NBER WORKING PAPER SERIES

CONDITIONING INFORMATION AND VARIANCE BOUNDS ON PRICING KERNELS

Geert Bekaert
Jun Liu

Working Paper 6880
http://www.nber.org/papers/w6880

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1999

We would like to thank Andrew Ang, Quang Dai, Darrell Duffie, Mark Garmaise, Ming Huang, and Jun Pan for valuable suggestions. Keneth Singleton provided extensive guidance, comments, and insights. Geert Bekaert thanks the NSF for research support. The views expressed here are those of the author and do not reflect those of the National Bureau of Economic Research.

Conditioning Information and Variance
Bounds on Pricing Kernels
Geert Bekaert and Jun Liu
NBER Working Paper No. 6880
January 1999

ABSTRACT

We show how to use conditioning information optimally to construct a sharper unconditional Hansen-Jagannathan (1991) bound. The approach in this paper is different from that of Gallant, Hansen and Tauchen (1990), but both approaches yield the same bound when the conditional moments are known. Unlike Gallant, Hansen and Tauchen, our approach is robust to misspecification of the first and second conditional moments. Potential applications include testing dynamic asset pricing models, studying the predictability of asset returns, diagnosing the accuracy of competing models for the first and second conditional moments of asset returns, dynamic asset allocation, and mutual fund performance measurement. The illustration in this article starts with the familiar Hansen-Singleton (1983) setup of an autoregressive model for consumption growth and bond and stock returns. Our innovation is to add time-varying volatility to the model. Both an unconstrained version and a version with the restrictions of the standard consumption-based asset pricing model imposed serve as the data-generating processes to illustrate the behavior of the bounds. In the process, we discover and explore an interesting empirical phenomenon: asymmetric volatility in consumption growth.

Geert Bekaert
Graduate School of Business
Stanford University
Stanford, CA 94305-5015
and NBER
bekaert_geert@gsb.stanford.edu

Jun Liu
Graduate School of Business
Stanford University
Stanford, CA 94305-5015

1 Introduction

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 (Bekaert and Urias (1996), DeSantis (1996), Snow (1991)), market integration tests (Chen and Knez (1995)), mutual fund performance measurement (Chen and Knez (1996), Ferson and Schadt (1996), Dahlquist and Söderlind (1997)) and more.

The pricing kernels implied by standard consumption-based asset pricing models with time-additive preferences dramatically fail to lie inside the feasible region defined by HJ bounds computed using a variety of asset returns. These results guided researchers to develop models that perform better (that is, imply more variable pricing kernels)[1], whereas others show that the bounds themselves are weaker when long-horizon returns are considered (Daniel and Marshall (1996)), or when market frictions are explicitly accommodated (He and Modest (1995), Luttmer (1996)). Of course, the tighter the HJ bound, the stronger the restrictions on asset pricing models. While a number of methods have been proposed to improve HJ bounds[2], the most obvious and promising one is the use of conditioning information in the computation of the bounds.

HJ bounds are computed by projecting the pricing kernel unconditionally on the space of available asset payoffs and computing the standard deviation of the projection. Obviously, when agents can use conditioning information to form portfolios, the dimension of the space of the available asset payoffs can be increased. Gallant, Hansen and Tauchen (1990, henceforth 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 pay-

[1]Changes in the preference structure have proved reasonably successful, see for example Bekaert (1996), Campbell and Cochrane (1997), Constantinides (1990) and Heaton (1995) for the effects of habit formation, Epstein and Zin (1991a) for non-expected utility preferences that disentangle risk-aversion from intertemporal substitution, Bekaert, Hodrick and Marshall (1997), Bonomo and Garcia (1993), and Epstein and Zin (1991b) for disappointment aversion preferences. Constantinides and Duffie (1996) and Heaton and Lucas (1996) examine incomplete-market economies with heterogeneous consumers.

[2]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.

offs. 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[3]. The GHT bound depends on the first and second conditional moments of the asset payoffs. To estimate these conditional moments, GHT (1990) use the semi-nonparametric technique (SNP) developed by Gallant and Tauchen (1989).

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, for example, Hansen and Jagannathan (1990), Cochrane and Hansen (1992), Bekaert and Hodrick (1992), He and Modest (1995) and many others). This procedure is much simpler to implement than GHT since it does not require knowledge of conditional moments at all. Many studies have found dramatic improvements in HJ bounds with the use of scaled returns and have interpreted this improvement as an indication of predictability in returns (see for example Bekaert and Hodrick (1992)). This seems 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.

In this article, we first find the optimal scaling factor. That is, we answer the question: when scaling a return with a function of the conditioning information, what is the function that maximizes the Hansen-Jagannathan bound? The solution is an application of functional analysis. As an important side result, we explore the relation between improvements in HJ bounds due to conditioning information and the presence of return predictability. Second, we show that our bound, which we term the *optimal bound*, can be as tight as the GHT bound when the conditional moments are known. When the conditional moments are not known, our bound has an interesting robustness property. The GHT bound is only correct when the conditional moments are accurate, if they are mis-specified 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, we use data on U.S. stock and bond returns to illustrate the relation between the GHT, the optimal and standard (scaled) HJ bounds in the context of a generalization of the Hansen and Singleton (1983) log-normal model for consumption and asset returns.

Contemporaneous with our work, Ferson and Siegel (1997a) derive and study the optimal scaling factor in the setting of mean-variance frontiers.

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

Since there is a well-known duality between Hansen-Jagannathan frontiers and the mean-variance frontier, these results are similar and a final section of Ferson and Siegel (1997a) examines the link with HJ bounds. Our proof is different however, and both the focus and applications of our work are very different as well. Ferson and Siegel (1997b) provide an alternative proof of the optimality of the GHT bound and compare empirically its performance to the use of scaled returns.

The remainder of the paper is organized as follows. In section 2, after briefly reviewing the construction of standard HJ bounds, we derive the optimal scaling factor in the one asset case and discuss some of its characteristics. In section 3, we generalize to the multiple asset case, and explore formally the relation with the GHT bound. Section 4 contains our empirical illustration. We estimate an asymmetric GARCH-in-Mean model on US consumption growth, bond and stock returns and test the restrictions of the standard consumption-based asset pricing model. Section 5 uses the constrained and unconstrained versions of the model to illustrate the behavior of the various bounds. We discuss future potential applications of our results briefly in the conclusions.

2 The Optimal Bound in the One-Asset Case

In this section, we first review the standard HJ bound while setting up notation. We then prove our main proposition regarding the form of the optimal scaling factor. Finally we examine under what conditions scaling improves the HJ bound.

2.1 Notation and Review

Let there be an asset with payoff r_{t+1} and price 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

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

By the law of iterated expectations, this implies

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

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 immediately from (2) by noting that

$$\mathrm{E}\Big[m_{t+1}r_{t+1}\Big] = \mathrm{cov}\Big[m_{t+1}, r_{t+1}\Big] + \mathrm{E}\Big[m_{t+1}\Big]\mathrm{E}\Big[r_{t+1}\Big]$$

and applying the Cauchy-Schwarz inequality:

$$\mathrm{Var}\Big[m_{t+1}\Big] \geq \frac{\Big(q - \mathrm{E}\Big[m_{t+1}\Big]\mathrm{E}\Big[r_{t+1}\Big]\Big)^2}{\mathrm{Var}\Big[r_{t+1}\Big]} \tag{3}$$

To simplify notation, let's denote the unconditional first and second moments of r_{t+1} by $\mu = \mathrm{E}\Big[r_{t+1}\Big]$ and $\sigma^2 = \mathrm{Var}\Big[r_{t+1}^2\Big]$. The HJ bound is a function of $v = \mathrm{E}\Big[m_{t+1}\Big]$ and of the first two moments (μ, σ^2) of r_{t+1} (if q is 1), hence, we denote the bound as $\sigma^2(v|r_{t+1})$. Re-writing from (3), we find:

$$\sigma^2(v|r_{t+1}) = d\,v^2 - 2b\,v + a, \tag{4}$$

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

2.2 The optimal HJ bound

Although only one asset is available, the presence of the conditioning variables y_t allows construction of a much larger payoff space. Let $z_t = f(y_t)$, where f is a measurable function, then the space of scaled payoffs, $z_t r_{t+1}$, can in principle be infinite dimensional (see Hansen and Richard (1987)). Such scaling has an intuitive interpretation when excess returns, $r_{t+1}^e = r_{t+1} - r^f$, are scaled as in Bekaert and Hodrick (1992) and Cochrane (1996). The gross "scaled" return, $r_{t+1} = z_t r_{t+1}^e + r^f = z_t r_{t+1} + (1 - z_t) r^f$, can then be interpreted as a "managed" portfolio with z_t being the time-varying proportion of the investment allocated to the risky asset.

Scaling will likely only improve HJ bounds if the weight z_t has information on future returns. In the literature, one sets $z_t = e'y_t$, where e is an indicator vector selecting the variable in y_t believed to predict r_{t+1} or to capture the time-variation in the expected return. But is this the optimal way to select from the set of information variables?

To pose the problem more formally, by varying z_t, we obtain different HJ bounds that only depend on the unconditional moments of $z_t r_{t+1}$,

$$\sigma^2(v|z_t r_{t+1}) = \frac{\Big(\mathrm{E}\Big[z_t p_t\Big] - v\mathrm{E}\Big[z_t r_{t+1}\Big]\Big)^2}{\mathrm{Var}\Big[z_t r_{t+1}\Big]} \tag{5}$$

The formal question our proposition answers is: What z_t yields the best (largest) HJ bound? Since $z_t = f(y_t)$, this is a problem in functional analysis.

