulated in Bakshi and Madan (2000), any payoff function with a bounded expectation can be spanned by a set of out-of-money European call and put. Since the payoff $\left(r_{it+1} - r_f\right)^2$ has a bounded expectation, it can be spanned by a collection of put and call. We then build on the Agarwal and Naik (2004), Bakshi and Madan (2000), and Bakshi, Kapadia, and Madan (2003) frameworks and specify a flexible piecewise linear involving the market return, the square of the market return, and the call option on the market index:

$$\left(r_{it+1} - r_{ft}\right)^2 = \beta_0 + \beta_1 R_{Mt+1} + \beta_2 R_{Mt+1}^2 + \beta_3 \max\left(R_{Mt+1} - k_1, 0\right) + \eta_{t+1,k_1}, \tag{46}$$

with some residual risk, but that residual risk will not be priced. The coefficients β_i's are constant. R_{Mt+1} represents the market return. We let the data determine the level k_1: this level is chosen to minimize the sum of the squared errors η_{t+1,k_1}^2. Since the square of the market return has a bounded expectation, it can be spanned by a set of put and call. Hence, specification (46) is consistent with the theoretical findings in Bakshi and Madan (2000) and Bakshi, Kapadia, and Madan (2003). The advandage of this specification is that it allows us to capture the contribution of linear (covariance), quadratic (co-skewness), and non-linear payoffs (call option) to the price of the volatility contract risk. The coefficient β_1 represents the volatility contract beta. Following the literature, this coefficient is expected to be negative. This should be attributed to the negative correlation between the volatility contract ϑ_{t+1} and the market return. The coefficient β_2 is closely related to the covariance between the volatility contract and the square of the market return. According to Harvey and Siddique (2000), this coefficient is the co-skewness of the volatility contract with the market. If this co-skewness is economically important, it will manifest through β_2. Following the empirical evidence provided by Carr and Wu (2004), this coefficient is expected to be positive. The coefficient β_3 captures the covariance between the volatility contract and the call option payoff. The sign of this coefficient is determined by the correlation between the volatility contract return and the call option return. The risks captured by β_1, β_2, and β_3 are important to effectively manage risk and allocate assets, to accurately price and hedge derivative securities and understand the behaviour of financial asset prices in general. Specification (46) provides a method to retrieve the price of the volatility risk. Applying the Hansen and Richard (1987) pricing formula to this specification, we deduce the risk neutral-variance of asset i:

$$\sigma_{it}^{*2} = \alpha_\sigma + \Lambda_\sigma \sigma_{mt}^{*2}, \tag{47}$$

with $\alpha_\sigma = \beta_0 + r_{ft}\beta_1 + \Lambda_\sigma r_{ft}^2 + r_{ft}\beta_3 Call_{k_1}$ and $\Lambda_\sigma = \beta_2$, where $Call_{k_1}$ represents the price of the call option with moneyness k_1, and σ_{mt}^{*2} represents the variance of the market return under the risk-neutral measure. To compute the price of the call option with moneyness k_1, a reasonable benchmark to start is to assume that R_{Mt+1} is lognormally distributed; then the price of the European call option is given by the Black-Scholes formula. The risk-neutral variance σ_{it}^{*2} can be substituted back into (45) to obtain a closed-form expression for the price of the volatility contract. Once the price of the volatility contract is calculated, it is easy to derive the return on the volatility contract ϑ_{t+1}/p_t^{v}.

5.3 Application to hedge funds and options

5.3.1 *Data*

We use hedge funds obtained from the TASS database.[8] It covers over 4,606 funds from February 1977 to March 2004. Our sample starts in January 1996 and ends in March 2004. The data provide monthly hedge fund returns. We use three types of indexes: 1) the Standard & Poor's Hedge Fund Index (SP); 2) the Hedge Fund Research (HFR) indexes, and 3) the Credit Suisse First Boston/Tremont (TREMONT). The conditioning variable used to proxy z_{1t}^* is the yield spread between 20-year Treasury bonds and 1-month Treasury bills. This variable has been used in the

[8] We would like to thank the referee for suggesting that we investigate this issue.

literature as a proxy for the changes of risk in the market. It is shown to be correlated with the business cycle.[9] The conditioning variable used to proxy z^*_{2t} is the lag of the volatility index, VIX, which measures the market expectation of 30-day volatility.[10] We also use hedge fund data after correcting for the backfilling (or instant-history) and the survivorship bias. The results with the hedge fund indexes are similar. Therefore, we present the results for TREMONT indexes without bias correction.

5.3.2 *Can we explain the pricing of the volatility contract with non-linear risk factors?*

Table 1 presents the piecewise linear fit for the volatility contract. The TREMONT indexes without bias correction are used. As shown in Table 1, the intercept β_0 comes out statistically significant (at the 5 per cent level) for all categories, except for Equity Market Neutral. The coefficient β_1, which captures the beta of the volatility contract, comes out statistically significant (at the 5 per cent level) for all categories, except for Fixed Income Arbitrage, Equity Market Neutral, Global Macro, and Managed Futures. As expected, this coefficient is negative for all categories, except for Equity Market Neutral. The CAPM argues that the expected excess return on an asset is proportional to the beta of the asset, or the covariance of the return on the asset with the market portfolio return. Qualitatively, the negative coefficient β_1 is consistent with the CAPM, given the well-documented negative correlation between the index returns and index volatility. The coefficient β_1 ranges from -10.01 to -0.03. The highest beta is obtained for Dedicated Short Bias, while the lowest beta is obtained for Fixed-Income Arbitrage. These results indicate that the market return is an important risk factor for the volatility contract. However, does the market factor matter only for the volatility contract? To investigate this issue, we look at the contribution of non-linear risk factors that appears in specification (46). The significance of the coefficients β_2 and β_3 reveals that the market return factor cannot fully explain the volatility contract. There are other economically interesting factors, such as the square of the market return and the call option payoff. As shown in Table 1, the coefficient β_2, which captures the co-skewness of the volatility contact with the market return, is positive and statistically significant at the 5 per cent level for most of the hedge fund categories. To see the economic impact of the squared market return factor, consider the Dedicated Short Bias category, which has some of the largest β_2 by magnitude. For a 1 per cent increase in the squared market return, the volatility contract based on the Dedicated Short Bias category changes

[9] For a robustness check, we use the yield on the three-month Treasury bill in excess of the yield on the one-month Treasury bill to proxy z^*_{1t}. The conclusions about our estimation do not change.

[10] The VIX measure is based on the S&P 500 index option prices and incorporates information from the volatility "skew" by using a wider range of strike prices, rather than just at the money series.

by 5.26 per cent. In this case, the squared market return has a larger economic impact on this volatility contract. The positive coefficient β_2 ranges from 0.02 to 5.26 for all hedge fund categories. Furthermore, the coefficient β_3 of the option factor is negative and statistically significant at the 5 per cent level for most of the hedge fund categories. To see the economic impact of the option return factor, consider again the Dedicated Short Bias category, which has some of the largest β_3 by magnitude. For a 1 per cent increase in the option return, the volatility contract based on the Dedicated Short Bias category changes by -0.99 per cent. This shows that the option return has a slightly larger economic impact on this volatility contract. The coefficient β_3 ranges from -0.99 to -0.01 for all hedge fund categories. This suggests that the non-linear factors, in addition to the market return, might be useful for explaining the volatility contract, and hence the price of the volatility contract. In addition, note that the specification (46) provides a reasonable estimate of the call option moneyness level k_1 (they are all significant at the 5 per cent level and they range from 0.96 to 1.03).[11]

5.3.3 *Performance of asset-pricing models*

We discuss the performance of asset-pricing models when the pricing kernel is expressed with constant and time-varying coefficients, as in equations (42), (44), and (43). The results are presented in Tables 2, 3, and 4. Table 2 presents the Hansen and Jagannathan (1997) distance measure, δ_{HJ}; the distance measure with higher moments, δ_{HM}; the Bekaert and Liu distance measure, δ_{BL}; and the distance measure with conditioning information and higher moments, δ_{OHM}. The standard errors for the distance measures are labelled se(δ). As described in section 3, the standard errors are calculated under the alternative hypothesis that the true distance is not equal to zero. These standard errors allow an assessment of the precision with which the distance measure is estimated. The p-values of the test $\delta = 0$ as calculated in section 3 under the null hypothesis that the true distance is zero are labelled P($\delta = 0$). The p-values of the J-statistics from optimal GMM estimates of the models are labelled P(J). δ_{err} is the maximum expected return error for a portfolio of basis-asset returns only. Tables 3 and 4 present the value and standard errors of constant and time-varying coefficients of pricing kernels. In the following, we first discuss the HJ distance results. Second, we discuss the HM distance results. We thereafter compare these two distances. Lastly, we introduce conditioning information into the distance measures and discuss the results.

[11] As a robustness check, we repeat the analysis for Standard & Poor's Hedge Fund Index (SP) and Hedge Fund Research indexes. We also do the analysis using hedge fund indexes after correcting the two well-known biases: backfilling and survivorship biases. We find similar conclusions. The results are untabulated, but are available from the author on request.

