Conditioning Information and Variance Bounds on Pricing Kernels

Geert Bekaert
Columbia University and NBER

Jun Liu
University of California, Los Angeles

Gallant, Hansen, and Tauchen (1990) show how to use conditioning information optimally to construct a sharper unconditional variance bound (the GHT bound) on pricing kernels. The literature predominantly resorts to a simple but suboptimal procedure that scales returns with predictive instruments and computes standard bounds using the original and scaled returns. This article provides a formal bridge between the two approaches. We propose an optimally scaled bound that coincides with the GHT bound when the first and second conditional moments are known. When these moments are misspecified, our optimally scaled bound yields a valid lower bound for the standard deviation of pricing kernels, whereas the GHT bound does not. We illustrate the behavior of the bounds using a number of linear and nonlinear models for consumption growth and bond and stock returns. We also illustrate how the optimally scaled bound can be used as a diagnostic for the specification of the first two conditional moments of asset returns.

Hansen and Jagannathan (1991) derive a lower bound (the HJ bound) on the standard deviation of the pricing kernel or the intertemporal marginal rate of substitution as a function of its mean. Using only unconditional first and second moments of available asset returns, the HJ bound defines a feasible region on the mean-standard deviation plane of pricing kernels. Whereas initially HJ bounds primarily served as informal diagnostic tools for consumption-based asset pricing models [see Cochrane and Hansen (1992) for a survey], its applications have rapidly multiplied in recent years. They now include formal asset pricing tests [Burnside (1994), Cecchetti, Lam, and Mark (1994), Hansen, Heaton, and Luttmer (1995)], predictability studies [Bekaert and Hodrick (1992)], mean variance spanning tests [Snow (1991), Bekaert and Urias (1996), De Santis (1995)], market integration tests (Chen and Knez (1995)), mutual fund

We would like to thank Andrew Ang, Luca Benzoni, Michael Brandt, Qiang Dai, Darrell Duffie, Wayne Ferson, Mark Garmaise, Ming Huang, Jun Pan, John Scruggs, seminar participants at Stanford University, the European Finance Association meeting in Barcelona, the American Finance Association meeting in Washington, D.C., and especially John Heaton (the editor) for valuable suggestions. Kenneth Singleton and an anonymous referee provided extensive guidance, comments, and insights. Geert Bekaert thanks the National Science Foundation for financial support. Address correspondence to Geert Bekaert, Columbia Business School, Uris Hall, Room 802, 3022 Broadway, New York, NY 10027, or e-mail: gb241@columbia.edu.

The Review of Financial Studies Vol. 17, No. 2, pp. 339–378 DOI: 10.1093/rfs/hhg052
 Advance Access publication October 15, 2003

performance measurement (Chen and Knez (1996), Ferson and Schadt (1996), Dahlquist and Söderlind (1999)), and others.

HJ bounds are computed by projecting the pricing kernel unconditionally on the space of available asset payoffs and computing the standard deviation of the projected pricing kernel. The larger this standard deviation, the stronger the restrictions on asset pricing models. The standard consumption-based asset pricing model with time-additive preferences dramatically fails to lie inside the feasible region defined by the HJ bounds computed using a variety of asset returns. However, the pricing kernels in more recent models, such as the nonseparable utility model in Heaton (1995) or incomplete markets model of Constantinides and Duffie (1996), satisfy the bounds.

In this article we study the use of conditioning information to effectively increase the dimension of the available asset payoffs and hence to improve the bounds.[1] Gallant, Hansen, and Tauchen (1990; hereafter GHT) show how to use conditioning information efficiently. The procedure is in principle straightforward. They construct an infinite space of available payoffs combining conditioning information and a primitive set of asset payoffs. The variance of the unconditional projection of the pricing kernel onto that space is the efficient HJ bound, which we will term the GHT bound.[2]

The GHT bound depends on the first and second conditional moments of the asset payoffs. The GHT procedure has not been used very much in practice, and researchers have mostly resorted to a simpler technique of embedding conditioning information in the computation of HJ bounds. They simply scale returns with predictive variables in the information set, augment the space of available payoffs (and corresponding prices) with the relevant scaled payoffs or returns, and compute a standard HJ bound for the augmented space [see, e.g., Hansen and Jagannathan (1991), Cochrane and Hansen (1992), Bekaert and Hodrick (1992), and many others]. This procedure is much simpler to implement than GHT since it does not require knowledge of conditional moments at all.

In this article we provide a formal bridge between the optimal but relatively unknown GHT bound and the ad hoc scaling methods prevalent in the literature. We prove two main results. First, we answer the following question: When scaling a return with a function of the conditioning information, what is the function that maximizes the HJ bound? The solution is an application of variational calculus. The resultant optimal

[1] Other methods have been proposed to improve HJ bounds. Snow (1991) studies the restrictions on the higher moments of the pricing kernel. Balduzzi and Kallal (1997) tighten the bounds by using the risk premiums that the pricing kernel assigns to arbitrary sources of risk.

[2] While GHT study both conditional as well as unconditional projections, we will only study unconditional projections.

scaling factor is decreasing in the conditional variance but is not monotonic in the conditional mean. Second, we show that our bound, which we term the optimally scaled bound, is as tight as the GHT bound when the conditional moments are known.

The optimally scaled bound has three important properties. First, it is efficient. Rather than arbitrarily scaling returns with an instrument, our procedure optimally exploits conditioning information leading to sharper bounds. We also use this property to explore the relation between improvements in HJ bounds due to conditioning information and the presence of return predictability.

Second, it is robust to misspecification of the conditional mean and variance. Whereas the GHT bound is also efficient, it is only correct when the conditional moments are accurate. If they are misspecified, the resulting bound may be larger than the variance of the true pricing kernel. Since the optimal bound we derive is a standard HJ bound, it always provides a bound to the variance of the true pricing kernel even if incorrect proxies to the conditional moments are used.

Third, the optimally scaled bound is a useful diagnostic for the specification of the first and second moments of asset returns. Our bound only attains the maximum when the first and second conditional moments are correctly specified. If they are not, the Hansen-Jagannathan frontier is not even a parabola, so that misspecification is visually clear. We also suggest a diagnostic test that can be used to formally compare the fit of alternative specifications of the conditional mean and variance. Given the nonnegligible modeling and parameter uncertainty regarding the first and second conditional moments of asset returns, this property of our bound is likely to be important in many finance applications.

We organize the article into three parts. Section 1 starts by clarifying the relation between standard HJ bounds, the GHT bound, ad hoc scaled bounds, and our optimally scaled bound. We then prove our two main results, deriving an optimal scaling function and showing that the resulting bound reaches the GHT bound when the conditional moments are correctly specified. Section 2 discusses the three main properties of our optimally scaled bound. We end the section by comparing our work to that of Ferson and Siegel (2001). They derive and study the optimal scaling factor in the setting of mean-variance frontiers. Since there is a well-known duality between Hansen-Jagannathan frontiers and the mean-variance frontier, their results are similar to ours but there are also some important differences.

Section 3 contains an empirical illustration. We estimate both an asymmetric GARCH-in-mean model and a regime-switching model on U.S. consumption growth, and bond and stock returns, and test the restrictions of the standard consumption-based asset pricing model. This generalization of the Hansen and Singleton (1983) model provides a natural null and

alternative model for the first and second moments, whereas the GARCH and regime-switching models provide two nonnested specifications for the conditional mean and variance. We use these models to explore the role of misspecification and robustness in the behavior of the various bounds. We briefly discuss future potential applications of our results in a concluding section.

1. Incorporating Conditioning Information into Variance Bounds

In this section we first review the standard HJ bound while setting up notation in Section 1.1. In Section 1.2 we briefly review the standard way of using conditioning information. Section 1.3 reviews the GHT bound. Section 1.4 introduces the optimally scaled bound.

1.1. Unconditional variance bounds

Let there be a set of assets with payoff vector r_{t+1} and price vector p_t. When the payoff is a (gross) return, the price equals one. Let the vector y_t denote the set of conditioning variables in the economy and let I_t be the σ algebra of the measurable functions of y_t, that is, I_t is the information set. The pricing kernel m_{t+1} prices the payoffs correctly if

$$\mathrm{E}[m_{t+1}r_{t+1}|I_t] = p_t. \tag{1}$$

By the law of iterated expectations, this implies

$$\mathrm{E}[m_{t+1}r_{t+1}] = \mathrm{E}[p_t] \equiv q. \tag{2}$$

Hansen and Jagannathan (1991) derive a bound on the volatility of m_{t+1} that can be computed from asset payoffs and prices alone. This bound follows from projecting the kernel onto the set of payoffs and a constant payoff:

$$m^*_{t+1} = v + \beta'(r_{t+1} - \mu) \tag{3}$$

$$= v + (q - v\mu)'\Sigma^{-1}(r_{t+1} - \mu), \tag{4}$$

where

$$v = \mathrm{E}[m_{t+1}] = \mathrm{E}[m^*_{t+1}], \quad \mu = \mathrm{E}[r_{t+1}] \tag{5}$$

and

$$\Sigma = \mathrm{E}[(r_{t+1} - \mu)(r_{t+1} - \mu)']. \tag{6}$$

The variance bound follows from realizing that $\mathrm{var}(m_{t+1}) \geq \mathrm{var}(m^*_{t+1})$. We denote the bound $\mathrm{var}(m^*_{t+1})$ as $\sigma^2(v; r_{t+1})$ (or $\sigma^2(v)$), since it depends

on the mean of the kernel and the first two moments of r_{t+1}:

$$\sigma^2(v; r_{t+1}) = (q - v\mu)'\Sigma^{-1}(q - v\mu) \tag{7}$$

$$= A - 2Bv + Dv^2, \tag{8}$$

where

$$A = q'\Sigma^{-1}q, \quad B = \mu'\Sigma^{-1}q, \quad D = \mu'\Sigma^{-1}\mu. \tag{9}$$

The parabola $(v, \sigma^2(v))$ is the HJ frontier. Note that if q equals one and there exists a risk-free asset such that $r^f = (\mathrm{E}[m_{t+1}])^{-1}$, then $\sigma^2(v; r_{t+1})$ is proportional to the square of the Sharpe ratio on the set of assets. Hence a sharper HJ bound corresponds to a better risk-return trade-off on the available assets.

To facilitate comparison with the derivations in GHT (1990), we provide an alternative formulation of m^*_{t+1} in terms of the uncentered moments of r_{t+1}:

$$m^*_{t+1} = (q - w\mu)'(\mu\mu' + \Sigma)^{-1}r_{t+1} + w \tag{10}$$

with

$$w = \frac{v - q'(\mu\mu' + \Sigma)^{-1}\mu}{1 - \mu'(\mu\mu' + \Sigma)^{-1}\mu} = \frac{v - b}{1 - d}, \tag{11}$$

where the definition of b and d is implicit.[3] That m^*_{t+1} unconditionally prices the returns follows immediately by substituting Equation (10) into Equation (1). The relation between w and v is apparent from taking the expected value of m^*_{t+1} in Equation (10):

$$v = q'(\mu\mu' + \Sigma)^{-1}\mu + (1 - \mu'(\mu\mu' + \Sigma)^{-1}\mu)w. \tag{12}$$

The intuition behind Equation (10) is rather straightforward. Rewrite the equation as

$$m^*_{t+1} = q'(\mu\mu' + \Sigma)^{-1}r_{t+1} + (1 - \mu'(\mu\mu' + \Sigma)^{-1}r_{t+1})w.$$

The first part of the right-hand side expression is the projection of m_{t+1} onto the original asset payoff space (not augmented with a constant payoff). However, we would like to project m^*_{t+1} on this space augmented with a constant payoff. The coefficient multiplying w is the residual of the projection of a unit payoff onto the r_{t+1} space and hence orthogonal to that space. Consequently, w is the projection coefficient of m^*_{t+1} onto that

[3] Equation (10) can be derived from Equation (4) directly using the identity $(F + gg')^{-1} = F^{-1} - F^{-1}g(\mathcal{I} + g'F^{-1}g)^{-1}g'F^{-1}$ with $F = \mu\mu' + \Sigma$, $g = \mu$, and $\mathcal{I}$ the identity matrix.

residual. The two parts together constitute the projection of m^*_{t+1} onto the r_{t+1} space augmented with a constant payoff.

1.2 Scaled variance bounds

The presence of the conditioning variable y_t allows construction of an infinite-dimensional payoff space [see Hansen and Richard (1987)]. Let $z_t = f(y_t)$, where f is a measurable function and z_t is an $n \times 1$ vector. Scaled returns are simply assets with payoffs equal to $z_t' r_{t+1}$ and prices $z_t' J$ (where J is an $n \times 1$ vector of ones), and do not raise any difficulty in computing standard HJ bounds.

Such scaling has an intuitive interpretation when excess returns, $r^e_{t+1} = r_{t+1} - r^f$, are scaled as in Bekaert and Hodrick (1992) and Cochrane (1996). The gross "scaled" return, $r_{t+1} = z_t' r^e_{t+1} + r^f = z_t' r_{t+1} + (1 - z_t' J) r^f$, can then be interpreted as a "managed" portfolio with $z_t' J$ being the time-varying proportion of the investment allocated to the risky assets.

Scaling likely only improves the HJ bound if the weight z_t has information on future returns. In the literature, one sets $z_t = G y_t$, where G is a selector matrix of ones and zeros selecting the variables in y_t believed to predict r_{t+1} or to capture the time variation in the expected returns.