Proposition 2.2: The solution z_t^* to the maximization problem

$$\max_{z_t \in I_t} \sigma^2(v|z_t r_{t+1}) \tag{6}$$

is given by

$$z_t^* = \frac{p_t + \lambda \mu_t}{\mu_t^2 + \sigma_t^2}, \tag{7}$$

where

$$\mu_t = \mathrm{E}\Big[r_{t+1}|I_t\Big], \ \mu_t^2 + \sigma_t^2 = \mathrm{E}\Big[r_{t+1}^2|I_t\Big], \tag{8}$$

$$\lambda = \frac{b - v}{1 - d} \tag{9}$$

$$b = \mathrm{E}\Big[\frac{\mu_t p_t}{\mu_t^2 + \sigma_t^2}\Big] \tag{10}$$

$$d = \mathrm{E}\Big[\frac{\mu_t^2}{\mu_t^2 + \sigma_t^2}\Big] \tag{11}$$

Furthermore, the maximum bound is given by

$$\sigma^{*2}(v|z_t r_{t+1}) \equiv \sigma^2(v|z_t^* r_{t+1}) = \frac{a(1-d) + b^2 - 2b\,v + d\,v^2}{1-d} \tag{12}$$

Where

$$a = \mathrm{E}\Big[\frac{p_t^2}{\mu_t^2 + \sigma_t^2}\Big] \tag{13}$$

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 then 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^2, it is not monotonic in the conditional mean, μ_t. The non-monotinicity 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, this 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 return 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.

Ferson and Siegel (1997a) provide a detailed characterization of the scaling factor in a mean-variance setting.

2.3 When does scaling improve the HJ bound?

Since the conditional moments are usually unknown, 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. If the scaling factor z_t is uncorrelated with the first and second conditional moment of r_{t+1} (that is, $\text{cov}(p_t, z_t) = \text{cov}(r_{t+1}, z_t) = \text{cov}(r_{t+1}^2, z_t^2) = 0$, then scaling the return with z_t will decrease the HJ bound. To see this, note that

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

where we omitted the time subscripts. The last inequality follows since $\text{E}[z_t^2]/\text{E}^2[z_t] \geq 1$. 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 (for example, a stock return with its lagged dividend yield) the variable must predict the return in order for the HJ bound to improve.

Bekaert and Hodrick (1992) use this intuition to interpret the dramatic improvement of the HJ bound when foreign exchange returns are scaled by the forward premium as evidence of strong predictability in the foreign exchange market. As in other studies (see Cochrane and Hansen (1992) and Cochrane (1996), for example), they consider two-dimensional spaces of the form

$$\begin{pmatrix} r_{t+1} \\ z_t r_{t+1} \end{pmatrix}, \tag{14}$$

where $z_t = e'y_t$. In this case, since $r_{t+1} \in \{r_{t+1}, z_t r_{t+1}\}$, we know for sure $\sigma^2(v|r_{t+1}) \leq \sigma^2(v|r_{t+1}, z_t r_{t+1}), \forall z_t$.

Even in this case, for the bound to strictly improve, the 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 2.3: Suppose there is an asset 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

$$\max_{z_t \in I_t} \sigma^2(v|r_{t+1}, z_t r_{t+1})$$

is given by

$$z_t^* = \frac{p_t + \lambda \mu_t}{\mu_t^2 + \sigma_t^2}.$$

The proof is given in the appendix.

Now, suppose μ_t and σ_t are constants (that is, there is no predictable variation in conditional means or variances), then z_t^* is a constant and r_{t+1} and $z_t^* r_{t+1}$ are linearly dependent. It follows that $\sigma^2(v|r_{t+1}, z_t^* r_{t+1}) = \sigma^2(v|r_{t+1})$. 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 and we will sometimes refer to the resulting bounds as "naive" bounds. Apart from our optimal bounds, we will also report "stacked" optimal bounds, $\sigma^2(v|r_{t+1}, z_t^* r_{t+1})$, which ought to be identical to the optimal bounds when the conditional moments are known (see below).

3 The Optimal and GHT Bound in the Multi-Asset Case

In this section, we extend the results to the multi-asset case and explain the links with the GHT bound. We now let r_{t+1} and p_t be n-dimensional vectors. A scaled asset is a one dimensional asset, $\tilde{r}_{t+1} = z_t' r_{t+1}$, where z_t is a 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 $P = \{z_t' r_{t+1} : \forall z_t\}$. Gallant, Hansen and Tauchen directly project the pricing kernel onto this space. They show that the projected pricing kernel is

$$m_{t+1}^* = (p_t + \lambda \mu_t)'(\mu_t \mu_t' + \Sigma_t)^{-1} r_{t+1} + \lambda,$$

where μ_t is the conditional mean vector and Σ_t the conditional variance-covariance matrix of the returns (the above equation is obtained from equation (20) of GHT noting that w is a constant). λ is given by

$$\lambda = \frac{b - v}{1 - d} \tag{15}$$

where

$$b = \mathrm{E}\Big[\mu_t'(\mu_t \mu_t' + \Sigma_t)^{-1} p_t\Big], \tag{16}$$

$$d = \mathrm{E}\Big[\mu_t'(\mu_t \mu_t' + \Sigma_t)^{-1} \mu_t\Big], \tag{17}$$

$$a = \mathrm{E}\Big[p_t'(\mu_t \mu_t' + \Sigma_t)^{-1} p_t\Big]. \tag{18}$$

The GHT bound by definition is

$$\sigma^{*2}(v) = \mathrm{var}(m^*_{t+1}). \tag{19}$$

It is a lower bound to the variance of all valid pricing kernels.

The approach in this paper 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 a Hansen-Jagannathan bound $\sigma^2(v|z_t' r_{t+1})$ associated with each scaling vector z_t. The optimal bound is the highest such Hansen-Jagannathan bound

$$\tilde{\sigma}^{*2}(v) \equiv \sup_z \sigma^2(v|z_t' r_{t+1}).$$

Both bounds $\sigma^{*2}(v)$ and $\tilde{\sigma}^{*2}(v)$ depend on the conditional mean and the conditional variance of the payoffs which may or may not be known to researchers. We will discuss the relation between $\sigma^{*2}(v)$ and $\tilde{\sigma}^{*2}(v)$ in both situations.

3.1 When conditional moments are known

Proposition 3.1: For a n-dimensional payoff r_{t+1} with price vector p_t, conditional mean μ_t, and conditional variance-covariance matrix Σ_t,

$$\sigma^{*^2}(v) = \tilde{\sigma}^{*2}(v) = \frac{a - ad + b^2 - 2bv + dv^2}{1-d}.$$

where a, b, and d are defined in equations (16), (17), and (18).

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

$$\sigma^2(v|z_t' r_{t+1}) \le \sup_z \sigma^2(v|z_t' r_{t+1}) = \tilde{\sigma}^{*2}(v) \le \sigma^{*2}(v). \tag{20}$$

The optimal scaling vector, z_t^*, follows from the multivariate extension to proposition 2.2. That is,

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

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

$$\sigma^{*2}(v) = \mathrm{var}(m^*_{t+1}) = \mathrm{E}[(z_t^{*\prime} r_{t+1} + \lambda)^2] - (\mathrm{E}[z_t^{*\prime} r_{t+1} + \lambda])^2. \tag{21}$$

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

$$\begin{aligned} \sigma^{*2}(v) &= \mathrm{E}[(p_t + \lambda\mu_t)'(\mu_t\mu_t' + \Sigma_t)^{-1}(p_t + \lambda\mu_t)] \\ &- \mathrm{E}([(p_t + \lambda\mu_t)'(\mu_t\mu_t' + \Sigma_t)^{-1}(p_t + \lambda\mu_t)])^2. \end{aligned} \tag{22}$$

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

This result is at first surprising. Our optimal bound is a standard HJ bound for a scaled return. Since the scaling factor depends on v, the mean of the pricing kernel, the optimal 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 becomes proportional to the square of the quadratic polynomial, and the optimal bound becomes quadratic in v and the optimal scaled frontier becomes a parabola, identical to the GHT frontier.

3.2 When conditional moments are not known

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^2_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 optimal bound is $\sigma^2(v|z^{*\prime}_t r_{t+1})$, where z^*_t depends on the first two conditional moments. When the conditional moments are unknown, z^*_t is unknown and so is $\sigma^2(v|z^{*\prime}_t r_{t+1})$. 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 a HJ bound. Hence, even when using a proxy for the conditional moments to get a proxy $\hat{z}^*_t$ for z^*_t, the resultant optimal 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 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 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 processes for returns contain jumps or regime - switches, it is not clear that the SNP approach provides a good approximation. Moreover, the current empirical evidence suggests that stock returns are predictable by a variety of variables, such as dividend yields, term spreads, forward premiums (see Bekaert and Hodrick (1992) and Ferson and Harvey (1993) for international evidence), and it is not clear how accurate SNP estimation is in such a complex multivariate setting.

The risk of over-estimating the variance bound can be avoided by applying our method. Given an empirical specification for the conditional moments, our "optimal" bound is as easy to implement as the original Hansen-Jagannathan 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 Σ_t by the unconditional variance. The resulting bound will not necessarily be the tightest bound and if there truly is time-variation in the conditional variance it will not be optimal. Nevertheless it is hoped that if the time-variation in conditional variances is minimal, our bound may still be sharper than using arbitrary scaling, a conjecture we will examine in the empirical section of this paper.

The fact that optimal bounds computed from mis-specified conditional moments remain valid bounds which are best when the true conditional moments are used, suggests an interesting application of our procedure. We can use the optimal bound to study not only which predictive variables yield the sharpest HJ bounds (as in Bekaert and Hodrick (1992)), but also to diagnose the accuracy of competing models for the first two conditional moments. To see this, note that when conditional moments are mis-specified, it need not be the case that $\sigma^2(v|r_{t+1}) \leq \sigma^2(v|z_t^* r_{t+1})$. Hence, mis-specified conditional moments may reveal themselves through poorly performing optimal bounds relative to the conditional, "naively" scaled or stacked optimal bounds. They may also result in the optimal HJ bound failing to be a parabola. We will illustrate the use of the optimal bound as a diagnostic tool in our empirical illustration below.

4 Empirical Application: The Model

4.1 The Econometric Model

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 $Y_t = [X_t, R_t^s, R_t^b]$. In the seminal work of Hansen and Singleton (1983, henceforth HS), it is assumed that y_t follows a vector-autoregressive (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.

Given that there is ample evidence of predictability in excess returns,[4] a natural extension of the HS framework is to allow for heteroskedasticity using the GARCH-in-Mean framework of Engle, Lilien and Robins (1987). Surprisingly, apart from an application to international data[5], there is little

[4]See for example, Campbell (1987) and Bekaert and Hodrick (1992).

[5]See Kaminsky and Peruga (1990).

work in this area. Two reasons may be the parameter proliferation that occurs with the multivariate GARCH models and the lack of heteroskedasticity in consumption growth (which may be due to a temporal aggregation bias[6]). Nevertheless, we will use this familiar framework to illustrate the properties of our "optimal bound".