The HJ distance The p-values of the HJ distance indicate that the linear and quadratic pricing kernels and their time-varying extensions are all rejected at the 5 per cent significance level. The HJ distance measure and p-value suggest marginal improvement in moving from a linear pricing kernel to a quadratic pricing kernel. Interestingly, the HJ distance measure and p-value suggest significant improvement in moving from pricing kernels with constant coefficients to pricing kernels with time-varying coefficients. The linear pricing kernel with time-varying coefficients CP(3) reduces the distance measure by 10.10 per cent relative to the linear pricing kernel with constant coefficients CP(1). The quadratic pricing kernel with time-varying coefficients HS(3) reduces the distance measure by 13.24 per cent relative to the quadratic pricing kernel with constant coefficients HS(1). In addition, the quadratic pricing kernel HS(3) reduces the distance measure by 3.55 per cent relative to the linear pricing kernel CP(3). These results indicate that incorporation of the quadratic term in the pricing kernel and the use of time-varying coefficients in the pricing kernel improve the fit of the model. These results are consistent with the finding of Harvey and Siddique (2000) and Dittmar (2002). The p-values of the HJ distance indicate that the Fama and French pricing kernel and its time-varying extensions cannot be rejected at the 5 per cent significance level. The Fama and French pricing kernel with time-varying coefficients FF(3) reduces the distance measure from 0.3396 to 0.1571, a drop of 53.74 per cent relative to the Fama and French pricing kernel with constant coefficients FF(1). Thus, the results suggest that the Fama and French pricing kernel and its time-varying extensions outperform the linear and quadratic pricing kernel and their time-varying extensions in pricing the cross-section of hedge fund returns. Furthermore, Table 3 presents the value and standard errors of the coefficients b_i and c_i, $i = 0, 1, 2, 3$. The coefficients of the linear and quadratic pricing kernels have the right sign and magnitude. Some coefficients are statistically significant at the 5 per cent level. Moreover, the coefficients b_i of the Fama and French pricing kernel and their time-varying extensions have reasonable magnitude and are, in majority, statistically significant at the 5 per cent level. Note that there is no sign restriction on the coefficients of Fama and French pricing kernels.

The HM distance As shown in Table 2, the p-values of the HM distance indicate that the linear and quadratic pricing kernels and their time-varying extensions are all rejected at the 5 per cent significance level. The HM distance measure suggests marginal improvement in moving from a linear specification of the pricing kernel to a non-linear specification. The HM distance measure also suggests significant improvement in moving from pricing kernels with constant coefficients to pricing kernels with time-varying coefficients. The linear pricing kernel with time-varying coefficients CP(3) reduces the distance measure from 2.6280 to 2.4413, a drop of 7.10 per cent relative to the linear pricing kernel with constant coefficients CP(1). The quadratic pricing kernel with time-varying coefficients HS(3) reduces the distance measure from 2.6027 to 1.9132, a drop of 26.49 per cent relative to the quadratic pricing kernel with constant coefficients HS(1). Moreover, the quadratic pricing kernel HS(3) reduces the distance measure by 21.63 per cent relative to the linear pricing kernel CP(3) with time-varying coefficients. These results suggest that the incorporation of the quadratic term in the pricing kernel, and the use of a time-varying coefficient in the pricing kernel, improve the fit of the model. Contrary to the HJ distance measure, the p-value of the HM distance measure indicates that the Fama and French pricing kernel and its time-varying extensions are all rejected at the 5 per cent significance level. The time-varying extension of the Fama and French pricing kernel FF(3) reduces the distance measure from 2.5847 to 2.0064, a drop of 22.37 per cent relative to the Fama and French pricing kernel with constant coefficients FF(1). We further investigate the sign of the pricing kernel coefficients. Table 3 presents the value and standard errors of coefficients b_i and c_i, $i = 0, 1, 2, 3$. The coefficients b_i of the linear and quadratic pricing kernels have the right sign and magnitude, and are all statistically significant. This is particularly interesting, since the signs of the coefficients are restricted by preference theory. In addition, the coefficients of the Fama and French pricing kernels have a reasonable magnitude and are all statistically significant.

Comparing the HJ with the HM distance As shown by the Fama and French pricing kernel results (see Table 2), the HJ and HM distances and their p-values lead to different conclusions about asset-pricing models. These results show that some existing pricing models are able to describe returns ignoring the impact of higher-order moments. When accounting for the impact of higher moments or non-linearities, these same models have difficulty in pricing asset non-linearities or higher moments, or have difficulty in explaining returns on the assets. The HM distance measure is always higher than the HJ distance measure. As pointed out by Hansen and Jagannathan (1997), $r_f \delta_{HJ}$ can be interpreted as the maximum possible expected return error for a portfolio of basis assets (only) per unit of standard deviation under the assumption that the true pricing kernel and the proxy pricing kernel have the same mean. As discussed in section 3, $r_f \delta_{HM}$ represents the maximum possible expected return error for a portfolio of basis assets and derivatives per unit of standard deviation under the assumption that the true pricing kernel and the proxy pricing kernel have the same mean. Table 2 shows that the maximum possible expected return error for a portfolio of basis assets and derivatives is considerable. This error ranges from 1.9132 to 2.6280 if we assume a risk-free return, $r_f = 1$. When we allow the coefficients of the pricing kernels to be time varying, the quadratic pricing kernel has the lowest maximum expected return error ($\delta_{HM} = 1.9132$). These results suggest that the existing pricing kernels are unable to correctly price asset returns and derivatives.

Table 2 also shows the maximum expected return error for a portfolio of basis-asset returns only (δ_{err}). As shown in this table, when accounting for higher moments, the maximum expected return error for a portfolio of basis-asset returns only is lower than the one obtained when higher moments are ignored. For example, when accounting for higher moments, the time-varying quadratic pricing kernel HS(3) reduces the maximum expected error from 0.0108 to 0.0006, a decline of 94.44 per cent relative to the case where higher moments are ignored. Indeed, when accounting for higher moments, the time-varying Fama and French pricing kernel HS(3) reduces the maximum expected error from 0.0059 to 0.0014, a decline of 76.27 per cent relative to the case where higher moments are ignored. These results are consistent with the findings of Harvey and Siddique (2000), who argue that the pricing error of a portfolio of basis asset (only) can be partially explained by skewness. Thus, incorporating higher moments in the distance measure helps provide an accurate measure of the expected return of a portfolio of basis assets only. This conclusion is reinforced by the implied variance of the estimated pricing kernels. Recall that both the HJ and HM distance measures can be express as:

$$\|p\| = \sqrt{E\left(p\right)^2 + Var\left(p\right)},$$

where p is the adjustment to the pricing kernel necessary to reduce the distance to an admissible pricing kernel to zero. The distance measure has two components: it is a function of the expected deviation from some admissible pricing kernel and the variance of that deviation. A proxy pricing kernel with a small distance measure tends to reduce the volatility of the adjustment necessary to make the proxy admissible. Graphs A and B of Figure 3 present the estimated pricing kernels. Each pricing kernel is represented by its mean and standard deviation. Graph A shows pricing kernels estimated with the HJ distance, and Graph B shows pricing kernels estimated with the HM distance. As shown in Graph B, when accounting for higher moments, the variance of the estimated pricing kernels is higher than the variance of pricing kernels estimated with the HJ distance, rendering the pricing kernel admissible to the Hansen and Jagannathan variance bound. This may explain why higher moments help provide an accurate measure of the expected excess return.

The BL distance We discuss the performance of asset-pricing models with conditioning information. As shown in Table 2, the outcome of the distance measures with conditioning information differs markedly from the results of the distance measures without conditioning information. All pricing kernels except the linear one improve substantially relative to the case in which the conditioning information is not included in the distance measure. For example, the BL distance measure implied by the linear pricing kernel with time-varying coefficients CP(2) falls to 0.4433, a decline of 11.69 per cent relative to the same pricing kernel estimated with the HJ distance. In addition, the BL distance measure implied by the linear pricing kernel with time-varying coefficients CP(3) falls to 0.4430, a decline of 2.25 per cent relative to the same pricing kernel estimated with the HJ distance. However, the linear pricing kernels with constant and time-varying coefficients are all rejected at the 5 per cent significance level. Considerable further improvement is observed in moving from linear to quadratic pricing kernels. The BL distance measure also indicates that quadratic pricing kernels result in an additional decrease in the distance measure relative to the linear pricing kernels. For example, the quadratic pricing kernel HS(3) reduces the BL distance from 0.4430 to 0.3860, a drop of 12.87 per cent relative to the linear pricing kernel with time-varying extensions CP(3). However, the quadratic pricing kernel is rejected at the 5 per cent significance level. We also investigate the ability of the Fama and French pricing kernel and its time-varying extensions to price the cross-section of hedge fund returns when accounting for conditioning information. When accounting for conditioning information in the HJ distance measure (i.e., by using the BL distance), the Fama and French pricing kernel and its time-varying extensions outperform the linear and quadratic pricing kernels and their time-varying extensions. For example, the BL distance measure implied by the Fama and French pricing kernel FF(3) falls to 0.2431, a decline of 37.02 per cent relative to the quadratic pricing kernel HS(3), and a decline of 45.12 per cent relative to the linear pricing kernel CP(3). Moreover, the specification test cannot reject the Fama and French pricing kernels at the 5 per cent significance level. Thus, incorporating conditioning information in the HJ distance (the BL distance) appears to have a significant impact on the fit of the pricing kernel. We further investigate the sign of the pricing kernel coefficients. Table 4 presents the value and standard errors of coefficients b_i and c_i, $i = 0, 1, 2, 3$. These coefficients have the right magnitude and most are statistically significant at the 5 per cent level. In addition, the coefficients b_i of the linear and quadratic pricing kernels have the right sign.

The OHM distance We use the OHM distance to estimate the pricing kernels. As shown in Table 2, the OHM distance measure implied by the linear pricing kernels CP(1), CP(2), and CP(3) falls to 1.2357, 1.1766, and 1.1762, respectively, a decline of 52.98, 55.22, and 51.82 per cent relative to the results obtained with the HM distance (i.e., when accounting for higher moments and ignoring conditioning information). The linear pricing kernels and their time-varying extensions are all rejected at the 5 per cent significance level. Marginal improvement is observed in moving from linear to quadratic pricing kernels. The results in Table 2 also indicate that quadratic pricing kernels slightly reduce the distance measure relative to the linear pricing kernels. For example, the quadratic pricing kernel HS(3) reduces the distance measure from 1.1762 to 1.1588, a drop of 1.48 per cent. However, the quadratic pricing kernels are rejected at the 5 per cent significance level. The performance of the Fama and French pricing kernel and its time-varying extensions is enhanced by incorporating conditioning information and higher moments in the distance measure (i.e., by using the OHM distance measure). For example, the time-varying Fama and French pricing kernel FF(3) falls from 2.0064 to 1.0322, a considerable decline relative to the case in which conditioning information is not included in the distance measure with higher moments (i.e., using the HM distance). In addition, the OHM distance implied by the Fama and French pricing kernel FF(3) falls to 1.0322, a decline of 10.93 per cent relative to the quadratic pricing kernel HS(3) and a decline of 12.24 per cent relative to the linear pricing kernel CP(3). In contrast to the BL distance, the specification test rejects the Fama and French pricing kernel and its time-varying extension at the 5 per cent significance level. These results suggest that the BL and the OHM distances lead to different conclusions about asset-pricing models. Further, Table 4 presents the value and standard errors of the pricing kernel coefficients. These coefficients have the right magnitude and most are statistically significant at the 5 per cent level. In addition, the coefficients b_i of the linear and quadratic pricing kernels, except CP(2), have the right sign.