Most studies stack actual returns with scaled returns [see, e.g., Bekaert and Hodrick (1992) and Cochrane and Hansen (1992)], considering the system $\begin{bmatrix} r_{t+1} \\ r_{t+1} \otimes y_t \end{bmatrix}$. This amounts to considering many different z_t's where each z_t is represented by a selection matrix with only one nonzero element, selecting a particular instrument out of the available instruments. It is fairly unlikely that this is the optimal way to select from the set of information variables. Therefore we sometimes refer to the bounds resulting from this ad hoc approach to scaling as "naive bounds."

1.3 The GHT variance bound

GHT (1990) show how to use conditioning information efficiently. Recall that a scaled asset is a one-dimensional asset, $\tilde{r}_{t+1} = z_t' r_{t+1}$, where z_t is an n-dimensional vector whose entries are measurable functions of y_t (so they belong to I_t). The space of all such scaled payoffs is an infinite-dimensional conditional Hilbert space $P = \{z_t' r_{t+1} : \forall z_t\}$. GHT directly project the pricing kernel onto this space augmented by a riskless payoff. They show that the projected pricing kernel is[4]

$$m^*_{t+1} = (p_t - w\mu_t)'(\mu_t \mu_t' + \Sigma_t)^{-1} r_{t+1} + w, \tag{13}$$

[4] Note that the projection is an unconditional not a conditional projection.

where μ_t is the conditional mean vector, Σ_t is the conditional variance-covariance matrix of the returns, and w is given by

$$w = \frac{v - b}{1 - d}, \tag{14}$$

Here the symbols b and d are the conditional analogues of the definitions in Section 1.1.:

$$b = \mathrm{E}[\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1}p_t] \tag{15}$$

and

$$d = \mathrm{E}[\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1}\mu_t]. \tag{16}$$

The GHT bound by definition is

$$\sigma^2_{GHT}(v) = \mathrm{var}(m^*_{t+1}). \tag{17}$$

It is a lower bound to the variance of all valid pricing kernels. The result in GHT is not surprising given our alternative derivation of the standard pricing kernels in Section 1.1. The GHT kernel is identical, replacing unconditional with conditional moments and expected prices with actual prices [compare Equations (10) and (13)]. This is because the kernel now prices all payoffs conditionally. There is an equivalent representation of the GHT kernel to the standard kernel representation in Equation (4), but it involves the conditional mean of the pricing kernel, $v_t = E_t[m^*_{t+1}]$,

$$m^*_{t+1} = (p_t - v_t\mu_t)'\Sigma_t^{-1}(r_{t+1} - \mu_t) + v_t. \tag{18}$$

Hence v_t is the price of a conditionally risk-free asset and $v = \mathrm{E}[v_t]$.

1.4 The optimally scaled variance bound

The approach in this article is different. Consider the family of infinitely many one-dimensional scaled payoff spaces $P_z = \{\alpha z_t' r_{t+1} : \alpha \in R^1\}$ indexed by z_t. There is an HJ bound $\sigma^2(v; r_{t+1})$ associated with each scaling vector z_t, which only depends on the unconditional moments of $z_t' r_{t+1}$,

$$\sigma^2(v;\, z_t' r_{t+1}) = \frac{(\mathrm{E}[z_t' p_t] - v\mathrm{E}[z_t' r_{t+1}])^2}{\mathrm{var}(z_t' r_{t+1})}. \tag{19}$$

Equation (19) simply applies Equation (7) to the single scaled return $z_t' r_{t+1}$. The optimally scaled bound is the highest such HJ bound:

$$\sigma^2_{OSB}(v;\, z_t' r_{t+1}) = \sup_{z_t} \sigma^2(v;\, z_t' r_{t+1}). \tag{20}$$

The question we answer is: What z_t yields the best (largest) HJ bound? Since $z_t = f(y_t)$, this is a problem of variational calculus.

Proposition 1. *The solution to the maximization problem*

$$\sigma^2_{OSB}(v;\, z_t' r_{t+1}) = \sup_{z_t} \sigma^2(v;\, z_t' r_{t+1}). \tag{21}$$

is given by

$$z_t^* = (\mu_t \mu_t' + \Sigma_t)^{-1}(p_t - w\mu_t) \tag{22}$$

where

$$w = \frac{v - b}{1 - d}, \tag{23}$$

with b and d are as defined in Equations (15) *and* (16). *Furthermore, the maximum bound is given by*

$$\sigma^2_{OSB}(v;\, z_t' r_{t+1}) \equiv \sigma^2(v;\, z_t^{*\prime} r_{t+1}) \tag{24}$$

$$= \frac{a(1-d) + b^2 - 2bv + dv^2}{1 - d}, \tag{25}$$

where a is as defined as follows:

$$a = \mathrm{E}[p_t'(\mu_t \mu_t' + \Sigma_t)^{-1} p_t]. \tag{26}$$

Proof. The appendix contains a formal proof. The proof proceeds in two steps. First, the optimal functional form is solved for. Second, the remaining constant parameter characterizing the function is solved for in a separate maximization.

Not surprisingly, the optimal scaling factor depends on the conditional distribution function only through the first and second conditional moments. Whereas the optimal scaling factor is decreasing in the conditional variance Σ_t, it is not monotonic in the conditional mean μ_t. The nonmonotonicity is easy to understand using the duality with the mean-variance frontier. Consider two independent risky assets with a different expected return but identical variance. In this case, the minimum variance portfolio is the equally weighted portfolio. Also, the inefficient part of the frontier goes through a point where the expected return is the return on the lowest yielding asset and all funds are invested in that asset. When, without loss of generality, the expected return on the best-yielding asset is raised, the minimum variance point is raised as well, but the inefficient part of the frontier still intersects the point where all is invested in the lowest-yielding asset. The part of the new frontier beyond that point is below the old frontier.

Both bounds $\sigma^2_{GHT}(v)$ and $\sigma^2_{OSB}(v)$ depend on the conditional mean and the conditional variance of the payoffs. When these moments are known

to researchers, the relation between $\sigma^2_{GHT}(v)$ and and $\sigma^2_{OSB}(v)$ is described by the following proposition:

Proposition 2. *For an n-dimensional payoff r_{t+1} with price vector p_t, conditional mean μ_t, and conditional variance-covariance matrix Σ_t,*

$$\sigma^2_{OSB}(v) = \sigma^2_{GHT}(v) = \frac{a(1-d) + b^2 - 2bv + dv^2}{1-d}, \tag{27}$$

where b, d, and a are defined in Equations (15), (16), *and* (26).

Proof. Since $P_z \in P$ (the GHT bound represents the most efficient way of using conditional information), it follows that

$$\sigma^2(v; z_t' r_{t+1}) \leq \sup_z \sigma^2(v; z_t' r_{t+1}) = \sigma^2_{OSB}(v) \leq \sigma^2_{GHT}(v). \tag{28}$$

From Proposition 1, we know that $\sigma^2(v; z_t^{*\prime} r_{t+1})$ has the form described in the proposition. Now consider the variance of m^*_{t+1}:

$$\sigma^2_{GHT}(v) = \mathrm{var}(m^*_{t+1}) = \mathrm{E}[(z_t^{*\prime} r_{t+1})^2] - (\mathrm{E}[z_t^{*\prime} r_{t+1}])^2. \tag{29}$$

Using the expression for z^*_t, the law of iterated expectations, and simplifying algebra, it follows that

$$\begin{aligned}\sigma^2_{GHT}(v) = \mathrm{var}(m^*_{t+1}) = {} & \mathrm{E}[(p_t - w\mu_t)'(\mu_t\mu_t' + \Sigma_t)^{-1}(p_t - w\mu_t)] \\ & - \mathrm{E}[(p_t - w\mu_t)'(\mu_t\mu_t' + \Sigma_t)^{-1}\mu_t].\end{aligned} \tag{30}$$

Using the definition for a, b, and d, the result follows.

This result is at first surprising. Our optimally scaled bound is a standard HJ bound for a scaled return. Since the scaling factor depends on v, the mean of the pricing kernel, the optimally scaled bound is the ratio of a quartic polynomial in v over a quadratic polynomial in v which is generally not a quadratic polynomial in v. Nevertheless, when evaluated at the true conditional moments, the quartic polynomial in the numerator becomes the square of the quadratic polynomial in the denominator, and the optimally scaled bound becomes quadratic in v. The optimal scaled frontier becomes a parabola, identical to the GHT frontier. Since this insight is useful later on, we prove it explicitly.

Corollary 1. *Let*

$$\sigma^2_{OSB}(v) = \sigma^2(v; z_t^{*\prime} r_{t+1}) = \frac{A^2}{B}, \tag{31}$$

with

$$A = E[z_t^{*\prime} p_t] - vE[z_t^{*\prime} r_{t+1}]$$

and

$$B = \mathrm{var}(z_t^{*\prime} r_{t+1}).$$

If the conditional moments are known, then $A = B$.

Proof. First note that $B = \sigma^2_{GHT}(v)$, the GHT bound. Thus from Proposition 2, we know that

$$B = \frac{a(1-d) + b^2 - 2bv + dv^2}{1-d}.$$

But A is given by

$$\begin{aligned} \mathrm{E}[z_t^{*\prime} p_t] - v\mathrm{E}[z_t^{*\prime} r_{t+1}] &= \mathrm{E}[(p_t - w\mu_t)'(\mu_t\mu_t' + \Sigma_t)^{-1}(p_t - w\mu_t)] \\ &= a - w(b - dv) - bv. \end{aligned} \tag{32}$$

Substituting for $w = \frac{v-b}{1-b}$ and collecting terms, the result follows. This corollary provides the basis for a diagnostic test in Section 2.3.

2. The Optimally Scaled Bound: Discussion

Three important properties make the optimally scaled bound very useful in applied work. First, if there is time variation in expected returns and volatility, the optimally scaled bound should be sharper than standard ad hoc bounds. In Section 2.1 we explore the relation between predictability and the optimally scaled bound. Second, Section 2.2 discusses how the optimally scaled bound is robust in that it always is a valid lower bound to the pricing kernel, which is not the case for the GHT bound. Third, Section 2.3 suggests how the optimally scaled bound could form the basis of a diagnostic test for the correct specification of the first and second moments. Finally, we discuss how our work relates to two recent articles by Ferson and Siegel (2001, 2003).

2.1 Efficiency and predictability

Whereas the optimally scaled bound uses conditioning information efficiently, it would be useful to derive conditions under which scaling improves the bound. In particular, one would hope that predictable variation in returns would result in sharper HJ bounds. Unfortunately it is difficult to derive sufficient conditions, but it is straightforward to derive a necessary condition. Let us, without loss of generality, focus on a univariate return space. If the scaling factor z_t is uncorrelated with the first and second conditional moment of r_{t+1} (i.e., $\mathrm{cov}(p_t, z_t) = \mathrm{cov}(r_{t+1}, z_t) = \mathrm{cov}(r_{t+1}^2, z_t^2) = 0$), then scaling the return with z_t will

decrease the HJ bound. To see this, note that

$$\begin{aligned}\sigma^2(v;zr) &= \frac{\mathrm{E}^2(z)(\mathrm{E}(p)-v\mathrm{E}(r))^2}{\mathrm{E}(z^2)\mathrm{E}(r^2)-\mathrm{E}^2(z)\mathrm{E}^2(r)} \\ &= \frac{(\mathrm{E}(p)-v\mathrm{E}(r))^2}{\mathrm{E}(r^2)-\mathrm{E}^2(r)} \times \frac{\mathrm{E}^2(z)(\mathrm{E}(r^2)-\mathrm{E}^2(r))}{\mathrm{E}(z^2)\mathrm{E}(r^2)-\mathrm{E}^2(z)\mathrm{E}^2(r)} \\ &= \sigma^2(v;r) \times \frac{\mathrm{E}(r^2)-\mathrm{E}^2(r)}{(\mathrm{E}(z^2)/\mathrm{E}^2(z))\mathrm{E}(r^2)-\mathrm{E}^2(r)} \leq \sigma^2(v;r),\end{aligned}$$

where we omitted the time subscripts. The last inequality follows since $\frac{\mathrm{E}[z_t^2]}{\mathrm{E}^2[z_t]} \geq 1$. Intuitively, scaling by an independent random variable just adds noise to the return. Conversely the scaling factor has to be correlated with the future return for the scaled HJ bound to improve relative to the standard bound. In other words, when the return is scaled with a conditioning variable (e.g., a stock return with its lagged dividend yield), the variable must predict the return in order for the HJ bound to improve. This is intuitively clear: when a variable predicts an asset return, it may be possible to create managed portfolios that improve the risk-return trade-off as measured by the Sharpe ratio, and it is well-known that HJ bounds and Sharpe ratios are closely related.

This intuition remains intact for the case where two-dimensional spaces of the form

$$\begin{pmatrix} r_{t+1} \\ z_t' r_{t+1} \end{pmatrix}, \tag{33}$$

where $z_t = Gy_t$, are considered. In this case, since $r_{t+1} \in \{r_{t+1}, z_t' r_{t+1}\}$, we know for sure that

$$\sigma^2(v;r_{t+1}) \leq \sigma^2(v;(r_{t+1}, z_t' r_{t+1})), \quad \forall z_t. \tag{34}$$

Even in this case, for the bound to strictly improve, predictable variation in the conditional mean or variance is a necessary condition. To see this, first note that the optimal scaling factor remains the same for this "stacked" return and scaled return case, which we show in the next proposition.

