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

Fousseni CHABI-YO*  
Financial Markets Department, Bank of Canada  
March 16, 2005

Abstract

We show how to use conditioning information optimally to construct a more restrictive unconditional variance bound on pricing kernels that incorporates skewness, and propose a simple approach to test whether asset pricing models price (conditional) skewness. This bound depends on the first four conditional moments of asset returns and the conditional price of derivatives whose payoffs are defined as non-linear functions of the underlying asset payoffs. When skewness is priced, this bound is higher than the Gallant, Hansen and Tauchen (1990) bound (the GHT bound). When skewness is not priced, this bound reaches the GHT bound. We also propose an optimally scaled bound with higher-order moments (the OHM bound) which coincides with this sharper unconditional variance bound when the assets’ first four conditional moments and the conditional prices of derivatives are correctly specified. When skewness is priced, the OHM bound is higher than the Bekaert and Liu (2004, BL) optimally scaled bound. When skewness is not priced, the OHM bound reaches the BL bound. We also show how the OHM bound can be used as a diagnostic for the specification of the first four conditional moments of asset returns when the conditional prices of derivatives are correctly specified. We illustrate the behaviour of the bounds using a number of linear and non-linear models for consumption growth, bond and stock returns proposed in Bekaert and Liu (2004). The empirical results indicate a significant difference between the bounds that incorporate conditional higher-order moments and both the GHT and BL bounds.

JEL Classifications: G11, G12, C61.

Keywords: Conditioning Information, Pricing Kernels, Skewness, Kurtosis, Variance Bounds

* I would like to thank Rene Garcia and Eric Renault for helpful comments. I also thank Geert Bekaert and Jun Liu for providing the data set and parameter estimates. This article represents the views of the author, not the Bank of Canada. Address for correspondence: Fousseni Chabi-Yo, Financial Markets Department, 4 East Tower, Bank of Canada, 234 Wellington Street, Ottawa, Ontario. E-mail: fchabiyo@bankofcanada.ca.
1. Introduction

Hansen and Jagannathan (1991) derive a lower variance bound (the HJ bound) on any admissible pricing kernel (or intertemporal marginal rate of substitution) that correctly prices a given set of asset returns. Their bound is a function of the pricing kernel mean and the first two moments of asset returns. It is obtained by projecting the pricing kernel unconditionally on the space of asset payoffs and computing the variance of this projected pricing kernel. The HJ bound has many applications in finance. It can be used to test whether a particular consumption-based model produces a pricing kernel that prices correctly a given set of asset returns (see Cochrane and Hansen 1992). Others applications include asset pricing tests (see Hansen, Heaton and Luttmer 1994), market integration tests (see Chen and Knez 1995) and mean-variance-spanning tests (see Bekaert and Urias 1996).

The HJ bound is still unconditional. Gallant, Hansen and Tauchen (1990, hereafter GHT) show how to use conditional information efficiently to improve the HJ bound. They project the pricing kernel unconditionally on the space that combines asset payoff and conditioning information. The variance of this projected pricing kernel is called the GHT bound. The GHT bound depends on the pricing kernel mean and the first two conditional moments of the asset payoffs. But in practice, it is not easy to compute the conditional moments of asset returns. Of great use to researchers would be a simple technique that incorporates conditioning information into the HJ bound. Researchers scale returns with predictive variables in the information set, and then compute the HJ bound based on the space of available scaled payoffs (see Bekaert and Hodrick 1992). We call this bound the “scaled HJ bound”. If the conditioning variables are believed to predict future returns, the scaled HJ bound is an improvement on the original HJ bound. The question of whether there is a procedure allowing for sharper HJ bounds when scaled returns are used is investigated by Bekaert and Liu (2004). They provide a formal bridge between the optimal but unknown GHT bound and the ad hoc scaling methods used in the literature. Their bound has three important properties. First, their procedure optimally exploits conditioning information leading to sharper bounds. Second, it is robust to misspecification of the conditional mean and variance. Third, it can be used as a diagnostic tool for the specification of the first two conditional moments of asset returns. The Bekaert and Liu’s (2004, hereafter BL) optimally scaled bound depends on the scaled payoffs of the first two moments. However, there is increasing use of asset pricing models that incorporate skewness (see Harvey and Siddique 2000 and Dittmar 2002).

Recently, Chabi-Yo, Garcia and Renault (2004) augment the available asset payoffs space with derivatives whose payoffs are defined as non-linear functions of the underlying asset payoffs (these non-linear functions are approximated by their linear regression onto the asset payoffs and squared asset payoffs). They project the pricing kernel unconditionally on this augmented space. They term the variance of this projected pricing kernel, the “CGR bound”. The CGR bound depends not only on the first two moments of asset payoffs, but also on the higher moments of asset payoffs in particular on the skewness of asset payoff. Their CGR bound improves on the unconditional HJ bound. Their paper is the first to show that one can improve on the HJ bound by using derivatives instead of using conditioning variables.

This paper makes three contributions to asset pricing. First, we derive an efficient CGR bound. To do this, we construct an infinite space of available payoffs (or returns) and derivatives on the same payoffs.
We combine conditioning information, a set of asset payoffs and derivatives on the same asset payoffs. The variance of the unconditional projection of the pricing kernel onto that space is the efficient CGR bound. We call this variance, the “CY bound”. The CY bound depends on the first four conditional moments of the asset payoffs and the conditional price of the squared asset payoffs. When there is strong evidence that the (conditional) skewness is priced in the market, this bound is sharper than the GHT bound. However, when there is no evidence that the conditional skewness is priced, this bound reaches the GHT bound. The CY bound is easy to compute if the conditional moments are easy to derive but is not easy to compute these moments.

Second, we resort to a simple technique of incorporating conditioning information in the computation of the CGR bound. To do this, we first simply scale assets and the payoff of derivatives with predictive variables in the information set. We augment the space of available assets and derivatives with the relevant scaled asset payoffs and scaled derivative payoffs. We then project the pricing kernel unconditionally on this augmented space and derive the variance of this projected pricing kernel. This procedure is easy to implement and does not require knowledge of the first four conditional moments. We call this bound the “scaled CGR bound”.

Third, this article provides a formal bridge between the optimal but unknown CY bound and the scaled CGR bound. We prove four main results: i) We propose a simple way to test whether a particular asset pricing model prices the (conditional) skewness. ii) We assume that skewness is priced and answer the following question: When scaling asset payoffs and the payoff of derivatives with functions of the conditioning information, what is the function that maximizes the CGR bound? We call this bound, the optimally scaled bound with higher moments, “the OHM bound”. iii) We show that our bound is as tight as the CY bound when the first four conditional moments of asset returns and the conditional price of derivatives (whose payoffs are defined as non-linear functions of the asset payoffs) are correctly specified. iv) We give the conditions under which the OHM bound reaches the Bekaert and Liu (2004) optimally scaled bound.

The OHM bound has some advantageous features. First, the OHM bound is efficient. Our approach optimally exploits conditioning information with higher moments leading to a sharper bound. Second, the OHM bound is robust to 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 use the OHM bound to propose a diagnostic test for the first four conditional moments of asset returns when the conditional price of derivatives are correctly specified.

This article is organized in four sections. Section 2 describes the CGR bound in a one-asset case. In this section, the CY bound and the OHM bound are derived and we show that the OHM bound reaches the CY bound when the first four conditional moments of asset returns and the conditional price of derivatives are correctly specified. Section 3 proposes a simple approach to test whether asset pricing models price conditional skewness. Section 4 discusses the three important properties of the OHM bound, and compares the OHM bound to the Bekaert and Liu (2004) optimally scaled bound, as well as the CGR bound to the Bekaert and Liu (2004) optimally scaled bound. Section 5 derives the results of Section 2 in a multiple-assets case. Section 6 contains an empirical illustration based on the estimated results of Bekaert and Liu (2004).
2. The variance bounds in the one-asset case

In section 2.1, we review the CGR bound and set up notations. In section 2.2, we derive the conditional counterpart to the results reported in section 2.1. In section 2.3, we investigate a means of using conditioning information to derive an unconditional variance bound on pricing kernels. In section 2.4, we investigate the conditions under which scaling improves the CGR variance bound. We also examine the conditions under which our results mimic the Bekaert and Liu (2004) optimally scaled variance bound.

2.1 Unconditional variance bound

Hansen and Jagannathan (1991; hereafter HJ) consider a set of primitive assets and derive the variance bound on pricing kernels. CGR suppose that economic agents include in their portfolio asset returns and derivatives whose payoffs are approximated by their linear regression onto asset returns and squared asset returns. Intuitively, they augment the available asset space with derivatives. They project the pricing kernel on this augmented space and derive the variance of the projected pricing kernel. This variance is called the CGR variance bound. This bound is sharper than the HJ bound. CGR also define the conditions under which their variance bound coincides with the HJ bound. In this section, we briefly review the variance bound.

Let there be an asset with payoff $r_{t+1}$ and price $p$ and denote $c$ the price of the squared asset payoff. If the payoff $r_{t+1}$ is a return, $p$ is equal to 1 but $c$ is still different from 1. Consider now the set

$$F(\bar{m}, c) = \{ m_{t+1} \in L^2 : E(m_{t+1}) = \bar{m}, E(m_{t+1}r_{t+1}) = p, E(m_{t+1}r_{t+1}^2) = c \},$$

where $L^2$ represents the set of random variables with a finite second moment. CGR assume that $p$ is equal to one and solves the optimization problem

$$\min_{m \in F(\bar{m}, c)} \sigma^2(m),$$

(2.1)

although problem (2.1) can be solved for $p \neq 1$. The way to motivate (2.1) is to look at the pricing of derivatives, considering for example the return $r_{t+1}$ and derivative with payoff $h(r_{t+1})$. Assume that this payoff could be approximated by the linear regression of $h(r_{t+1})$ onto $r_{t+1}$ and $r_{t+1}^2$:

$$h(r_{t+1}) \approx EL[h(r_{t+1}) | r_{t+1}, r_{t+1}^2].$$

Thus, any pricing kernel that aims to price the payoff $h$ has to correctly price, at a minimum, the payoffs $r_{t+1}$ and $r_{t+1}^2$. If $m_{t+1}^{\text{us}}$ denotes the solution to problem (2.1), CGR show that:

$$m_{t+1}^{\text{us}} = m_{HJ} + \gamma \left( r_{t+1}^2 - (\mu^2 + \sigma^2) - \sigma^{-2}s(r_{t+1} - \mu) \right),$$

(2.2)

with

$$m_{HJ} = \bar{m} + \beta (r_{t+1} - \mu),$$

and

$$\beta = \sigma^{-2}(p - \bar{m}\mu),$$

$$\gamma = \left( \kappa - (\sigma^2 + \mu^2)^2 - s^2\sigma^{-2} \right)^{-1}(c - \bar{c}),$$

$$\bar{c} = \bar{m}(\mu^2 + \sigma^2 - s \sigma^{-2}\mu) + s \sigma^{-2}p,$$
where
\[ \kappa = E r_{t+1}^4, \quad s = Cov \left( r_{t+1}, r_{t+1}^2 \right), \quad \sigma^2 = Var \left( r_{t+1} \right), \quad \mu = E r_{t+1}. \]

The quantity \( \kappa \) will term the asset payoff fourth moment (kurtosis). For the sake of simplicity, our definition of skewness and kurtosis are not identical to standard definitions of skewness and kurtosis. The parameter \( s \) is related to the notion of skewness (see Ingersoll (1987) and more recently Harvey and Siddique 2000). The quantity \( \kappa - (\sigma^2 + \mu^2)^2 - s^2 \sigma^2 \) denotes the residual variance in the regression of \( r_{t+1}^2 \) onto \( r_{t+1} \). This quantity is assumed to be different from zero. This assumption will be maintained in the rest of this paper. The parameter \( \beta \) is determined by the correlation between the pricing kernel and the asset payoff, whereas \( \gamma \) is determined by the correlation between the pricing kernel and non-linear functions of the same asset payoff. To interpret \( \gamma \), assume there is a risk-free asset and that \( r_{t+1} \) is the return on the market portfolio, that is, \( r_{t+1} = r_{mt+1} \). Equation (2.2) is quadratic in the market return. This pricing kernel specification is used in Harvey and Siddique (2000) and more recently in Dittmar (2002) and Barone-Adesi et al. (2004) to investigate the role of “co-skewness” in asset-pricing models. The quadratic specification of the pricing kernel implies an asset pricing model where the expected excess return on an asset is determined by its covariance with both the market return and the square of the market return (co-skewness). When there is evidence that skewness is not important in an investment decision, \( \gamma \) is equal to zero and equation (2.2) reduces to a linear function of the market return (that is the Capital Asset Pricing Model). In that case, we say that skewness is not priced into the market.