Comparing the BL with the OHM distance As shown in Table 2, the p-values of the BL and OHM distance measures implied by the Fama and French pricing kernel lead to different conclusions about asset-pricing models. These results reinforce the conclusion that some existing pricing models are able to describe returns ignoring the impact of higher-order moments. When accounting for the impact of conditioning information and higher moments, these same models have difficulty in explaining returns on the assets and are unable to price non-linearities or higher moments. Table 2 also shows that the maximum possible expected return error for a portfolio of basis assets and derivatives, $r_f \delta_{OHM}$, is considerable. This error ranges from 1.0322 to 1.2368 if we assume a risk-free return, $r_f = 1$. Although the existing pricing kernels are unable to correctly price asset returns and derivatives, these results suggest that conditioning information improves the ability of pricing kernels to price asset returns and derivatives. Table 2 also shows the maximum expected return error for a portfolio of basis assets only (δ_{err}); considerable improvement is observed in δ_{err} when accounting for higher moments and conditioning information. For example, when accounting for conditioning information and ignoring higher moments, the maximum expected return error for a portfolio of basis assets, δ_{err}, implied by the Fama and French pricing kernel FF(3) is 0.0059. When accounting for higher moments and conditioning information, δ_{err} reduces from 0.0059 to 0.001271×10^{-5}.

5.4 Application to industry portfolios

5.4.1 *Data*

Industry portfolios have been used in the empirical asset-pricing literature for tests of candidates' asset-pricing models. We utilize the return on 20 industry-sorted portfolios, where the industry definitions follow the two-digit SIC codes used in Moskowitz and Grinblatt (1999). The sample starts from January 1990 and ends in December 2005. Industry groupings proxy the investment opportunity set well. These groupings maximize intragroup and minimize intergroup correlations. The data used to compute the industry portfolio returns, value-weighted index return, and risk-free return were obtained from CRSP.

5.4.2 *Can we explain the price of the volatility contract with non-linear risk factors?*

Table 5 presents the piecewise linear fit for the volatility contract using industry portfolios. As shown in this table, the intercept β_0 is positive and statistically significant at the 5 per cent level for all industry portfolio returns, except for Electrical Equipment and Utilities. The coefficient β_1, which captures the volatility contract beta, comes out statistically significant (at the 5 per cent level) for all industries, except for Electrical Equipment and Utilities. The β_1's have the expected sign and range from -5.41 to -2.05. The significance of β_2 and β_3 indicates that the market factor cannot fully explain the price of the volatility contract. As shown in Table 5, the coefficient β_2, which captures the co-skewness of the volatility contract with the market return, is positive and statistically significant (at the 5 per cent level) for all industry portfolios, except for Electrical Equipment and Utilities.

To see the economic impact of the squared market return factor, consider the Primary Metals portfolio, which has some of the largest β_2 by magnitude. For a 1 per cent increase in the squared market return, the volatility contract based on the Primary Metals portfolio changes by 2.75 per cent. In this case, the squared market return has a larger economic impact on this volatility contract. The positive coefficient β_2 ranges from 1.04 to 2.75. Furthermore, the coefficient β_3 is negative and statistically significant (at the 5 per cent level) for most of the industry portfolios, except for Electrical Equipment and Utilities, which has a significant (at the 5 per cent level) and positive coefficient β_3. To see the economic impact of the option return factor on the volatility contract, consider again the Primary Metals portfolio, for which the coefficient β_3 is -0.63. For a 1 per cent increase in the option return factor, the volatility contract based on the Primary Metals portfolio changes by -0.63 per cent. This shows that the option return has a slightly larger economic impact on this volatility contract. These results suggest that the non-linear factors such as the square of the market return and the call option payoff, in addition to the market return, might be useful for explaining the volatility contract, and hence the price of the volatility contract. Note that the specification (46) provides a reasonable estimate of the call option moneyness level, k_1. They are all significant at the 5 per cent level, and they range from 0.98 to 1.05.

5.4.3 *Performance of asset-pricing models*

We use industry portfolio returns and discuss the performance of asset-pricing models when the pricing kernel is expressed with constant and time-varying coefficients, as in equations (42), (43), and (44). The results are presented in Tables 6, 7, and 8. Table 6 presents the distance measures, their standard errors, and p-values. It also presents the maximum expected return error for a portfolio of basis asset returns only, δ_{err}. Tables 7 and 8 present the value and standard errors of the constant and time-varying coefficients of pricing kernels. In the following, we first discuss the HJ distance results. Second, we discuss the HM distance results. We then compare these two distances. Lastly, we introduce conditioning information into the distance measures and discuss the results.

The HJ distance The p-values of the HJ distance indicate that the linear and quadratic pricing kernels and their time-varying extensions are all rejected at the 5 per cent significance level. The HJ distance measure suggests significant improvement in moving from the linear pricing kernel to the quadratic pricing kernel. For example, the quadratic time-varying pricing kernel HS(3) reduces the HJ distance from 0.4533 to 0.4085, a drop of 9.88 per cent relative to the linear time-varying pricing kernel CP(3). The HJ distance suggests marginal improvement in moving from the linear pricing kernel to its time-varying extensions. However, the HJ distance suggests significant improvement in moving from the quadratic pricing kernel HS(1) to its time-varying extension HS(3). The quadratic pricing kernel with time-varying coefficient HS(3) reduces the HJ distance from 0.4413 to 0.4085, a decline of 7.43 per cent relative to the quadratic pricing kernel with constant coefficients HS(1). These results indicate that the use of a time-varying coefficient and the incorporation of the quadratic term in the pricing kernel improves the fit of the model. We also investigate the ability of the Fama and French pricing kernel to explain industry returns. As shown in Table 6, the Fama and French pricing kernel, FF(3), reduces the HJ distance to 0.3784, a decline of 7.37 per cent relative to the quadratic pricing kernel HS(3), and a decline of 16.52 per cent relative to the linear pricing kernel CP(3). Thus, these results suggest that the Fama and French pricing kernel outperforms the linear and the quadratic pricing kernels in pricing the cross-section of industry returns.[12] Furthermore, we investigate the sign, magnitude, and significance

[12] Dittmar (2002) finds that the quadratic pricing kernel outperforms the Fama and French pricing kernel in pricing the cross-section of industry returns. Note that Dittmar (2002) assumes that the coefficients of the quadratic pricing

of the pricing kernel coefficients. Table 7 presents the value and standard errors of the pricing kernel coefficients. Most of the coefficients are statistically significant at the 5 per cent level. The coefficients of the linear and quadratic pricing kernels have the right sign and magnitude.

The HM distance As shown in Table 6, the HM distance and its p-value indicate that the pricing kernels and their time-varying extensions are all rejected at the 5 per cent significance level. The HM distance measure and its p-value suggest marginal improvement in moving from the linear pricing kernels to the quadratic pricing kernels. The HM distance measure and p-value also suggest significant improvement in moving from pricing kernels with constant coefficients to pricing kernels with time-varying coefficients. The linear pricing kernel with time-varying coefficients CP(3) reduces the distance measure from 5.7505 to 5.5374, a decline of 3.71 per cent relative to the linear pricing kernel with constant coefficients CP(1). The quadratic pricing kernel with time-varying coefficients HS(3) reduces the distance measure from 5.7276 to 5.4909, a decline of 4.13 per cent relative to the quadratic pricing kernel with constant coefficients HS(1). Thus, the quadratic pricing kernel with time-varying coefficients improves the fit of the model.

We also investigate the ability of the Fama and French pricing kernel to price the cross-section of industry returns. The time-varying extension of the Fama and French pricing kernel FF(3) reduces the distance measure from 5.7253 to 5.1330, a drop of 10.35 per cent relative to the Fama and French pricing kernel with constant coefficients FF(1). Furthermore, the Fama and French pricing kernel FF(3) reduces the HM distance measure from 5.5374 to 5.1330, a drop of 7.30 per cent relative to the time-varying linear pricing kernel CP(3). Indeed, the Fama and French pricing kernel FF(3) reduces the HM distance measure from 5.4909 to 5.1330, a drop of 6.52 per cent relative to the time-varying quadratic pricing kernel HS(3). These results suggest that incorporation of the time-varying Fama and French pricing kernel improves the fit of the model. The Fama and French pricing kernel outperforms the linear and quadratic pricing kernels and their time-varying extensions. We further investigate the sign of the pricing kernel coefficients. Table 7 presents the value and standard errors of the pricing kernel coefficients. The coefficients are all statistically significant at the 5 per cent level. It is particularly interesting to see that the linear and quadratic pricing kernels have the right sign.

---

kernel are a quadratic function of the conditioning variable while the coefficients of the Fama and French pricing kernel are a linear function of the conditioning variable. In Table 6, the HJ distance indicates that the quadratic pricing kernel HS(3), with linear time-varying coefficients, outperforms the Fama and French pricing kernel with constant coefficients FF(1). We do not investigate the case when the coefficients of the quadratic pricing kernel are a quadratic function of the conditioning variables.