Proposition 3. *Suppose there is an asset vector with payoff r_{t+1}, price p_t. Let I_t denote the σ algebra of the measurable functions of the conditioning variables y_t. Then the solution z_t^* to the maximization problem*

$$\sup_z \sigma^2(v;(r_{t+1}, z_t' r_{t+1})) \tag{35}$$

is given by

$$z_t^* = (\mu_t \mu_t' + \Sigma_t)^{-1}(p_t - w\mu_t). \tag{36}$$

The proof is given in the appendix.

Now suppose μ_t and σ_t are constants (i.e., there is no predictable variation in conditional means or variances), then z_t^* is a constant and r_{t+1} and $z_t^{*\prime} r_{t+1}$ are linearly dependent. It follows that $\sigma^2(v;(r_{t+1}, z_t^{*\prime} r_{t+1})) = \sigma^2(v; r_{t+1})$. But since our bound is optimal, this implies $\sigma^2(v;(r_{t+1}, z_t' r_{t+1})) \leq \sigma^2(v; r_{t+1})$. Conversely, for the bound to improve, z_t must predict r_{t+1}. In the empirical illustrations below, we will use standard scaling in the "stacked" space as indicated above. Apart from our optimally scaled bounds, we will also report "stacked" optimally scaled bounds, $\sigma^2(v;(r_{t+1}, z_t^{*\prime} r_{t+1}))$, which ought to be identical to the optimally scaled bounds when the conditional moments are known.

Our work here is related to Kirby (1998), which is the only article we are aware of that provides an explicit link between linear predictability and HJ bounds. More specifically, he shows that the Wald test of the null of no predictability in a linear regression is proportional to the standard HJ measure. He then uses this insight to investigate whether several asset pricing models are consistent with the evidence on predictability. Our work suggests that if the predictability is correctly described by a linear predictive model, our optimally scaled return should lead to a sharper HJ bound, and hence sharper restrictions on these asset pricing models. Furthermore, our framework can also accommodate nonlinear predictive relations.

2.2 Robustness

The GHT bound is given by $\text{var}(m_{t+1}^*)$, where m_{t+1} depends on the conditional mean μ_t and the conditional variance Σ_t of the returns. In practice, these conditional moments are not known. We use a proxy for them and thus a proxy $\hat{m}_{t+1}^*$ for m_{t+1}^*. In that case, the proxy for the GHT bound, $\text{var}(\hat{m}_{t+1}^*)$, may either underestimate or overestimate $\text{var}(m_{t+1}^*)$. When it overestimates, $\text{var}(\hat{m}_{t+1}^*)$ fails to be a lower bound for the variance of valid pricing kernels. On the other hand, the optimally scaled bound is $\sigma^2(v, z_t^{*\prime} r_{t+1})$, where z_t depends on the first two conditional moments. When the conditional moments are unknown, z_t^* is unknown and so is $z_t^{*\prime} r_{t+1}$. However, for every z_t, $\sigma^2(v, z_t' r_{t+1})$, remains a lower bound to the variance of all pricing kernels, since $\sigma^2(v, z_t' r_{t+1})$ is an HJ bound. Hence, even when using a proxy for the conditional moments to get a proxy $\hat{z}_t^*$ for z_t^*, the resultant optimally scaled bound remains a valid lower bound to the variance of pricing kernels.

This robustness property is important since conditional moments are notoriously difficult to estimate from the data. GHT (1990) propose to use the seminonparametric (SNP) method to estimate conditional moments. The SNP method approximates the conditional density using a Hermite expansion, where a standardized Gaussian density is multiplied with a squared polynomial. In their preferred model, the leading term is a linear

vector-autoregressive (VAR) model with autoregressive conditionally heteroscedastic (ARCH) volatility. In GHT's application on stock and bond returns, the conditioning set is restricted to contain only past returns, and SNP estimation may be adequate. However, when the data-generating process for returns contains jumps or regime switches and involves other predictive variables, such as dividend yields or term spreads, it is not clear that the SNP approach provides a good approximation.[5] The risk of overestimating the variance bound can be avoided by applying our method.

Given an empirical specification for the conditional moments, our "optimally" scaled bound is as easy to implement as the original HJ bounds, since we only need to compute unconditional moments. For example, if we deem the time variation in the conditional mean to be more important than the time variation in the conditional variance, we obtain valid bounds by just replacing the conditional variance with the unconditional variance. The resulting bound will not be optimal if there truly is time variation in the conditional variance. However, it may still be sharper than using arbitrary scaling.

2.3 Diagnostics

The fact that optimally scaled bounds computed from misspecified conditional moments remain valid bounds that are best when the true conditional moments are used suggests an interesting application of our procedure. We can use the optimally scaled bound to diagnose the accuracy of competing models for the first two conditional moments. There are several ways in which misspecification of the conditional moments may manifest itself. First, it need not be the case that $\sigma^2(v; r_{t+1}) \leq \sigma^2(v; z_t^{*\prime} r_{t+1})$. Hence, misspecified conditional moments may reveal themselves through poorly performing optimally scaled bounds relative to the conditional, "naively" scaled or stacked optimally scaled bounds.

Second, and most strikingly, the HJ bound need not be a parabola, since its numerator is a quartic in v and its denominator a quadratic in v. That is, misspecification should be visibly clear from graphing the optimal bound and we will illustrate this behavior in the empirical section below.

This reasoning also makes it possible to develop a general diagnostic test for the first and second conditional moments of asset returns.[6] To develop such a test, let's revisit Corollary 1 in Section 1.4. The optimally scaled bound can be written as $\frac{A^2}{B}$, where B is the GHT bound, and correct moment specification implies $A = B$. This suggests a simple diagnostic test.

[5] More and more research reveals that some of the predictable patterns detected in returns, even in linear settings, may be spurious, [e.g., see Kirby (1997)].

[6] We thank the referee for stimulating our thinking on this issue. The test in Kirby (1998) diagnoses the performance of several asset pricing models with respect to linear predictability but does not accommodate heteroscedasticity.

The GHT bound is a quadratic in v where the coefficients are nonlinear functions of the three unconditional moments a, b, and d, defined above. For the parabola A to coincide with B for all v's, it should be the case that its coefficients are equal to the coefficients in B. Rewrite $A = \mathrm{E}[z_t^{*\prime} p_t] - v\mathrm{E}[z_t^{*\prime} r_{t+1}]$ as $\mathrm{E}[f_{1t} + f_{2t}v + f_{3t}v^2]$ and denote the estimated constants a, b, and d by $\hat{a}$, $\hat{b}$, and $\hat{d}$. It is straightforward to derive

$$
\begin{aligned}
f_{1t} &= p_t'(\mu_t\mu_t' + \Sigma_t)^{-1} p_t + \frac{b}{1-d}\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} p_t, \\
f_{2t} &= -\frac{\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} p_t}{1-d} - p_t'(\mu_t\mu_t' + \Sigma_t)^{-1} r_{t+1} \\
&\quad - \frac{b}{1-d}\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} r_{t+1}, \\
f_{3t} &= -\frac{\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} r_{t+1}}{1-d}.
\end{aligned}
$$

To test the equality of A and B, we use the following orthogonality conditions:

$$
g_t = \begin{bmatrix}
p_t'(\mu_t\mu_t' + \Sigma_t)^{-1} p_t - a \\
\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} p_t - b \\
\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} \mu_t - d \\
p_t'(\mu_t\mu_t' + \Sigma_t)^{-1} r_{t+1}(1-d) + (\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} r_{t+1})b - b \\
\mu_t'(\mu_t\mu_t' + \Sigma_t)^{-1} r_{t+1} - d
\end{bmatrix}, \qquad (37)
$$

where the first three conditions estimate and define the fundamental constants, the fourth condition is a rewrite of $\mathrm{E}[f_{2t}] = -2b$, including a rescaling by $1 - \hat{d}$ that ensures that all orthogonality conditions are of a similar order of magnitude and the fifth condition is the rescaled version of $\mathrm{E}[f_{3t}] = \hat{d}/(1-\hat{d})$. The restriction $\mathrm{E}[f_{1t}] = \hat{a} + \hat{b}^2/(1-\hat{b})$ does not yield any conditional moments restrictions since returns do not enter this expression.

There are three parameters to be estimated, so that there are two overidentifying restrictions, which can be tested using the usual statistic $Tg_T' W g_T$, where g_T is the mean of g_t, T is the number of observations, and W is a suitable weighting matrix; for example, obtained from a Newey and West estimate (1987) of the inverse of the spectral density matrix of g_t at frequency zero. Note that whatever the dimensionality of returns, the test is always a $\chi^2(2)$ and can be used to compare the performance of nonnested models for the first and second conditional moments. Of course, in a formal application, the sampling error in the parameters

generating μ_t and Σ_t should be taking into account. In a generalized method of moments (GMM) context, this can be easily accomplished employing a sequential GMM procedure, as in Bekaert (1994), Burnside (1994), and Heaton (1995).

The economic intuition for the test is straightforward. In a standard unconditional HJ framework, the HJ bound, which is the variability of the projected pricing kernel, can be viewed as a quadratic form in the deviations from risk-neutral pricing [see Hansen and Jagannathan (1991)]. Let's consider $A: A = \mathrm{E}[z_t' p_t] - v\mathrm{E}[z_t' r_{t+1}]$. If a risk-free asset exists, then v is the inverse of the risk-free rate and A can be seen as the deviation from risk-neutral pricing for the portfolio with weights z_t (the optimally scaled portfolio), since the first term is the expected actual price and the second term is the risk-neutral price. Note that the portfolio weights do not need to add up to one. B on the other hand is simply the variability of the optimally scaled portfolio and at the same time the variability of the GHT kernel. If the scaling is done with the correct moments, the variability of the scaled return exactly equals the deviation of risk-neutral pricing.

This suggests another useful diagnostic statistic that could be used to compare alternative models. One could simply select two economically relevant v's (v_1 and v_2, say) and create a quadratic form using the following orthogonality conditions:

$$g_t = \begin{bmatrix} z_t(v_1)' p_t - v_1 z_t(v_1)' r_{t+1} - B(v_1) \\ z_t(v_2)' p_t - v_2 z_t(v_2)' r_{t+1} - B(v_2) \end{bmatrix}, \tag{38}$$

where $B(v_i)$ is the GHT bound evaluated at v_i and $z_t(v_i)$ is the optimal scaling function evaluated at v_i. This statistic ignores the sampling error in a, b, and d and the original model parameters, but can be viewed as a distance measure to rank alternative models.

To fully explore the properties of diagnostic tests based on the optimally scaled bound is beyond the scope of the present article. In our empirical illustration we report the test developed in Equation (37) for a number of different cases, including cases with simulated data where the true first and second moments model is known.

2.4 Relation to Ferson and Siegel (2001, 2003)

Ferson and Siegel have two contemporaneous articles that are related to the present article. Ferson and Siegel (2001) solve for unconditionally minimum variance portfolios while using conditioning information efficiently. They provide explicit solutions for the portfolio weights as a function of the conditional means and volatilities of the available asset returns, both when a risk-free asset exists and when it does not. Since there

is a duality between HJ frontiers and mean standard deviation frontiers, the Ferson and Siegel portfolios have some similarities to our optimally scaled returns, as Ferson and Siegel note in a final section. However, there are also many differences between our analyses so that our respective articles should really be viewed as complements rather than substitutes.

First, the HJ bounds derived from the Ferson and Siegel procedure are not as sharp as our bounds, because of the restriction that the portfolio weights have to sum to one. To appreciate the potential effect of this restriction, consider the extreme case where a researcher examines HJ bounds using one asset return (the equity return, for example) and multiple instruments (dividend yields, default spreads, and short rates, for example). One would imagine that conditioning information should be very valuable in increasing the HJ bound, but the Ferson and Siegel bound would equal the unconditional bound, since the conditioning information is useless with only one return, which forces the weight to be one for all t. Second, the optimality proof in Ferson and Siegel basically guesses the right solution for the optimal portfolio weight and verifies that it is correct, so no variational analysis is used. Third, Ferson and Siegel do not attempt to link their results to the GHT optimal HJ bound, and finally, they assume the conditional moments to be correctly specified.

Whereas our focus is mostly on the relation between GHT bounds and our optimally scaled bound, Ferson and Siegel extensively analyze the form of the weight function, providing extensive intuition on the nonlinear relation between the optimal portfolio weight and the magnitude of the expected return. In particular, extreme values for the expected return for a risky asset decrease the optimal weight on that asset, providing an interesting form of conservativeness to optimal scaling. This result applies to our bounds too, since it derives from the influence of the expected return on the uncentered second moment.

Ferson and Siegel (2003) is a purely empirical article that provides useful information about the small sample properties of alternative methods to embed conditioning information into HJ bounds. They compare the naively scaled (multiplicative) bounds, the GHT bound, and a bound based on their unconditionally efficient portfolios. Perhaps not surprisingly, all bounds suffer from significant biases that increase the bounds relative to their true values. They conclude that the parsimony of the Ferson and Siegel (2001) bounds enhances their attractiveness in small samples and that they are often close to optimal (i.e., close to the GHT bound). The analysis in Ferson and Siegel (2003) also assumes correct specification of the conditional moments. In a sense, our results strengthen their conclusions since we show that the use of optimally scaled bounds is robust to misspecification. We show that this remains true even in the presence of significant nonlinearities as generated by a regime-switching model.