If \( \sigma^2 \left( m_{t+1}^{mv} \right) \) denotes the variance of \( m_{t+1}^{mv} \) and \( \sigma^2 \left( m_{HJ} \right) \) the HJ variance bound, it can be shown that:
\[ \sigma^2 \left( m_{t+1}^{mv} \right) = \sigma^2 \left( m_{HJ} \right) + \gamma^2 \left( \kappa - (\sigma^2 + \mu^2)^2 - s^2 \sigma^2 \right). \]

with \( \sigma^2 \left( m_{HJ} \right) = (\sigma^2)^{-1} \left( p - \mu \sigma \right)^2 \). If skewness is priced, \( \gamma \neq 0 \), the CGR variance bound is more restrictive than the HJ variance bound.

Notice that the CGR variance bound is a function of \( \overline{m}, c \) and the risky asset’s first four moments \((\mu, \sigma^2, s, \kappa)\), whereas the HJ bound depends on \( \overline{m} \) and the first two moments of the risky asset \((\mu, \sigma^2)\). The set of points \((\overline{m}, c, \sigma^2 \left( m_{t+1}^{mv} \right))\) represents the CGR variance bound surface frontier. For each value \( c \), the set of points \((\overline{m}, \sigma^2 \left( m_{t+1}^{mv} \right))\) is a parabola. Note that if \( p = 1 \) and there exists a risk-free asset \( r_f \) such that \( \overline{m} = 1/r_f \) with \( c \neq \overline{c} \), the variance \( \sigma^2 \left( m_{t+1}^{mv} \right) \) is not proportional to the square of the Sharpe ratio on the risky asset:
\[ \sigma^2 \left( m_{t+1}^{mv} \right) = \left( \frac{1}{r_f} \right)^2 \frac{\left( r_f - \mu \right)^2}{\sigma^2} + \left( \frac{\sigma^2 \left( \kappa - (\sigma^2 + \mu^2)^2 - s^2 \sigma^2 \right)}{\sigma^2} \right)^{-1} \frac{1}{(c - \overline{c})^2}. \]

Intuitively, this last equation shows that the cost of the squared portfolio return is relevant for selecting a portfolio [see Chabi-Yo, Garcia and Renault 2004]. Note that the CGR variance bound is unconditional. Gallant, Hansen and Tauchen (1990; hereafter GHT) show how conditioning information can be used to tighten the HJ bound. In the next section, we derive the conditional counterpart to the CGR variance bound.
2.2 Conditional Variance Bound

Now, we assume there is a relevant information set available to economic agents and econometricians and derive the conditional counterpart to the results obtained in section 2.1. Let $I_t$ be the conditioning information set available to economic agents and econometrician at a given point in time. Investors are presumed to use this set to form portfolios of asset payoffs and derivatives in the same asset. The payoffs are approximated by their (conditional) linear regression on the asset payoffs and squared asset payoffs. Then, economic agents have a larger set of assets to form their portfolios than in GHT. GHT presume that economic agents use their information set to form portfolios of only risky assets. If $I_t^{GHT}$ represents the GHT information set, we have $I_t \supset I_t^{GHT}$.

We consider the set of admissible pricing kernels that price conditionally the bond, the asset payoff and derivatives whose payoffs are approximated by their linear regression on asset payoffs and squared asset payoffs.

$$\mathcal{F}(\mathbf{m}_t,c_t) = \{ m_{t+1} \in L^2 : E(m_{t+1}|I_t) = \mathbf{m}_t, E(m_{t+1}r_{t+1}|I_t) = p_t, E(m_{t+1}r_{t+1}^2|I_t) = c_t \}.$$ 

To derive the counterpart to problem (2.1) using conditioning information, we solve

$$\min_{m \in \mathcal{F}(\mathbf{m}_t,c_t)} \sigma^2(m|I_t).$$

and show:

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

$$m_{t+1}^{MSC} = m_{GHT} + \gamma_t \left( r_{t+1}^2 - \left( \mu_t^2 + \sigma_t^2 \right) - \sigma_t^{-2} s_t \left( r_{t+1} - \mu_t \right) \right)$$

with

$$m_{GHT} = \mathbf{m}_t + \beta_t \left( r_{t+1} - \mu_t \right)$$

and

$$\beta_t = \sigma_t^{-2} (p_t - \mathbf{m}_t \mu_t),$$

$$\gamma_t = \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right) - s_t \sigma_t^{-2} \right)^{-1} (c_t - \mathbf{m}_t),$$

$$\mathbf{c}_t = \mathbf{m}_t \left( \mu_t^2 + \sigma_t^2 - s_t \sigma_t^{-2} \mu_t \right) + s_t \sigma_t^{-2} p_t,$$

where $m_{GHT}$ represents the GHT pricing kernel, $\kappa_t = E \left( r_{t+1}^4 | I_t \right)$, $s_t = Cov \left( r_{t+1}, r_{t+1}^2 | I_t \right)$, $\sigma_t^2 = Var \left( r_{t+1} | I_t \right)$ and $\mu_t = E \left( r_{t+1} | I_t \right)$.

Equation (2.5) says that the pricing kernel with minimum variance for its conditional expectation $\mathbf{m}_t$ is the conditional projection of $m_{t+1}$ onto the $\{ z_{1t} r_{t+1}, z_{2t} r_{t+1}^2 ; \forall z_{1t}, z_{2t} \}$ space augmented with a constant payoff. The conditional variance of the pricing kernel displayed in (2.5) is a function of asset payoff conditional mean, conditional variance, conditional skewness and conditional kurtosis, whereas the variance of the GHT pricing kernel depends on the first two conditional moments of asset payoffs. It is possible to define the conditions under which the conditional variance of this pricing kernel reaches the GHT bound.
Proposition 2.2 Given the information set $I_t$, if the price of the residual obtained when regressing the squared asset payoff onto the asset payoff is null, the conditional variance of $m_{t+1}^{\text{MVX}}$ (see equation (2.5)) reaches the GHT bound.

Proof. Given the information set $I_t$, the residual obtained when regressing the squared asset payoff onto the asset payoff is

$$\varepsilon_{t+1} = r_{t+1}^2 - (\mu_t^2 + \sigma_t^2) - \sigma_t^{-2}s_t(r_{t+1} - \mu_t) \quad (2.6)$$

The price of this residual is $E(m_{t+1}^{\text{MVX}} \varepsilon_{t+1} | I_t) = c_t - \overline{m}_t(\mu_t^2 + \sigma_t^2) - \sigma_t^{-2}s_t(p_t - \overline{m}_t \mu_t) = c_t - \overline{c}_t$. Then, $E(m_{t+1}^{\text{MVX}} \varepsilon_{t+1} | I_t) = 0$ is equivalent to $c_t - \overline{c}_t = 0$ which is also equivalent to $\gamma_t = 0$. This ends the proof. □

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.

2.3 Unconditional Variance Bound with Conditioning Information

Our strategy in this section is to replicate the analysis in section 2.2 using $\mathcal{F}(\mu, c_t)$ in place of $\mathcal{F}(\overline{m}, c_t)$ and using an unconditional projection in place of the conditional projection. To proceed with our development, we consider the problem:

$$\min_{m \in \mathcal{F}(\mu, c_t)} \sigma^2(m). \quad (2.7)$$

Similarly to proposition 2.1, we show:

Proposition 2.3 The pricing kernel solution to (2.7) is

$$m_{t+1}^{\text{MVX}} = m_{GHT}^* + \gamma_t \varepsilon_{t+1} \quad (2.8)$$

with

$$m_{GHT}^* = \overline{m} + \beta_t (r_{t+1} - \mu_t)$$

and

$$\beta_t = \sigma_t^{-2} (p_t - \overline{m}_t \mu_t),$$

$$\gamma_t = \left( \nu_t - (\sigma_t^2 + \mu_t^2) - s_t \sigma_t^{-2} \right)^{-1} (c_t - \overline{c}_t),$$

$$\overline{c}_t = \overline{m} \left( \mu_t^2 + \sigma_t^2 - s_t \sigma_t^{-2} \mu_t \right) + s_t \sigma_t^{-2} p_t,$$

where $\varepsilon_{t+1}$ is defined in (2.6). So $\sigma^2(m_{t+1}^{\text{MVX}}) = \text{Var}(m_{t+1}^{\text{MVX}})$ is the CGR variance bound with conditioning information. This bound will be termed the CY variance bound.

When $\gamma_t$ is equal to 0, the CY variance bound reaches the GHT variance bound. For further convenience, we rewrite (2.8) as:

$$m_{t+1}^{\text{MVX}} = m_{GHT}^* + \gamma_t \varepsilon_{t+1}, \quad (2.9)$$

with

$$m_{GHT}^* = (p_t - \omega \mu_t) \left( \sigma_t^2 \right)^{-1} r_{t+1} + \omega,$$
where
\[ \omega = \frac{\bar{y} - b_1}{1 - d_1} \]
\[ b_1 = E \left( \mu_t^2 + \sigma_t^2 \right)^{-1} \mu_t, \quad (2.10) \]
\[ d_1 = E \mu_t \left( \mu_t^2 + \sigma_t^2 \right)^{-1} \mu_t. \quad (2.11) \]

We also denote
\[ a_1 = E \rho_t^2 \left( \mu_t^2 + \sigma_t^2 \right)^{-1}, \quad (2.12) \]
\[ a_2 = E \left( c_t - s_t \sigma_t^{-2} m_t \right)^2 \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^{-2} s_t^2 \right)^{-1}, \quad (2.13) \]
\[ b_2 = E \left( c_t - s_t \sigma_t^{-2} m_t \right) \left( \mu_t^2 + \sigma_t^2 - s_t \sigma_t^{-2} \mu_t \right) \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^{-2} s_t^2 \right)^{-1}, \quad (2.14) \]
\[ d_2 = E \left( \mu_t^2 + \sigma_t^2 - s_t \sigma_t^{-2} \mu_t \right)^2 \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^{-2} s_t^2 \right)^{-1}. \quad (2.15) \]

In section 2.4, we introduce conditioning information using scaled returns (see Cochrane 1996). The conditional moments are not easy to derive, so to incorporate conditioning information into the variance bound on pricing kernels, most studies use scaled returns. Specifically, in the section below, we resort to a simple approach that incorporates conditioning information into the CGR bound.