Comparing the HJ with the HM distance As shown in Table 6, the maximum possible expected return error for a portfolio of basis assets and derivatives per unit of standard deviation, $r_f \delta_{HM}$, is considerable. This error ranges from 5.1330 to 5.7505 if we assume a risk-free return, $r_f = 1$. If we allow the coefficients of the pricing kernels to be time varying, the quadratic and the Fama and French pricing kernel FF(3) has the lowest maximum expected return error ($\delta_{HM} =$ 5.1330). However, the maximum possible expected return error for a portfolio of basis assets only, $r_f \delta_{HJ}$, ranges from 0.3784 to 0.4534 if we assume a risk-free return, $r_f = 1$. These results show that some existing pricing models are able to describe industry returns ignoring the impact of higher-order moments. When accounting for the impact of higher moments or non-linearities, these same models have difficulty in explaining returns on the assets and derivatives.

Table 6 also shows the maximum expected return error for a portfolio of basis asset returns only (δ_{err}). As shown in this table, when accounting for higher moments, the maximum expected return error is lower than the one obtained when higher moments are ignored. For example, when accounting for higher moments, the time-varying quadratic pricing kernel HS(3) reduces the maximum expected error from 0.0104 to 0.0004, a decline of 96.15 per cent relative to the case where higher moments are ignored. Indeed, when accounting for higher moments, the time-varying Fama and French pricing kernel HS(3) reduces the maximum expected error from 0.0113 to 0.0004, a drop of 96.46 per cent relative to the case where higher moments are ignored. Thus, incorporating higher moments in the distance measure helps provide an accurate measure of the expected return of a portfolio of basis assets only. This conclusion is reinforced by the implied variance of the estimated pricing kernels. Figure 4 presents the estimated pricing kernels. Each pricing kernel is represented by its mean and standard deviation. Graph A shows pricing kernels estimated with the HJ distance, and Graph B shows pricing kernels estimated with the HM distance. As shown in Graph B, when accounting for higher moments, the variance of the estimated pricing kernels is higher than the variance of pricing kernels estimated with the HJ distance, rendering the pricing kernel admissible to the Hansen and Jagannathan variance bound. This supports the argument that higher moments help provide an accurate measure of the expected excess return.

The BL distance We next discuss the performance of asset-pricing models with conditioning information. As shown in Table 6, the distance measures with conditioning information differ from the distance measures when the conditioning information is ignored. All pricing kernels with time-varying coefficients improve substantially relative to the case in which conditioning information is not included in the distance measure. For example, the BL distance measure implied by the linear pricing kernel with time-varying coefficient CP(2) falls to 0.2403, a drop of 46.99 per cent relative to the same pricing kernel estimated without conditioning information (i.e., using the HJ distance). In addition, the BL distance measure implied by the linear pricing kernel with time-varying coefficients CP(3) falls to 0.2392, a decline of 47.23 per cent relative to the same pricing kernel estimated with the HJ distance. The linear pricing kernels with constant coefficients are rejected at the 5 per cent significance level. However, the linear pricing kernels with time-varying coefficients cannot be rejected at the 5 per cent significance level. Further, marginal improvement is observed in moving from linear to quadratic pricing kernels. The BL distance measure indicates that quadratic pricing kernels result in an additional decrease in the distance measure relative to the linear pricing kernels. For example, the quadratic pricing kernel HS(3) reduces the BL distance from 0.2392 to 0.2268, a drop of 5.18 per cent relative to the linear pricing kernel with time-varying extensions CP(3). The quadratic pricing kernel with constant coefficient is rejected at the 5 per cent significance level. However, the quadratic pricing kernel with time-varying coefficients cannot be rejected at the 5 per cent significance level.

We also investigate the ability of the Fama and French pricing kernel and its time-varying extensions to price the cross-section of industry returns. The performance of the Fama and French pricing kernel and its time-varying extensions is improved by incorporating conditioning information in the distance measure (i.e., by using the BL distance). For example, the Fama and French pricing kernel FF(3) falls to 0.1974, a drop of 47.83 per cent relative to the same pricing kernel estimated without conditioning information (i.e., using the HJ distance). In addition, the Fama and French pricing kernels outperform the linear and quadratic pricing kernels. For example, the BL distance implied by the Fama and French pricing kernel FF(3) falls to 0.1974, a decline of 12.96 per cent relative to the quadratic pricing kernel HS(3), and a decline of 17.47 per cent relative to the linear pricing kernel CP(3). Moreover, the specification test cannot reject the Fama and French pricing kernel with time-varying coefficients at the 5 per cent significance level. Thus, incorporating conditioning information in the HJ distance (i.e., using the BL distance) appears to have a significant impact on the fit of the pricing kernel. We also investigate the sign of the pricing kernel coefficients. Table 8 presents the value and standard errors of the pricing kernel coefficients. The coefficients b_i of the linear and quadratic pricing kernels, except for CP(2) and HS(2), have the right sign.

The OHM distance We use the OHM distance to estimate the pricing kernels. As shown in Table 6, the OHM distance measure implied by the linear pricing kernels CP(1), CP(2), and CP(3) falls to 0.8538, 0.8313, and 0.8265, respectively, a considerable decline relative to the results obtained with the HM distance (i.e., when accounting for higher moments and ignoring conditioning information). The linear pricing kernels and its time-varying extensions are all rejected at the 5 per cent significance level. Marginal improvement is observed in moving from linear to quadratic pricing kernels. The results in Table 2 also indicate that quadratic pricing kernels cause a small decrease in the distance measure relative to the linear pricing kernels. For example, the quadratic pricing kernel HS(3) reduces the distance measure from 0.8265 to 0.7990, a drop of 3.33 per cent relative to the linear pricing kernel CP(3). However, the quadratic pricing kernels are rejected at the 5 per cent significance level. The performance of the Fama and French pricing kernel and its time-varying extensions is improved by incorporating conditioning information and higher moments in the distance measure (i.e., by using the OHM distance measure). For example, the time-varying Fama and French pricing kernel FF(3) falls from 5.1330 to 0.7197, a considerable drop relative to the case in which conditioning information is not included in the distance measure with higher moments (i.e., using the HM distance). In addition, the OHM distance implied by the Fama and French pricing kernel FF(3) falls to 0.7197, a decline of 9.92 per cent relative to the quadratic pricing kernel HS(3), and a decline of 12.92 per cent relative to the linear pricing kernel CP(3). In contrast to the BL distance, the specification test rejects the linear, the quadratic, and the Fama and French pricing kernel with time-varying coefficients at the 5 per cent significance level. These results suggest that the BL and the OHM distances lead to different conclusions about asset-pricing models. Furthermore, Table 8 presents the value and standard errors of the pricing kernel coefficients. Most of the coefficients of the linear and quadratic pricing kernels are statistically significant at the 5 per cent level. The sign of the coefficients of the linear pricing kernel is wrong. However, the coefficients b_i of the time-varying quadratic pricing kernels HS(3) have the right sign. The coefficients of the Fama and French pricing kernel and its time-varying extension are all statistically significant at the 5 per cent level.

Comparing the BL with the OHM distance As shown in Table 6, the p-values of the BL and OHM distance measures implied by the time-varying extension of the linear, the quadratic, and the Fama and French pricing kernel lead to different conclusions about asset-pricing models. These results show that, when accounting for the impact of conditioning information and higher moments, existing asset-pricing models have difficulty in explaining returns on the assets and are unable to price non-linearities or higher moments. Table 6 also presents the maximum possible expected return error for a portfolio of basis assets and derivatives, $r_f \delta_{OHM}$. This error ranges from 0.7197 to 0.8538 if we assume a risk-free return, $r_f = 1$. In addition, Table 6 presents the maximum expected return error for a portfolio of basis assets only (δ_{err}). As shown in this table, considerable improvement is observed in δ_{err} when accounting for higher moments and conditioning information. For example, when accounting for conditioning information and ignoring higher moments, the maximum expected return error for a portfolio of basis assets, δ_{err}, implied by the Fama and French pricing kernel FF(3) is 0.0001. When accounting for higher moments and conditioning information, δ_{err} reduces from 0.0001 to 0.008×10^{-5}.

6. Concluding Remarks

The finance profession is showing an increasing interest in building asset-pricing models that incorporate time-varying higher moments and variance risk premia. To compare asset-pricing models, it is critical to optimally incorporate higher moments and variance risk premia in the variance bound on pricing kernels. To evaluate the performance of asset-pricing models, it is also important to derive a distance measure that incorporates conditioning information, higher moments, and time-varying variance risk premia.

Our paper provides three variance bounds on pricing kernels. First, we derive an efficient variance bound on pricing kernels, which we call the UCHM bound. It incorporates time-varying higher moments and variance risk premia. Second, we derive a variance bound on pricing kernels, which we call the HM bound. It incorporates unconditional higher moments and variance risk premia. Third, we derive the best possible variance bound, which we call the OHM bound. It incorporates time-varying higher moments and variance risk premia. We show that the OHM bound is robust to the misspecification of the first four conditional moments of asset returns. There are interesting applications of this work. In a simulation exercise, we use these bounds to examine the predictability of asset returns when non-linearities in returns are priced. Important results stand out. First, the difference between the bounds derived in this paper and existing variance bounds reveals considerable predictability. Moreover, the OHM bound is significantly higher than the Bekaert and Liu (2004) optimally scaled bound. This result suggests that conditional higher moments contribute to better predict future returns. Second, while the Bekaert and Liu (2004) bound is robust to the misspecification of the first two moments of asset returns, the OHM bound is robust to the misspecification of the first four conditional moments of asset returns. Third, we show how the OHM bound can be used to propose a GMM-based specification test for the conditional first four moments.

Our paper also provides distance measures to evaluate asset-pricing models. We propose two distance measures. First, we propose an unconditional distance measure, which we call the HM distance. It incorporates higher moments and variance risk premia. When non-linearities in returns are not priced, the HM distance is reduced to the Hansen and Jagannathan distance (the HJ distance). We also propose the best (largest) distance measure, which we call the OHM distance, to evaluate pricing models. The OHM distance is a function of time-varying higher moments and time-varying variance risk premia. When non-linearities in returns are not priced, the OHM distance is reduced to the distance measure obtained using the Bekaert and Liu (2004) scaling approach (the BL distance).