3. Empirical Application

3.1 The econometric model

3.1.1 An extension of Hansen and Singleton (1983). Let R_t^i be the logarithm of the stock return ($i=s$) and the bond return ($i=b$) and let X_t be the logarithm of gross consumption growth. Define

$$Y_t = [X_t, R_t^s, R_t^b]'.$$

Hansen and Singleton (1983, henceforth HS) assume that y_t follows a VAR process with normal disturbances. HS then examine the restrictions imposed by the standard consumption-based asset pricing model with time-additive constant relative risk aversion (CRRA) preferences on the joint dynamics of the variables. A critical assumption is the time invariance of the conditional covariance matrix of y_t. It is well known that in this lognormal version of the consumption-based asset pricing model, time variation in expected excess returns is driven by the time variation in this covariance matrix. To accommodate predictability in excess returns, a natural extension of the HS framework is to allow for heteroscedasticity using the GARCH-in-mean framework of Engle, Lilien, and Robins (1987). Surprisingly, apart from an application to international data [Kaminsky and Peruga (1990)], there is little work in this area. Two reasons may be the parameter proliferation that occurs with multivariate GARCH models and the lack of heteroscedasticity in consumption growth (which may be due to a temporal aggregation bias[7]). Nevertheless, we will use this familiar framework to illustrate the properties of our "optimally scaled bound."

Our first specification has two important features. First, we impose a parsimonious factor structure on the conditional covariance matrix inspired by Engle, Ng, and Rothschild (1990). Second, we allow negative shocks to have a different effect on the conditional variance than positive shocks, that is, we accommodate asymmetric volatility as in Glosten, Jagannathan, and Runkle (1993) and Bekaert and Wu (2000). The presence of asymmetry in stock return volatility is well known, but in a previous version of this article [Bekaert and Liu (1998)] we also document asymmetry in the conditional variance of quarterly consumption growth. While it is intuitively plausible that uncertainty about future consumption growth is higher in a recession than in a boom, we could not find articles in the business cycle literature that document this phenomenon. In the finance literature, the available empirical evidence is mixed. Ferson and Merrick (1987) report U.S. consumption volatility to be higher in a nonrecession sample relative to a recession sample. Kandel and Stambaugh

[7] See Bekaert (1996) for an elaboration of this point.

(1990) find peaks in the standard deviation of U.S. consumption growth to occur at the end of recessions or immediately after them.

For the multivariate setup, we begin by parameterizing an unconstrained model:

$$Y_t = c_{t-1} + A Y_{t-1} + \Omega_{t-1} e_t \tag{39}$$

where

$$c_t = \begin{pmatrix} c_{xt} \\ c_{st} \\ c_{bt} \end{pmatrix}, \tag{40}$$

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

$$h_{iit} = \delta_i + \alpha_i h_{iit-1} + \kappa_i e^2_{iit-1} + \eta_i (\max(0, -e_{iit-1}))^2. \tag{41}$$

If $\eta_i > 0$, volatility displays the well-known asymmetric property. The e_t vector contains the fundamental shocks to the system. The error terms of the system are linked to e_t through Ω_t. A parsimonious factor structure arises by assuming that Ω_t is time invariant and upper triangular:

$$\Omega_t = \Omega = \begin{pmatrix} 1 & 0 & 0 \\ f_{xb} & 1 & 0 \\ f_{xs} & f_{bs} & 1 \end{pmatrix}. \tag{42}$$

To further limit parameter proliferation, we set $f_{bs} = 0$ and let the consumption shock be the only factor. This is consistent with the standard consumption-based asset pricing model, where consumption growth is the only state variable. In addition, we set

$$\alpha_b = \kappa_b = \eta_b = \alpha_s = \kappa_s = \eta_s = 0. \tag{43}$$

All the time variation in volatility of the Y_t system is driven by time-varying uncertainty in consumption growth. The covariance of the error terms becomes

$$\Sigma_t = \Omega H_t \Omega'. \tag{44}$$

We denote its elements by σ_{ijt} with $i, j = x, b, s$. Since the consumption-based asset pricing model introduces elements of the conditional variance-covariance matrix in the conditional mean, the unconstrained model should allow the conditional covariance matrix to affect the conditional mean as well. Therefore we let

$$c_{it} = v_i h_{xxt} + c_i, \tag{45}$$

where i is either b or s. This simple expression for the constant arises because of the one-factor structure of the conditional covariance matrix. The parameter vector to be estimated is

$$\Theta = [\text{vec}(A)', c_x, c_b, c_s, v_b, v_s, f_{xb}, f_{xs}, \delta_x, \alpha_x, \kappa_x, \eta_x, \delta_b, \delta_s]'.$$

Hence there are a total of 22 parameters and it is clear that relaxation of some of the parameter restrictions we impose would be stretching the data too far. This unconstrained model serves as a natural alternative to the model constrained by the consumption-based asset pricing model. Let γ be the CRRA and let β be the discount factor. The model implies

$$\text{E}_t[R^i_{t+1}] = -1/2\sigma_{iit} - 1/2\gamma^2\sigma_{xxt} + \gamma\sigma_{ixt} + \gamma\text{E}_t[x_{t+1}] - \ln\beta.$$

If conditional variances are constant, the time variation in the conditional means of asset returns and consumption growth is proportional and the proportionality constant is the CRRA. The restriction also shows the role of γ as the price of risk, with the risk being the covariance with consumption. With our particular GARCH structure, the model further simplifies to

$$\text{E}_t[R^i_{t+1}] = -(\ln\beta + 1/2h_{ii}) - 1/2(\gamma - f_{xi})^2 h_{xxt} + \gamma\text{E}_t[x_{t+1}].$$

Note that h_{ii} does not depend on t for $i = b, s$ because of Equation (43). Our particular parameterization has the implication that increased uncertainty about future consumption growth always decreases expected returns. This seems at odds with the data where the price of risk has been shown to move countercyclically. The model does predict that, if shocks to returns depend positively on consumption shocks, an increased covariance with consumption will drive up expected returns. Furthermore, the covariance with consumption increases when consumption volatility increases because of the factor structure. However, this effect is swamped by the Jensen's inequality terms, which depend negatively on consumption volatility. As a result, this comparative static is not necessarily true for gross returns:

$$\text{E}_t[\exp[R^i_{t+1})] = \exp(-\ln\beta - \gamma/2(\gamma - 2f_{xi})h_{xxt} + \gamma\text{E}_t[x_{t+1}]).$$

Depending on the relative size of the sensitivity to consumption shocks, f_{xi} and the CRRA, higher consumption volatility may now increase the gross expected asset return. Empirically our unconstrained model potentially allows for a positive relation between consumption volatility and expected log returns and so we can test whether this feature of the model is a source for rejection. The restricted parameter vector Θ^R contains 14 parameters,

$$\Theta^R = [c_x, A_{11}, A_{12}, A_{13}, \beta, \gamma, f_{xb}, f_{xs}, \delta_i, \alpha_x, \kappa_x, \eta_x]', \quad i = x, b, s.$$

3.1.2 A nonlinear dynamic model. Although the above unconstrained model features nonlinearities in the volatility dynamics, the conditional mean is linear in the information variables. There are two reasons to explore nonlinear models more explicitly. First, empirical research has documented regime-switching behavior in both consumption growth and equity return data [see, Whitelaw (2000) and Ang and Bekaert (2002)]. Second, if nonlinear predictability is present, it can be easily accommodated in our optimally scaled bound, whereas naively scaled bounds are not likely to reflect it.

Consequently we formulate a regime-switching version of the unconstrained VAR model of Section 3.1.1. With S_t a discrete regime variable that can take on the values of one or two, we assume

$$X_t = \mu_x(S_t) + \phi_x(S_t) R^b_{t-1} + \sigma_x(S_t)\epsilon^x_t, \tag{46}$$

$$R^i_t = \mu_i(S_t) + \phi_i(S_t) R^b_{t-1} + b_i \sigma_x(S_t)\epsilon^x_t + \sigma_i \epsilon^i_t, \tag{47}$$

where $i = b, s$. Note that to avoid parameter proliferation, we constrained the conditional mean dynamics to only depend on the past bond return, but not on past consumption growth or stock returns. The correlation between bond and stock returns in this model stems either from conditional mean dynamics or from their joint dependence on consumption shocks. The S_t variable follows a Markov chain with either constant transition probabilities or transition probabilities that depend on the past bond return. Ang and Bekaert (2002) find evidence of nonlinear predictability in monthly equity returns using the short rate as an instrument:

$$P = \Pr(S_t = 1 | S_{t-1} = 1, I_{t-1}) = \frac{\exp(a_1 + d_1 R^b_{t-1})}{1 + \exp(a_1 + d_1 R^b_{t-1})}, \tag{48}$$

$$Q = \Pr(S_t = 2 | S_{t-1} = 2, I_{t-1}) = \frac{\exp(a_2 + d_2 R^b_{t-1})}{1 + \exp(a_2 + d_2 R^b_{t-1})}. \tag{49}$$

The parameter vector for this system contains 20 elements, with two additional parameters for the case of time-varying transition probabilities:

$$\Theta^{RS} = [\mu_i(j), \phi_i(j), \sigma_x(j), \sigma_k, b_k, a_j, d_j],$$

with $j = 1, 2$ (denoting regime dependence), $i = x, s, b$, and $k = s, b$. It is straightforward to derive a representation of this model that imposes the restrictions of the consumption-based asset pricing model, as we did for the model in Section 3.1.1. This requires imposing the restrictions "within" each regime. For reasons explained in the next section, we do not use this model in the empirical illustration of the bounds.

3.2 Data and estimation results

Our consumption measure is the sum of per capita real nondurables and services consumption in the United States. These data were downloaded from DATASTREAM. The stock return is the quarterly value-weighted dividend-inclusive index return on the New York Stock Exchange (NYSE), taken from Wharton's website (http://wrdsx.wharton.upenn.edu). The interest rate is the U.S. three-month Treasury-bill rate taken from the Federal Reserve website. We use a dataset on weekly secondary market rates (averages of daily) and use the rate closest to the end of the month. All data run from the second quarter of 1959 to the end of 1996.

Table 1 shows the results from the unconstrained estimation. Despite the presence of very large coefficients on the GARCH-in-mean term, consumption growth and bond returns show strong autocorrelation as they do univariately. Although the standard errors for the GARCH-in-mean coefficients seem very small, they should be interpreted with much caution. Standard errors computed from the cross-product of the first derivatives of the likelihood are quite large and more adequately represent the uncertainty regarding these parameter estimates. In fact, the likelihood function is very flat with respect to these parameters, and a number of locals exist where the GARCH-in-mean parameters are in fact positive. This is not that surprising. Much work on GARCH-in-mean models for stock returns [see Bekaert and Wu (2000) for a survey] has stressed the weakness of a positive relation between stock return volatility and its conditional mean. In this model, stock and bond returns are linked to consumption volatility, which in turn drives asset return volatility. The much smaller magnitude of consumption volatility relative to stock return

Table 1
Unconstrained GARCH-in-mean model

Equations	Coefficients			
	Constant	X_{t-1}	R^b_{t-1}	R^s_{t-1}
X_t	0.0030	0.361	−0.029	0.008
	(0.0005)	(0.033)	(0.022)	(0.005)
R^b_t	0.0056–162.65h_{xxt}	−0.198	0.738	−0.0002
	(0.0006)(0.0001)	(0.031)	(0.037)	(0.0043)
R^s_t	0.0188–58.02h_{xxt}	−1.734	1.029	0.077
	(0.0083)(0.0003)	(0.005)	(0.014)	(0.034)
	Constant	α_i	κ_i	η_i
h_{11t}	0.000019	−0.0265	0.0008	0.2705
	(0.000018)	(0.0807)	(0.7898)	(0.0426)
h_{22t}	0.000014	0	0	0
	(0.000002)			
h_{33t}	0.006134	0	0	0
	(0.00103)			
	$f_{xb} = -0.0564$		$f_{xs} = 3.182$	
	(0.1425)		(0.003)	

The model estimated is described in Equations (39)–(45). Standard errors are in parentheses and are robust to misspecification of the error distribution in the sense of White (1982). Parameter values without standard errors reflect constrained parameters.

volatility explains the large coefficients we find relative to the GARCH-in-mean literature for stock returns. When we estimate a univariate GARCH-in-mean model for stock returns, we find a GARCH-in-mean parameter of 6.29 with a large standard error of 5.23. Note that there is virtually no GARCH in the volatility dynamics, but strong asymmetry, with the coefficient on positive shocks being slightly negative. This is somewhat problematic since the conditional variance may theoretically become negative, although it never does in-sample.

Not surprisingly, the constrained model (see Table 2) is rejected by a likelihood ratio test. The chi-square test statistic is 75.32 with a p-value of 0.000 (there are eight restrictions). The CRRA is estimated to be 14.675 and the discount factor β is 1.071. Although the latter is greater than one, we know from Kocherlakota's (1996) work that the economy remains well defined and, in fact, our parameter values are quite close to the ones he uses to explain the equity premium puzzle. The estimation results reveal that the key parameter the model attempts to match is the autoregressive coefficient in the bond equation, which is almost perfectly matched. Given the proportionality restrictions imposed by the model on expected returns, this causes a bad fit for both stock returns and especially consumption dynamics. Because the GARCH-in-mean parameters are pretty similar, and are imprecisely estimated, it is very likely that the model rejection is driven by this phenomenon.