### 2.4 Optimally Scaled Variance Bound Under Higher Moments

#### 2.4.1 Scaled Variance Bounds

The GHT variance bound is difficult to implement because of the set \( I_t \). To implement the HJ bound, most studies use scaled returns (see Cochrane 1996). An effective way to implement the HJ bound is to exploit the information in a set of lagged instruments \( y_t \in I_t \), while using unconditional moments to describe this bound. Let \( p_t = 1 \) and \( r_{t+1} = r_{t+1} - r_t \) be the excess return. The HJ bound with conditioning information uses the information
\[ E \left[ m_{t+1} r_{t+1} I_t \right] = 0, \quad (2.16) \]
which implies that for any instrument, \( E \left[ m_{t+1} r_{t+1} y_{1t} | y_{1t} \right] = 0 \) and therefore \( E \left[ m_{t+1} r_{t+1} y_{1t} \right] = 0 \). The excess return \( r_{t+1} y_{1t} \) is interpreted as a dynamic trading strategy (see Cochrane 1996). To see how alternative approaches to conditioning information can refine the HJ bound, equation (2.16) implies
\[ E \left[ m_{t+1} r_{t+1} h (y_{1t}) \right] = 0 \]
for all functions \( h \).

In that case, if \( r_{t+1} h (y_{1t}) \) represents a dynamic trading strategy, then (2.17) says that \( m_{t+1} \) should price correctly all the dynamic trading strategies. Most studies, instead of using only scaled returns, stack actual returns with scaled returns (see Bekaert and Hodrick 1992), considering the system
\[
\begin{bmatrix}
  r_{t+1} \\
  r_{t+1} y_{1t}
\end{bmatrix}
\]
. Recently, BL provide a bridge between this scaling approach and the GHT bound. BL find the scaling factor \( z_{1t} = h (y_{1t}) \) which yields the best largest HJ bound. Scaling only improves the HJ bound if the scaling factor \( z_{1t} \) has information on future returns. In the literature, this factor is believed to predict future returns or to capture time variation in the expected return.
Since the CY variance bound depends on the first four conditional moments of asset returns, we expect that time variation in these moments will improve the CY variance bound. To investigate this issue more closely, we consider the conditional price of the derivative with payoff \( g(r_{t+1}) \):

\[
E[m_{t+1}g(r_{t+1})|I_t] = \text{price}(g(r_{t+1})).
\]

If \( g(r_{t+1}) \) is approximated by its (conditional) linear regression on \( r_{t+1} \) and \( r_{t+1}^2 \), the conditional price of \( g \) can be computed if the conditional price of \( r_{t+1}^2 \),

\[
E[m_{t+1}r_{t+1}^2|I_t] = \text{known},
\]

is known. Therefore, this last equality implies that for any instrument,

\[
E[m_{t+1}r_{t+1}^2z_{2t}] = E[z_{2t}c_t] \text{ with } z_{2t} = h(y_{2t}) \text{ for all function } h.
\] (2.18)

Equations (2.17) and (2.18) can be combined together to determine the CGR variance bound that incorporates conditioning information. This bound depends on \((z_{1t}, z_{2t})\). In the next section, we find the scaling factor \((z_{1t}, z_{2t})\) which yields the largest CGR variance bound.

### 2.4.2 Optimally Scaled Variance Bound Under Higher Moments

The approach pursued in section 2.3 uses \( E(m_{t+1}r_{t+1}|I_t) = p_t \) and \( E(m_{t+1}r_{t+1}^2|I_t) = c_t \) to determine the variance bound on pricing kernels (see proposition 2.1 and proposition 2.3). For the sake of simplicity, we rewrite these two equations as:

\[
E(m_{t+1}r_{t+1}|I_t) = p_t \text{ and } E(m_{t+1}\varepsilon_{t+1}|I_t) = (c_t - \bar{c}_t)
\]

where \( \varepsilon_{t+1} \) is defined in (2.6). The CY bound is difficult to implement because of the set \( I_t \). To implement this bound, we scale the risky asset return by the conditioning random variable \( z_{1t} \in I_t \):

\[
E(m_{t+1}r_{t+1}z_{1t}|z_{1t}) = p_t z_{1t},
\]

and the residual \( \varepsilon_{t+1} \) by \( z_{2t} \in I_t \),

\[
E(m_{t+1}\varepsilon_{t+1}z_{2t}|z_{2t}) = (c_t - \bar{c}_t) z_{2t}.
\]

The conditioning variable \( z_{1t} \) is believed to capture time variation in expected returns, whereas the variable \( z_{2t} \) captures, at least, time variation in the expected return, variance and skewness. To derive an unconditional variance bound on pricing kernels, we consider the family of infinitely many one-dimensional scaled payoff space:

\[
P_z = \left\{ az'_t g_{t+1} : \ g_{t+1} = (r_{t+1}, \varepsilon_{t+1}), \ a \in \mathbb{R} \right\}
\]

indexed by \( z_t = h(y_t) \) where \( h \) is a measurable function. Intuitively, this payoff space contains not only asset payoff but derivatives. To see this, assume an investor who invests not only to stock markets but also into derivatives markets. If \( r_p \) represents his portfolio payoff,

\[
r_p = \omega_0 r_{t+1} + \sum_{i=1}^{i=K} \omega_i f_i (r_{t+1}),
\]
where \( f_t (r_{t+1}) \) represents the payoff for derivative \( i \) and \( K \) is the number of derivatives. We assume that the payoffs of derivatives can be approximated by their linear regression onto \( r_{t+1} \) and \( \varepsilon_{t+1} \). In that case, \( f_t (r_{t+1}) = \psi_{i1} r_{t+1} + \psi_{i2} \varepsilon_{t+1} \) and the portfolio payoff is \( r_p = \phi' g_{t+1} \), where \( \phi \) is a vector whose components are:

\[
\phi_1 = \omega_0 + \sum_{i=1}^{i=K} \omega_i \psi_{i1} \quad \text{and} \quad \phi_2 = \sum_{i=1}^{i=K} \omega_i \psi_{i2}.
\]

To identify time-varying portfolio weights we scale the return \( r_p \) by the conditioning random variable \( z_t \). This scaled portfolio payoff, \( z_t r_p \), belongs to \( P_z \).

We now define the return \( r^*_t g_{t+1} = r_{t+1}/price \) and the hypothetical return \( \varepsilon^*_t g_{t+1} = \varepsilon_{t+1}/price \) and consider the vector of returns \( g^t_{t+1} = (r^*_t g_{t+1}, \varepsilon^*_t g_{t+1}) \). The scaled return \( z_t^* g_{t+1}^t \) has an intuitive interpretation (considering the excess return \( g^t_{t+1} = g_{t+1} - r^f J \) where \( J \) is an \( 2 \times 1 \) vector of ones). The scaled return

\[
g_{t+1} = z_t^* g_{t+1} + r^f = z_t^* g_{t+1} + (1 - z_t^* J) r^f
\]

can be interpreted as a managed portfolio, where \( z_t \) is the (time-varying) proportion of the investment allocated to a portfolio of risky asset \( r_{t+1} \) and derivatives in \( r_{t+1} \).

If one considers the payoff \( z_t^* g_{t+1} \) in \( P_z \) with \( \varepsilon^*_t g_{t+1} \) defined in (2.6), there exists a CGR variance bound associated with each scaling factor \( z_t = (z_{t1}, z_{t2}) \):

\[
\sigma^2 (\overline{m}, z^*_t g_{t+1}) = \frac{\left( E (z_t^i \pi_t) - \overline{m} E (z_t^i g_{t+1}) \right)^2}{\text{Var} (z_t^i g_{t+1})}
\]

where \( \pi_t = (p_t, c_t - \overline{r}) \). The relevant question is: what \( z_t = f (y_t) \) yields the best (largest) CGR variance bound? This is a problem of variational calculus. This bound is called the “optimally scaled bound under higher moments” (the OHM bound):

\[
\sigma^2_{OHM} (\overline{m}, z^*_t g_{t+1}) = \sup_{z_t \in I_t} \sigma^2 (\overline{m}, z^*_t g_{t+1}).
\]

This bound is the highest CGR variance bound when conditioning information is used. Proposition 2.4 gives the solution to (2.19).

**Proposition 2.4** The solution \( z^*_t \) to the maximization problem (2.19) is given by \( z^*_t = (z_{t1}^*, z_{t2}^*) \) with

\[
z_{t1}^* = (\sigma_t^2 + \mu_t^2)^{-1} (p_t - \omega \mu_t),
\]

\[
z_{t2}^* = (\kappa_t - (\sigma_t^2 + \mu_t^2)^2 - \sigma_t^{-2} \sigma_t^2)^{-1} (c_t - \overline{r}),
\]

where \( \omega = \overline{m} \sqrt{d_1} \). Thus, the optimally scaled payoff is \( z^*_t g_{t+1} = z_{t1}^* r_{t+1} + z_{t2}^* \varepsilon_{t+1} \) and the optimally scaled bound under higher moments is

\[
\sigma^2_{OHM} (\overline{m}, z^*_t g_{t+1}) = \frac{a_1 (1 - d_1) + \overline{m} d_1 - 2 \overline{m} b_1 + b_1^2}{1 - d_1} + \frac{a_2}{(\overline{m}^2 d_2 - 2 \overline{m} b_2 + a_2)}
\]

where \( a_1, b_1, \) and \( d_1 \) are defined in (2.12), (2.10) and (2.11), and \( a_2, b_2, d_2 \) are defined in (2.13), (2.14) and (2.15).

\[1\] The return \( \varepsilon^*_t g_{t+1} \) is not observed in the market, but could be inferred from derivatives available in derivatives markets.
The optimal scaling factor $z_t$ depends on the conditional distribution function through the first four conditional moments $(\mu_t, \sigma_t^2, s_t, \kappa_t)$. Its first component $z_t^1$ depends only on the conditional mean and conditional variance, and is decreasing in the conditional mean but not monotonic in the conditional variance. However, the second component $z_t^2$ is increasing in the conditional mean, not monotonic in the conditional variance, increasing in the conditional skewness and decreasing in the conditional kurtosis. When these moments are known to econometricians or researchers and if $c_t$ and $p_t$ are correctly specified, we show the relation between the OHM bound and the CY bound.

**Proposition 2.5** Consider the payoffs $r_{t+1}$ and $r_{t+1}^2$ with conditional prices $p_t$ and $c_t$ and assume the first four conditional moments of asset payoffs are $(\mu_t, \sigma_t^2, s_t, \kappa_t)$, then the OHM bound is:

$$\sigma_{OHM}^2(\mu, z_t^t g_{t+1}) = \sigma^2(\mu_{t+1}^{mu})$$

where $g_{t+1} = (r_{t+1}, \varepsilon_t)$ and $\varepsilon_t$ is defined in (2.6).

The OHM bound is the CGR bound for a scaled payoff. If the first four conditional moments of $r_{t+1}$ are known and if $c_t$ and $p_t$ are correctly specified, the OHM bound reaches the CY variance bound. We prove this in the next proposition.

**Proposition 2.6** The OHM bound can be rewritten as

$$\sigma_{OHM}^2(\mu, z_t^t g_{t+1}) = \frac{A^2}{B}$$

with $A = E z_t^t \pi_t - \mu E z_t^t g_{t+1}$, $B = Var(z_t^t g_{t+1})$, where $g_{t+1} = (r_{t+1}, \varepsilon_t)$, with the random variable $\varepsilon_t$ defined in (2.6). If the first four conditional moments of $r_{t+1}$ are known and if $c_t$ and $p_t$ are correctly specified, $A = B$.