We test the linear, the quadratic, and the Fama and French pricing kernel, and their time-varying extensions. To do this, we use hedge fund indexes and industry portfolio returns. When accounting for the impact of higher moments and variance risk premia (ignoring the conditioning information), tests of models show that the HM distance rejects all models at the 5 per cent significance level, while the HJ distance does not. These results indicate that some existing pricing kernels are able to describe returns ignoring the impact of higher-order moments and variance risk premia. When accounting for the impact of higher moments and variance risk premia, these same pricing kernels have difficulty in explaining returns on the assets and are unable to price non-linearities or higher moments. However, the maximum expected return error for a portfolio of basis assets only is reduced when accounting for higher moments and variance risk premia. This result is consistent with the findings of Harvey and Siddique (2000), who argue that the pricing error of a portfolio of basis assets (only) can be partially explained by skewness. Our results show that the pricing error of a portfolio of basis assets (only) can be partially explained by higher moments and variance risk premia. Moreover, the pricing kernels estimated with HJ distance often lie outside the region defined by the HJ bound. Although the pricing kernels estimated with the HM distance do not lie inside the HM bound, they generate sufficient volatility to be inside the region defined by the HJ bound. These results indicate that the HM distance contains information about the distribution of the pricing kernels that is not contained in the HJ distance. Further, when using the HM distance measure, we find that the Fama and French pricing kernel and its time-varying extensions are able to price the cross-section of return better than the linear and quadratic pricing kernels and their time-varying extensions.

When the conditioning information is used, tests of models show that the OHM and BL distances also yield different conclusions about asset-pricing models. The OHM distance rejects all pricing kernels at the 5 per cent significance level, while the BL distance does not. Further, the pricing kernels estimated with the OHM distance are able to price the cross-section of returns substantially better than the pricing kernels estimated with the HM distance. This suggests that time-varying higher moments and variance risk premia are important to price the cross-section of returns.

References

Agarwal, V. and N.Y. Naik. 2004. "Risks and Portfolio Decisions Involving Hedge Funds." *Review of Financial Studies* 17: 63–98.

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: 101–43.

Bakshi, G. and D. Madan. 2000. "Spanning and Derivative Security Valuation." *Journal of Financial Economics* 55: 205–38.

Bansal, R., D.A. Hsieh, and S. Viswanathan. 1993. "No-Arbitrage and Arbitrage Pricing: A New Approach." *Journal of Finance* 48: 1231–62.

Bekaert, G. and J. Liu. 2004. "Conditioning Information and Variance Bounds on Pricing Kernels." *Review of Financial Studies* 17: 339–78.

Bondareko, O. 2004. "Market Price of Variance Risk and Performance of Hedge Funds." Unpublished manuscript. University of Illinois at Chicago.

Campbell, J. and J. Cochrane. 2000. "Explaining the Poor Performance of Consumption-Based Asset Pricing Models." *Journal of Finance* 55: 2863–78.

Carr, P. and L. Wu. 2004. "Variance Risk Premia." Unpublished manuscript. New York University.

Chan, K.S. and R.S. Tsay. 1998. "Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model." *Biometrika* 85: 413–26.

Chapman, D. 1997. "Approximating the Asset Pricing Kernel." *Journal of Finance* 52: 1383–1410.

Constantinides, G.M. and D. Duffie. 1996. "Asset Pricing with Heterogeneous Consumers." *Journal of Political Economy* 104: 219–40.

Dittmar, R.F. 2002. "Non-Linear Pricing Kernels, Kurtosis Preference, and Evidence from Cross Section of Equity Returns." *Journal of Finance* 57: 368–403.

Fama, E. and K. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33: 3-56.

Ferson, W. and A.F. Siegel. 2001. "The Efficient Use of Conditioning Information in Portfolios." *Journal of Finance* 56: 967–82.

——. 2003. "Stochastic Discount Factor Bounds with Conditioning Information." *Review of Financial Studies* 16: 567–95.

Gallant, R., L.P. Hansen, and G. Tauchen. 1990. "Using Conditional Moments of Asset Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution." *Journal of Econometrics* 45: 141–79.

Hansen, L. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50: 1269-86.

Hansen, L.P. and R. Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy* 91: 249–65.

——. 1997. "Assessing Specification Errors in Stochastic Discount Factor Models." *Journal of Finance* 52: 557–90.

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.

Harvey, C.R. and A. Siddique. 2000. "Conditional Skewness in Asset Pricing Tests." *Journal of Finance* LV(3): 1263–95.

Heaton, J.C. 1995. "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications." *Econometrica* 63: 681–717.

Jagannathan, R. and Z. Wang. 1996. "The Conditional CAPM and the Cross-Section of Expected Returns." *Journal of Finance* 51: 3-53.

Moskowitz, T. and M. Grinblatt. 1999. "Do Industries Explain Momentum?" *Journal of Finance* 54: 1249–90.

Snow, K.N. 1991. "Diagnosing Asset Pricing Models Using the Distribution of Asset Returns." *Journal of Finance* 46: 955-83.

Piecewise Linear Fit: TREMONT indexes

	β_0	β_1	β_2	β_3	k_1	R^2
Convertible Arbitrage	0.1476	-0.3086	0.1612	-0.0243	0.9988	52.31%
	(0.0351)	(0.0739)	(0.0389)	(0.0072)	(0.0029)	
Fixed-Income Arbitrage	0.0154	-0.0327	0.0174	-0.0134	1.0535	0.73%
	(0.0101)	(0.0217)	(0.0116)	(0.0131)	(0.0223)	
Event Driven	0.1971	-0.4050	0.2079	-0.0305	1.0317	37.66%
	(0.0328)	(0.0676)	(0.0347)	(0.0124)	(0.0108)	
Equity Market Neutral	-0.2671	0.5847	-0.3180	0.0837	0.9988	2.19%
	(0.2495	(0.5436)	(0.2947)	(0.0717)	(0.0028)	
Long/Short Equity Hedge	0.4521	-0.9255	0.4739	-0.0779	1.0317	12.17%
	(0.0540)	(0.1134)	(0.0594)	(0.0681)	(0.0257)	
Global Macro	0.8284	-1.8397	1.0200	-0.2899	0.9767	1.32%
	(0.9133)	(1.9891)	(1.0814)	(0.2859)	(0.0124)	
Emerging Markets	2.1279	-4.4532	2.3301	-0.3894	0.9988	25.96%
	(0.6644)	(1.3984)	(0.7352)	(0.1380)	(0.0038)	
Dedicated Short Bias	4.7624	-10.0124	5.2636	-0.9949	0.9988	29.24%
	(1.2228)	(2.6208)	(1.4036)	(0.3187)	(0.0053)	
Managed Futures	0.3897	-0.8444	0.4585	-0.1031	0.9771	-0.72%
	(0.4064)	(0.8835)	(0.4785)	(0.1238)	(0.0160)	
Funds of Funds	0.2761	-0.5759	0.3003	-0.0452	0.9988	44.41%
	(0.0754)	(0.1596)	(0.0843)	(0.0162)	(0.0037)	
Other	1.1324	-2.3747	1.2443	-0.2175	0.9966	71.40%
	(0.2756)	(0.5788)	(0.3037)	(0.0560)	(0.0016)	

Table 1: This table shows the result of the following piecewise linear fit for TREMONT indexes from January 1996 to March 2004:

$$(r_{it+1} - r_{ft})^2 = \beta_0 + \beta_1 R_{Mt+1} + \beta_2 R_{Mt+1}^2 + \beta_3 \max\left(R_{Mt+1} - k_1, 0\right) + \eta_{t+1,k_1}.$$

Standard errors in parentheses are computed using Chan and Tsay (1998).

		δ_{HJ}	δ_{HM}		δ_{HJ}	δ_{HM}		δ_{HJ}	δ_{HM}
	CP(1)	0.5041	2.6280	HS(1)	0.5038	2.6027	FF(1)	0.3396	2.5847
se(δ)		0.1042	0.2191		0.1044	0.2271		0.1484	0.2213
P($\delta = 0$)		0.0061	0.0001		0.0033	0.0001		0.1843	0.0001
P(J)		0.0055	0.0000		0.0031	0.0000		0.1851	0.0000
δ_{err}		0.0097	0.0001		0.0096	0.0001		0.0074	0.0010
	CP(2)	0.5020	2.6278	HS(2)	0.5017	2.6018	FF(2)	0.1921	2.5327
se(δ)		0.1045	0.2199		0.1049	0.2263		0.1340	0.2354
P($\delta = 0$)		0.0037	0.0001		0.0019	0.0001		0.8289	0.0001
P(J)		0.0033	0.0000		0.0018	0.0000		0.8227	0.0000
δ_{err}		0.0092	0.0001		0.0091	0.0001		0.0049	0.0009
	CP(3)	0.4532	2.4413	HS(3)	0.4371	1.9132	FF(3)	0.1571	2.0064
se(δ)		0.1126	0.2902		0.1133	0.4064		0.1721	0.3218
P($\delta = 0$)		0.0103	0.0001		0.0039	0.0001		0.6600	0.0001
P(J)		0.0099	0.0000		0.0046	0.0000		0.6592	0.0000
δ_{err}		0.0112	0.0002		0.0108	0.0006		0.0059	0.0014
		δ_{BL}	δ_{OHM}		δ_{BL}	δ_{OHM}		δ_{BL}	δ_{OHM}
	CP(1)	0.5907	1.2357	HS(1)	0.5469	1.2368	FF(1)	0.3021	1.1189
se(δ)		0.1002	0.1446		0.1216	0.1421		0.1303	0.1798
P($\delta = 0$)		0.0004	0.0001		0.0006	0.0001		0.3614	0.0001
P(J)		0.0002	0.0000		0.0006	0.0000		0.3550	0.0000
δ_{err}		0.0002	0.0000		0.0002	0.0000		0.0001	0.0000
	CP(2)	0.4433	1.1766	HS(2)	0.4106	1.1769	FF(2)	0.3010	1.0471
se(δ)		0.1064	0.1596		0.1088	0.1606		0.1364	0.1797
P($\delta = 0$)		0.0261	0.0001		0.0385	0.0001		0.2692	0.0001
P(J)		0.0247	0.0000		0.0376	0.0000		0.2682	0.0000
δ_{err}		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000
	CP(3)	0.4430	1.1762	HS(3)	0.3860	1.1588	FF(3)	0.2431	1.0322
se(δ)		0.1069	0.1584		0.1140	0.1411		0.1214	0.1836
P($\delta = 0$)		0.0122	0.0001		0.0254	0.0001		0.2167	0.0001
P(J)		0.0146	0.0000		0.0249	0.0000		0.2201	0.0000
δ_{err}		0.0001	0.0000		0.0001	0.0000		0.0001	0.0000