Table 3 contains the estimation results for the regime-switching model. The construction of the likelihood function for such models is by now

Table 2
Constrained GARCH-in-mean model

Equations	Coefficients			
	Constant	X_{t-1}	R^b_{t-1}	R^s_{t-1}
X_t	0.005	−0.018	0.050	0.0001
	(0.0005)	(0.005)	(0.005)	(0.0003)
R^b_t	0.0053–108.97h_{xxt}	−0.264	0.734	0.0012
R^s_t	0.0021–82.086h_{xxt}	−0.264	0.734	0.0012
	$\gamma =$	14.675	$\beta =$	1.071
		(0.0376)		(0.0082)
	Constant	α_i	κ_i	η_i
h_{11t}	0.000022	−0.0652	0.00	0.3907
	(0.000006)	(0.0208)	(0.00)	(0.0876)
h_{22t}	0.000013	0	0	0
	(0.000002)			
h_{33t}	0.006457	0	0	0
	(0.001009)			
	$f_{xb} = -0.0877$		$f_{xs} = 1.847$	
	(0.0813)		(0.0872)	

The model estimated imposes the following constraint on the unconstrained model reported in Table 1;

$$E_t[R^i_{t+1}] = -\left(\log\beta + \frac{1}{2}h_{ii}\right) - \frac{1}{2}[\gamma - f_{xi}]^2 h_{xxt} + \gamma E_t[X^i_{t+1}].$$

The table reports all parameters, including parameters constrained by the model. Robust standard errors are in parentheses.

Table 3
Regime-switching model

Equations	Coefficients						
	Constant		R^b_{t-1}		σ_i		b_i
	$S_t=1$	$S_t=2$	$S_t=1$	$S_t=2$	$S_t=1$	$S_t=2$	
X_t	0.0055	−0.0103	0.005	0.813	0.0042	0.0042	
	(0.0005)	(0.0020)	(0.064)	(0.306)	(0.0017)	(0.0003)	
R^b_t	0.0009	0.0087	0.802	−0.721	0.0034		−0.0044
	(0.0004)	(0.0018)	(0.051)	(0.269)	(0.0002)		(0.0663)
R^s_t	0.0071	0.0332	1.31	1.11	0.0773		4.62
	(0.0084)	(0.0385)	(1.19)	(5.70)	(0.0045)		(1.52)

The model estimated is described in Equations (46)–(49). Standard errors are in parentheses and are robust to misspecification of the error distribution in the sense of White (1982).

standard [see, for instance, Hamilton (1994)]. Identifying a global maximum in a regime-switching model is difficult and we followed an elaborate procedure to ensure that we indeed identified the global maximum. We first used 20 different sets of starting values, covering the parameter space as widely as possible. We identified a number of local maxima and then, for each local maximum, ran three to four estimations with starting values randomized around the converged local optimum parameter values. Finally, we ran some 20 estimations with starting values randomized around the global maximum converged values. The global maximum we report in Table 3 has been confirmed more than 20 times in our different estimation experiments.

Whereas the identification of regimes its typically driven by differences in volatilities across regimes, we find the volatility of consumption shocks to be very similar in the two regimes. However, regime 1 is a regime with overall high consumption growth, whereas consumption growth in regime 2 is often negative, although it depends positively and significantly on the bond return. In this recession regime, expected asset returns are high, consistent with the conventional wisdom. In the recession regime, bond returns are negatively serially correlated. Bond returns have a small, insignificant consumption beta, whereas stock returns have a large, positive, and statistically significant consumption beta. We verified by simulation that this model matches the first and second moments of the data very well.

When we allow the transition probabilities to depend on the past bond return, we find $d_1 = 0.3785$ with a standard error of 1.0048 and $d_2 = 13.08$ with a standard error of 8.65. A likelihood ratio test rejects the restriction of constant transition probabilities at the 5% level.

For the sake of completeness, we should mention that we estimated two other regime-switching models that we decided not to use in the empirical illustration. First, we estimated the model in Table 3, subject to a set of restrictions imposed by the consumption-based asset pricing model with CRRA preferences. This model was strongly rejected by the data and the

CRRA coefficient was negative. Hansen and Singleton (1983) also found convex utility for some of their specifications. Furthermore, we estimated a model using only data on bond and stock returns. By freeing the regime variable from having to fit regimes in consumption growth, we conjectured that this model would provide a better fit with the return data. However, this was not the case: the joint model estimated in Table 3 provides sharper HJ bounds!

3.3 The HJ bounds

This section illustrates the performance of our optimally scaled bound along the three dimensions that we discussed in Section 2: efficiency, robustness, and diagnostics.

We have four candidates for the computation of the conditional moments we need in deriving the optimally scaled and GHT bounds: the VAR model for stock and bond returns and consumption growth in its unconstrained and constrained form, and the regime-switching model with constant and time-varying transition probabilities. We will also use these models as data-generating processes in simulations. Simulations both serve to illustrate the effect of misspecifications where the conditional moments are known and to help interpret data results that may be sensitive to sampling error in our short sample. Simulations use 10,000 observations.[8] Table 4 provides a complete guide to the figures we produce. Of importance is that we always focus on both stock returns, and bond returns, and naive scaling uses the past bond and stock returns as instruments for both returns.

3.3.1 Efficiency. Figure 1 uses the unconstrained VAR model and the two regime-switching models to compute the conditional moments in the optimally scaled bounds. Also on the graph are the unconditional and naively scaled HJ bounds. Three results stand out. First of all, the difference between the unconditional and scaled bounds reveals considerable predictability. The main source of the predictability is the autoregressive component in bond returns.

Second, the difference between the various scaled bounds is small, but the arbitrarily scaled bound is even somewhat sharper for small v's than is the optimally scaled bound computed from the unconstrained VAR. This can be due to either misspecification of the conditional moments or chance (sampling error). To examine this issue more closely, we first produce the same graphs for a long simulated sample from the unconstrained model in Figure 2. We also show the GHT bound. As should be

[8] We simulate 10,100 observations but discard the first 100 observations to reduce dependence on initial conditions. Such dependence is unavoidable in the graphs using short sample data. Our sample estimates of the HJ bounds may also be subject to the finite sample bias documented in Ferson and Siegel (2003), but the number of asset returns we use is much smaller than theirs.

Table 4
Guide to figures

Figure No.	Data-generating process	Model for (μ_t, Σ_t)	Illustration
1	Data	UC VAR, CP RS, TP RS	Efficiency
2	Simulation, UC VAR	UC VAR	Efficiency
3	Simulation, TP RS	TP RS	Efficiency
4	Simulation, TP RS	TP RS	Efficiency
5	Data	CO VAR	Diagnostic
6	Simulation, CO VAR	CO VAR	Diagnostic
7	Simulation, TP RS	all	Diagnostic/Robustness
8	Simulation, TP RS	all	Diagnostic/Robustness
9	Simulation, CO VAR	all	Diagnostic/Robustness
10	Simulation, TP RS	all	Robustness
11	Simulation, CO VAR	all	Robustness

The data-generating process column records the origin of the data in the construction of the optimally scaled and GHT bounds: actual data ("data"), simulated data from one of the models: UC VAR (unconstrained VAR), CO VAR (constrained VAR), CP RS (regime-switching model with constant transition probabilities), TP RS (regime-switching model with time-varying transition probabilities). The simulated samples are of length 10,000. The model column records the model used in constructing the conditional moments. The last column identifies the property of the optimally scaled bound the figure purports to illustrate. Figures 10 and 11 graph GHT bounds; Figures 7–9 graph optimally scaled bounds. The first six figures graph a variety of bounds, including unconditional HJ bounds (denoted by HJ in the figures) and naively scaled bounds. For the scaled bounds, the instruments are the past returns for both returns.

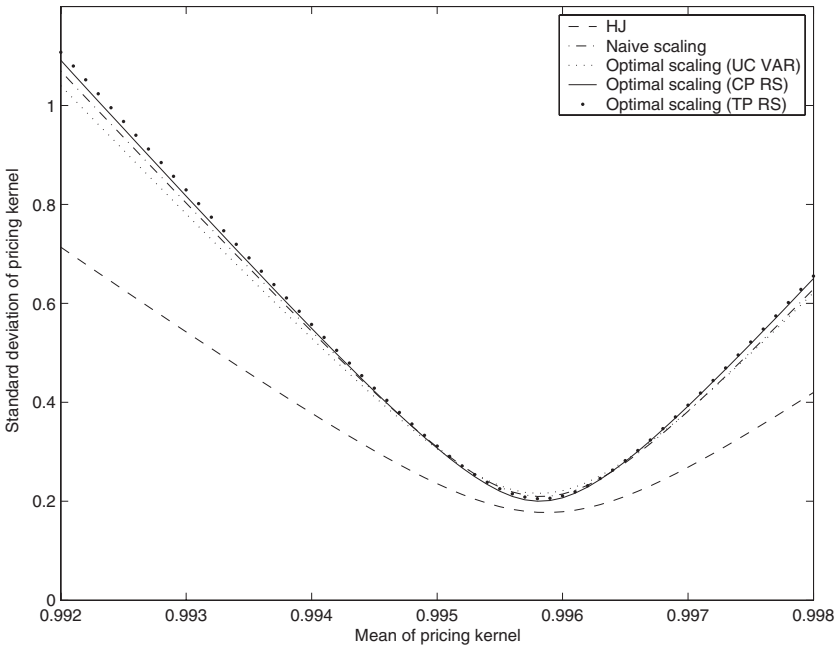

Figure 1
HJ bounds for real data with conditional moments calculated from three models.

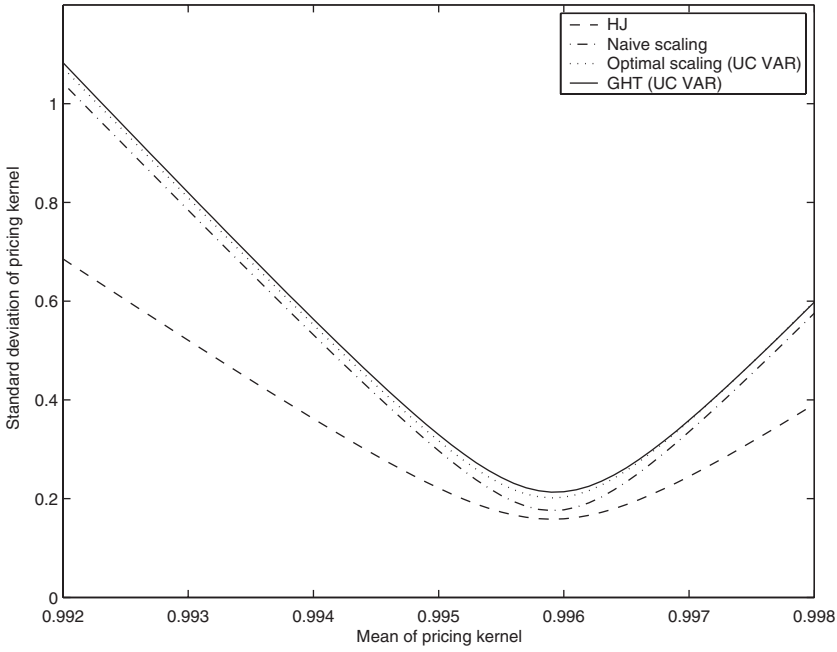

Figure 2
HJ bounds for simulated data according to the unconstrained VAR model with conditional moments calculated from the unconstrained VAR model.

the case, the GHT and optimally scaled bound are virtually on top of one another and dominate ad hoc scaling, but only slightly. In other words, in a world where the unconstrained model generates the data, naive scaling will closely approximate the efficient use of the conditioning information. In fact, since the unconstrained model describes the data rather well, the dominance of the naively scaled bound in Figure 1 may be simply due to sampling error, which we confirmed by performing simulations using 151 data points only.