### 2.4.3 Relation to Bekaert and Liu (2004)

The Bekaert and Liu (2004; hereafter BL) article is related to the present article. BL find the scaling factor that yields the largest HJ bound. Their variance bound depends on the first two moments of asset returns. Their bound uses only asset payoffs whereas, in the present article we use asset payoffs and derivatives whose payoffs are non-linear functions of the same asset payoffs. Recall that the BL optimally scaled bound is:

$$\sigma_{OSB}^2(\mu, z_{t+1}^t r_{t+1}) = \frac{a_1(1 - d_1) + \mu^2 d_1 - 2 \mu b_1 + b_1^2}{1 - d_1}.$$

where $a_1$, $b_1$, and $d_1$ are defined in (2.12), (2.10) and (2.11). To highlight the relationship between the OHM bound and the BL bound, we show:

**Proposition 2.7** The optimally scaled bound under higher moments can be rewritten as

$$\sigma_{OHM}^2(\mu, z_{t+1}^t g_{t+1}) = \frac{(A_1 + A_2)^2}{B_1 + B_2}.$$

with $A_1 = E z_{t+1}^t p_t - \mu E z_{t+1}^t r_{t+1}$, $A_2 = E z_{t+1}^t(c_t - \pi_t)$, $B_1 = Var(z_{t+1}^t r_{t+1})$, $B_2 = Var(z_{t+1}^t \varepsilon_{t+1})$, where $\varepsilon_{t+1}$ is defined in (2.6). If the conditional skewness is not priced, $c_t = \pi_t$, then

$$\sigma_{OHM}^2(\mu, z_{t+1}^t g_{t+1}) = \frac{A_1^2}{B_1}.$$
so the scaled bound under higher moments coincides with the Bekaert and Liu (2004) scaled bound.

3. Testing whether asset pricing models price skewness

As mentioned above, skewness is not priced if and only if \( \gamma = 0 \). This is equivalent to \( c = \overline{c} \). Since \( c = E m_{t+1} r^2_{t+1} \), we derive:

\[
c - \overline{c} = E m_{t+1} \left( \frac{r^2_{t+1}}{m} - \frac{1}{m} c \right) = 0.
\]

However, it is known that \( E m_{t+1} = \overline{m} \). This suggests the following test:

\[
H_0 : E f_{t+1}(\overline{m}, c) = 0
\]

where

\[
f_{t+1}(\overline{m}, c) = \begin{bmatrix}
    m_{t+1} \left( \frac{r^2_{t+1}}{m} - \frac{1}{m} c \right) \\
    m_{t+1} - \overline{m}
\end{bmatrix}.
\]

Given the functional form for \( m_{t+1} \), the GMM approach developed by Hansen (1982) can be used to test the null hypothesis \( H_0 \). To test whether conditional skewness is priced, we consider the equality \( \gamma_{m} = 0 \):

\[
c_t - \overline{c}_t = E \left[ m_{t+1} \left( r^2_{t+1} - \frac{1}{m} c_t \right) \bigg| I_t \right] = 0.
\]

We then scale this last expression with the conditioning information \( z_{2t} \) and write the null hypothesis \( H_0 \) with:

\[
f_{t+1}(\overline{m}, c_t) = \begin{bmatrix}
    m_{t+1} \left( \frac{r^2_{t+1}}{m} - \frac{1}{m} c_t \right) z_{2t} \\
    m_{t+1} - \overline{m}
\end{bmatrix}.
\]

This test is more general than the proposed test of Barone-Adesi et al. (2004). Their test allows the researcher to know whether a quadratic pricing kernel in the market return prices skewness. But the approach proposed in this section could be used to test whether any pricing kernel prices skewness.

4. Optimally scaled bound under higher moments: Discussion

In section 4.1, we investigate the relationship between predictability and the OHM bound. Section 4.2 discusses how the OHM bound represents a valid lower bound to the pricing kernels. In section 4.3 we discuss how the OHM bound can be used as the basis of a diagnostic test for the correct specification of the first, second, third and fourth conditional moments.

4.1 Efficiency and predictability with higher moments

As shown in the previous section, because the OHM bound uses conditioning information efficiently, we should expect that predictable variation in expected return, variance, skewness and kurtosis leads to a sharper CGR bound. The correlation between the conditioning variable \( z_{1t} \) and the future return \( r_{t+1} \) captures time variation in expected returns. This conditioning variable is believed to predict the future return. Looking at equation (2.6), it is clear that the correlation between the conditioning variable, \( z_{2t} \), and
the residual obtained when regressing the squared future return onto the future return, $\varepsilon_{t+1}$, captures time variation in expected return, variance, skewness and kurtosis. To derive the conditions under which scaling cannot improve the CGR variance bound, we make two assumptions:

- **Assumption 1** (no prediction through $z_{1t}$) the scaling factor $z_{1t}$ is uncorrelated with $r_{t+1}$ and $p_t$ and the squared scaling factor $z_{1t}^2$ is uncorrelated with $r_{t+1}^2$.

- **Assumption 2** (no prediction through $z_{2t}$) the scaling factor $z_{2t}$ is uncorrelated with the conditional price of $\varepsilon_{t+1}$ and the squared scaling factor $z_{2t}^2$ is uncorrelated with $\varepsilon_{t+1}^2$.

**Proposition 4.1** Under Assumptions 1 and 2, if the premium $E(z_{1t}r_{t+1}) - r_f$ is positive,

$$\sigma^2 (\overline{m}, z_{1t}r_{t+1}) \leq \sigma^2 (\overline{m}, z_{1t}^*g_{t+1}) \leq \sigma^2 (m_{t+1}^{ms}) , \quad (4.1)$$

where $\sigma^2 (\overline{m}, z_{1t}r_{t+1})$ represents the BL optimally scaled bound, $g_{t+1} = (r_{t+1}, \varepsilon_{t+1})$ with $\varepsilon_{t+1}$ defined in (2.6).

When the payoff $g_{t+1}$ is stacked with the scaled payoff $z_{1t}^*g_{t+1}$, the optimal scaling factor $z_{1t}^*$ remains the same for this stacked payoff. We show this explicitly in the next proposition.

**Proposition 4.2** Assume there is an asset with payoff $r_{t+1}$ and price $p_t$, and that there are also derivatives whose payoffs are non-linear functions of the same asset payoffs. Assume the payoff of these derivatives could be approximated by their linear regression on $r_{t+1}$ and $r_{t+1}^2$ and denote $c_t$ the conditional price of $r_{t+1}^2$.

Let $I_t$ denote the sigma algebra of the measurable functions of the conditioning variable $z_t$. Then the solution $z_{*t}$ to the maximization problem

$$\sup_{z_t \in I_t} \sigma^2 (\overline{m}, (g_{t+1}, z_{*t}^*g_{t+1}))$$

is given by $z_{*t} = (z_{1t}^*, z_{2t}^*)$ with $z_{1t}^*$ and $z_{2t}^*$ defined in (2.20) and (2.21), where $g_{t+1} = (r_{t+1}, \varepsilon_{t+1})$, the random variable $\varepsilon_{t+1}$ is defined in (2.6). We term $\sigma^2 (\overline{m}, (g_{t+1}, z_{*t}^*g_{t+1}))$ the stacked optimally scaled bound under higher moments.

### 4.2 Robustness

In practice, it is difficult to estimate the GHT variance bound $\sigma^2 (m_{GHT}^*)$. This variance depends on the conditional mean $\mu_t$ and the conditional variance $\sigma_t^2$, of the return and these two moments are not known. GHT use a proxy for these moments to approximate $m_{GHT}^*$. They propose the use of the seminonparametric method to estimate conditional moments. As highlighted in Bekaert and Liu (2004), if $\hat{m}_{GHT}^*$ is the proxy for $m_{GHT}^*$, the variance $\sigma^2 (\hat{m}_{GHT}^*)$ may overestimate the true variance $\sigma^2 (m_{GHT}^*)$. In that case, the GHT fails to be a lower bound for the variance of the pricing kernels.

Recall that the OHM bound is $\sigma^2 (\overline{m}, z_{*t}^*g_{t+1})$ where $z_{*t}^*$ depends on the first four conditional moments of asset returns, $\mu_t, \sigma_t^2$, $s_t$ and $\kappa_t$. When these moments are unknown, the optimal scaling factor $z_{*t}^*$ is unknown, and the scaled payoff $z_{*t}^*g_{t+1}$ is unknown. However, for any $z_t$, $\sigma^2 (\overline{m}, z_{*t}^*g_{t+1})$ remains a lower bound to the variance of all pricing kernels that price correctly asset payoffs and derivatives. In that case, even if a

---

2 Bekaert and Liu (2004) use this assumption to investigate whether the scaled HJ bound improves on the standard HJ bound. Specifically under assumption 1, the scaled HJ bound is below the standard HJ bound.
proxy \( \hat{z}_t^* \) for \( z_t^* \) is used, the variance \( \sigma^2 \left( \hat{m}_t, \hat{z}_t^*, g_{t+1} \right) \) there remains a valid lower bound to the variance of all pricing kernels that price correctly asset payoff \( r_{t+1} \) and derivatives of \( r_{t+1} \). When the empirical specification of the first four conditional moments of asset returns are known, the OHM bound is easy to implement. If there is suspicion that time variation in the asset conditional mean and variance is more important than time variation in the conditional skewness and kurtosis, we obtain a valid OHM bound by replacing the conditional skewness and kurtosis by the unconditional skewness and kurtosis. But this bound will not be optimal if there is evidence of time variation in the conditional skewness and kurtosis.

### 4.3 Diagnostics

In this section, we develop a general diagnostic test for the first, second, third and fourth conditional moments of asset returns. From proposition 2.6, the optimally scaled bound under higher moments can be written as \( A_f^2 / \bar{f} \). If the first four conditional moments are correctly specified and if \( c_t \) and \( p_t \) are correctly specified, \( A = B \). This suggests a diagnostic test for the first, second, third and fourth conditional moments of asset return when \( c_t \) and \( p_t \) are correctly specified. To test the equality of \( A \) and \( B \), the following orthogonality conditions can be used:

\[
E f_t = 0 \quad (4.2)
\]

where,

\[
f_t = \begin{bmatrix}
\mu_t^2 \left( \sigma_t^2 + \mu_t^2 \right)^{-1} - a_1 \\
\mu_t \left( \sigma_t^2 + \mu_t^2 \right)^{-1} p_t - b_1 \\
\mu_t \left( \sigma_t^2 + \mu_t^2 \right)^{-1} - \mu_t - d_1 \\
\left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^2 s_t^2 \right)^{-1} \left( c_t - s_t \sigma_t^{-2} p_t \right)^2 - a_2 \\
\left( c_t - s_t \sigma_t^{-2} p_t \right) \left( \mu_t^2 + \sigma_t^2 - s_t \sigma_t^{-2} \mu_t \right) \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^2 s_t^2 \right)^{-1} - b_2 \\
\left( \mu_t^2 + \sigma_t^2 - s_t \sigma_t^{-2} \mu_t \right) \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^2 s_t^2 \right)^{-1} - d_2 \\
p_t \left( 1 - d_1 \right) \left( \sigma_t^2 + \mu_t^2 \right)^{-1} r_{t+1} + b_1 \mu_t \left( \sigma_t^2 + \mu_t^2 \right)^{-1} r_{t+1} - b_1 + \nu_{1t} \\
\mu_t \left( \sigma_t^2 + \mu_t^2 \right)^{-1} - r_{t+1} - d_1 + \nu_{2t}
\end{bmatrix}
\]

and

\[
\nu_{1t} = 2 \left( 1 - d_1 \right) \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^{-2} s_t^2 \right)^{-1} \left( c_t - s_t \sigma_t^{-2} p_t \right) \left( \left( \mu_t^2 + \sigma_t^2 \right) - \mu_t s_t \sigma_t^{-2} \right) - 2 b_2 \left( 1 - d_1 \right)
\]

\[
\nu_{2t} = \left( 1 - d_1 \right) \left( \kappa_t - \left( \sigma_t^2 + \mu_t^2 \right)^2 - \sigma_t^{-2} s_t^2 \right)^{-1} \left( \left( \mu_t^2 + \sigma_t^2 \right) - \mu_t \sigma_t^{-2} \right)^2 - d_2 \left( 1 - d_1 \right)
\]

The first six conditions estimate and define the constants (2.9), (2.10), (2.11), (2.12), (2.13) and (2.14). The last two conditions result from equating \( A \) and \( B \). There are six parameters to be estimated, then there are two overidentifying restrictions. This can be tested using the statistics \( T f_T W f_T \) where \( f_T \) is the mean of \( f_t \), \( T \) is the number of observations and \( W \) is the weighting matrix (see Newey and West 1987). This statistic follows a \( \chi^2 \) (2) and can be used to compare the performance of nonnested models for the first, second, third and fourth conditional moments when \( p_t \) and \( c_t \) are correctly specified. The sampling error in the parameters \( \mu_t, \sigma_t^2, s_t \) and \( \kappa_t \) should be taken into account in the estimation procedure, employing a sequential GMM procedure (see Ogaki 1993).

---

3 This orthogonality condition is easy to prove by equating \( A \) and \( B \). The prove is available from the author upon request.
5. The optimally scaled bound in the multiple-assets case

Let us now consider a multiple assets case in which we replicate some results of the previous section. Let \( r_{t+1} \) denote the set of assets payoff. In that case, \( p \) is a set of prices. Since \( r_{t+1} \) is not a scalar, we need to redefine the squared asset payoff. Instead of using \( r_{t+1}^2 \), we denote by \( r_{t+1}^{(2)} \) the squared asset payoff, that is a vector column whose components are of the form \( r_{i+1} r_{j+1} \) with \( i \leq j \). We also denote:

\[
\begin{align*}
\sigma_t^2 &= E_t (r_{t+1} - E_t r_{t+1})^2, \\
\gamma_t &= E_t (r_{t+1}^{(2)} - E_t r_{t+1}^{(2)}) r_{t+1}^t, \\
\kappa_t &= E_t r_{t+1}^{(2)} (r_{t+1}^{(2)})', \\
\nu_t^2 &= \kappa_t - \left( E_t r_{t+1}^{(2)} \right) \left( E_t r_{t+1}^{(2)} \right)', \\
\varepsilon_{t+1} &= r_{t+1}^{(2)} - E_t r_{t+1}^{(2)} - s_t \left( \sigma_t^2 \right)^{-1} (r_{t+1} - \mu_t).
\end{align*}
\]

Solving problem (2.4), we show:

Proposition 5.1 The pricing kernel solution to (2.4) is:

\[
m_{t+1}^{mv \ast} = m_{GHT} + \gamma_t \varepsilon_{t+1},
\]

with

\[
m_{GHT} = (p_t - \overline{m}_t \mu_t) \left( \sigma_t^2 \right)^{-1} (r_{t+1} - \mu_t) + \overline{m}_t
\]

and

\[
\gamma_t = \left( \nu_t^2 - s_t \left( \sigma_t^2 \right)^{-1} s_t \right)^{-1} (c_t - \overline{m}),
\]

(5.3)

\[
r_t = \overline{m}_t E_t r_{t+1}^{(2)} + s_t \left( \sigma_t^2 \right)^{-1} (p_t - \overline{m}_t E_t r_{t+1}^{(2)}),
\]

(5.4)

where \( c_t \) is the conditional price of the squared payoff \( r_{t+1}^{(2)} \).

This proposition shows that the pricing kernel \( m_{t+1}^{mv \ast} \) depends on the asset returns conditional covariance matrix, \( \sigma_t^2 \), conditional co-skewness matrix, \( s_t \) and conditional co-kurtosis matrix \( \kappa_t \). The conditional co-skewness matrix elements are of the form \( E [r_{i+1} r_{j+1} r_{k+1} | I_t] \) and the conditional co-kurtosis matrix elements are of the form \( E [r_{i+1} r_{j+1} r_{k+1} r_{l+1} | I_t] \). The next proposition provides the solution to problem (2.7).

Proposition 5.2 The pricing kernel solution to (2.7) is:

\[
m_{t+1}^{mv \ast \ast} = m_{GHT}^\ast + \gamma_t \varepsilon_{t+1}
\]

(5.5)

with

\[
m_{GHT}^\ast = (p_t - \omega \mu_t) \left( \mu_t^2 + \sigma_t^2 \right)^{-1} r_{t+1}^t + \omega,
\]

with

\[
\omega = \frac{\overline{m} - b_1}{1 - d_1}
\]

(5.6)
and
\[ b_1 = E \left( \mu_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} p_t \right), \quad (5.6) \]
\[ d_1 = E \left( \mu_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} \mu_t \right), \quad (5.7) \]
where \( \gamma_t = \left( v_t^2 - s_t' (\sigma_t^2)^{-1} s_t \right)^{-1} (c_t - \bar{c}_t) \) and \( \bar{c}_t = \bar{m} E_t r_{t+1}^{(2)} + s_t' (\sigma_t^2)^{-1} (p_t - \bar{m} E_t r_{t+1}). \)

To derive the OHM bound in the multiple-assets case, we consider the following notation:
\[ a_1 = E \left( p_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} p_t \right), \quad (5.8) \]
\[ a_2 = E \left( c_t - s_t' (\sigma_t^2)^{-1} p_t \right)' \left( v_t^2 - s_t' (\sigma_t^2)^{-1} s_t \right)^{-1} \left( c_t - s_t' (\sigma_t^2)^{-1} p_t \right), \quad (5.9) \]
\[ b_2 = E \left( E_t r_{t+1}^{(2)} - s_t' (\sigma_t^2)^{-1} r_{t+1} \right)' \left( v_t^2 - s_t' (\sigma_t^2)^{-1} s_t \right)^{-1} \left( E_t r_{t+1}^{(2)} - s_t' (\sigma_t^2)^{-1} E_t r_{t+1} \right), \quad (5.10) \]
\[ d_2 = E \left( E_t r_{t+1}^{(2)} - s_t' (\sigma_t^2)^{-1} E_t r_{t+1} \right)' \left( v_t^2 - s_t' (\sigma_t^2)^{-1} s_t \right)^{-1} \left( E_t r_{t+1}^{(2)} - s_t' (\sigma_t^2)^{-1} E_t r_{t+1} \right), \quad (5.11) \]
and show:

**Proposition 5.3** The solution \( z_t' \) to the maximization problem (2.19) is given by
\[ z_t' = (z_{1t}', z_{2t}'), \]
with
\[ z_{1t}' = \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} (p_t - \omega \mu_t), \]
\[ z_{2t}' = \left( v_t^2 - s_t' (\sigma_t^2)^{-1} s_t \right)^{-1} (c_t - \bar{c}_t). \]

So the optimally scaled payoff is \( z_t' g_{t+1} = z_{1t}' r_{t+1} + z_{2t}' \varepsilon_{t+1} \) and the optimally scaled bound under higher moments is
\[ \sigma_{OHM}^2 (\bar{m}, z_t' g_{t+1}) = \frac{a_1 (1 - d_1) + \bar{m} d_1 - 2 \bar{m} b_1 + b_2^2}{1 - d_1} + \left( \bar{m}^2 d_2 - 2 \bar{m} b_2 + a_2 \right) \]
where \( a_1, b_1, \) and \( d_1 \) are defined in (5.8), (5.6) and (5.7) and \( a_2, b_2, d_2 \) are defined in (5.9), (5.10) and (5.11).
Similarly to the previous section, this statistic follows a where the pricing kernel is obtained by setting

\[ p_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} p_t - a \]

\[ \mu_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} p_t - b \]

\[ \mu_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} \mu_t - d \]

\[ \left( c_t - s_t' \left( \sigma_t^2 \right)^{-1} p_t \right)' \left( v_t^2 - s_t' \left( \sigma_t^2 \right)^{-1} s_t \right)^{-1} \left( c_t - s_t' \left( \sigma_t^2 \right)^{-1} p_t \right) - a_2 \]

\[ \left( E_t r_{t+1}^{(2)} - s_t' \left( \sigma_t^2 \right)^{-1} E_t r_{t+1} \right)' \left( v_t^2 - s_t' \left( \sigma_t^2 \right)^{-1} s_t \right)^{-1} \left( E_t r_{t+1}^{(2)} - s_t' \left( \sigma_t^2 \right)^{-1} E_t r_{t+1} \right) - a_2 \]

\[ (1 - d_1) p_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} r_{t+1} + b_1 \mu_t' \left( \mu_t \mu_t' + \sigma_t^2 \right)^{-1} r_{t+1} - b_1 + \nu_{1t} \]

\[ \nu_{1t} = 2 (1 - d_1) \left( c_t - s_t' \left( \sigma_t^2 \right)^{-1} p_t \right)' \left( v_t^2 - s_t' \left( \sigma_t^2 \right)^{-1} s_t \right)^{-1} \left( E_t r_{t+1}^{(2)} - s_t' \left( \sigma_t^2 \right)^{-1} E_t r_{t+1} \right) - 2 b_2 (1 - d_1), \]

\[ \nu_{2t} = (1 - d_1) \left( E_t r_{t+1}^{(2)} - s_t' \left( \sigma_t^2 \right)^{-1} E_t r_{t+1} \right)' \left( v_t^2 - s_t' \left( \sigma_t^2 \right)^{-1} s_t \right)^{-1} \left( E_t r_{t+1}^{(2)} - s_t' \left( \sigma_t^2 \right)^{-1} E_t r_{t+1} \right) - d_2 (1 - d_1). \]

Similarly to the previous section, this statistic follows a \( \chi^2 \) (2).

6. Empirical section

6.1 The econometric model

Implementation of the variance bound on pricing kernels that incorporates conditioning information and higher moments requires knowledge of the conditional price of squared asset returns. To compute this conditional price, we assume that we live in a world where the pricing kernels are of the form:

\[ m_{t+1} = \phi_t \left( X_{t+1} \right)^{\theta_1} \left( R_{mt+1} \right)^{\theta_2}, \tag{6.1} \]

where \( X_{t+1} \) is the gross consumption growth, \( R_{mt+1} \) is the return on the market portfolio, \( \theta_1, \theta_2 \) are constant parameters, and \( \phi_t \) may be constant or a time \( t \) parameter. Most consumption-based models produce a pricing kernel of the form (6.1), for example the CRRA consumption-based model, the Epstein and Zin (1989) recursive preference model, the Melino and Yang (2003) state-dependent recursive preference model and the Gordon and Saint-Amour (2000) state-dependent preference model. For example, the CRRA pricing kernel is obtained by setting \( (\theta_1, \theta_2, \phi_t) = (-\alpha, 0, \beta) \) where \( \alpha \) is the constant relative risk aversion parameter and \( \beta \) is the subjective discount parameter. To compute conditional moments, we simulate asset returns based on several interesting econometric models estimated by Bekaert and Liu (2004).
6.1.1 The Bekaert and Liu (2004) VAR model with heteroskedasticity

Let \( r^b_t \) denote the logarithm of the bond return, \( r^s_t \) the logarithm of the stock return, \( x_t \) the logarithm of gross consumption growth. Define

\[
y_t = [x_t, r^b_t, r^s_t]'.
\]

Bekaert and Liu (2004) assume that this variable follows a VAR process with normal disturbance where the conditional covariance matrix of \( y_t \) is time varying. It is known that time variation in the expected return is driven by time variation in this covariance matrix. To accommodate predictability in excess returns, Bekaert and Liu (2004) allow for heteroskedasticity using the GARCH-in-mean framework of Engle, Lilien and Robins (1987), parameterizing an unconstrained model as

\[
y_t = \chi_{t-1} + ay_{t-1} + \Omega_{t-1}e_t
\]

where \( e_t | I_{t-1} \sim N(0, H_t) \) with \( H_t \) a diagonal matrix where the diagonal elements, \( h_{iit} \), follow

\[
h_{iit} = \delta_i + \alpha_i h_{iit-1} + \tau_i e^2_{iit-1} + \eta_i (\max (0, -e_{iit-1}))^2, \quad i \in \{b, s\}.
\]

(6.2)

The vector \( e_t \) contains the fundamental shocks to the system, and the matrix \( \Omega_t \) is assumed to be time invariant and upper triangular. To limit the number of parameters, BL assume that the consumption shock is the only factor in this model because in a standard consumption-based model, the consumption growth is the only state variable. BL also assume that time-variation in the volatility of \( y_t \) is driven by time varying uncertainty in consumption growth. Since it is well-known that consumption-based pricing models introduce elements of the conditional covariance matrix in the conditional mean, BL allow the conditional mean to depend on the elements of the conditional covariance matrix. They estimate this model in its unconstrained and constrained form. The constrained model is the VAR with a restriction of CRRA preference. The unconstrained VAR is termed “UCO VAR” while the constrained VAR is termed “CO VAR”.