Table 2: This table reports the Hansen and Jagannathan (1997) distance measure (δ_{HJ}), the distance measure implied by the Bekaert and Liu (2004) scaled variance bound (δ_{BL}), the distance measure with higher moments (δ_{HM}), and the distance measure with higher moments that incorporates conditioning information (δ_{OHM}). The asset returns considered are monthly TREMONT indexes returns and Treasury bills. The sample size is from January 1996 to March 2004. The standard errors for the distance are labelled se(δ). $p\,(\delta = 0)$ is the p-value for the test $\delta = 0$ calculated under the null $\delta = 0$. The p-value for the optimal GMM test is $p(J)$. δ_{err} is the maximum expected return error for a portfolio of basis assets only.

			δ_{HJ}	
	b_0	b_1	c_0	c_1
CP(1)	1.00	-3.09		
	(0.10)	(2.48)		
CP(2)	1.00	-3.25	7.63	
	(0.10)	(2.50)	(16.71)	
CP(3)	1.00	5.24	0.86	-682.21
	(0.10)	(4.69)	(17.01)	(319.27)

				δ_{HJ}		
	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	-12.69	4.85			
	(0.10)	(136.69)	(69.27)			
HS(2)	1.00	-21.47	9.23	7.65		
	(0.10)	(138.05)	(69.94)	(16.85)		
HS(3)	1.00	-252.52	130.75	34.85	-892.69	-11781.63
	(0.10)	(217.60)	(110.36)	(39.69)	(366.80)	(13089.44)

					δ_{HJ}			
	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	-23.61	-17.75	-223.34				
	(0.10)	(6.10)	(5.79)	(185.38)				
FF(2)	1.00	-29.88	-23.97	1636.18	188.91			
	(0.10)	(6.50)	(6.21)	(695.84)	(68.14)			
FF(3)	0.70	-36.92	-31.08	2658.24	302.38	510.72	427.63	-27151.00
	(0.33)	(24.55)	(23.50)	(1205.09)	(128.43)	(1862.84)	(2237.88)	(29160.84)

			δ_{HM}	
	b_0	b_1	c_0	c_1
CP(1)	1.00	-3.72		
	(0.10)	(2.32)		
CP(2)	1.00	-3.82	4.07	
	(0.10)	(2.35)	(14.47)	
CP(3)	1.00	-35.07	20.17	2392.58
	(0.10)	(4.01)	(14.57)	(248.58)

				δ_{HM}		
	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	292.00	-149.48			
	(0.10)	(82.19)	(41.53)			
HS(2)	1.00	307.72	-157.31	-10.49		
	(0.10)	(85.19)	(43.00)	(14.95)		
HS(3)	1.00	-1241.71	619.94	216.63	31.91	-78772.95
	(0.10)	(142.37)	(72.63)	(21.43)	(305.84)	(5537.63)

					δ_{HM}			
	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	14.11	19.92	400.41				
	(0.10)	(4.20)	(4.28)	(145.12)				
FF(2)	1.00	9.54	15.32	2852.64	264.23			
	(0.10)	(4.29)	(4.37)	(501.87)	(51.77)			
FF(3)	0.42	-79.28	-59.80	6810.17	574.54	7977.63	6262.67	-53662.00
	(0.23	(7.35)	(8.00)	(765.61)	(86.11)	(598.68)	(629.18)	(19129.10)

Table 3: This table reports the parameter estimates and standard errors when the HJ and HM distance measures are used. The sample size of the TREMONT indexes is from January 1996 to March 2004.

δ_{BL}

	b_0	b_1	c_0	c_1
CP(1)	1.00	-5.78		
	(0.10)	(2.80)		
CP(2)	1.00	-3.89	-49.76	
	(0.10)	(2.84)	(12.94)	
CP(3)	1.00	-2.87	-49.77	-47.34
	(0.10)	(7.83)	(12.94)	(336.24)

δ_{BL}

	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	-480.24	239.19			
	(0.10)	(208.23)	(104.91)			
HS(2)	1.00	-351.21	174.94	-46.53		
	(0.10)	(211.36)	(106.46)	(13.08)		
HS(3)	1.00	-773.89	390.66	-14.89	-392.54	-10818.92
	(0.10)	(384.67)	(194.61)	(28.34)	(377.75)	(8832.20)

δ_{BL}

	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	-30.83	-24.30	190.31				
	(0.10)	(9.21)	(9.15)	(145.21)				
FF(2)	1.00	-30.88	-24.78	384.50	21.89			
	(0.10)	(9.22)	(9.35)	(793.63)	(87.96)			
FF(3)	0.46	-32.31	-19.60	1425.61	146.83	175.46	-186.93	-60472.36
	(0.33)	(34.66)	(21.74)	(1086.12)	(128.68)	(1564.38)	(1081.38)	(35551.73)

δ_{OHM}

	b_0	b_1	c_0	c_1
CP(1)	1.00	-5.46		
	(0.10)	(2.36)		
CP(2)	1.00	-4.85	-41.63	
	(0.10)	(2.37)	(11.20)	
CP(3)	1.00	-3.98	-41.11	-56.97
	(0.10)	(3.78)	(11.33)	(192.59)

δ_{OHM}

	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	-77.63	36.57			
	(0.10)	(148.93)	(75.20)			
HS(2)	1.00	21.26	-13.16	-42.50		
	(0.10)	(151.25)	(76.37)	(11.35)		
HS(3)	1.00	-379.36	189.76	-11.14	-254.52	-11857.37
	(0.10)	(249.56)	(126.39)	(19.19)	(225.36)	(5905.36)

δ_{OHM}

	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	5.21	10.17	534.02				
	(0.10)	(4.71)	(4.59)	(100.27)				
FF(2)	1.00	1.29	5.88	2015.00	175.97			
	(0.10)	(4.82)	(4.72)	(394.28)	(45.31)			
FF(3)	1.13	8.47	13.81	1630.94	140.56	-474.73	-578.26	15178.53
	(0.19)	(7.55)	(6.86)	(548.33)	(67.04)	(462.38)	(510.78)	(17755.93)

Table 4: This table reports the parameter estimates and standard errors when the BL and OHM distance measures are used. The sample size of the TREMONT indexes is from January 1996 to March 2004.

	β_0	β_1	β_2	β_3	k_1	R^2
Mining	2.1842	-4.4598	2.2778	-0.3286	1.0227	26.61%
	(0.7988)	(1.6375)	(0.8387)	(0.1400)	(0.0045)	
Food and Beverage	1.0081	-2.0478	1.0400	-0.1268	1.0231	52.26%
	(0.2519)	(0.5145)	(0.2625)	(0.0416)	(0.0030)	
Textile and Apparel	1.1980	-2.3936	1.1970	-0.1879	1.0447	22.19%
	(0.2196)	(0.4541)	(0.2348)	(0.0790)	(0.0113)	
Paper Products	1.2800	-2.5686	1.2893	-0.2173	1.0490	41.75%
	(0.1919)	(0.3863)	(0.1944)	(0.0812)	(0.0103)	
Chemicals	2.5648	-5.2143	2.6519	-0.32	1.0317	10.43%
	(0.3021	(0.6271)	(0.3249)	(0.2701)	(0.0240)	
Petroleum	1.0585	-2.1539	1.0963	-0.2479	1.0490	19.03%
	(0.2845)	(0.5770)	(0.2924)	(0.1072)	(0.0105)	
Construction	1.2566	-2.5297	1.2739	-0.2013	1.0490	32.72%
	(0.1628)	(0.3299)	(0.1671)	(0.1113)	(0.0147)	
Primary Metals	2.6698	-5.4139	2.7453	-0.6326	1.0482	28.37%
	(0.3736)	(0.7784)	(0.4052)	(0.1607)	(0.0072)	
Fabricated Metals	1.3071	-2.6421	1.3357	-0.2720	1.0482	31.54%
	(0.2168)	(0.4381)	(0.2212)	(0.1054)	(0.0099)	
Machinery	2.1617	-4.3537	2.1942	-0.3151	1.0490	22.84%
	(0.3244)	(0.6751)	(0.3510)	(0.1358)	(0.0126)	
Electrical Equipment	0.1490	0.1679	-0.3253	0.6987	0.9839	17.40%
	(0.8844)	(1.9062)	(1.0259)	(0.2244)	(0.0028)	
Transport Equipment	1.5201	-3.0518	1.5326	-0.2346	1.0482	34.29%
	(0.2288)	(0.4636)	(0.2350)	(0.0884)	(0.0083)	
Manufacturing	1.7204	-3.4842	1.7659	-0.2948	1.0395	16.52%
	(0.2329)	(0.4907)	(0.2581)	(0.1187)	(0.0115)	
Railroads	1.2273	-2.4802	1.2537	-0.2877	1.049	34.47%
	(0.2330)	(0.4705)	(0.2374)	(0.0672)	(0.0058)	
Other Transportation	1.5542	-3.1209	1.5672	-0.2215	1.0482	41.16%
	(0.2704)	(0.5429)	(0.2725)	(0.0794)	(0.0082)	
Utilities	0.0158	0.0104	-0.0265	0.0674	0.9941	17.42%
	(0.1124)	(0.2453)	(0.1337)	(0.0347)	(0.0053)	
Department Stores	1.4568	-2.9355	1.4806	-0.2607	1.049	17.03%
	(0.2768)	(0.5708)	(0.2943)	(0.1005)	(0.0107)	
Other Retail	1.6882	-3.408	1.7208	-0.264	1.0395	25.39%
	(0.2483)	(0.5022)	(0.2538)	(0.1177)	(0.0123)	
Finance, Real Estate	1.0248	-2.1251	1.1023	-0.1546	0.9988	27.88%
	(0.2937)	(0.6187)	(0.3257)	(0.0635)	(0.0041)	
Other	2.688	-5.4201	2.734	-0.3643	1.0395	19.48%
	(0.3677)	(0.7516)	(0.3840)	(0.2289)	(0.0177)	

Table 5: This table shows the result of the following piecewise linear fit for industry portfolios from January 1990 to December 2005:

$$(r_{it+1} - r_{ft})^2 = \beta_0 + \beta_1 R_{Mt+1} + \beta_2 R_{Mt+1}^2 + \beta_3 \max\left(R_{Mt+1} - k_1, 0\right) + \eta_{t+1,k_1}.$$

Standard errors in parentheses are computed using Chan and Tsay (1998).