Our 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[7]. Table 1 demonstrates the importance of this feature. We report estimates of a simple univariate GARCH model with asymmetry, applied to the residuals of a first-order VAR on Y_t. The asymmetry in stock returns is no surprise, but we also find some evidence of asymmetry in the conditional variance of quarterly consumption growth.

Whereas this evidence is economically and statistically weak, it is strongly suggestive of an asymmetric component in the volatility of consumption growth. First, it is intuitively plausible that uncertainty about future consumption growth is higher in a recession than in a boom.[8] Second, we know that there is strong asymmetry in stock returns that may help accommodate the negative skewness we observe, which a standard GARCH model with normally distributed innovations cannot match. It is likely that temporal aggregation and the lack of data considerably weaken the results. We will see below that in the multivariate GARCH model we estimate, the asymmetry patterns become stronger.

For the multivariate set-up, we begin by parameterizing an unconstrained model:

$$Y_t = \mu_{t-1} + AY_{t-1} + \Omega_{t-1}e_t, \tag{23}$$

where

$$\mu_t = \begin{bmatrix} \mu_{xt} \\ \mu_{bt} \\ \mu_{st} \end{bmatrix}, \tag{24}$$

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

$$h_{iit} = d_i + a_i h_{iit-1} + b_i e_{iit-1}^2 + \eta_i\{\max[0, -e_{iit-1}]\}^2. \tag{25}$$

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

[7]See Glosten, Jagannathan and Runkle (1993) and Bekaert and Wu (1997).

[8]We could not find articles in the business cycle literature that document this phenomenon. Most empirical articles (see, e.g., Filardo (1994)) focus on the conditional mean dynamics (the duration and steepness of recessions and expansions.) Sichel (1983) does report evidence of "deepness", troughs are further below the trend than peaks are above. An interesting implication of this feature is that the stationary component of the series should exhibit negative skewness, which we observe in consumption growth and which may be accommodated by volatility asymmetry.

If $\eta_i > 0$, volatility displays the well-known asymmetric property.

The e_t-vector constitutes 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:

$$\Omega_t = \Omega = \begin{bmatrix} 1 & 0 & 0 \\ f_{xb} & 1 & 0 \\ f_{xs} & f_{bs} & 1 \end{bmatrix} \tag{26}$$

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

$$a_b = b_b = \eta_b = a_s = b_s = \eta_s = 0. \tag{27}$$

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

$$\Sigma_t = \Omega H_t \Omega'. \tag{28}$$

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 should allow the conditional covariance matrix to affect the conditional mean as well. Therefore, we let

$$\mu_{it} = v_i h_{xxt} + \mu_i, \tag{29}$$

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

$$\Theta = [\text{vec}(A)', \mu_x, \mu_b, \mu_s, v_b, v_s, f_{xb}, f_{xs}, d_x, a_x, b_x, \eta_x, d_b, d_s]'.$$

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

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

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

$$\mathrm{E}_t[R^i_{t+1}] = -(\ln\beta + \frac{1}{2}h_{ii}) - \frac{1}{2}[\gamma - f_{xi}]^2 h_{xxt} + \gamma \mathrm{E}_t[X_{t+1}] \quad (30)$$

Note h_{ii} does not depend on t for $i = b, s$ because of equation (27). 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 empirically shown to move countercyclically (Fama and French (1989), Bekaert and Harvey (1995)).

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.

$$\mathrm{E}_t[\exp(R^i_{t+1})] = \exp\Big(-\ln\beta - \frac{\gamma}{2}[\gamma - 2f_{xi}]h_{xxt} + \gamma \mathrm{E}_t[X_{t+1}]\Big) \quad (31)$$

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,

$$\Theta^R = [\mu_x, A_{11}, A_{12}, A_{13}, \beta, \gamma, f_{xb}, f_{xs}, d_i, a_i, b_i, \eta_i]', \quad i = b, s.$$

4.2 Data

Our consumption measure is the sum of per capita real non-durables and services consumption in the US. These data were downloaded from DATASTREAM, The stock return is the quarterly value - weighted dividend-inclusive index return on the NYSE, taken from Wharton's web site (http://wrdsx.wharton.upenn.edu). The interest rate is the U.S. 3 month Treasury Bill rate taken the Federal reserve web site. We used a data set on weekly secondary market rates (averages of daily) and used the rate closest to the end of the month. All data run from the second quarter in 1959 to the end of 1996.

Table 1 summarizes some of the data properties. Consumption growth and real bond returns have about the same mean and volatility, dwarfed by the mean and volatility of stock returns. All series show leptokurtosis, but only consumption growth and stock returns show negative skewness. Consumption growth is more highly autocorrelated than could be explained by time-averaging and bond returns have an autocorrelation of 0.734. Univariate GARCH processes reveal "strong" asymmetry (in that positive shocks reduce the variance) for both stock returns and consumption growth. Bond returns show very weak asymmetry. The high standard errors for all estimated systems are hardly a surprise given that we are working with 151 quarterly data points. Nevertheless, we know that use of high frequency stock return data leads to the finding of strong asymmetric GARCH patterns which remain preserved here in the quarterly data but are necessarily statistically weak. Similarly, our results suggest (although do not prove) that if we were able to use high frequency consumption growth data we might find similar strong asymmetric volatility patterns.

4.3 Estimation Results

To lead into our GARCH-in-mean models, Table 2 first presents the autoregressive dynamics implied by a first-order VAR. Except for autocorrelation coefficients in the consumption growth and bond return equations, there are no highly significant coefficients. Table 3 shows the results from the unconstrained estimation. The conditional mean parameters mimic the coefficients of the unconstrained VAR rather well, despite the presence of very large coefficients on the GARCH-in-mean term. Although the standard errors 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 for example French, Schwert and Stambaugh (1987), Glosten, Jagannathan and Runkle (1993)) 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 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. This is somewhat problematic since the conditional variance may theoretically become negative although it never does in sample. It should be noted that the local optima with positive GARCH-in-mean parameters typically display larger volatility persistence as does the univariate stock return model.

The constrained model (see Table 4) is not surprisingly rejected by a likelihood ratio test. The chi-square test statistic is 75.32 with a p-value of 0.000 (there are 8 restrictions). The CRRA is estimated to be 14.675 and the discount factor β is 1.071. Although the latter is above 1, 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 used to explain the equity premium puzzle and they are less extreme than the ones reported by Kandel and Stambaugh (1990). 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. Given that 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. This confirms the importance of the autoregressive dynamics in the performance of the consumption-based asset pricing model, first noted by Singleton (1990). He pointed out that in a GMM estimation, the Euler equation residual simply inherited the serial correlation properties of the original return series. In our set-up, the model matches the bond dynamics, but fails to match the autoregressive dynamics in consumption growth.

5 Empirical Application: the Bounds

This section illustrates the performance of our optimal bound along three dimensions. First, we show the role of the predictability of returns on HJ bounds, by comparing unconditional HJ bounds with the different bounds embedding conditioning information. In particular, our optimal bound should yield sharper HJ bounds than the standard method of arbitrarily scaling the returns with instruments. Second, we demonstrate the robustness of our optimal bound relative to the GHT bound. That is, we give an economically interesting case in which the GHT bound over-estimates the variance of the true pricing kernel. Third, we show that the difference between the our optimal and the GHT bound in certain settings can be used as diagnostic tool for dynamic asset pricing models.

The setting is the log-normal model for stock and bond returns and consumption growth estimated before. The model, in its unconstrained and constrained form, yields two candidates for the computation of the conditional moments we need in deriving the optimal and GHT bounds. We will also use these models as data generating processes in simulations. Simulations will both serve to illustrate the effect of mis-specifications 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.[9] Generally, we defer formal econometric issues and the formal handing of sampling error to future work (see Bekaert and Liu (1998)).

[9] 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 the short data sample.

5.1 Predictability

Figure 1 graphs the unconditional bound, the naively scaled bound, the optimal bound, the optimal stacked and the GHT bound for the two returns, assuming the unconstrained model for the conditional moments. Naive scaling uses the past bond and stock returns as instruments for both returns. First of all, the difference between the unconditional and scaled bounds reveals considerable predictability. By varying the instrument set, it is straightforward to establish that 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 than the optimally scaled bounds and the GHT bounds. This can be due to either mis-specification of the conditional moments or chance (sampling error). In any case, for this particular example, the naive scaling method suffices to get a sharp, valid bound.

To examine this issue closer, we first produce the same graphs for a long simulated sample from the unconstrained model in Figure 2. As should be the case, the GHT and optimal bounds are now on top of one another and dominate arbitrarily 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 our model describes the data rather well, the dominance of th · 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, but that is exactly quite weak in quarterly data. In fact, a linear VAR with constant variance-covariance matrix describes the data rather well. When we replace the conditional moments generated by the unconstrained model by the moments generated by a simple linear VAR, the optimal and GHT bounds change very little. Clearly, the importance of optimal scaling in generating sharper Hansen-Jagannathan bounds is potentially more dramatic when strong non-linearities are present, as in regime-switching models or high frequency GARCH or stochastic volatility models.

5.2 Diagnostics

Figure 3 graphs the naively scaled bound, the optimal bound and the stacked optimal bound, but this time the constrained model yields the conditional moments. Two factors stand out. First, the stacked optimal bound gets pretty close to the naively scaled bound, despite the mis-specification of the conditional moments. Of course, the constrained model managed 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 optimal 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 optimal 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 4 produces the optimal, stacked optimal and naively scaled bounds for data simulated from the constrained model. Since the model for conditional moments is correctly specified in this case, we now do obtain 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 optimal bound more starkly, we can use simulations and our two data generating processes to generate mis-specified bounds. Figure 5 shows the optimal and naively scaled bound for data simulated from the unconstrained model, but conditional moments erroneously generated from the constrained model. Figure 6 reverses the roles of the unconstrained and constrained model, generating data satisfying the constrained model and computing the optimal bound using moments according to the unconstrained model. In both cases, the optimal and naively scaled bounds are close and the bounds are uniformly higher when the data satisfy the unconstrained model (that is, the constrained model misses some of the predictable components the unconstrained model generates). Strikingly, in both cases, the optimal bound does show non-parabolic behavior near the trough of the graph.