Implementation of our variance bound requires knowledge of the conditional price, \( c_t \), of the squared return. Under the above mentioned assumption that \( y_t \) is conditionally normally distributed, the joint process pricing kernel-asset return is also conditionally lognormally distributed. In that case, it is straightforward to show that the components of \( c_t \) are

\[
c_{ijt} = \frac{\exp (r^b_t) E_t \left( \exp \left( r^b_{i+1} + r^s_{i+1} \right) \right)}{E_t \left( \exp r^b_{i+1} \right) E_t \left( \exp r^s_{i+1} \right)} \quad \text{for} \quad i, j \in \{b, s\}.
\]

(6.3)

To illustrate the CGR bound, we need to compute the unconditional price of the squared return \( c \). Its components are \( c_{ij} = Ec_{ijt} \).

6.1.2 The Bekaert and Liu (2004) regime switching model

As mentioned in Bekaert and Liu (2004), there are many reasons to explore non-linear models, e.g., there is empirical evidence of regime-switching behaviour in both the consumption growth and equity return data [see Ang and Bekaert 2002]. Another reason is that the regime-switching behaviour of these data can

---

4 See the appendix for the proof of (6.3).
be accommodated in the optimally scaled bound under higher moments. Bekaert and Liu (2004) formulate a regime-switching version of the unconstrained VAR model of the previous section. Using a discrete regime variable \((U_t)\) that can take on the values of zero or one, they suppose

\[
x_t = \mu_x(U_t) + \psi_x(U_t) r_{t-1}^b + \sigma_x(U_t) \epsilon_t^x,
\]

\[
r_t^i = \mu_i(U_t) + \psi_i(U_t) r_{t-1}^b + b_i \sigma_x(U_t) \epsilon_t^x + \sigma_i \epsilon_t^i,
\]

where \(i = b, s\). In this model, BL constrained the conditional mean dynamics to only depend on the past bond return but not on the past stock return. The \(U_t\) variable follows a Markov chain with either constant transition probabilities or transition probabilities that depend on the past bond return:

\[
\Pr (U_t = 0 | U_{t-1} = 0, I_{t-1}) = \frac{\exp (a_0 + d_0 r_{t-1}^b)}{1 + \exp (a_0 + d_0 r_{t-1}^b)},
\]

\[
\Pr (U_t = 1 | U_{t-1} = 1, I_{t-1}) = \frac{\exp (a_1 + d_1 r_{t-1}^b)}{1 + \exp (a_1 + d_1 r_{t-1}^b)}.
\]

The parameter vector \(\Theta_{RS}\) to be estimated contains 20 parameters

\[
\Theta_{RS} = \left[ \begin{array}{cccccc}
\mu_i (j) & \psi_i (j) & \sigma_x (j) & \sigma_k & b_k & a_j & d_j
\end{array} \right], \ k = s, b \text{ and } i = x, s, b.
\]

In this regime-switching framework, using (6.1), it is straightforward to show that the joint process of the pricing kernel and asset returns is conditionally lognormally distributed. Consequently, the expression (6.3) remains valid.

When \(d_0 = 0\) and \(d_1 = 0\), we term this model the “CP RS” model. When the transition probabilities are not constant, we term this model the “TP RS” model.

6.2 Data simulation

To illustrate the OHM and CY bounds, we simulate asset and bond returns. We use the econometric models proposed by Bekaert and Liu (2004), see details in the previous section. The estimated parameters (see the Appendix) used in the simulations are taken from Bekaert and Liu (2004). Simulations use 15,500 observations where the first 500 observations are discarded. Recall that Bekaert and Liu (2004) reject the constrained VAR model with a likelihood ratio test with a p-value of 0.000, while the other models are not rejected at the 5% level.

6.3 Empirical results

6.3.1 Efficiency

The purpose of this section is to empirically investigate whether higher-order moments may account for predictability in asset returns. Figures 1, 2, 3, and 4 present the variance bounds when data are simulated from the UCO VAR, CO VAR, TP RS and the CP RS models. Four important results stand out in these figures. First, the difference between the HJ and the CGR bounds reveal little predictability, although the difference between these bounds is sharper for small \(\overline{m}\)’s. Second, the difference between the OHM bound and the BL bound reveals considerable predictability. The difference between the CY bound and the GHT
The OHM bound also reveals considerable predictability. When the pricing kernel mean is in the neighborhood of 0.995, the OHM bound is 40% higher than the BL bound, while the CY bound is 25% 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% higher than the HJ bound, while the BL bound is 20% higher than the HJ bound. As well, the CY bound is 40% higher than the HJ bound while the GHT is 20% higher than the HJ bound. This predictability is due to the conditional higher-order moments of asset returns. Intuitively, this result shows that conditioning variables whose distributions are characterized by the mean, the variance, the skewness and the kurtosis may help to better predict future returns than conditioning variables whose distributions are characterized by the mean and variance alone.

Fourth, surprisingly, the difference between the CY bound and the larger OHM bound is huge with the OHM bound being larger. There are two potential explanations for this. First, parameter uncertainty risk and second the approach used to compute the conditional price of the squared asset return may account for this difference. Actually, the conditional price of the squared return is computed from the assumption that we live in a world where the pricing kernels are of the form (6.1). To examine this issue more closely, we simulate data according to the UCO VAR model and TP RS model. We then plot in Figures 5 and 7 the OHM bound with the conditional moments calculated from the UCO VAR model, and the conditional price of the squared stock return calculated from four different models. When the conditional price of the squared return is misspecified, Figures 5 and 7 reveal that the OHM bounds are below the bound calculated with the true conditional price (i.e., the price calculated from the UCO VAR using (6.1)). The difference between the bounds is large when data are simulated from the TP RS (see Figure 7). When returns are simulated from the CO VAR model and conditional moments calculated from the CO VAR model, the CY and OHM bounds are above the GHT and BL bounds even if the OHM bound is not a parabola. The CO VAR model is rejected by Bekaert and Liu (2004) with a p-value of 0.0000. This potentially explains why the OHM bound is not a parabola. We now illustrate the CY bound when the conditional price of the squared return is misspecified. In this case, Figures 6 and 8 reveal that the CY bound underestimates the true lower bound on the variance of pricing kernels. But the difference between the CY bound calculated with the misspecified conditional price and true conditional price is quite small. Overall, it stands out clearly that conditioning variables with significant conditional higher-order moments contribute to better predicting future returns.

6.3.2 Diagnostics

The goal of this section is to show that the OHM bound remains a lower bound to the variance of pricing kernels if the conditional moments of asset return are not correctly specified. Figure 9 presents the bounds with data simulated according to the UCO VAR model and conditional moments calculated from the CO VAR. Three results stand out. First, the GHT and CY bounds fail to highlight the misspecification of the first four conditional moments. When the first four conditional moments are misspecified, the CY and GHT bounds quite overestimate the bound calculated with the true conditional moments. Second, when the first four conditional moments are misspecified, the OHM bound underestimates the bound calculated with the true conditional moments. Third, the OHM bound highlights the misspecification of the first four conditional moments while the BL bound does not. To examine more closely how the misspecification of higher-order conditional moments affects the variance bounds, we investigate several cases. We simulate data according
Figure 1: Pricing kernel bounds for simulated data according to the UCO VAR model with conditional moments calculated from the UCO VAR model.

Figure 2: Pricing kernel bounds for simulated data according to the CO VAR model with conditional moments calculated from the CO VAR model.
Figure 3: Pricing kernel bounds for simulated data according to the TP RS model with conditional moments calculated from the TP RS model.

Figure 4: Pricing kernel bounds for simulated data according to the CP RS model with conditional moments calculated from the CP RS.
Figure 5: OHM bounds for simulated data according to the UCO VAR model with conditional moments calculated from the UCO VAR model and the conditional prices of squared returns calculated from four different models.

Figure 6: CY bounds for simulated data according to the UCO VAR model with conditional moments calculated from the UCO VAR model and the conditional prices of squared returns calculated from four different models.
Figure 7: OHM bounds for simulated data according to the TP RS model and conditional moments calculated from the TP RS model and the conditional prices of squared returns calculated from four different models.

Figure 8: GHT bounds for simulated data according to the TP RS model and conditional moments calculated from the TP RS model and the conditional prices of squared returns calculated from four different models.
to the UCO VAR and TP RS models and document several misspecification cases. The first case assumes that the conditional skewness and conditional kurtosis are misspecified, the second case assumes that the conditional kurtosis is misspecified, and the last case assumes that the conditional skewness is misspecified. The results are displayed in Figures 10, 11, 12 and 13.

6.3.3 Robustness

In Figures 14 and 16, we generate data according to the regime-switching model with time-varying probabilities and the unconstrained VAR model. For each case, when the first four conditional moments are not correctly specified, the OHM bound underestimates the true lower bound on the variance of pricing kernels. However, when the four conditional moments are misspecified, Figures 15 and 17 reveal that the CY bound quite overestimates the true lower bound on the variance of pricing kernels. This lack of robustness is a drawback to the CY.
Figure 10: We simulate data according to the UCO VAR model. CY is the bound with conditional moments calculated from the UCO VAR model. CY-K is the CY bound with the conditional kurtosis calculated from the CO VAR model. CY-S is the CY bound with the conditional skewness calculated from the CO VAR model. CY-SK is the CY bound with the conditional skewness and conditional kurtosis calculated from the CO VAR model.

Figure 11: We simulate data according to the unconstrained VAR. OHM is the bound with conditional moments calculated from the unconstrained VAR. OHM-K is the OHM bound with the conditional kurtosis calculated from the constrained VAR. OHM-S is the OHM bound with the conditional skewness calculated from the constrained VAR. OHM-SK is the OHM bound with the conditional skewness and conditional kurtosis calculated from the constrained VAR.
Figure 12: We simulate data according to the TP RS model. CY is the bound with conditional moments calculated from the TP RS. CY-K is the CY bound with the conditional kurtosis calculated from the CO VAR model. CY-S is the CY bound with the conditional skewness calculated from the CO VAR model. CY-SK is the CY bound with the conditional skewness and conditional kurtosis calculated from the CO VAR model.