		δ_{HJ}	δ_{HM}		δ_{HJ}	δ_{HM}		δ_{HJ}	δ_{HM}
	CP(1)	0.4534	5.7505	HS(1)	0.4413	5.7276	FF(1)	0.4120	5.7253
se(δ)		0.0827	0.4640		0.0897	0.4513		0.0862	0.4669
P($\delta = 0$)		0.0036	0.0001		0.0049	0.0001		0.0150	0.0001
P(J)		0.0043	0.0000		0.0052	0.0000		0.0140	0.0000
δ_{err}		0.0135	0.0001		0.0136	0.0004		0.0134	0.0009
	CP(2)	0.4533	5.7429	HS(2)	0.4294	5.7261	FF(2)	0.4116	5.7219
se(δ)		0.0826	0.4606		0.0973	0.4506		0.0861	0.4649
P($\delta = 0$)		0.0024	0.0001		0.0064	0.0001		0.0095	0.0001
P(J)		0.0028	0.0000		0.0062	0.0000		0.0095	0.0000
δ_{err}		0.0135	0.0001		0.0120	0.0002		0.0132	0.0009
	CP(3)	0.4533	5.5374	HS(3)	0.4085	5.4909	FF(3)	0.3784	5.1330
se(δ)		0.0826	0.4467		0.1170	0.4305		0.0873	0.4733
P($\delta = 0$)		0.0020	0.0001		0.0067	0.0001		0.0132	0.0001
P(J)		0.0018	0.0000		0.0071	0.0000		0.0117	0.0000
δ_{err}		0.0135	0.0002		0.0104	0.0004		0.0113	0.0004
		δ_{BL}	δ_{OHM}		δ_{BL}	δ_{OHM}		δ_{BL}	δ_{OHM}
	CP(1)	0.4760	0.8538	HS(1)	0.4743	0.8460	FF(1)	0.4056	0.7814
se(δ)		0.0670	0.0682		0.0672	0.0682		0.0724	0.0774
P($\delta = 0$)		0.0026	0.0001		0.0012	0.0001		0.0171	0.0001
P(J)		0.0014	0.0000		0.0009	0.0000		0.0195	0.0000
δ_{err}		0.0006	0.0000		0.0006	0.0000		0.0002	0.0000
	CP(2)	0.2403	0.8313	HS(2)	0.2311	0.8176	FF(2)	0.2082	0.7562
se(δ)		0.0508	0.0721		0.0517	0.0711		0.0590	0.0890
P($\delta = 0$)		0.8872	0.0001		0.9039	0.0001		0.9421	0.0001
P(J)		0.8979	0.0000		0.8997	0.0000		0.9429	0.0000
δ_{err}		0.0004	0.0000		0.0003	0.0000		0.0001	0.0000
	CP(3)	0.2392	0.8265	HS(3)	0.2268	0.7990	FF(3)	0.1974	0.7197
se(δ)		0.0522	0.0720		0.0519	0.0799		0.0606	0.1015
P($\delta = 0$)		0.8685	0.0001		0.8337	0.0001		0.8824	0.0001
P(J)		0.8662	0.0000		0.8369	0.0000		0.8823	0.0000
δ_{err}		0.0004	0.0000		0.0003	0.0000		0.0001	0.0000

Table 6: This table reports the Hansen and Jagannathan (1997) distance measure (δ_{HJ}), the distance measure implied by the Bekaert and Liu (2004) scaled variance bound (δ_{BL}), the distance measure with higher moments (δ_{HM}), and the distance measure with higher moments that incorporates conditioning information (δ_{OHM}). The asset returns considered are industry portfolio returns and Treasury bills. The sample size is from January 1990 to December 2005. The standard errors for the distance are labelled se(δ). $p\,(\delta = 0)$ is the p-value for the test $\delta = 0$ calculated under the null $\delta = 0$. The p-value for the optimal GMM test is $p(J)$. δ_{err} is the maximum expected return error for a portfolio of basis assets only.

			δ_{HJ}	
	b_0	b_1	c_0	c_1
CP(1)	1.03	-5.51		
	(0.07)	(1.93)		
CP(2)	1.03	-5.53	2.32	
	(0.07)	(1.94)	(16.12)	
CP(3)	1.03	-5.48	6.32	-4.03
	(0.08)	(5.15)	(382.66)	(385.02)

				δ_{HJ}		
	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	-182.94	89.19			
	(0.07)	(123.82)	(62.23)			
HS(2)	1.00	-330.11	163.05	29.6		
	(0.07)	(162.33)	(81.54)	(21.12)		
HS(3)	1.00	-729.08	369.66	-10593.6	22236.41	-11589.7
	(0.07)	(278.99)	(142.63)	(7665.85)	(15652.08)	(7980.07)

					δ_{HJ}			
	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	-9.13	3.67	-7.15				
	(0.07)	(2.41)	(3.90)	(3.62)				
FF(2)	1.00	-9.10	3.05	-7.58	4.93			
	(0.07)	(2.42)	(4.60)	(3.99)	(19.51)			
FF(3)	1.00	-17.31	17.55	-3.19	765.63	703.96	-1190.34	-263.1
	(0.07)	(7.34)	(9.63)	(7.69)	(1172.27)	(518.98)	(669.26)	(698.97)

			δ_{HM}	
	b_0	b_1	c_0	c_1
CP(1)	1.05	-8.8		
	(0.07)	(1.87)		
CP(2)	1.04	-8.05	-49.95	
	(0.07)	(1.88)	(12.29)	
CP(3)	1.49	-79.38	-5928.13	5914.58
	(0.08)	(3.88)	(280.35)	(281.81)

				δ_{HM}		
	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	-441.4	217.96			
	(0.07)	(61.29)	(30.87)			
HS(2)	1.00	-401.77	198.17	-23.62		
	(0.07)	(65.08)	(32.74)	(13.04)		
HS(3)	1.00	-586.01	261.28	-41548.6	78304.95	-36704.4
	(0.07)	(113.55)	(58.52)	(3898.87)	(7974.29)	(4073.01)

					δ_{HM}			
	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	-15.15	-7.65	-23.28				
	(0.07)	(2.17)	(3.06)	(3.18)				
FF(2)	1.00	-14.64	-4.97	-21.11	-35.24			
	(0.07)	(2.18)	(3.22)	(3.28)	(12.99)			
FF(3)	0.76	-154.85	27.98	-127.1	-16186	10531.73	-4175.29	9604.98
	(0.07)	(4.86)	(6.03)	(5.54)	(696.43)	(328.21)	(433.20)	(431.54)

Table 7: This table reports the parameter estimates and standard errors when the HJ and HM distance measures are used. The sample size of industry portfolio is from January 1990 to December 2005.

δ_{BL}

	b_0	b_1	c_0	c_1
CP(1)	1.04	-6.49		
	(0.08)	(4.71)		
CP(2)	0.99	1.55	-80.60	
	(0.08)	(4.92)	(14.27)	
CP(3)	1.02	-4.34	-353.91	271.29
	(0.14)	(18.88)	(845.69)	(839.31)

δ_{BL}

	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	-83.89	38.38			
	(0.07)	(141.3)	(70.02)			
HS(2)	1.00	134.29	-65.65	-83.98		
	(0.07)	(146.4)	(72.37)	(14.75)		
HS(3)	1.00	-124.52	61.90	-5999.87	11692.74	-5766.09
	(0.07)	(669.35)	(340.40)	(13877.02)	(28363.8)	(14465.55)

δ_{BL}

	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	-7.15	-4.03	-13.70				
	(0.07)	(3.17)	(4.63)	(5.03)				
FF(2)	1.00	-2.52	0.96	6.34	-88.03			
	(0.07)	(3.31)	(4.75)	(6.55)	(18.40)			
FF(3)	0.99	-13.35	11.44	4.67	-213.97	510.21	-488.76	98.57
	(0.08)	(15.96)	(20.30)	(27.52)	(2066.29)	(735.29)	(956.3)	(1168.52)

δ_{OHM}

	b_0	b_1	c_0	c_1
CP(1)	0.96	6.61		
	(0.07)	(2.70)		
CP(2)	0.95	8.02	-31.42	
	(0.07)	(2.75)	(11.74)	
CP(3)	0.93	11.37	285.62	-314.39
	(0.08)	(3.88)	(259.53)	(257.10)

δ_{OHM}

	b_0	b_1	b_2	c_0	c_1	c_2
HS(1)	1.00	118.83	-57.19			
	(0.07)	(70.83)	(36.08)			
HS(2)	1.00	156.46	-75.56	-35.55		
	(0.07)	(71.95)	(36.60)	(11.91)		
HS(3)	1.00	-36.28	24.26	-8257.75	16837.57	-8595.71
	(0.07)	(109.20)	(56.30)	(3568.55)	(7235.26)	(3664.82)

δ_{OHM}

	b_0	b_1	b_2	b_3	c_0	c_1	c_2	c_3
FF(1)	1.00	-10.07	7.40	-10.18				
	(0.07)	(2.33)	(3.19)	(3.41)				
FF(2)	1.00	-10.02	11.91	-6.46	-35.86			
	(0.07)	(2.33)	(3.60)	(3.67)	(13.25)			
FF(3)	1.02	-10.16	26.00	3.33	1356.40	112.08	-989.19	-511.68
	(0.07)	(4.33)	(5.86)	(5.82)	(591.92)	(251.59)	(330.44)	(322.77)

Table 8: This table reports the parameter estimates and standard errors when the BL and OHM distance measures are used. The sample size of industry portfolio is from January 1990 to December 2005.