It is no mystery why the use of the true conditional moments adds little in this setting. The feature of the data that arbitrary scaling would most likely fail to capture is the GARCH-in-mean feature, which happens to be weak in quarterly data. The importance of optimal scaling in generating sharper HJ bounds is likely more dramatic when strong non-linearities are present. This brings us to the third important result captured by Figure 1: the bounds generated by both regime-switching models are indeed sharper than the naively scaled bound, with the sharpest bound delivered by the most nonlinear model, the model with time-varying transition probabilities. Hence there appear to be nonlinear conditional mean effects in the data, but it is not surprising that they are not terribly strong in this

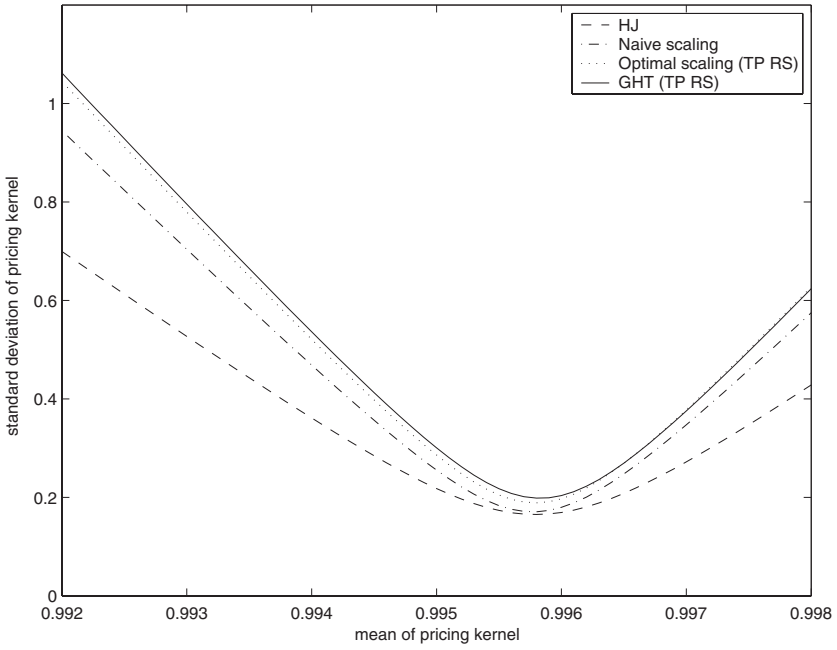

Figure 3
HJ bounds for simulated data according to the regime-switching model with time-varying transition probabilities with conditional moments calculated from the regime-switching model with time-varying transition probabilities.

quarterly dataset. To investigate this further, Figure 3 simulates data from the regime-switching model with time-varying transition probabilities and produces the unconditional, naively scaled, optimally scaled, and GHT bounds, the latter two using the true model to compute conditional moments. As expected, the optimally scaled bound and the GHT bound are practically on top of one another, but there is now a bit more of a wedge between naive and optimal scaling. To further demonstrate that naive scaling produces inefficient bounds when nonlinear predictability is present, we simulate data from a regime-switching model with stronger nonlinear predictability. In Figure 4, we use the parameter estimates from Table 3, but multiply the parameters governing the state dependence of the transition probabilities (d_1 and d_2) by 10. The wedge between the naively scaled and the efficient optimally scaled and GHT bounds now gets much larger.

3.3.2 Diagnostics. In Figure 5, which uses real data and the constrained VAR model to generate the conditional moments, two results stand out. First, the stacked optimally scaled bound gets pretty close to the naively scaled bound, despite the misspecification of the conditional moments.

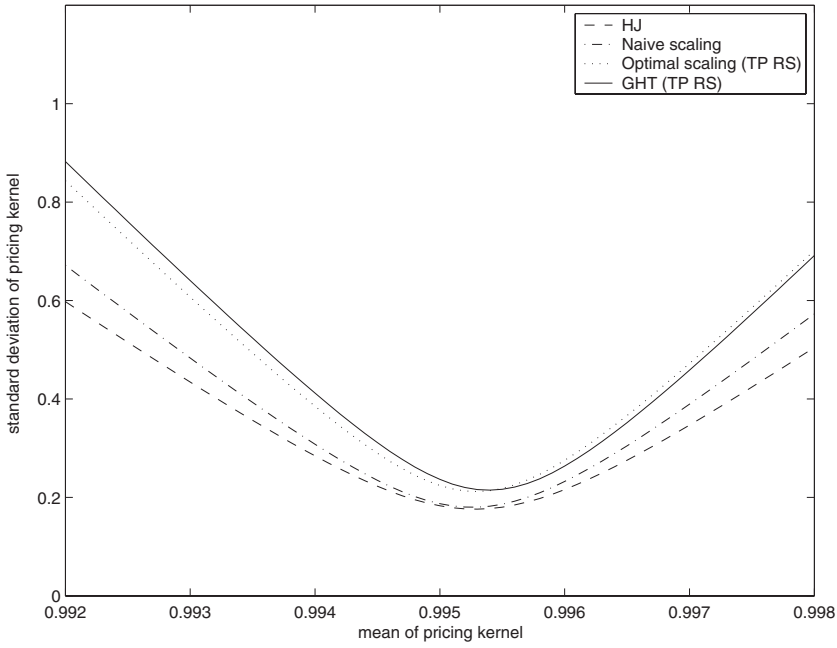

Figure 4
HJ bounds for simulated data according to the time-varying transition probabilities (TP) regime-switching model with conditional moments calculated from the TP regime-switching model (stronger predictability).

Of course, the constrained model manages to reproduce the most important aspect of the predictability, namely the autoregressive component in bond returns, so this result is not so surprising. What may strike some readers as surprising is the second main fact: the optimally scaled bound is not a parabola. As we indicated above, if the moments are correctly specified it ought to be. Since we know the model is rejected, the optimally scaled bounds seem to provide a striking alternative specification test. Of course, it is again possible that some quirk in the constrained model, coupled with sampling error, generates this result. This is not the case. Figure 6 uses data simulated from the constrained model. Since the model for conditional moments is correctly specified in this case, we now obtain a smooth parabola. We also produced these bounds for a number of simulated samples of length 151 and never found the same "strange" behavior.

To illustrate the diagnostic power of the optimally scaled bound more starkly, we can use simulations and our estimated data-generating processes to generate misspecified bounds. Figure 7 uses observations simulated from the model that best fits the data: the regime-switching model with time-varying transition probabilities. We show the optimally scaled bounds using the true model and the three other models to compute the conditional moments. In the last three cases, the moments are of course

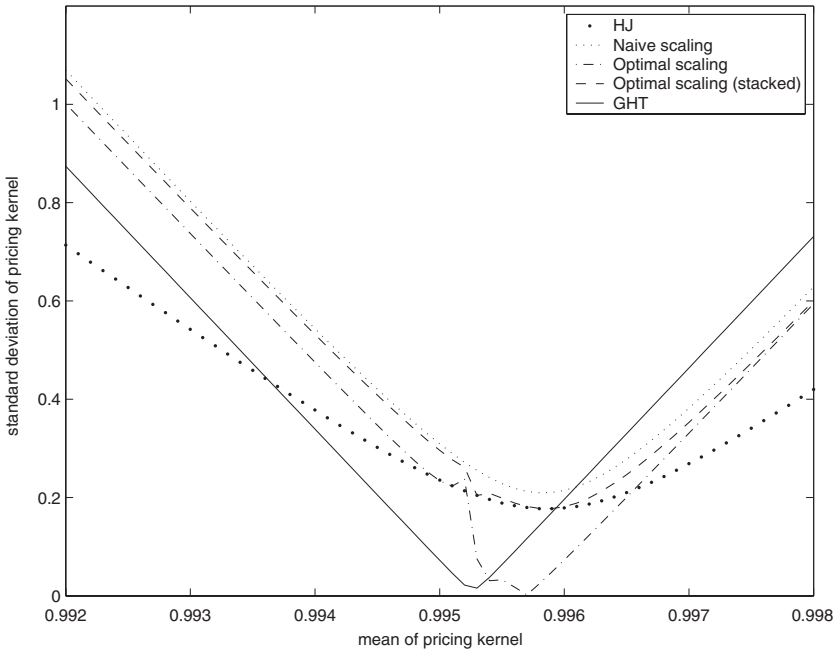

Figure 5
HJ bounds for real data with conditional moments calculated from the constrained VAR model.

misspecified. The three models not constrained by the consumption-based asset pricing model generate similar optimally scaled bounds with no obviously visible misspecification. Of course, the bound from the true model is the sharpest. However, the optimally scaled bound computed using constrained VAR moments shows striking nonparabolic behavior near the trough of the graph. The bound is also far from efficient, which may suggest that misspecification may lead to very inefficient bounds. In this rather stylized example, this is of course true by construction, but the graph also shows that misspecified models (the unconstrained VAR and the regime-switching model with constant transition probabilities) that get the dynamics "almost" right yield rather tight bounds. Moreover, in the case of misspecification, one can always do better by using the stacked optimal bound. Figure 8 illustrates this by repeating Figure 7 with optimally scaled stacked bounds. Of course, for the true model, the optimally scaled and the optimally scaled stacked bound are identical. The constrained VAR model still generates a nonparabolic optimally scaled bound, but the bound is now much closer to the true bounds.

As another illustration, Figure 9 generates data satisfying our worst model, the constrained VAR model, but computes the optimally scaled bound using moments according to the true and our three other, now

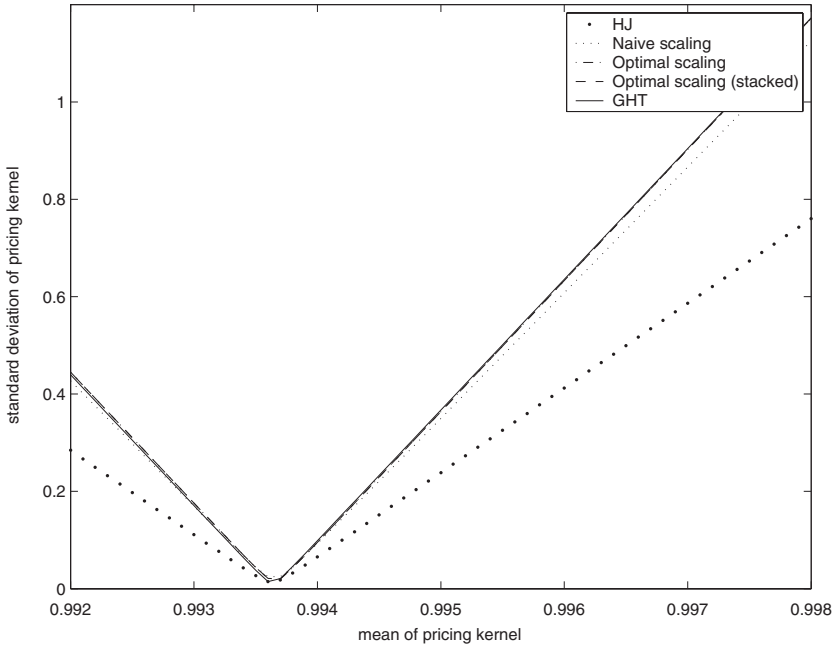

Figure 6
HJ bounds for simulated data according to the constrained VAR model with conditional moments calculated from the constrained VAR model.

misspecified models. Now, all misspecified models generate nonparabolic optimally scaled bounds.

Finally, Table 5 produces the diagnostic test of Section 2.3 [see Equation (37)], ignoring the sampling error in the original parameters, but taking the sampling error in estimating the a, b, and d constants into account. All test values are $\chi^2(2)$ and the p-values are in parentheses. We first produce the test for the data, using all four of our models to compute the conditional moments. The test rejects the constrained model, as did the likelihood ratio test, at the 1% level. However, the diagnostic test also provides a test of the first and second moment specification embedded in the other models, including the regime-switching models. Here the test fails to reject in each case with p-values of over 90%. Testing the constrained or unconstrained VAR models versus the regime-switching models would be a difficult task because of the presence of nuisance parameters under the null. Our specification test clearly shows that the models that do not impose the restrictions of the consumption-based asset pricing model provide a reasonable specification of the first and second moments, but that the constrained VAR model does not.

Our simulated samples offer a controlled environment to examine the performance of the test. We consider 16 cases simulating data from

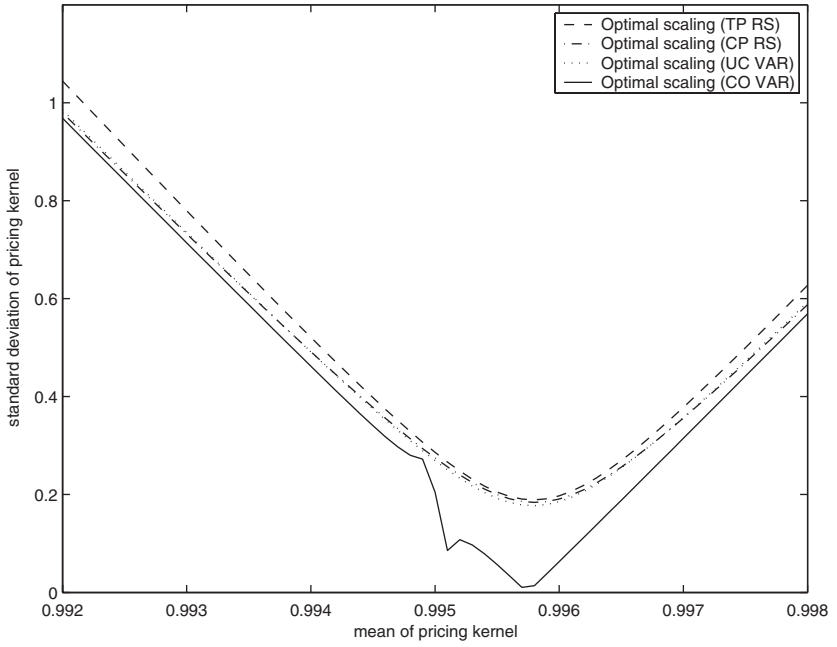

Figure 7
Optimally scaled HJ bounds for simulated data according to the TP regime-switching model with conditional moments calculated from different models.

the four different models and computing the moments according to each model (which will be misspecifying the conditional moments in three of the four cases). Given the size of the simulated samples (10,000 observations), we expect to reject when the moments are clearly misspecified, as in the constrained VAR model. The three other models all generate very similar first and second moments, but for the test to be useful (i.e., consistent) it must be able to distinguish these small differences in conditional moments and reject when given enough observations. This indeed happens. Focusing on the simulation rows and columns in Table 5, the moments are correctly specified along the diagonal, and there the test yields small, insignificant test values. For all off-diagonal elements, the moments are misspecified and the test should reject. It does in each and every case at less than the 1% significance level. For entries involving the constrained model, the test generates very large test values. We conclude that the test is well-behaved.