Figure 13: We simulate data according to the TP RS model. OHM is the bound with conditional moments calculated from the TP RS model. OHM-K is the OHM bound with the conditional kurtosis calculated from the CO VAR model. OHM-S is the OHM bound with the conditional skewness calculated from the CO VAR model. OHM-SK is the OHM bound with the conditional skewness and conditional kurtosis calculated from the CO VAR model.
Figure 14: Optimally scaled pricing kernel bounds for simulated data according to TP RS model with conditional moments calculated from four models.

Figure 15: CY and GHT bounds for simulated data according to the TP RS model with conditional moments calculated from four models.
Figure 16: Optimally scaled pricing kernel bounds for simulated data according to the UCO VAR model with conditional moments calculated from four models.

Figure 17: CY and GHT bounds for simulated data according to the UCO VAR model with conditional moments calculated from four models.
7. Conclusion

There is growing use of the HJ bound in finance. And there is increasing interest shown by the finance profession (both academic and practitioners) in building asset pricing models that incorporate skewness (see Harvey and Siddique 2000). With evidence of time variation in the conditional mean, conditional variance, conditional skewness and conditional kurtosis, it becomes critical to optimally incorporate not only the conditional mean and variance in the HJ bound, but to find a bound on the variance of pricing kernels that incorporates conditional skewness and kurtosis.

Our paper provides an efficient variance bound on pricing kernels that incorporates the conditional higher moments. It also provides a bridge between this bound and a variance bound on pricing kernels with higher moments that use optimal scaling instruments. The advantage of the optimal scaling procedure is that it often produces a valid lower bound to the variance of pricing kernels, whereas the efficient bound that incorporates higher conditional moments may overestimate the lower bound on the variance of pricing kernels when the higher moments are misspecified. In this paper, we derive the best possible scaled bound with higher moments. But this bound requires specifying the conditional mean, variance, skewness and kurtosis and the squared asset returns’ conditional prices. When these conditional inputs are correctly specified, the scaled bound under higher moments coincides with the efficient variance bound that incorporates higher moments. Even, when these inputs are misspecified, this bound quite often produces a valid lower bound to the variance of the pricing kernels.

There are interesting applications of this work. Our bound can be used to examine the predictability of asset returns when there is strong evidence that skewness is priced into the market. It can also be used to examine which instruments yield the sharpest restrictions on asset return dynamics when these returns display higher moments. When skewness is priced into the market, the OHM bound is significantly higher than the Bekaert and Liu (2004) optimally scaled bound.

Second, our optimally scaled bound can produce information on expected return, conditional variance, conditional skewness, conditional kurtosis, and conditional price of derivatives and serve as a diagnostic tool to judge the performance of dynamic asset-pricing models (in this article we approximate the payoff of derivatives by their linear regression on the returns and squared returns). In fact, when the conditional moments of asset returns and conditional prices of derivatives are correctly specified, the optimally scaled bound under higher moments is the best. This property is used to propose a GMM-based specification test for the conditional mean, the conditional variance, the conditional skewness and the conditional kurtosis when the conditional prices of derivatives are correctly specified.

Third, the optimally scaled bound under higher moments can be used in dynamic models of optimal asset and derivative allocations.

Fourth, this bound can also be used in developing performance measures for portfolio managers when they employ a dynamic framework.
References


Table 1: Unconstrained GARCH-in-mean model

<table>
<thead>
<tr>
<th>Equations</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_t )</td>
<td>Constant: 0.0030, ( X_{t-1} = 0.361 ), ( R^b_{t-1} = -0.029 ), ( R^s_{t-1} = 0.008 ) ((0.0005)) ((0.033)) ((0.022)) ((0.005))</td>
</tr>
<tr>
<td>( R^b_t )</td>
<td>0.0056 (- 162.65 h_{xxt} ), ( R^s_t ) (0.0188 - 58.02 h_{xxt} ) ((0.0006)) ((0.0001)) ((0.031)) ((0.031)) ((0.037)) ((0.0043))</td>
</tr>
<tr>
<td>( h_{11t} )</td>
<td>(\alpha_i = 0.000019), (\kappa_i = -0.0265), (\eta_i = 0.0008), (\eta_i = 0.2705) ((0.000018)) ((0.0807)) ((0.7898)) ((0.0426))</td>
</tr>
<tr>
<td>( h_{22t} )</td>
<td>(0.000014), (0), (0), (0) ((0.000002))</td>
</tr>
<tr>
<td>( h_{33t} )</td>
<td>(0.006134), (0), (0), (0) ((0.00103))</td>
</tr>
<tr>
<td>( f_{xb} = -0.0564 ), ( f_{xs} = 3.182 ) ((0.1425)) ((0.003))</td>
<td></td>
</tr>
</tbody>
</table>

Notes: In this table, we reproduce the results of the unconstrained GARCH-in-mean model estimated by Bekaert and Liu (2004). Standard errors are in parentheses.
Table 2: Constrained GARCH-in-mean model

<table>
<thead>
<tr>
<th>Equations</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_t$</td>
<td>Constant</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
</tr>
<tr>
<td>$R^b_t$</td>
<td>0.0053 - 108.97$h_{xxt}$</td>
</tr>
<tr>
<td>$R^s_t$</td>
<td>0.0021 - 82.086$h_{xxt}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>14.675</td>
</tr>
<tr>
<td>$\alpha_i$</td>
<td>$\kappa_i$</td>
</tr>
<tr>
<td>$h_{11t}$</td>
<td>0.000022</td>
</tr>
<tr>
<td></td>
<td>(0.000006)</td>
</tr>
<tr>
<td>$h_{22t}$</td>
<td>0.000013</td>
</tr>
<tr>
<td></td>
<td>(0.000002)</td>
</tr>
<tr>
<td>$h_{33t}$</td>
<td>0.006457</td>
</tr>
<tr>
<td></td>
<td>(0.001009)</td>
</tr>
<tr>
<td>$f_{xb}$</td>
<td>-0.0877</td>
</tr>
<tr>
<td></td>
<td>(0.0813)</td>
</tr>
</tbody>
</table>

Notes: In this table, we reproduce the results of the constrained GARCH-in-mean model estimated by Bekaert and Liu (2004). Standard errors are in parentheses.
Table 3: Regime-switching model with constant transition probabilities

<table>
<thead>
<tr>
<th>Equations</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
</tr>
<tr>
<td>$U_t = 1$</td>
<td>$U_t = 2$</td>
</tr>
<tr>
<td>$U_t = 1$</td>
<td>$U_t = 2$</td>
</tr>
<tr>
<td>$U_t = 1$</td>
<td>$U_t = 2$</td>
</tr>
<tr>
<td>$X_t$</td>
<td>$R^b_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_i$</td>
</tr>
<tr>
<td></td>
<td>$b_i$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: In this table, we reproduce the results of the regime switching model with constant transition probabilities estimated by Bekaert and Liu (2004). The parameters estimates for the constant transition probabilities are $a_0 = 3.4493$ and $d_0 = 1.2569$. Standard errors are in parentheses.

Table 4: Regime-switching model with time varying transition probabilities

<table>
<thead>
<tr>
<th>Equations</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
</tr>
<tr>
<td>$U_t = 1$</td>
<td>$U_t = 2$</td>
</tr>
<tr>
<td>$U_t = 1$</td>
<td>$U_t = 2$</td>
</tr>
<tr>
<td>$U_t = 1$</td>
<td>$U_t = 2$</td>
</tr>
<tr>
<td>$X_t$</td>
<td>$R^b_{t-1}$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_i$</td>
</tr>
<tr>
<td></td>
<td>$b_i$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: In this table, we reproduce the results of the regime switching model with time varying transition probabilities estimated by Bekaert and Liu (2004). The parameters estimates for the time varying transition probabilities are $a_0 = 3.5003$, $a_1 = -1.8830$, $d_0 = 0.3785$ and $d_1 = 13.0799$. Standard errors are in parentheses.
8. Appendix

**Proof of Proposition 2.1.** The proof is similar to the proof of (2.2).

**Proof of Proposition 2.4.** Bekaert and Liu (2004) give the solution to \( \sup_{z_t \in I_t} \sigma^2 (\bar{m}, z_t g_{t+1}) \). In our case, it is easy to solve \( \sup_{z_t \in I_t} \sigma^2 (\bar{m}, z_t g_{t+1}) \) using the same approach as Bekaert and Liu (2004), see proposition 1 in Bekaert and Liu (2004).

**Proof of Proposition 2.5.** The CGR variance bound with conditioning information represents the efficient way of using conditional information. Then it follows that \( \sigma^2 (\bar{m}, z_t g_{t+1}) \leq \sup_{z_t \in I_t} \sigma^2 (\bar{m}, z_t g_{t+1}) \leq \sigma^2 (m_m^{mvs+t}) \). Using the expression for \( z_t^* \) the variance of \( z_t^* g_{t+1} \) is

\[
\text{Var} \left( z_t^* g_{t+1} \right) = \text{Var} (z_t^* r_{t+1}) + \text{Var} (z_t^* \varepsilon_{t+1})
\]

\[
= \sigma^2 (m_{GH}^{t+1}) + E \left( \kappa_t - (\sigma_i^2 + \mu_t^2 - \sigma_i^{-2} s_i^2)^{-1} (c_t - \bar{c}_t)^2 \right)
\]

Using the definition of \( a_1, b_1 \) and \( d_1 \) and \( a_2, b_2 \) and \( d_2 \), the result follows.

**Proof of Proposition 2.6.** We first note that

\[
B = \sigma^2 (m_{t+1}^{mvs}) = \frac{a (1 - d_1) + \overline{m} d_1 - 2 \overline{m} b_1 + b_1^2}{1 - d_1}
\]

where \( a = a_1 + \frac{a_2 + \overline{m} b_2 - 2 \overline{m} d_2}{1 - d_1} \).

But \( A \) is given by

\[
A = Ez_t^* \pi_t - \overline{m} Ez_t^* g_{t+1}
\]

\[
= Ez_t^* p_t - \overline{m} Ez_t^* r_{t+1} + Ez_t^* (c_t - \bar{c}_t)
\]

\[
= \text{Var} (z_t^* r_{t+1}) + Ez_t^* (c_t - \bar{c}_t)
\]

\[
= \text{Var} (z_t^* r_{t+1}) + E (u_t^2 - \sigma_i^{-2} s_i^2)^{-1} (c_t - \bar{c}_t)^2
\]

\[
= \text{Var} (z_t^* g_{t+1})
\]

Using proposition 2.5, the result follows.