5.3 Robustness

We have so far not focussed on the GHT bounds very much. Generally, optimal bounds do not much worse or better than the GHT bound. Moreover, our simulations reveal that the GHT bounds quite often over-estimate the variance of the true pricing kernel. A first example is in Figure 7. In Figure 7, we generate data from the unconstrained model. We show two GHT bounds, one bound uses the actual, true conditional moments, the other mis-specified moments from the constrained model. We also show our optimal bound, which uses the constrained moments. When the moments are mis-specified, the GHT bound generates too high values for the bounds on the right-hand side. When we reverse the roles of the unconstrained and constrained models in Figure 8, a similar phenomenon appears. This time, the bound over-estimates at the left hand side of the graph. The optimal bound never exceeds the true GHT bound but manages to be quite close to it.

6 Conclusions

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[10]. Our paper provides a bridge between the insightful but complex analysis of GHT (1990), and the simple but sub-optimal practice of arbitrarily scaling of 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 over-estimate the variance of the pricing kernel when the conditional moments are mis-specified. In this article, we derive the best possible scaled bound, the optimal bound. As does the GHT bound, this bound requires specifying the conditional mean and variance of the returns and we show that the optimal bound is as good as the GHT bound when these moments are correctly specified. When they are mis-specified 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 a HJ bound.

There are potentially many interesting applications of our framework. First, as we showed in section 5, the difference between the GHT bound or the stacked optimal bond and the optimal bound can sometimes serve as a diagnostic tool to judge the performance of dynamic asset pricing models. Although we restricted ourselves to the well-known world of the standard consumption-based asset pricing model, applications could extend to any other dynamic asset pricing model, for example affine term structure models.

Second, as partly illustrated in section 4 as well, the bounds can be used to re-examine the predictability of asset returns and to examine which instruments yield the sharpest restrictions on asset return dynamics. In Bekaert and Liu (1998), we repeat the analysis of Bekaert and Hodrick (1992) on international asset return predictability, with an expanded data set and incorporating conditioning information optimally.

Third, the bounds can also yield information on expected return and conditional variance modeling. The reason is t hat 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. There exists the danger that empirical models of conditional mean and variance fit quirks in the data that are of no statistical significance. It is therefore critical to develop statistical tests. In Bekaert and Liu (1998), we develop the econometrics needed to compare various models of time-variation in expected returns and variances and use the optimal bound to learn about various conditional moments specification of international stock, bond and

[10]One response may be to drop an unconditional framework all together, but in both financial practice and the recent academic literature (see especially Cochrane and Saa-Requejo (1997)) the importance of unconditional analysis remains prevalent.

foreign exchange returns.

Fourth, using the duality with the mean-variance frontier, the optimal bound can be used in dynamic models of optimal asset allocation that seek to maximize an unconditional mean-variance criterion. The groundwork for such an application has been laid in the work of Ferson and Siegel (1997a).

Fifth, 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. The role optimal bounds can play in this setting is also briefly discussed in Ferson and Siegel (1997a).

Appendix

Proof of proposition 2.2: The problem we would like to solve is

$$\sup_z \sigma^2(v|z_t r_{t+1}) = \sup_z \frac{(E(zp - vzr))^2}{E(zr)^2 - E^2(zr)}.$$

This is a well defined problem since $\sigma^2(v|z_t r_{t+1})$ is bounded from above by the GHT bound $\sigma^{*2}(v)$ and from below by 0. We will show the case of one asset, the general case of multiple payoffs is a straightforward extension.

For the case of one asset r, we have

$$E(z_t p_t) = E(p_t f(y_t)),$$

$$E(z_t r_{t+1}) = E(z_t E_t(r_{t+1})) = E(f(y_t)\mu_t),$$

and

$$E((z_t r_{t+1})^2) = E(z_t^2 E_t(r_{t+1}^2)) = E(f^2(y_t)(\mu_t^2 + \sigma_t^2)),$$

where μ_t and σ_t are the conditional mean and conditional variance of the return respectively. So the above problem is reduced to the problem (we omit the subscript t in the derivation),

$$\sup_{f(y)} \frac{[E((p - v\mu)f(y))]^2}{E(f^2(y)(\mu^2 + \sigma^2)) - E^2(f(y)\mu)}, \tag{32}$$

where

$$E((p - \mu)f(y)) = \int (p - \mu)f(y)\rho(y)dy,$$

and

$$E(\mu f(y)) = \int \mu f(y)\rho(y)dy,$$

and

$$E(f^2(y)(\mu^2 + \sigma^2)) = \int f^2(y)(\mu^2 + \sigma^2)\rho(y)dy.$$

$\rho(y)$ is the multi-variate distribution function of y, and y is a multi-dimensional vector. This is a variation-like problem and we adapt the calculus of variation technique to solve it.

Let $g(y) = f(y) + \epsilon h(y)$, the first order condition with respect to ϵ gives

$$E\Big[\frac{(p - v\mu)}{E(pf) - vE(\mu f)}h(y)\Big] = E\Big[\frac{(\mu^2 + \sigma^2)f - E(\mu f)\mu}{E((\mu^2 + \sigma^2)f) - E^2(\mu f)}h(y)\Big], \forall h(y)$$

So this implies that (we write f instead of $f(y)$ whenever there is no confusion),

$$\frac{(p - v\mu)}{E(pf) - vE(\mu f)} = \frac{(\mu^2 + \sigma^2)f - E(\mu f)\mu}{E((\mu^2 + \sigma^2)f) - E^2(\mu f)}. \tag{33}$$

Note that the probability density function of y does not appear explicitly. Solving for f from (33), we obtain:

$$f = \{\frac{E(f(\mu^2+\sigma^2)) - E^2(f\mu)}{E((p-v\mu)f)}(p-v\mu) + E(f\mu)\mu\}\frac{1}{\mu^2+\sigma^2}. \tag{34}$$

This completes our solution for the functional form of $f(y)$, since the expectations in right-hand side of (34) only depends on y through some constant parameters, representing unconditional moments. Hence, we obtain,

$$f = \frac{\alpha p + \lambda\mu}{\mu^2+\sigma^2}.$$

where α and λ are constants. Further, note that the scaling by a constant does not change the Hansen-Jagannathan 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 by solving a standard maximization problem (instead of a functional problem):

$$\max_{\lambda} g(\lambda) \equiv \max_{\lambda} \frac{[E((p-v\mu)(p+\lambda\mu)/(\mu^2+\sigma^2))]^2}{E((p+\lambda\mu)^2/(\mu^2+\sigma^2)) - E^2(\mu(p+\lambda\mu)/(\mu^2+\sigma^2))}. \tag{35}$$

So we have

$$g(\lambda) = \frac{(a - vb + \lambda b - \lambda vd)^2}{(a+2\lambda b+\lambda^2 d) - (b+\lambda d)^2}, \tag{36}$$

where

$$\begin{aligned} a &= E(p_t/(\mu_t^2+\sigma_t^2)), \\ b &= E(p_t\mu_t/(\mu_t^2+\sigma_t^2)), \\ d &= E(\mu_t^2/(\mu_t^2+\sigma_t^2)). \end{aligned} \tag{37}$$

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

$$\begin{aligned} 0 &= \frac{2(a-vb+\lambda b-\lambda vd)(b-vd)}{(a+2\lambda b+\lambda^2 d)-(b+\lambda d)^2)}, \\ &- \frac{(a-vb+\lambda b-\lambda vd)^2 2(b+\lambda d-(b+\lambda d)d)}{((a+2\lambda b+\lambda^2 d)-(b+\lambda d)^2)^2}. \end{aligned} \tag{38}$$

Factoring out $(a - vb + \lambda b - \lambda vd)$ ($\lambda = \frac{vb-a}{b-vd}$ is a minimum since it leads to $\sigma^{*2}(v) = 0$), we have

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

Simplifying this equation gives

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

So the optimal scaling factor is

$$z_t = \frac{p_t + \lambda\mu_t}{\mu_t^2 + \sigma_t^2} \tag{39}$$

and the optimal scaled asset is

$$\tilde{r}^*_{t+1} = \frac{p_t + \lambda\mu_t}{\mu_t^2 + \sigma_t^2} r_{t+1}. \tag{40}$$

Substituting the optimal scaled returns into equation (5), we obtain the optimal bound

$$\tilde{\sigma}^{*2}(v) = \sigma^2(v|z_t^* r_{t+1})) = \frac{a - ad + b^2 - 2bc + dc^2}{1-d}. \tag{41}$$

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 in the problem (32)

$$\max_{f(y)} \frac{[E((p - v\mu)f(y))]^2}{E(f^2(y)(\mu^2 + \sigma^2)) - E^2(f(y)\mu)},$$

is homogeneous of degree zero in $f(y)$, so it is equivalent to the problem[11]:

$$\begin{aligned} &\min_{f(y)} E(f^2(y)(\mu^2 + \sigma^2)) - E^2(f(y)\mu) \\ &\quad \text{s.t. } [E((p - v\mu)f(y))]^2 = 1. \end{aligned}$$

Because both $E(f^2(y)(\mu^2+\sigma^2)) - E^2(f(y)\mu)$ and $[E((p-v\mu)f(y))]^2$ are convex in $f(y)$ and there is 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 2.3: 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

$$\max_{z_t \in I_t} \sigma^2(v|r_{t+1}, z_t r_{t+1}) = \max_{z_t \in I_t} \sigma^2(v|z_t r_{t+1}) = \sigma^2(v|z_t^* r_{t+1})$$

But

$$\sigma^2(v|z_t^* r_{t+1}) \le \sigma^2(v|r_{t+1}, z_t^* r_{t+1})$$

Combining the above two expressions, we get

$$\max_{z_t \in I_t} \sigma^2(v|r_{t+1}, z_t r_{t+1}) = \sigma^2(v|z_t^* r_{t+1}) = \sigma^2(v|r_{t+1}, z_t^* r_{t+1}).$$

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

Table 1: Univariate Properties of the Data

	X_t	R_t^b	R_t^s
mean	0.449	0.421	1.51
volatility	0.547	0.561	8.05
Skewness	-0.491	0.559	-1.08
Kurtosis	1.01	1.18	2.46
ρ	0.489	0.734	0.077
a	-0.0348	0.1384	-0.0434
	(0.0583)	(0.1375)	(0.1141)
b	0.7847	0.8126	0.3763
	(0.5417)	(0.1047)	(0.1478)
η	0.0895	0.0266	0.4208
	(0.1885)	(0.1232)	(0.4606)

Notes: The sample period for X_t (log-consumption growth), R_t^s (log-real bond return) and R_t^b (log-real stock return) is from 2nd quarter of 1959 to the fourth quarter of 1996, for a total of 151 observations. The mean and volatility are expressed in percent (not annualized), ρ stands for first-order autocorrelation. The last three rows contain the estimated coefficients (with standard error in parentheses) of a univariate GARCH(1,1) model applied to the residuals of a first order VAR on $Y_t = [X_t, R_t^b, R_t^s]'$. That is, if h_t is the conditional variance and e_t the residual, the model for h_t is:

$$h_t = c + ae_{t-1}^2 + bh_{t-1} + \eta(\max[0, -e_{t-1}])^2.$$

Hence, the GARCH model accommodates asymmetry as in Glosten, Jagannathan and Runkle (1993).

Table 2: Unconstrained VAR

Equations	Coefficients			
	Constant	X_{t-1}	R^b_{t-1}	R^s_{t-1}
X_t	0.002	0.481	0.0058	0.0085
	(0.0006)	(0.081)	(0.0526)	(1.0053)
R^b_t	0.0013	-0.027	0.735	-0.0013
	(0.0005)	(0.086)	(0.070)	(0.0043)
R^s_t	0.015	-1.259	1.144	0.076
	(0.011)	(1.052)	(1.315)	(0.092)

Notes: A first-order VAR on $Y_t = [X_t, R^b_t, R^s_t]'$ is estimated using OLS. Standard errors are in paratheses and are heteroskeodasticity-consistent.

Table 3: Unconstrained GARCH-in-Mean Model

Equations	Coefficients			
	Constant	X_{t_1}	R^b_{t-1}	R^s_{t-1}
X_t	0.00295	0.361	-0.029	0.008
	(0.00047)	(0.033)	(0.022)	(0.005)
R^b_t	0.00555-162.65h_{xxt}	-0.198	0.738	-0.0002
	(0.00059) (0.00007)	(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	a_i	b_i	η_i
$h_{11,t}$	0.000019	-0.0265	0.0008	0.2705
	(0.000018)	(0.0807)	(0.7898)	(0.0426)
$h_{22,t}$	0.000014	0	0	0
	(0.000002)			
$h_{33,t}$	0.006134	0	0	0
	(0.00103)			
	f_{xb}=-0.0564		f_{xs}=3.182	
	(0.1425)		(0.003)	

Notes: The model estimated is described in equation (23) to (29). Standard errors are in paratheses and are robust to mis-specification of the error distribution in the sense of White (1982). Parameter values without standard errors reflect constrained parameters.

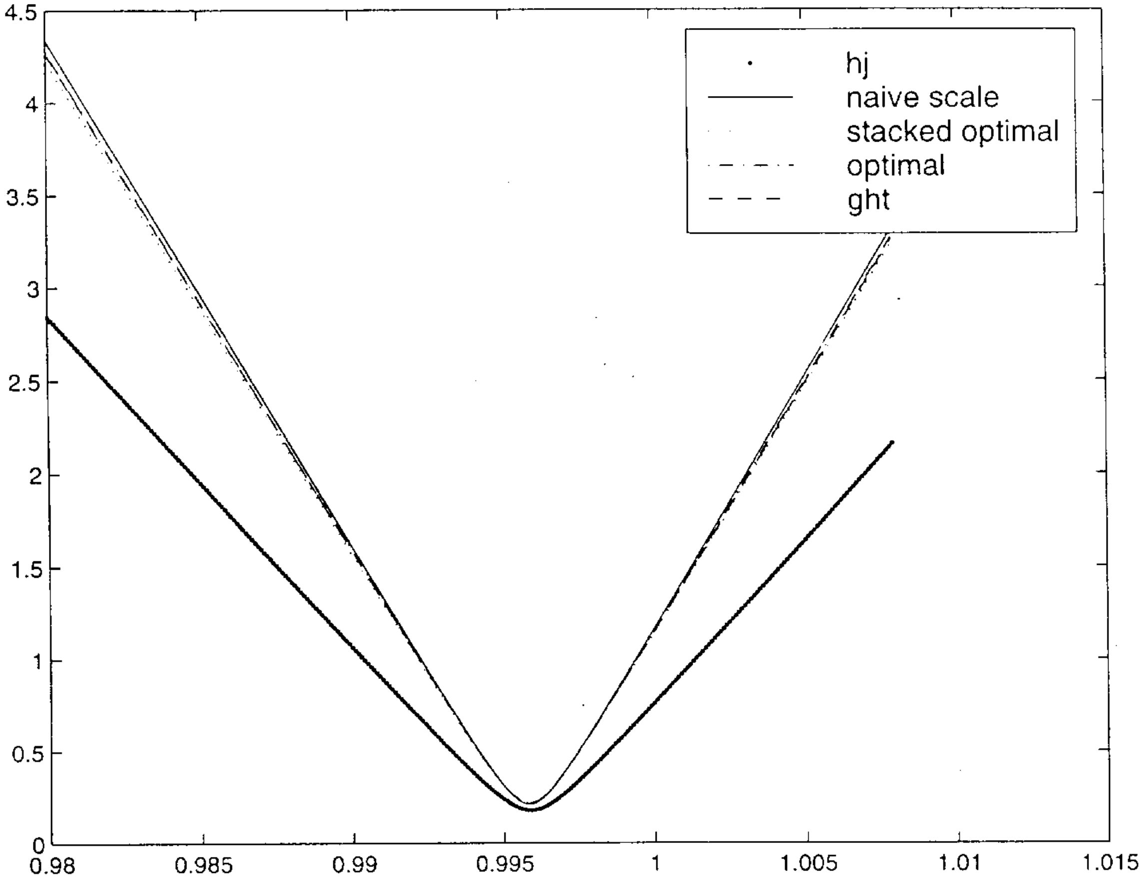
Table 4: Constrained GARCH-in-Mean Model

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	$0.0053\text{-}108.97h_{xxt}$	-0.264	0.734	0.0012
R^s_t	$0.0021\text{-}82.086h_{xxt}$	-0.264	0.734	0.0012
	$\gamma=$	14.675	$\beta=$	1.071
		(0.0376)		(0.0082)
	Constant	a_i	b_i	η_i
$h_{11,t}$	0.000022	-0.0652	0.00	0.3907
	(0.000006)	(0.0208)	(0.00)	(0.0876)
$h_{22,t}$	0.000013	0	0	0
	(0.000002)			
$h_{33,t}$	0.006457	0	0	0
	(0.001009)			
	f_{xb}=-0.0877		f_{xs}=1.847	
	(0.0813)		(0.0872)	

Notes: The model estimated imposes the following constraint on the unconstrained model reported in Table 3:

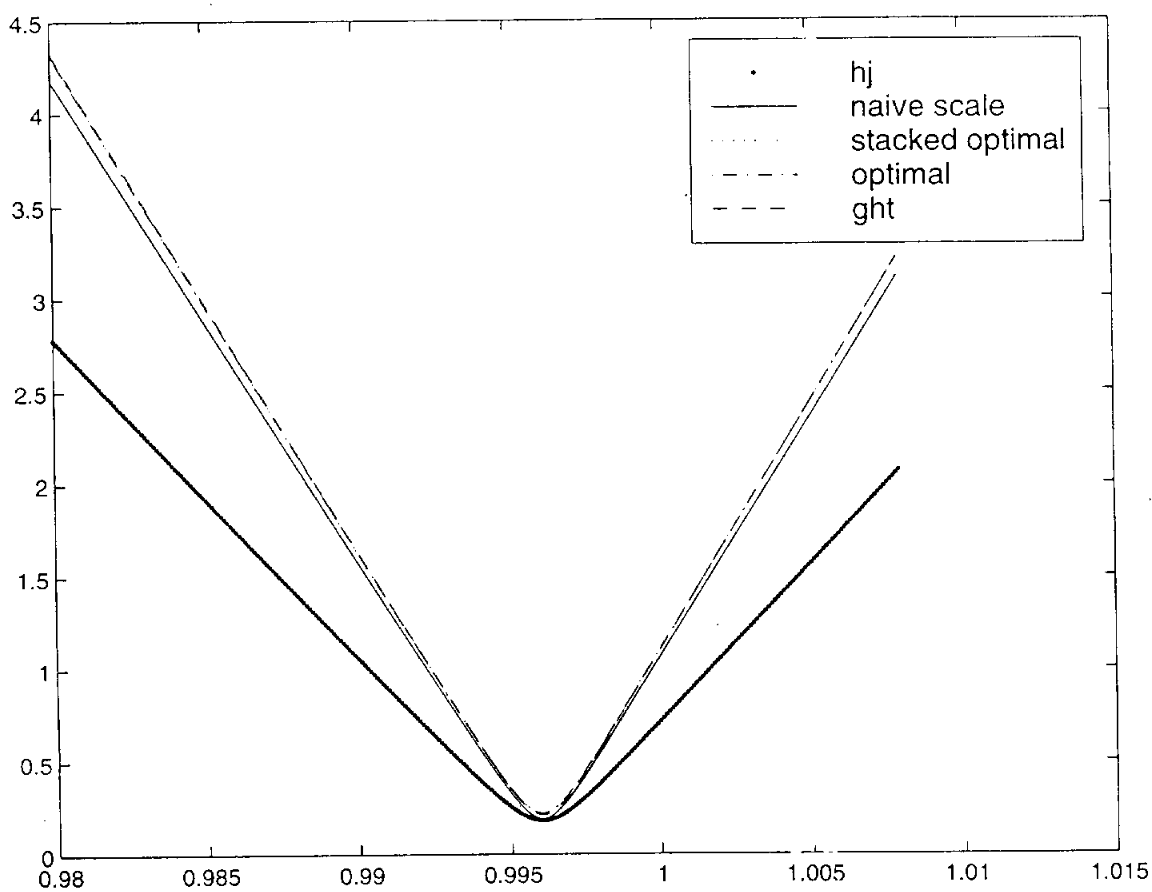
$$\mathrm{E}_t[R^i_{t+1}] = -(\log\beta + \frac{1}{2}h_{ii}) - \frac{1}{2}[\gamma - f_{xi}]^2 h_{xxt} + \gamma \mathrm{E}_t[X^i_{t+1}].$$

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



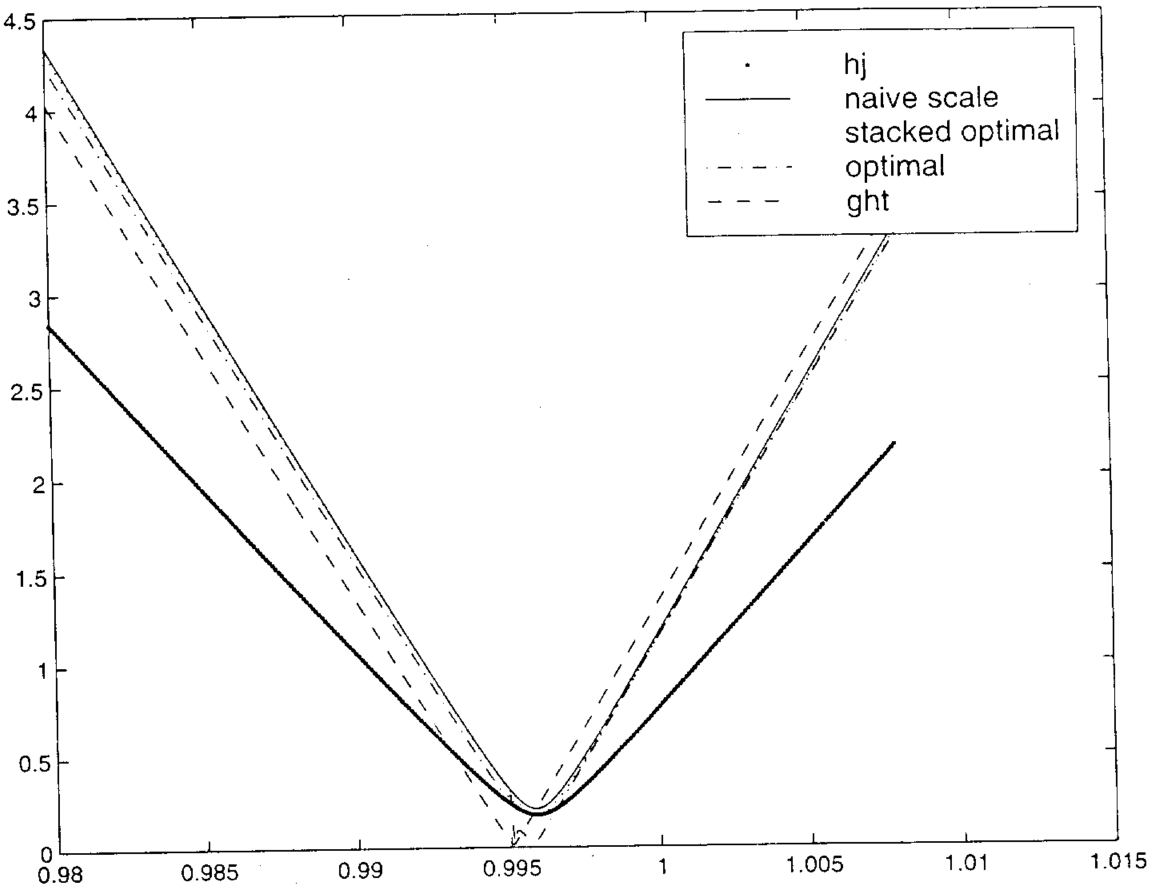
Conditional moments are calculated from the unconstrained model

Figure 1: Hansen-Jagannathan Bounds for Real Data



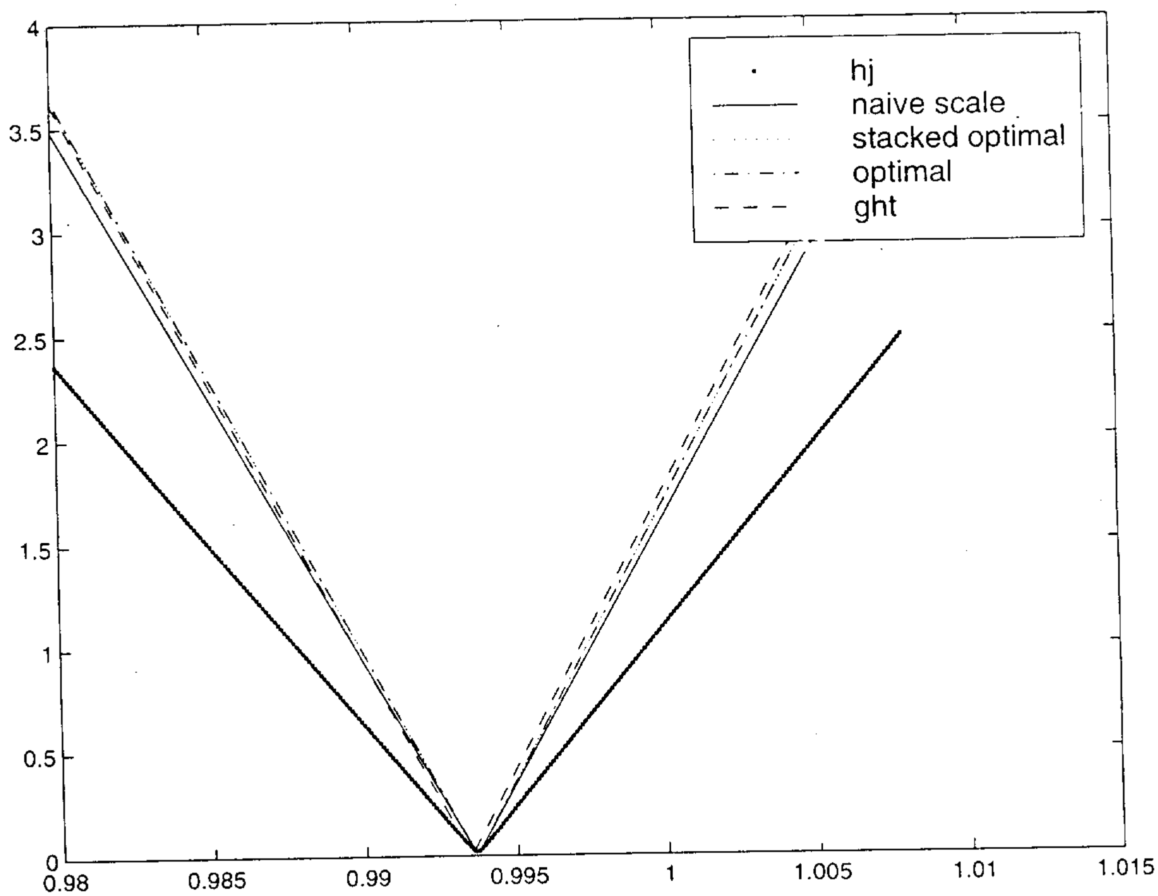
Conditional moments are calculated from the unconstrained model.

Figure 2: Hansen-Jagannathan Bounds for Simulated Data According to the Unconstrained Model



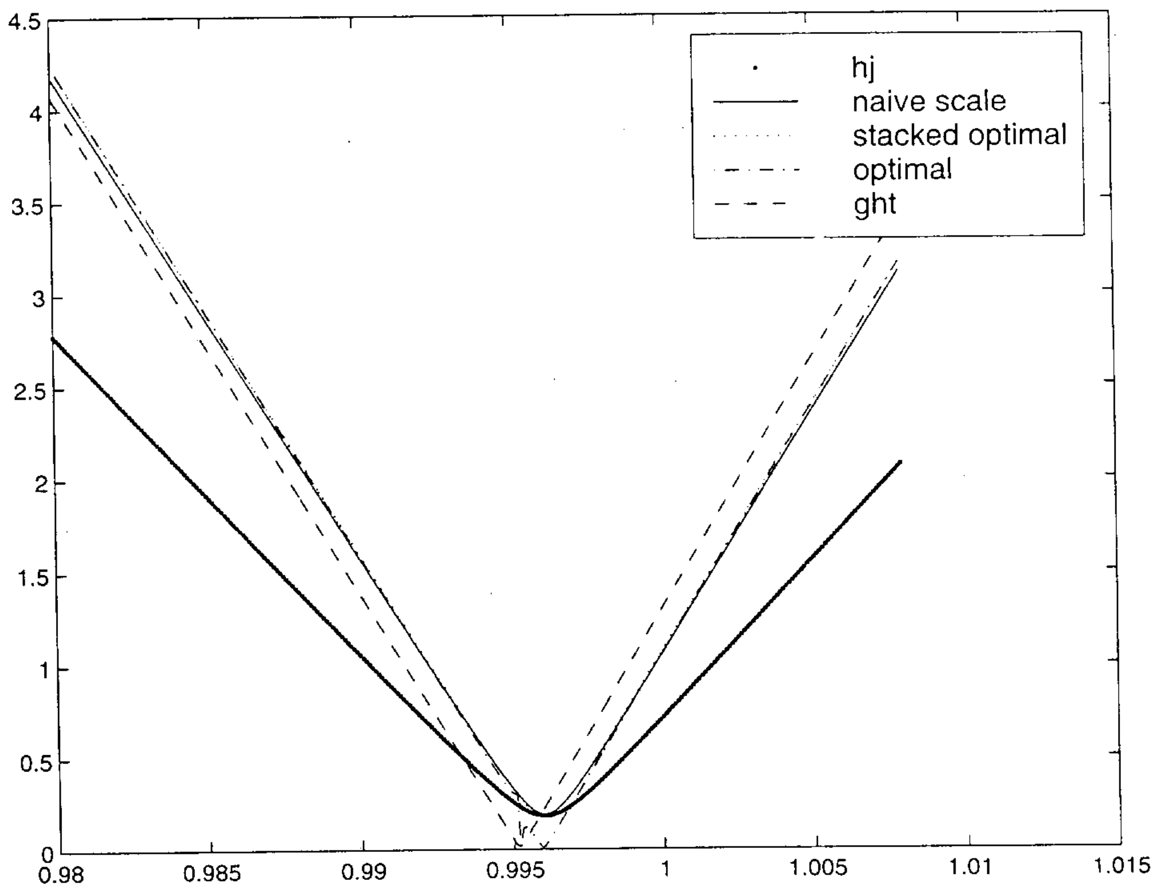
Conditional moments are calculated from the constrained model.