3.3.3 Robustness. So far we have not focused on the GHT bounds very much. Generally, optimally scaled bounds do not perform much worse or better than the GHT bound. Moreover, our simulations reveal that the GHT bounds quite often overestimate the variance of the true pricing

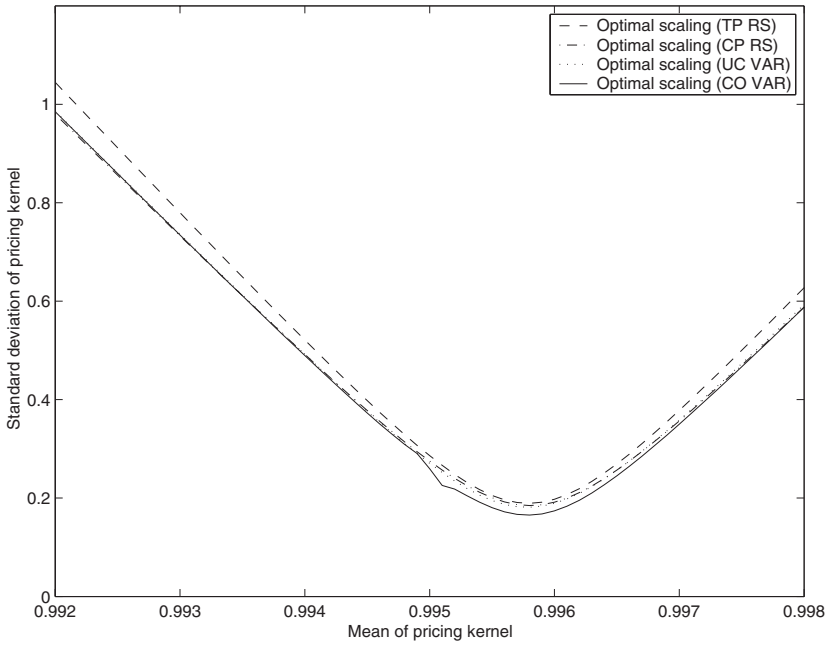

Figure 8
Optimal stacked HJ bounds for simulated data according to the TP regime-switching model with conditional moments calculated from different models.

kernel. A first example is in Figure 10. In Figure 10, we generate data from the regime-switching model with time-varying transition probabilities. When we correctly specify the conditional moments, the GHT bound mimics the behavior of the optimally scaled bound, which was graphed in Figure 7: it starts out high on the left (a little over 1) and ends up at a level of about 0.6 on the right. When we use the unconstrained VAR and the constant transition probabilities regime-switching model, the GHT bounds are not too different. Nevertheless, on the right-hand side of the graph they slightly overestimate the true variability of the pricing kernel. When we use misspecified moments from the constrained VAR model, the GHT bound generates values that are way too high for the bounds on the right-hand side. When we use the constrained VAR model to generate truth in Figure 11, a similar phenomenon appears. This time, the GHT bound overestimates to varying degrees on the left-hand side of the graph for all three misspecified models.

This lack of robustness is a serious drawback to the GHT bound. The optimally scaled bound never exceeds the true GHT bound, but manages to be quite close to it. Figures 7–9 amply demonstrate this fact. The top HJ bound in these figures corresponds to the GHT bound in either Figure 10

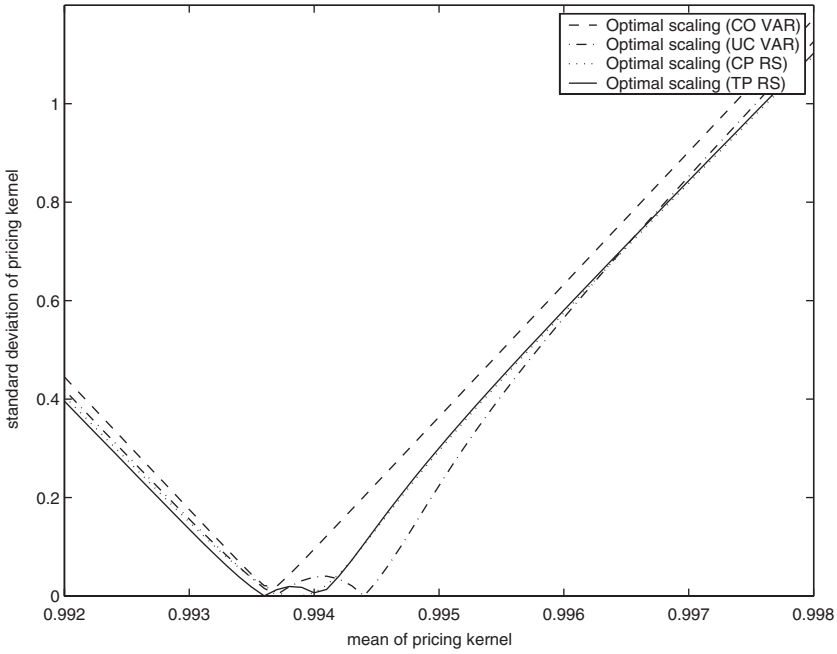

Figure 9
Optimally scaled HJ bounds for simulated data according to the constrained VAR model with conditional moments calculated from different models.

Table 5
Diagnostic test

	UC VAR	CO VAR	CP RS	TP RS
Data	0.117 (0.9433)	14.97 (0.0006)	0.295 (0.8630)	0.050 (0.9751)
Sim. (UC VAR)	1.77 (0.4128)	761.55 (0.0000)	12.98 (0.0015)	24.37 (0.0000)
Sim. (CO VAR)	510.19 (0.0000)	0.543 (0.7620)	438.56 (0.0000)	374.00 (0.0000)
Sim. (CP RS)	23.00 (0.0000)	517.73 (0.0000)	0.666 (0.7167)	46.09 (0.0000)
Sim. (TP RS)	27.34 (0.0000)	834.22 (0.0000)	9.47 (0.0088)	3.52 (0.1720)

This table produces the diagnostic test proposed in Section 2.3 for 20 different environments, depending on which model was used to construct the conditional moments and which data were used (actual data or simulated samples of 10,000 observations according to one of the models). The model nomenclature is as follows: UC VAR for unconstrained VAR, CO VAR for constrained VAR, CP RS for regime-switching model with constant transition probabilities, TP RS for regime-switching model with time-varying transition probabilities. All statistics are $\chi^2(2)$ distributed and p-values are in parentheses.

or Figure 11. Importantly, when the moments are misspecified, the optimally scaled bound always remains below the true bounds and the misspecification shows up in nonparabolic behavior of the bound. The latter is particularly apparent in Figure 7. This surprising robustness should make the optimally scaled bound the preferred method of incorporating conditioning information efficiently.

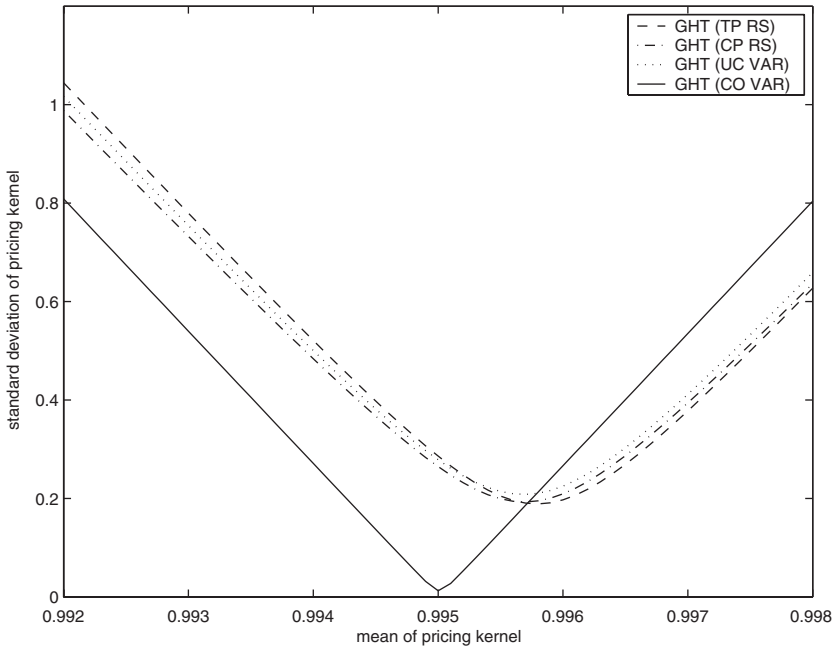

Figure 10
GHT bounds for simulated data according to the TP regime-switching model with conditional moments calculated from different models.

4. Conclusion

With the continued interest of the finance profession in the use of (unconditional) HJ bounds on the one hand, and the growing evidence of time variation in conditional means and variances of asset returns on the other hand, it becomes important to optimally incorporate conditioning information in these bounds. Our article provides a bridge between the insightful, but complex analysis of GHT (1990), and the simple, but suboptimal practice of arbitrarily scaling returns with instruments that predict them. The advantage of the latter approach is that it always produces valid bounds to the variance of the pricing kernel, whereas the GHT bound may overestimate the variance of the pricing kernel when the conditional moments are misspecified. In this article we derive the best possible scaled bound, the optimally scaled bound. As does the GHT bound, this bound requires specifying the conditional mean and variance of the returns and we show that the optimally scaled bound is as good as the GHT bound when these moments are correctly specified. When they are misspecified, our bound is robust, in the sense that it will always produce a valid bound to the variance of the pricing kernel since it is an HJ bound.

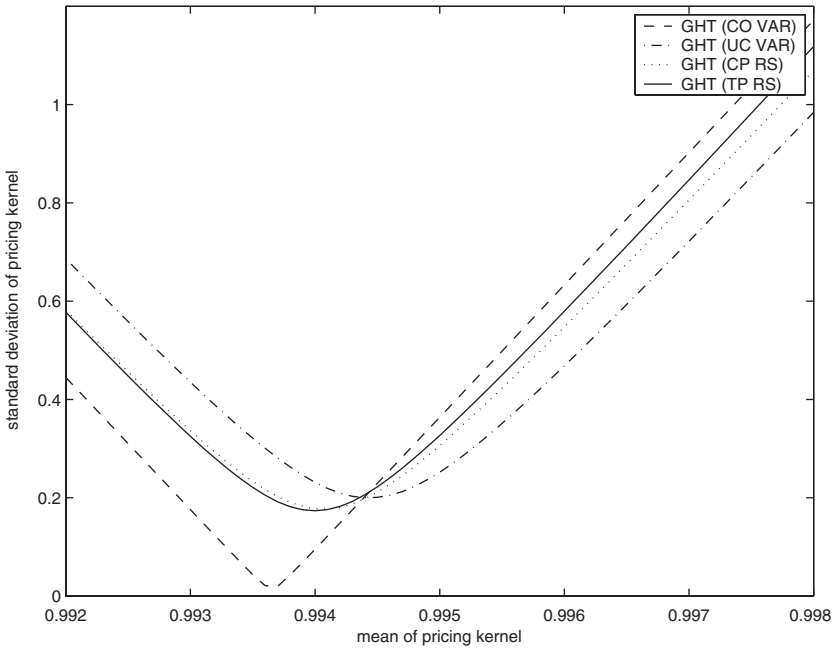

Figure 11
GHT bounds for simulated data according to the constrained VAR model with conditional moments calculated from different models.

There are potentially many interesting applications of our framework. First, the bounds can be used to reexamine the predictability of asset returns and to examine which instruments yield the sharpest restrictions on asset return dynamics. In our application here, using the optimally scaled bound does not sharpen the bounds dramatically. However, Ferson and Siegel (2003) show cases where the efficient use of conditioning information substantially increases the efficient volatility bound.

Second, the bounds can also yield information on expected return and conditional variance modeling and serve as a diagnostic tool to judge the performance of dynamic asset pricing models. The reason is that the optimal scaling function depends on the conditional mean and conditional variance of the returns and that the resulting HJ bound is best when they represent the true conditional moments. We use this property of the optimally scaled bound to develop a GMM-based specification test for the first and second moments, but much more needs to be done. We ignored the sampling error in the parameter estimates of the original models, and did not examine the small sample properties of the test.[9]

[9] See Hansen, Heaton, and Yaron (1996) for a study of the small sample properties of GMM estimators.

Third, using the duality with the mean-variance frontier, the optimally scaled bound can be used in dynamic models of optimal asset allocation that seek to maximize an unconditional mean-variance criterion. Fourth, the bounds could be used in developing performance measures for portfolio managers. In the standard mean-variance paradigm, there is no role for a portfolio manager since the optimal portfolio weights are fixed over time. In a dynamic setting, with changing conditional information, the role of the portfolio manager is to adjust the portfolio weights according to the arrival of information, preferably optimally.