**Proof of Proposition 2.7.** From proposition 2.6, using the expression of \( z_t^* \), the result follows.

**Proof of Proposition 4.1.** The CGR scaled bound can be written as:

\[
\sigma^2 (\bar{m}, z_t g_{t+1}) = \frac{(A_1 + A_2)^2}{B_1 + B_2}
\]

where

\[
A_1 = Ez_{1t} p_t - \overline{m} Ez_{1t} r_{t+1}, \quad A_2 = Ez_{2t} (c_t - \bar{c}_t), \quad B_1 = \text{Var} (z_{1t} r_{t+1}), \quad B_2 = \text{Var} (z_{2t} \varepsilon_{t+1})
\]
If the risk premium $Ez_{t+1}r_{t+1} - r_f > 0$, we should have $A_1 = Cov (m_{t+1}^{\text{mv}}, z_{t+1}r_{t+1}) < 0$. Since $A_2 > 0$, we have
\begin{equation}
\sigma^2 \left( \overline{m}, z_{t}g_{t+1} \right) \leq \frac{A_1^2}{B_1} + \frac{A_2^2}{B_2}.
\end{equation}
Under Assumption 1,
\begin{equation}
\frac{A_1^2}{B_1} = \left( \frac{Ep - \overline{m}Er_{t+1}}{Er_{t+1} - (Er_{t+1})^2} \right)^2 \frac{\left( Er_{t+1} - (Er_{t+1})^2 \right)^2}{\left( \frac{Ez_{t+1}r_{t+1}}{Er_{t+1} - (Er_{t+1})^2} \right)^2} \leq \frac{(Ep - \overline{m}Er_{t+1})^2}{\text{Var}(r_{t+1})}.
\end{equation}
The last inequality follows since $\frac{(Ez_{t+1}r_{t+1})^2}{Er_{t+1} - (Er_{t+1})^2} \geq 1$. Under Assumption 2,
\begin{equation}
\frac{A_2^2}{B_2} = \frac{(E(c_t - \overline{\tau}))^2}{Ez_{t+1}^2} \left[ \frac{(Ez_{t+1}^2)}{Ez_{t+1}^2} \right] \leq \frac{(E(c_t - \overline{\tau}))^2}{\text{Var}(\epsilon_{t+1})} = \frac{(\text{Price}(\epsilon_{t+1}))^2}{\text{Var}(\epsilon_{t+1})}
\end{equation}
where $\epsilon_{t+1}$ is
\begin{equation}
\epsilon_{t+1} = r_{t+1}^2 - (\mu_t^2 + \sigma_t^2) - \sigma_t^{-2}s_t(r_{t+1} - \mu_t).
\end{equation}
We could also regress the squared return onto the return without using conditioning information. In that case, one get
\begin{equation}
\epsilon_{t+1} = r_{t+1}^2 - (\mu_t^2 + \sigma_t^2) - \sigma_t^{-2}s_t(r_{t+1} - \mu)
\end{equation}
However, we claim that
\begin{equation}
\text{Price}(\epsilon_{t+1}) \leq \text{Price}(\epsilon_{t+1})
\end{equation}
\begin{equation}
\text{Var}(\epsilon_{t+1}) \geq \text{Var}(\epsilon_{t+1})
\end{equation}
Using the last two equalities, we argue that
\begin{equation}
\sigma^2 \left( \overline{m}, z_{t}g_{t+1} \right) \leq \frac{A_1^2}{B_1} + \frac{A_2^2}{B_2} \leq \frac{(Ep - \overline{m}Er_{t+1})^2}{\text{Var}(r_{t+1})} + \frac{(c - \overline{\tau})^2}{\text{Var}(\epsilon_{t+1})}.
\end{equation}
We remind that $\text{Price}(\epsilon_{t+1}) = c - \overline{\tau}$, therefore,
\begin{equation}
\sigma^2 \left( \overline{m}, z_{t}g_{t+1} \right) \leq \frac{(Ep - \overline{m}Er_{t+1})^2}{\text{Var}(r_{t+1})} + \frac{(c - \overline{\tau})^2}{\text{Var}(\epsilon_{t+1})}
\end{equation}
\begin{equation}
\leq \frac{(Ep - \overline{m}Er_{t+1})^2}{\text{Var}(r_{t+1})} + \left( (c - \overline{\tau})^2 \right)^{-1} \left( \frac{\left( \overline{m}^2 + \sigma_t^2 - \sigma_t^{-2}s_t \right)^2}{\text{Var}(\epsilon_{t+1})} \right)
\end{equation}
We remind that Bekaert and Liu (2004) scaled bound, $\sigma^2 (\overline{m}, z_{1t}r_{1t+1})$, is lower than the CGR scaled bound. To prove it explicitly, we remind that the scaled CGR variance bound is found using equalities
\begin{equation}
Em_{t+1}^{\text{mv}}z_{1t}r_{1t+1} = Ez_{1t}p_t,
\end{equation}
\begin{equation}
Em_{t+1}^{\text{mv}}z_{2t}r_{1t+1}^2 = Ez_{2t}c_t.
\end{equation}
Letting $\nu_{t+1}$ be the linear regression of $z_{2t}r_{1t+1}^2$ on $z_{1t}r_{1t+1}$:
\begin{equation}
\nu_{t+1} = z_{2t}r_{1t+1}^2 - \frac{Cov \left( z_{2t}r_{1t+1}^2, z_{1t}r_{1t+1} \right)}{\text{Var}(z_{1t}r_{1t+1})} \left( z_{1t}r_{1t+1} - Ez_{1t}r_{1t+1} \right)
\end{equation}
The scaled CGR pricing kernel can be written as

\[ m_{t+1}^{s_{vs}} = m + \alpha_1 (z_t r_{t+1} - E z_t p_t) + \alpha_2 \nu_{t+1}, \]  

(8.8)

where \( \alpha_1 \) and \( \alpha_2 \) are real numbers. Now we plug (8.8) in equalities (8.6) and (8.7) and solve the resultant equation for \( \alpha_1 \) and \( \alpha_2 \):

\[
\alpha_1 = \frac{E z_t p_t - \overline{m} E z_t r_{t+1}}{\text{Var}(z_t r_{t+1})}, \\
\alpha_2 = \frac{1}{\text{Var}(\nu_{t+1})} \left[ E z_t c_t - \frac{\text{Cov}(z_t r_{t+1}, z_t r_{t+1})}{\text{Var}(z_t r_{t+1})} (E z_t p_t - \overline{m} E z_t r_{t+1}) \right].
\]

Consequently, the variance of \( m_{t+1}^{s_{vs}} \) is

\[
\sigma^2(m, z_t) = \frac{(E z_t p_t - \overline{m} E z_t r_{t+1})^2}{\text{Var}(z_t r_{t+1})} + \alpha_2^2 \text{Var}(\nu_{t+1}) \\
= \sigma^2(m, z_t r_{t+1}) + \alpha_2^2 \text{Var}(\nu_{t+1}) \\
\geq \sigma^2(m, z_t r_{t+1}).
\]

Then,

\[
\sigma^2(m, z_t r_{t+1}) \leq \sigma^2(m, z'_t g_{t+1}) \leq \text{Var}(m_{t+1}^{s_{vs}}).
\]

This ends the proof. \[\Box\]

**Proof of Proposition 4.2.** The vector \( (g_{t+1}, z'_t g_{t+1}) \) can be written as \( z'_t g_{t+1} \), consequently

\[
\sup_{z_t \in I_t} \sigma^2(m, (g_{t+1}, z'_t g_{t+1})) = \sup_{z_t \in I_t} \sigma^2(m, z'_t g_{t+1})
\]

Using proposition 4, we have

\[
\sigma^2(m, z'_t g_{t+1}) \leq \sigma^2(m, z''_t g_{t+1})
\]

Then

\[
\sup_{z_t \in I_t} \sigma^2(m, z'_t g_{t+1}) = \sigma^2(m, (g_{t+1}, z'_t g_{t+1})) \leq \sigma^2(m, z''_t g_{t+1})
\]

However it is clear that

\[
\sigma^2(m, z''_t g_{t+1}) \leq \sigma^2(m, (g_{t+1}, z''_t g_{t+1}))
\]

Combining the last two equalities ends the proof. \[\Box\]

**Computing the Cost of the Squared Return.** Assume that the joint process \( (m_{t+1}, R'_{t+1}) \) is conditionally lognormal. Then,

\[
\begin{bmatrix} \log(m_{t+1}) \\ \log R_{t+1} \end{bmatrix} | I_t \sim N \left( \begin{bmatrix} \mu_m \\ \mu_s \end{bmatrix}, \begin{bmatrix} \sigma_m^2 & \Sigma_m r \\ \Sigma_m r & \Sigma_r \end{bmatrix} \right)
\]

with

\[ R_{t+1} = (R'_{t+1}, R^b_{t+1})'. \]
For the sake of notational convenience let’s denote

\[ \log R_{t+1} = (r_{i+1}^*, r_{t+1}^b)'. \]

We know that

\[ E_t m_t R_t^i = 1 \forall i \in \{b, s\}. \]

Let compute

\[ c_{ijt} = E_t m_{t+1} R^i_{t+1} R^j_{t+1} \]

Therefore,

\[ \log (m_{t+1} R^i_{t+1} R^j_{t+1}) = \log (m_{t+1}) + \log (R^i_{t+1}) + \log (R^j_{t+1}) \]

Let \( \mu_m \) and \( \sigma_m^2 \) denote the first two moments of \( \log (m_{t+1}) \) and \( \mu_i \) and \( \sigma_i^2 \) denote the first two moments of \( \log (R^i_{t+1}) \). As result,

\[
c_{ijt} = \exp \left[ \mu_m + \mu_i + \mu_j + \frac{1}{2} \left( \sigma_m^2 + \sigma_i^2 + \sigma_m^2 + 2\sigma_{im} + 2\sigma_{jm} \right) \right] \]
\[
= \exp \left[ \mu_m + \frac{1}{2} \sigma_m^2 + \mu_i + \frac{1}{2} \sigma_i^2 + \sigma_{im} \right] \exp \left[ \mu_j + \frac{1}{2} \left( \sigma_j^2 + 2\sigma_{ij} + 2\sigma_{jm} \right) \right] \]
\[
= E_t m_{t+1} R^i_{t+1} \exp \left[ \mu_j + \frac{1}{2} \left( \sigma_j^2 + 2\sigma_{ij} + 2\sigma_{jm} \right) \right] \]
\[
= \exp \left[ \mu_j + \frac{1}{2} \left( \sigma_j^2 + 2\sigma_{ij} + 2\sigma_{jm} \right) + \mu_m + \frac{1}{2} \sigma_m^2 \right] \exp \left[ -\mu_m - \frac{1}{2} \sigma_m^2 \right] \]
\[
= \exp \left[ \mu_m + \frac{1}{2} \sigma_m^2 + \mu_j + \frac{1}{2} \sigma_j^2 + \sigma_{jm} \right] \exp \left[ \frac{1}{2} (2\sigma_{ij}) \right] \exp \left[ -\mu_m - \frac{1}{2} \sigma_m^2 \right] \]
\[
= E_t m_{t+1} R^i_{t+1} \exp \left[ \frac{1}{2} (2\sigma_{ij}) \right] \exp \left[ -\mu_m - \frac{1}{2} \sigma_m^2 \right] \]

Then,

\[
c_{ijt} = \exp \left[ \frac{1}{2} (2\sigma_{ij}) \right] \exp \left[ -\mu_m - \frac{1}{2} \sigma_m^2 \right] \]
\[
= \frac{1}{E_t m_{t+1}} \exp \left[ \frac{1}{2} (2\sigma_{ij}) \right] \]
\[
= \frac{1}{E_t m_{t+1}} \exp \left[ \mu_i + \mu_j + \frac{1}{2} \left( \sigma_i^2 + \sigma_j^2 + 2\sigma_{ij} \right) \right] \exp \left[ -\mu_i - \frac{1}{2} \sigma_i^2 \right] \exp \left[ -\mu_j - \frac{1}{2} \sigma_j^2 \right] \]
\[
= \frac{1}{E_t m_{t+1}} \frac{E_t \left( R^i_{t+1} R^j_{t+1} \right)}{E_t \left( \exp r^i_{t+1} \right) E_t \left( \exp r^j_{t+1} \right)}. \]

But \( \frac{1}{E_t m_{t+1}} \) is the conditional risk-free return, \( R_{ft} \). Here the time series of Treasury bill return can be used as the conditionally risk-free return. Therefore

\[
c_{ijt} = \frac{\exp (r^i_t) E_t \left( \exp \left( r^i_{t+1} + r^j_{t+1} \right) \right)}{E_t \left( \exp r^i_{t+1} \right) E_t \left( \exp r^j_{t+1} \right)} \text{ for } i, j \in \{b, s\}. \]

This ends the proof. ■