Figure 3: Hansen-Jagannathan Bounds for Real Data



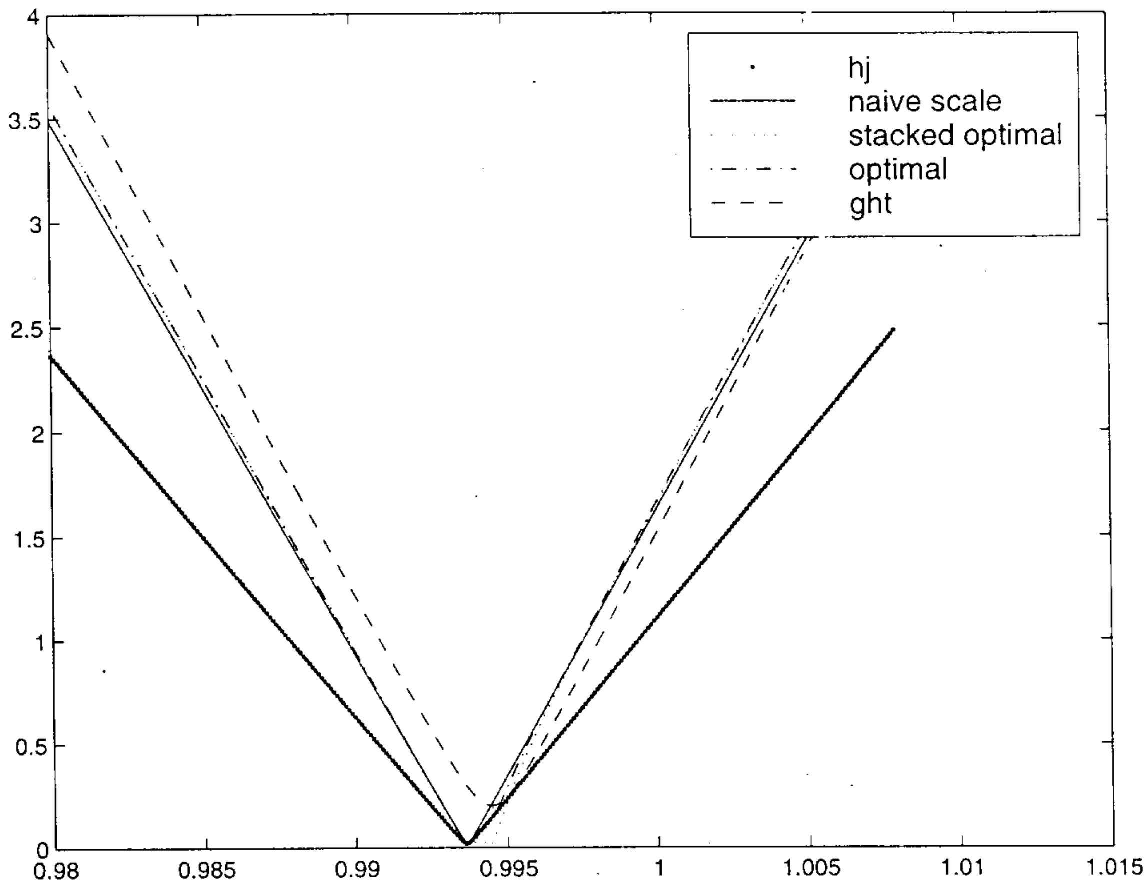
Conditional moments are calculated from the constrained model

Figure 4: Hansen-Jagannathan Bounds for Simulated Data According to the Constrained Model



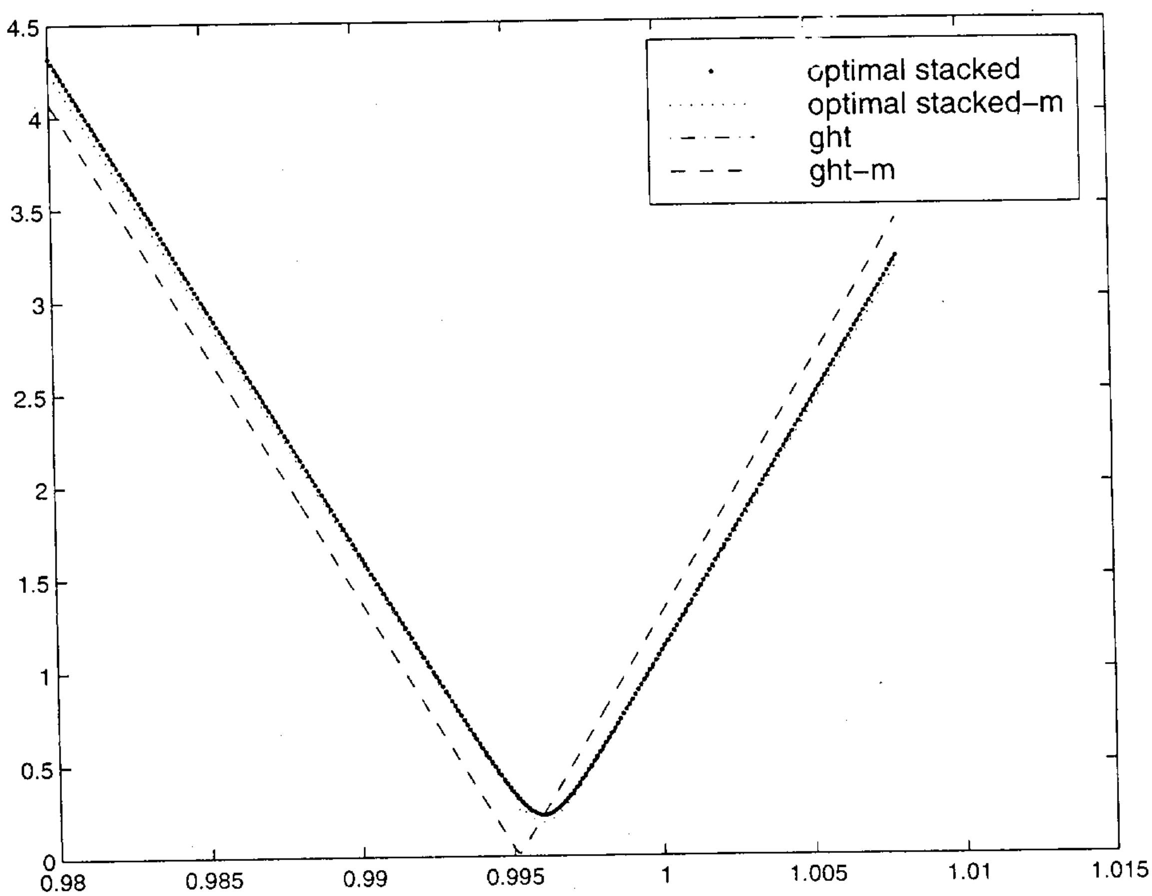
Conditional moments are calculated from the constrained model.

Figure 5: Hansen-Jagannathan Bounds for Simulated Data According to the Unconstrained Model



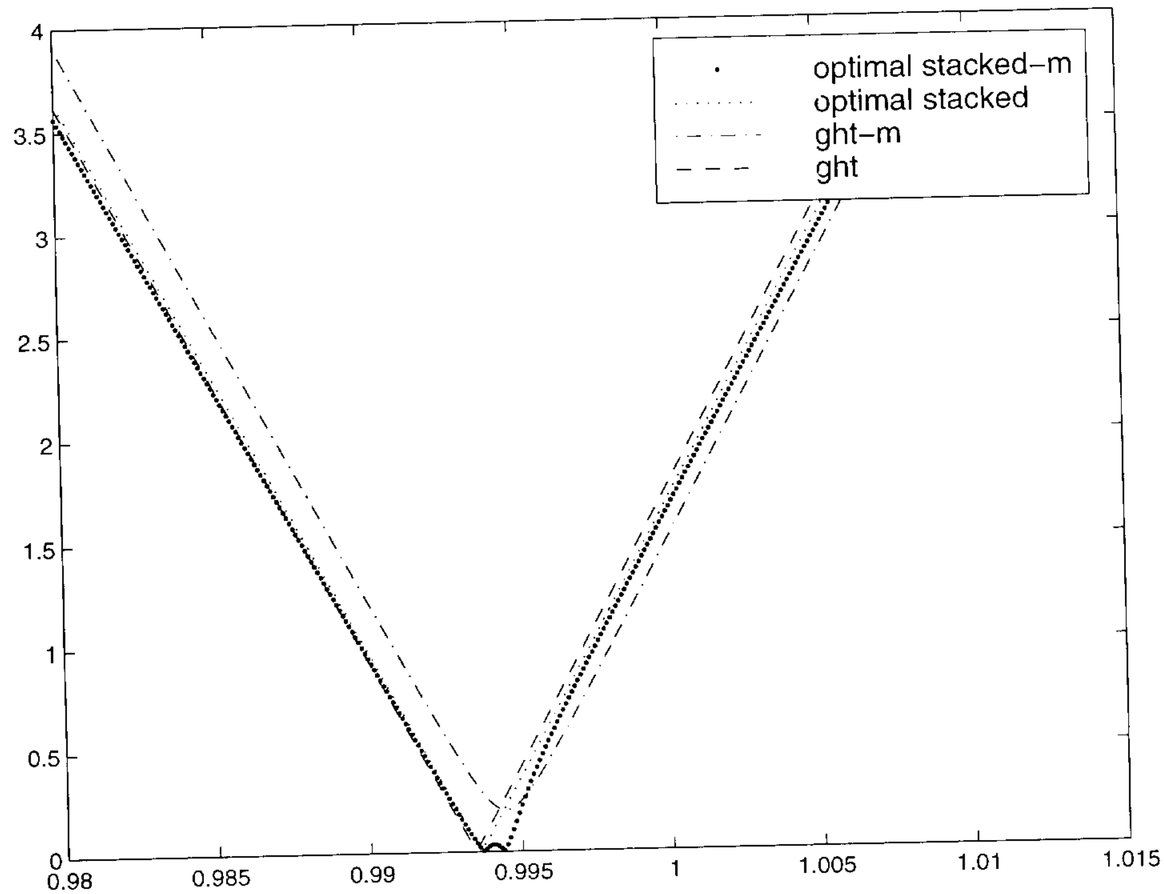
Conditional moments are calculated from the unconstrained model.

Figure 6: Hansen-Jagannathan Bounds for Simulated Data According to the Constrained Model



Conditional moments for "optimal stacked" and "ght" are calculated from the unconstrained model; conditional moments for "optimal stacked-m" and "ght-m" are calculated from the constrained model.

Figure 7: Hansen-Jagannathan Bounds for Simulated Data According to the Unconstrained Model



Conditional moments for "optimal stacked" and "ght" are calculated from the constrained model; conditional moments for "optimal stacked-m" and "ght-m" are calculated from the unconstrained model.

Figure 8: Hansen-Jagannathan Bounds for Simulated Data According to the Constrained Model

References

[1] Balduzzi, Pierluigi and Hédi Kallal (1997), "Risk Premiums and Variance Bound", Journal of Finance, v52, n5, 1913-49.

[2] Bekaert, Geert (1996), "The Time Variation of Risk and Return in Foreign Exchange Markets: A General Equilibrium Perspective", Review of Financial Studies, v9, n2, 427-70.

[3] Bekaert, Geert and Robert Hodrick (1992), "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets", Journal of Finance, v47, n2, 467-509.

[4] Bekaert, Geert and Campbell R. Harvey (1995), "Time-Varying World Market Integration", Journal of Finance, v50, n2, 403-44.

[5] Bekaert, Geert, Robert J. Hodrick and David A. Marshall (1997), "The Implications of First-Order Risk Aversion For Asset Market Risk Premiums", Journal of Monetary Economics, 40, 3-39.

[6] Bekaert, Geert and Jun Liu (1998), "Opitmally Exploiting International Return Predictability", work in progress, Stanford University.

[7] Bekaert, Geert and Michael Urias (1996), "Diversification, Integration and Emerging Market Closed-End Funds", Journal of Finance, v51, n3, 835-69.

[8] Bekaert, Geert and Guoju Wu (1998), "Asymmetric Volatility and Risk in Equity Returns", Working Paper, Stanford University.

[9] Burnside, Craig (1994), "Hansen-Jagannathan Bounds as Classical Tests of Asset-Pricing Models", Journal of Business and Economic Statistics, v12, n1, 57-79.

[10] Bonomo, Marco and Rene Garcia (1993), "Disappointment Aversion as a Solution to the Equity Premium and the Risk-Free Rate Puzzles", Cahier 2793, C.R.D.E., University de Montreal.

[11] Campbell, John and John H. Cochrane (1997), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior", Working Paper.

[12] Cecchetti, Stephen G., Pok-sang Lam, and Nelson C. Mark (1994), "Testing Volatility Restrictions on Intertemporal Marginal Rates of Substitution Implied by Euler Equations and Asset Returns", Journal of Finance, v49, n1, 123-52.

[13] Chen, Zhiwu and Peter J. Knez (1995), "Measurement of Market Integration and Arbitrage", Review of Financial Studies, v8, n2, 287-325.

[14] Chen, Zhiwu and Peter J. Knez (1996), "Portfolio Performance Measurement: Theory and Applications", Review of Financial Studies, v9, n2, 511-55.

[15] Cochrane, John H. (1996), "A Cross-Sectional Test of an Investment-Based Asset Pricing Model", Journal of Political Economy, v104, n3, 572-621.

[16] Cochrane, John H. and Lars. P. Hansen (1992), "Asset Pricing Explorations for Macroeconomics", Blanchard, Olivier Jean; Fischer, Stanley, eds.. 1992 NBER MACROECONOMICS, ANNUAL, 1992, MIT Press.

[17] Cochrane, John H. and Jesus Saa-Requejo (1996), "Beyond Arbitrage: "Good-Deal" Asset Price Bounds in Incomplete Markets", NBER Working Paper No. 5489.

[18] Constantinides, George M. (1990), "Habit Formation: A Resolution of the Equity Premium Puzzle", Journal of Political Economy, v98, n3, 519-43.

[19] Constantinides, G. M. and Darrell Duffie (1996), "Asset Pricing with Heterogeneous Consumers", Journal of Political Economy, v104, n2, 219-40.

[20] Dahlquist, Magnus and Paul Söderlind, 1997, "Evaluating Portfolio Performance with Stochastic Discount Factors", forthcoming, Journal of Business.

[21] Daniel, Kent and David Marshall (1996), "Consumption-Based Modeling of Long-Horizon Returns", Manuscript, University of Chicago.

[22] De Santis, Giorgio (1995), "Volatility Bounds for Stochastic Discount Factors: Tests and Implications from International Stock Returns", Working Paper, Univsersity of Southern California.

[23] Engle, Robert F., David M. Lilien, and Russell P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model", Econometrica, v55, n2, 391-407.

[24] Engle, Robert F., Victor K. Ng, and Michael Rothschild (1990), "Asset Pricing with a FACTOR-ARCH Covariance Structure: Empirical Estimates for Treasury Bills", Journal of Econometrics, v45, n1-2, 213-37.

[25] Epstein, Larry G. and Stanley Zin (1989), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", July 1989, ECONOMETRICA, v57, n4, 0012-9682.

[26] Epstein, Larry G. and Stanley Zin (1991a), "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis", Journal of Political Economy, v99, n2, 263-86.

[27] Epstein, Larry G. and Stanley Zin (1991b), "The Independence Axiom and Asset Returns", July 1991, National Bureau of Economic Research Technical Paper: 109.

[28] Fama, Eugene and Kenneth French (1989), "Business Conditions and Expected Returns on Stocks and Bonds", Journal of Financial Economics, 25, 23-49.

[29] Ferson, Wayne E. and George. M. Constantinides (1991), "Habit Persistence and Durability in Aggregate Consumption: Empirical Tests", Journal of Financial Economics, v29, n2, 199-240.

[30] Ferson, Wayne E. and Rudi W. Schadt (1996), "Measuring Fund Strategy and Performance in Changing Economic Conditions", Journal of Finance, v51, n2, 425-61.

[31] Ferson, Wayne and Andrew Siegel (1997a), "The Efficient Use of Conditioning Information in Portfolios", manuscript, University of Washington.

[32] Ferson, Wayne and Andrew Siegel (1997b), "Optimal Moment Restrictions on Stochastic Discount Factors", manuscript, University of Washington.

[33] Filardo, Andrew J. (1994), "Business-Cycle Phases and Their Transitional Dynamics", Journal of Business and Economic Statistics, v12, n3, 299-308.

[34] French, Kenneth R., G. William Schwert and Robert F. Stambaugh (1987), "Expected Stock Returns and Volatility", Journal of Financial Economics, v19, n1, 3-29.

[35] Gallant, Ronald and George Tauchen (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications" Econometrica, v57, n5, 1091-1120.

[36] Gallant, Ronald, Lars P. Hansen and George Tauchen (1990), "Using Conditional Moments of Asset Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution", Journal of Econometrics, v45, n1-2, 141-79.

[37] Glosten, Lawrence R. Ravi Jagannathan, and David E. Runkle, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", Journal of Finance, v48, n5, 1779-1801.

[38] Hansen, Lars Peter, John Heaton, Erzo G. J. Luttmer (1995), "Econometric Evaluation of Asset Pricing Models", Review of Financial Studies, v8, n2, 237-74.

[39] Hansen, Lars P. and Ravi Jagannathan (1991), "Implications of Security Market Data for Models of Dynamic Economies", Journal of Political Economy, v99, n2, 225-62.

[40] Hansen, Lars. P. and Scott. F. Richard (1987), "The Role of Conditioning Information in Deducing Testable", Econometrica, v55, n3, 587-613.

[41] Hansen, Lars P. and Kenneth Singleton (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns" Journal of Political Economy, v91, n2, 249-65.

[42] Hansen, Lars. P. and Kenneth Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", Econometrica, v50, n5, 1269-86.

[43] He, Hua and David. M. Modest (1995), "Market Frictions and Consumption-Based Asset Pricing", Journal of Political Economy, v103, n1, 94-117.

[44] Heaton, John (1995), "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications", Econometrica, v63, n3, 681-717.

[45] Heaton, John and Deborah J. Lucas (1996), "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing", Journal of Political Economy, v104, n3, 443-87.

[46] Kaminsky, Graciela and Rodrigo Peruga (1990), "Can a Time-Varying Risk Premium Explain Excess Returns in the Forward Market for Foreign Exchange?", Journal of International Economics, v28, n1/2, 47-70.

[47] Kandel, Shmuel and Robert F. Stambaugh, 1990, "Expectations and Volatility of Consumption and Asset Returns", Review of Financial Studies, v3 n2, 207-32.

[48] Kocherlakota, Narayana R. (1996), "The Equity Premium: It's Still a Puzzle", Journal of Economic Literature, v34, n1, 42-71.

[49] Mehra, Rajnish and Edward G. Prescott, "The Equity Premium: A Puzzle", Journal of Monetary Economics, v15, n2, 145-61.

[50] Luttmer, Erzo G. J. (1996), "Asset Pricing in Economies with Frictions" Econometrica, v64, n6, 1439-67.

[51] Rietz, Thomas A. (1988), "The Equity Risk Premium: A Solution" Journal of Monetary Economics, v22, n1, 117-31.

[52] Sichel, Daniel E. (1993), "Business Cycle Asymmetry: A Deeper Look", Economic Inquiry, v31, n2, 224-36.

[53] Singleton, Kenneth (1990), in *Advances in Economic Theory: World Congress of the Econometric Society*, Jean-Jacques Laffont ed., Cambridge University Press, 1995.

[54] Weil, Philippe (1989) "The Equity Premium Puzzle and the Risk-Free Rate Puzzle", Journal of Monetary Economics, v24, n3, 401-21.

[55] Snow, Karl N. (1991), "Diagnosing Asset Pricing Models Using the Distribution of Asset Returns", Journal of Finance, v46, n3, 955-83.

[56] White, Halbert (1982), " Maximum Likelihood Estimation of Misspecified Models", Econometrica, v50, n1, 1-25.