Appendix

Proof of Proposition 1. The problem we would like to solve is

$$\sup_z \sigma^2(v; z_t' r_{t+1}) = \sup_z \frac{(\mathrm{E}[z_t' p_t - v z_t' r_{t+1}])^2}{\mathrm{E}[(z_t' r_{t+1})^2] - \mathrm{E}^2[z_t' r_{t+1}]}.$$

This is a well-defined problem since $\sigma^2(v; z_t' r_{t+1})$ is bounded from above by the GHT bound $\sigma^2_{GHT}(v)$ and from below by zero. Note that

$$\mathrm{E}[z_t' p_t] = \mathrm{E}[f'(y_t) p_t],$$

$$\mathrm{E}[z_t' r_{t+1}] = \mathrm{E}[f'(y_t) r_{t+1}] = \mathrm{E}[f'(y_t) \mu_t],$$

$$\mathrm{E}[(z_t' r_{t+1})^2] = \mathrm{E}[f'(y_t) \mathrm{E}_t[r_{t+1} r_{t+1}'] f(y_t)] = \mathrm{E}[f'(y_t)(\mu_t \mu_t' + \Sigma_t) f(y_t)],$$

where μ_t and Σ_t are the conditional mean vector and conditional covariance matrix of the set of returns, respectively. So the above problem is reduced to the problem (we omit the subscript t in the derivation),

$$\sup_{f(y)} \frac{(\mathrm{E}[(p - v\mu)' f(y)])^2}{\mathrm{E}[f'(y)(\mu\mu' + \Sigma) f(y)] - \mathrm{E}^2[f'(y)\mu]}, \tag{50}$$

where

$$\mathrm{E}[(p - v\mu)' f(y)] = \int (p - v\mu)' f(y) \rho(y) dy, \tag{51}$$

$$\mathrm{E}[f'(y)(\mu\mu' + \Sigma) f(y)] = \int f'(y)(\mu\mu' + \Sigma) f(y) \rho(y) dy, \tag{52}$$

$$\mathrm{E}[f'(y)\mu] = \int \mu' f(y) \rho(y) dy, \tag{53}$$

where y is a multidimensional vector and $\rho(y)$ is the multivariate distribution function of y. This is a variation-like problem and we adapt the calculus of variation technique to solve it. Let $g(y) = f(y) + \epsilon h(y)$, where $\epsilon > 0$, the first-order condition with respect to ϵ gives

$$\mathrm{E}\left[\frac{(p - v\mu)' h}{\mathrm{E}[(p - v\mu)' f]}\right] = \mathrm{E}\left[\frac{(f'(\mu\mu' + \Sigma) - \mathrm{E}[\mu' f]\mu)' h}{\mathrm{E}[f'(\mu\mu' + \Sigma) f] - \mathrm{E}^2[\mu' f]}\right], \quad \forall h,$$

where we write f or h instead of $f(y)$ or $h(y)$ whenever there is no confusion. This implies that

$$\frac{(p - v\mu)}{\mathrm{E}[(p - v\mu)' f]} = \frac{((\mu\mu' + \Sigma) f - \mathrm{E}[\mu' f]\mu}{\mathrm{E}[f'(\mu\mu' + \Sigma) f] - \mathrm{E}^2[\mu' f]}. \tag{54}$$

Note that the probability density function $\rho(y)$ of y does not appear explicitly. Solving for f from Equation (54), we obtain

$$f = (\mu\mu' + \Sigma)^{-1}\left(\frac{\mathrm{E}[f'(\mu\mu' + \Sigma)f] - \mathrm{E}^2[\mu' f]}{\mathrm{E}[(p - v\mu)' f]}(p - v\mu) - \mathrm{E}[\mu' f]\mu\right). \tag{55}$$

This completes our solution for the functional form of $f(y)$, since the expectations on the right-hand side of Equation (55) only depend on y through some constant parameters, representing unconditional moments. Hence we obtain

$$f = (\mu\mu' + \Sigma)^{-1}(\alpha p + \lambda\mu),$$

where α and λ are constants. Further, note that the scaling by a constant does not change the HJ bound, so we can solve f only up to a constant. We can thus let $\alpha = 1$. With the functional form of the scaling factor known, we can determine the constant λ [note that $-\lambda$ is w in Equation (10)] by solving a standard maximization problem (instead of a functional problem):

$$\sup_{\lambda} g(\lambda) = \max_{\lambda} \frac{(\mathrm{E}[(p - v\mu)'(\mu\mu' + \Sigma)^{-1}(p + \lambda\mu)])^2}{\mathrm{E}[(p + \lambda\mu)'(\mu\mu' + \Sigma)^{-1}(p + \lambda\mu)] - \mathrm{E}^2[\mu'(\mu\mu' + \Sigma)^{-1}(p + \lambda\mu)]} \tag{56}$$

So we have

$$g(\lambda) = \frac{(a - vb + \lambda b - \lambda vd)^2}{(a + 2\lambda b + \lambda^2 d) - (b + \lambda d)^2} \tag{57}$$

where

$$\begin{aligned} a &= \mathrm{E}[p'(\mu\mu' + \Sigma)^{-1}p], \\ b &= \mathrm{E}[p'(\mu\mu' + \Sigma)^{-1}\mu], \\ d &= \mathrm{E}[\mu'(\mu\mu' + \Sigma)^{-1}\mu]. \end{aligned} \tag{58}$$

Now we can use the standard first-order conditions to determine λ. The first-order condition in λ gives

$$0 = \frac{2(a - vb + \lambda b - \lambda vd)(b - vd)}{(a + 2\lambda b + \lambda^2 d) - (b + \lambda d)^2} \tag{59}$$

$$- \frac{2(a - vb + \lambda b - \lambda vd)^2(b + \lambda d - (b + \lambda d)d)}{((a + 2\lambda b + \lambda^2 d) - (b + \lambda d)^2)^2}. \tag{60}$$

Factoring out $(a - vb + \lambda b - \lambda vd)$ (this is not a problem because $\lambda = \frac{vb - a}{b - vd}$ is a minimum since it leads to $\sigma^2_{OSB}(v) = 0$), we have

$$(b - vd)((a + 2\lambda b + \lambda^2 d) - (b + \lambda d)^2) - (a - vb + \lambda b - \lambda vd)(b + \lambda d - (b + \lambda d)d) = 0.$$

Solving this equation gives

$$\lambda = \frac{b - v}{1 - d}.$$

So the optimal scaling factor is

$$z^*_t = (\mu_t\mu_t' + \Sigma_t)^{-1}(p_t + \lambda\mu_t) \tag{61}$$

and the optimally scaled return is

$$r^*_{t+1} = (p_t + \lambda\mu_t)'(\mu_t\mu_t' + \Sigma_t)^{-1}r_{t+1}. \tag{62}$$

Substituting the optimally scaled returns into Equation (7), we obtain the optimally scaled bound

$$\sigma^2_{OSB} = \sigma^2(v; z_t^{*\prime} r_{t+1}) = \frac{a - ad + b^2 - 2bv + bv^2}{1 - d}. \tag{63}$$

We should remark that the above formulas constitute solutions to the first-order condition, which is only a necessary condition for optimality. We need to verify that the solution is a maximum. We can argue that the first-order condition is sufficient in the following way. Note that in the problem of Equation (50),

$$\sup_{f(y)} \frac{(\mathrm{E}[(p - v\mu)' f(y)])^2}{\mathrm{E}[f'(y)(\mu\mu' + \Sigma) f(y)] - \mathrm{E}^2[f'(y)\mu]}$$

is homogeneous of degree zero in $f(y)$, so it is equivalent to the problem:[10]

$$\inf_{f(y)} \mathrm{E}[f'(y)(\mu\mu' + \Sigma) f(y)] - \mathrm{E}^2[f'(y)\mu]$$
$$\text{s.t.} \quad (\mathrm{E}[(p - v\mu)' f(y)])^2 = 1.$$

Because both $\mathrm{E}[f'(y)(\mu\mu' + \Sigma) f(y)]$ and $(\mathrm{E}[(p - v\mu)' f(y)])^2$ are convex in $f(y)$ and there is an interior point, this is a convex programming problem and there is a minimum. In fact, one can easily verify that the solution is the one we obtained above.

Proof of Proposition 3. Note that the pricing kernel written in terms of scaled assets formed using r_{t+1} and $z_t' r_{t+1}$ can always be written as $\tilde{z}_t' r_{t+1}$ for some $\tilde{z}_t$. So we have

$$\max_{z_t \in I_t} \sigma^2(v; r_{t+1}, z_t' r_{t+1}) = \max_{z_t \in I_t} \sigma^2(v; z_t' r_{t+1}) = \max_{z_t \in I_t} \sigma^2(v; z_t^{*\prime} r_{t+1}),$$

but

$$\sigma^2(v; z_t^{*\prime} r_{t+1}) \leq \sigma^2(v; r_{t+1}, z_t^{*\prime} r_{t+1}).$$

Combining the above two expressions, we get

$$\sigma^2(v; r_{t+1}, z_t^{*\prime} r_{t+1}) = \sigma^2(v; z_t^{*\prime} r_{t+1}).$$

References

Ang, A., and G. Bekaert, 2002, "International Asset allocation with Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Balduzzi, P., and H. Kallal, 1997, "Risk Premia and Variance Bounds," *Journal of Finance*, 52, 1913–1949.

Bekaert, G., 1994, "Exchange-Rate Volatility and Deviations from Unbiasedness in a Cash-in-Advance Model," *Journal of International Economics*, 36, 29–52.

Bekaert, G., 1996, "The Time Variation of Risk and Return in Foreign Exchange Markets: A General Equilibrium Perspective," *Review of Financial Studies*, 9, 427–470.

Bekaert, G., and R. Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets," *Journal of Finance*, 47, 467–509.

Bekaert, G., and J. Liu, "Conditioning Information and Variance Bounds on Pricing Kernels," Working Paper 6880, NBER.

[10] We would like to thank Darrell Duffie for suggesting this proof.

Bekaert, G., and M. Urias, 1996, "Diversification, Integration and Emerging Market Closed-End Funds," *Journal of Finance*, 51, 835–869.

Bekaert, G., and G. Wu, 2000, "Asymmetric Volatility and Risk in Equity Returns," *Review of Financial Studies*, 13, 1–42.

Burnside, C., 1994, "Hansen-Jagannathan Bounds as Classical Tests of Asset-Pricing Models," *Journal of Business and Economic Statistics*, 12, 57–79.

Cecchetti, S. G., P.-S. Lam, and N. C. Mark, 1994, "Testing Volatility Restrictions on Intertemporal Marginal Rates of Substitution Implied by Euler Equations and Asset Returns," *Journal of Finance*, 49, 123–152.

Chen, Z., and P. J. Knez, 1995, "Measurement of Market Integration and Arbitrage," *Review of Financial Studies*, 5, 287–325.

Chen, Z., and P. J. Knez, 1996, "Portfolio Performance Measurement: Theory and Applications," *Review of Financial Studies*, 9, 511–555.

Cochrane, J. H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model," *Journal of Political Economy*, 104, 572–621.

Cochrane, J. H., and L. P. Hansen, 1992, "Asset Pricing Explorations for Macroeconomics," O. J. Blanchard and S. Fischer (eds.), *1992 NBER Macroeconomics Annual*, MIT Press, Cambridge, MA.

Constantinides, G. M., and D. Duffie, 1996, "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104, 219–240.

Dahlquist, M., and P. Söderlind, 1999, "Evaluating Portfolio Performance with Stochastic Discount Factors," *Journal of Business*, 72, 347–383.

De Santis, G., 1995, "Volatility Bounds for Stochastic Discount Factors: Tests and Implications from International Stock Returns," working paper, University of Southern California.

Engle, R. F., D. M. Lilien, and R. P. Robins, 1987, "Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model," *Econometrica*, 55, 391–407.

Engle, R. F., V. K. Ng, and M. Rothschild, 1990, "Asset Pricing with a FACTOR-ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, 45, 213–237.

Ferson, W. E., and J. J. Merrick, Jr., 1987, "Non-stationarity and Stage-of-the-Business-Cycle Effects in Consumption-Based Asset Pricing Relations," *Journal of Financial Economics*, 18, 127–146.

Ferson, W. E., and R. W. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance*, 51, 425–461.

Ferson, W., and A. Siegel, 2001, "The Efficient Use of Conditioning Information in Portfolios," *Journal of Finance*, 56, 967–982.

Ferson, W., and A. Siegel, 2003, "Stochastic Discount Factor Bounds with Conditioning Information," *Review of Financial Studies*, 16, 567–595.

French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–29.

Gallant, R., and G. Tauchen, 1989, "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

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, 141–179.

Glosten, L., R. R. Jagannathan, and D. E. Runkle, 1993, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Hamilton, J., 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hansen, L. P., J. Heaton, and E. G. J. Luttmer, 1995, "Econometric Evaluation of Asset Pricing Models," *Review of Financial Studies*, 8, 237–274.

Hansen, L. P., J. Heaton, and A. Yaron, 1996, "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

Hansen, L. P., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

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

Hansen, L. P., and K. Singleton, 1983, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–265.

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

Kaminsky, G., and R. Peruga, 1990, "Can a Time-Varying Risk Premium Explain Excess Returns in the Forward Market for Foreign Exchange?," *Journal of International Economics*, 28, 47–70.

Kandel, S., and R. F. Stambaugh, 1990, "Expectations and Volatility of Consumption and Asset Returns," *Review of Financial Studies*, 3, 207–232.

Kirby, C., 1997, "Measuring the Predictable Variation in Stock and Bond Returns," *Review of Finanical Studies*, 10, 579–630.

Kirby, C., 1998, "The Restrictions on Predictability Implied by Rational Asset Pricing Models," *Review of Financial Studies*, 11, 343–382.

Kocherlakota, N. R., 1996, "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, 42–71.

Newey, W. K., and K. D. West, 1987, "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Snow, K. N., 1991, "Diagnosing Asset Pricing Models Using the Distribution of Asset Returns," *Journal of Finance*, 46, 955–983.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Whitelaw, R. F., 2000, "Stock Market Risk and Return: An Equilibrium Approach," *Review of Financial Studies*, 13, 521–547.