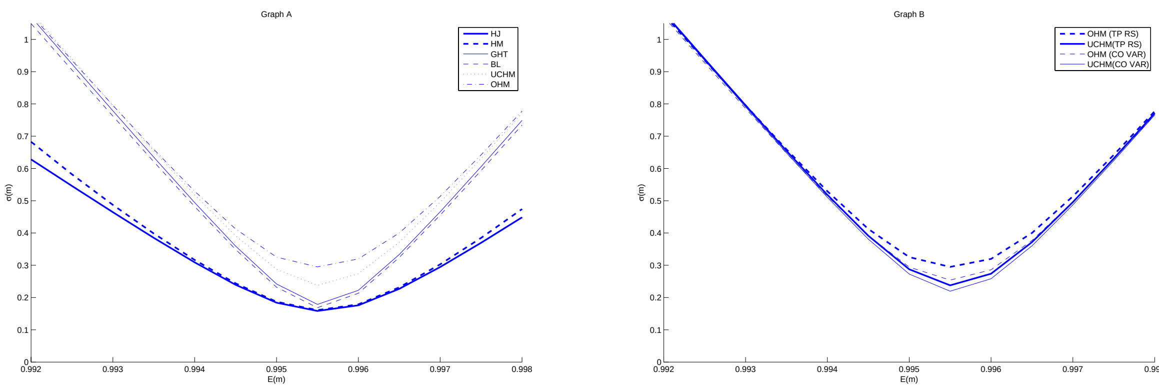


Figure 1: Graph A presents the variance bounds when data are simulated from the TP RS model, and when conditional moments and the conditional price of the volatility contract are calculated with the TP RS model. Graph B presents the OHM bound with conditional moments calculated from the TP RS model and the conditional price of the volatility contract calculated from the CO VAR model.

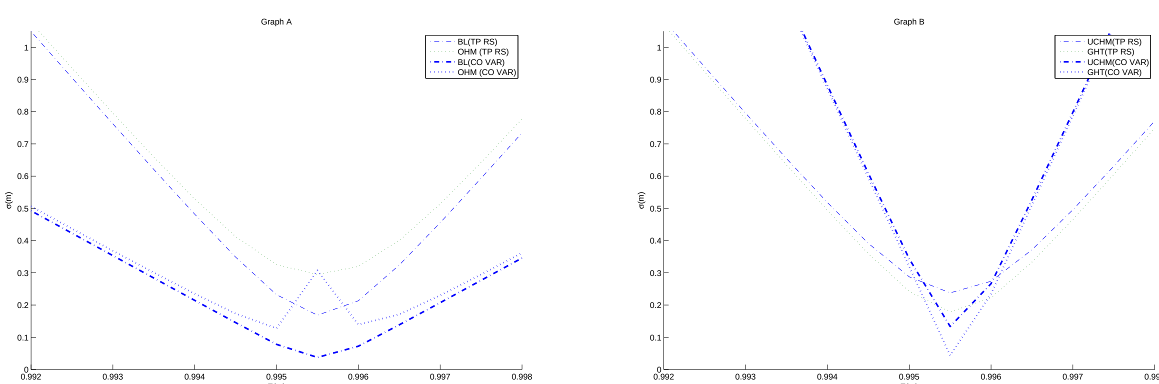


Figure 2: Graphs A and B present the bounds with data simulated according to the TP RS model and conditional moments calculated from the CO VAR model.

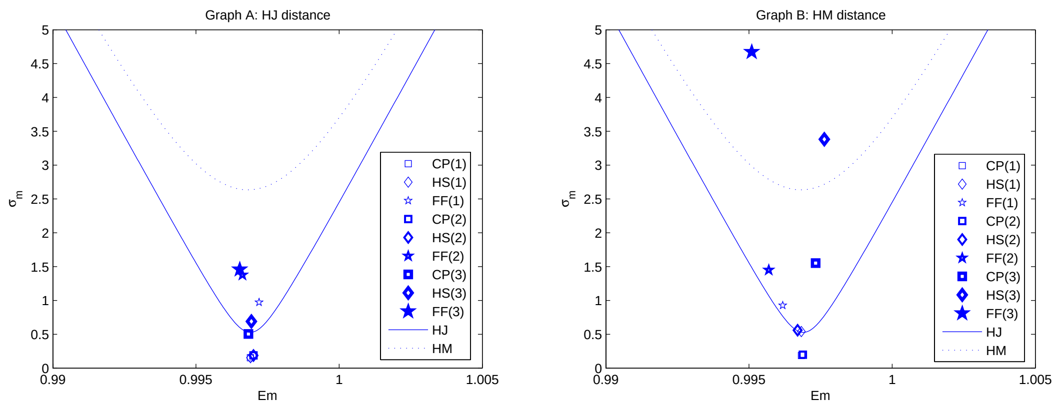


Figure 3: Graphs A and B present the HJ and HM bound. In addition, we plot in Graph A the pricing kernels estimated using the HJ distance. Graph B contains the pricing kernels estimated with the HM distance. We use TREMONT indexes without bias correction.

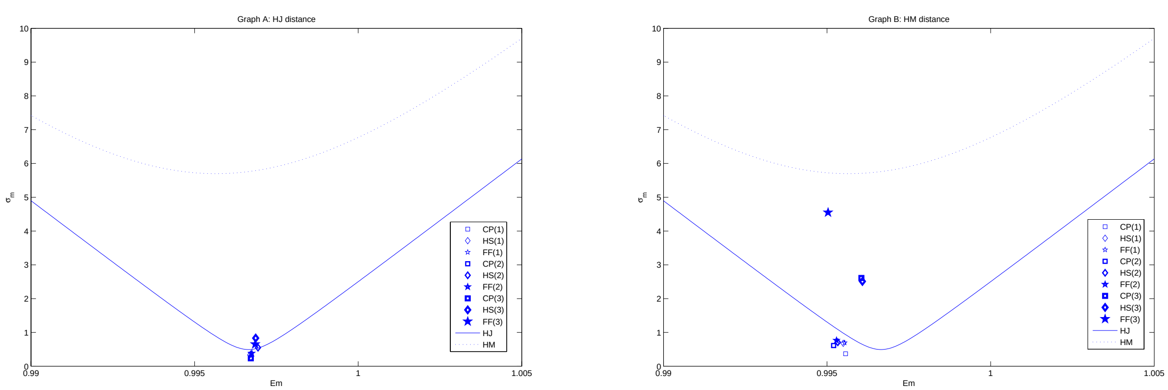


Figure 4: Graphs A and B present the HJ and HM bound. In addition, we plot in Graph A the pricing kernels estimated using the HJ distance. Graph B contains the pricing kernels estimated with the HM distance. We use industry portfolios.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages, and are available on the Bank's website (see bottom of page). ***Les documents de travail sont généralement publiés dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue. On peut les consulter dans le site Web de la Banque du Canada, dont l'adresse est indiquée au bas de la page.***

2006

2006-37	Endogenous Borrowing Constraints and Consumption Volatility in a Small Open Economy	C. de Resende
2006-36	Credit in a Tiered Payments System	A. Lai, N. Chande, and S. O'Connor
2006-35	Survey of Price-Setting Behaviour of Canadian Companies	D. Amirault, C. Kwan, and G. Wilkinson
2006-34	The Macroeconomic Effects of Non-Zero Trend Inflation	R. Amano, S. Ambler, and N. Rebei
2006-33	Are Canadian Banks Efficient? A Canada–U.S. Comparison	J. Allen, W. Engert, and Y. Liu
2006-32	Governance and the IMF: Does the Fund Follow Corporate Best Practice?	E. Santor
2006-31	Assessing and Valuing the Non-Linear Structure of Hedge Fund Returns	A. Diez de los Rios and R. Garcia
2006-30	Multinationals and Exchange Rate Pass-Through	A. Lai and O. Secrieru
2006-29	The Turning Black Tide: Energy Prices and the Canadian Dollar	R. Issa, R. Lafrance, and J. Murray
2006-28	Estimation of the Default Risk of Publicly Traded Canadian Companies	G. Dionne, S. Laajimi, S. Mejri, and M. Petrescu
2006-27	Can Affine Term Structure Models Help Us Predict Exchange Rates?	A. Diez de los Rios
2006-26	Using Monthly Indicators to Predict Quarterly GDP	I.Y. Zheng and J. Rossiter
2006-25	Linear and Threshold Forecasts of Output and Inflation with Stock and Housing Prices	G. Tkacz and C. Wilkins
2006-24	Are Average Growth Rate and Volatility Related?	P. Chatterjee and M. Shukayev
2006-23	Convergence in a Stochastic Dynamic Heckscher-Ohlin Model	P. Chatterjee and M. Shukayev
2006-22	Launching the NEUQ: The New European Union Quarterly Model, A Small Model of the Euro Area and the U.K. Economies	A. Piretti and C. St-Arnaud
2006-21	The International Monetary Fund's Balance-Sheet and Credit Risk	R. Felushko and E. Santor

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
Telephone: 1 877 782-8248
(toll free in North America)
Email: publications@bankofcanada.ca
Website: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Téléphone : 1 877 782-8248 (sans frais en Amérique du Nord)
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca