TECHNICAL WORKING PAPER SERIES

ECONOMETRIC EVALUATION
OF ASSET PRICING MODELS

Lars Peter Hansen
John Heaton
Erzo Luttmer

Technical Working Paper No. 145

ECONOMETRIC EVALUATION
OF ASSET PRICING MODELS

ABSTRACT

In this paper we provide econometric tools for the evaluation of intertemporal asset pricing models using specification-error and volatility bounds. We formulate analog estimators of these bounds, give conditions for consistency and derive the limiting distribution of these estimators. The analysis incorporates market frictions such as short-sale constraints and proportional transactions costs. Among several applications we show how to use the methods to assess specific asset pricing models and to provide nonparametric characterizations of asset pricing anomalies.

Lars Peter Hansen
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER

John Heaton
Sloan School of Management
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge, MA 02139
and NBER

Erzo Luttmer
Kellogg Graduate School of Management
Northwestern University
Evanston, IL 60208

I. Introduction

Frictionless market models of asset pricing imply that asset prices can be represented by a stochastic discount factor or pricing kernel. For example, in the Capital Asset Pricing Model (CAPM) the discount factor is given by a constant plus a scale multiple of the return on the market portfolio. In the Consumption-Based CAPM (CCAPM) the discount factor is given by the intertemporal marginal rate of substitution of an investor. If r is the net return on an asset and m is the marginal rate of substitution, then the CCAPM implies that:

$$(1.1) \qquad 1 \; = \; E[m(1+r) \mid \mathscr{I}]$$

where $\mathscr{I}$ is the information set of the investor today. More generally, if m is the stochastic discount factor, today's price, $\pi(p)$, of an asset payoff, p, tomorrow is given by:

$$(1.2) \qquad \pi(p) \; = \; E(mp \mid \mathscr{I}) \; .$$

Thus a stochastic discount factor m "discounts" payoffs in each state of the world and, as a consequence, adjusts the price according to the riskiness of the payoff. From the vantage point of an empirical analysis, we envision the stochastic discount factor as the vehicle linking a theoretical model to observable implications.

Given a particular model for the stochastic discount factor, the implications of (1.2) can be assessed by first taking unconditional expectations, yielding

1

$$(1.3) \qquad E\pi(p) \;=\; E(mp).$$

When m is observable (at least up to a finite-dimensional parameter vector) by the econometrician, a test of (1.2) can be performed using a time series of a vector of portfolio payoffs and prices by examining whether the sample analogs of the left and right sides of (1.2) are significantly different from each other. Examples of this type of procedure can be found in Hansen and Singleton (1982), Brown and Gibbons (1985), MacKinlay and Richardson (1991) and Epstein and Zin (1991).

While tests such as these can be informative, it is often difficult to interpret the resulting statistical rejections. Further, these tests are not directly applicable when there are market frictions such as transactions costs or short-sale constraints. For example, when an asset cannot be sold short, (1.2) is replaced with the pricing inequality:

$$(1.4) \qquad \pi(p) \;\geq\; E(mp\,|\,\mathscr{I}) \quad .$$

Finally, these tests can not be used when the candidate discount factor depends on variables unavailable to the econometrician.

As an alternative to testing directly pricing errors using (1.3), we consider a different set of tests and diagnostics using the specification-error bounds of Hansen and Jagannathan (1993), and the volatility bounds of Hansen and Jagannathan (1991). We also consider extensions of these tests and diagnostics, developed by He and Modest (1992) and Luttmer (1993), that handle transactions costs, short-sale restrictions and other market frictions. We develop an econometric methodology to provide

2

consistent estimators of the specification-error and volatility bounds. Further, we develop asymptotic distribution theory that is easy to implement and that can be used to make statistical inferences about asset pricing models and asset market data using the bounds. The specification-error and volatility bounds, along with the econometric methodology that we develop, can be applied to address several related issues.

The specification-error bounds of Hansen and Jagannathan (1993) can be used to examine a discount factor *proxy* that does not necessarily correctly price the assets under consideration (see also Bansal, Hsieh and Viswanathan 1992 for an application). This is important since formal statistical tests of many particular models of asset pricing imply that the hypothesis that their pricing errors are zero is a very low probability event. Since these models are typically very simple, it is perhaps not surprising that they do not completely capture the complexity of pricing in financial markets. The specification-error bounds give measures of the maximum pricing error made by the discount factor proxy. This provides a way to assess the usefulness of a model even when it is technically misspecified. Further, this tool can easily accommodate market frictions such as transactions costs and short-sale constraints.

Given a vector of asset payoffs and prices, (1.3) typically does not uniquely determine *m*. Instead there is a whole family of *m*'s that will work. Any parametric model for *m* imposes additional restrictions on that family, often sufficient to *identify* a unique stochastic discount factor. Rather than imposing these extra restrictions, Hansen and Jagannathan (1991) showed how asset market data on payoffs and prices can be used to construct feasible sets for means and standard deviations of stochastic discount factors. The boundary points of these regions provide lower bounds on the

3

volatility (standard deviation) indexed by the mean. He and Modest (1992) and Luttmer (1993) showed how to extend this analysis to the case where some of the assets are subject to transactions costs or short-sales constraints.

These feasible sets of means and standard deviations of the stochastic discount factor can be used to isolate those aspects of the asset market data that are most informative about the stochastic discount factor. One way to do this is to ask whether the volatility bound becomes significantly sharper as more asset market data is added to the analysis. This would help one assess the incremental importance of additional security market data in an econometric analysis without having to limit *a priori* the family of stochastic discount factors. More generally, it is valuable to have a characterization of the sense in which an asset market data set is puzzling without having to take a precise stand on the underlying valuation model.

When testing a particular model of asset pricing in which the candidate m is specified, it is often useful to examine whether the candidate is in the feasible region. Moreover, when diagnosing the failures of a specific model, it is valuable to determine whether the candidate discount factor is not sufficiently volatile or whether it is other aspects of the *joint* distribution of asset payoffs and the candidate discount factor that are problematic.

As we remarked previously, sometimes it is not possible to construct direct observations of m, making pricing-error tests infeasible. However, it may still be possible to calculate the moments of a stochastic discount factor implied by a model which can then be compared to the volatility bounds. For example in Heaton (1993) a consumption-based CAPM model is examined in which the consumption data is time averaged and preferences are such that a simple linearization of the utility function can not be done to

4

account for the time-averaging. Although it is possible to simulate the entire model and compare the moments of asset prices predicted by the model to those in the data, this is very difficult to do for a large cross-section of assets. However, it is a simple matter to calculate the implications of the mean and standard-deviation of a model for m and compare them to the bounds implied by a large number of assets.

Recently Burnside (1992) and Cecchetti, Lam and Mark (1992) showed how to test models of the discount factor using a different parameterization of the volatility bound. Our formulation and analysis can be applied to develop tests that are simpler to implement and do not require the calculation of numerical derivatives for statistical inference. Also, our tests can accommodate market frictions in a straightforward manner.

The rest of the paper is organized as follows. In Section II we review the specification and volatility bounds of Hansen and Jagannathan (1991, 1993), He and Modest (1992) and Luttmer (1993). We show formally that the volatility bound can be viewed as a special case of the specification-error bound. This permits us to developed the underlying econometric tools in a unified way. We then exploit duality theory to obtain simple maximization problems that can be used to compute the bounds and are amenable to econometric analysis. One of these maximization problems incorporates more fully the implications of the absence of arbitrage opportunities.

In Section III we provide consistency and asymptotic distribution results for estimators of the bounds. Section III.A presents the notation that we use throughout the rest of the paper. Section III.B presents our consistency results. Sections III.C and III.D present our results on the asymptotic distribution of the estimators of the bound and shows how to calculate standard errors for the estimators. Section III.E presents the

consistency result for estimators of the arbitrage bounds. Those uninterested in the consistency results, but who are interested in the calculations necessary for conducting statistical inference, need only read Sections III.A., III.C and III.D before moving to Section IV.

In Section IV we present several applications and extensions of our results each of which can be read independently after reading Sections II, III.A, III.C and III.D. In Section IV.A we discuss the sense in which the entire feasible set of means and standard deviations for the stochastic discount factor can be estimated. Section IV.B provides a discussion of tests of whether the volatility bound becomes sharper with additional asset market data. Section IV.C shows how to use the volatility bounds to test models of the discount factor. Finally in Section IV.D we extend the specification-error bound to a case where there are parameters of the discount factor proxy that are unknown and must be estimated. Section V contains some concluding remarks.

II. General Model and Bounds

Our starting point is a model in which asset prices are represented by a *stochastic discount factor* or pricing kernel. To accommodate security market pricing subject to transactions costs, we permit there to be short-sale constraints for a subset of the securities. Although a short-sale constraint is an extreme version of a transactions cost, other proportional transactions costs such as bid-ask spreads can also be handled with this formalism. This is done as in Foley (1970), Jouni and Kallal (1992) and Luttmer (1993) by constructing two payoffs according to whether a security is purchased or sold. A short-sale constraint is imposed on both

artificial securities to enforce the distinction between a buy and a sell, and a bid-ask spread is modeled by making the purchase price higher than the sale price.

Suppose the vector of security market payoffs used in an econometric analysis is denoted x with a corresponding price vector q. The vector of x is used to generate a collection of payoffs formed using portfolio weights in a closed convex cone C of R^n:

(2.1) $P = \{p : p = \alpha'x$ for some $\alpha \in C\}$.

The cone C is constructed to incorporate all of the short-sale constraints imposed in the econometric investigation. If there are no price distortions induced by market frictions, then C is R^n. More generally, partition x into two components: $x' = [x^n{}', x^s{}']$ where x^n contains the k components whose prices are not distorted by market frictions and x^s contains the ℓ components subject to short-sale constraints. Then the cone C is formed by taking the Cartesian product of R^k and the nonnegative orthant of R^ℓ.

Let q denote the random vector of prices corresponding to the vector x of securities payoffs. These prices are observed by investors at the time assets are traded and are permitted to be random because the prices may reflect conditioning information available to the investors. Since it is difficult to model empirically this conditioning information, we instead work with the average or expected price vector Eq.

While information may be lost in our failure to model explicitly the conditioning information of investors, some conditioning information can be incorporated in the following familiar *ad hoc* manner. Suppose some of the security payoffs used in an econometric analysis are one-period stock or

7

bond-market returns with prices equal to one by construction. Additional *synthetic payoffs* can be formed by an econometrician by taking one of the original returns, say x^1, and multiplying it by a random variable, say z, in the conditioning information set of economic agents. The corresponding constructed payoff is then x^1z with a price of z. Hence the price of the synthetic payoff is random even though the price of original security is constant. If x^1 is subject to a short-sale constraint, then z should be nonnegative.

The vehicle linking payoffs to average prices is a *stochastic discount factor*. To represent formally this link and provide a characterization of a stochastic discount factor, we introduce the *dual* of C, which we denote C^*. This dual consists of all vectors in R^n whose dot product with every element of C is nonnegative. For instance, when C is all of R^n, C^* consists only of the zero vector. More generally, if x can be partitioned in the manner described previously, the elements of C^* are of the form $(0,\beta')'$ where β is nonnegative.

A stochastic discount factor m is a random variable that satisfies the pricing relation:

(2.2) $Eq - Emx \in C^*$.

To interpret this relation, first consider the case in which C is R^n. Then there are no market frictions and we have linear pricing. In this case relation (2.2) is the familiar pricing equality because C^* has only one element, namely the zero vector. Consider next the case in which x can be partitioned into the two components described previously. Partition q comparably, and relation (2.2) becomes:

(2.3) $Eq^a - Emx^a = 0$

          $Eq^s - Emx^s \geq 0.$

The inequality restriction emerges because pricing the vector of payoffs x^s subject to short-sale constraints must allow for the possibility that these constraints bind and hence contribute positively to the market price vector.

II.A: Maintained Assumptions

There are three restrictions on the vector of payoffs and prices that are central to our analysis. The first is a moment restriction, the second is equivalent to the absence of arbitrage on the space of portfolio payoffs, and the third eliminates redundancy in the securities.

For pricing relation (2.2) to have content, we maintain:

Assumption 2.1: $E|x|^2 < \infty, E|q| < \infty.$

Assumption 2.2:   There exists an $m > 0$ satisfying (2.2) such that $Em^2 < \infty.$

The positivity component of Assumption 2.2 can often be derived from the Principle of No-Arbitrage (*e.g.*, see Kreps 1981, Prisman 1986, Jouni and Kallal 1992 and Luttmer 1993). The Principle of No-Arbitrage specifies that the smallest cost associated with any payoff that is nonnegative and not identically equal to zero must be strictly positive. Notice that from (2.2), a stochastic discount factor m satisfies:

(2.4) $\alpha' E(mx) \leq \alpha' Eq$ for any $\alpha \in C,$

9

which shows that Assumption 2.2 implies the Principle of No-Arbitrage (applied to expected prices).

Next we limit the construction of x by ruling out redundancies in the securities:

Assumption 2.3: If $\alpha'x = \alpha^{*\prime}x$ and $\alpha'Eq = \alpha^{*\prime}Eq$ for some α and α^* in C, then $\alpha = \alpha^*$.

In the absence of transaction costs, Assumption 2.3 precludes the possibility that the second moment matrix of x is singular. Otherwise, there would exist a nontrivial linear combination of the payoff vector x that is zero with probability one. In light of (2.2), the (expected) price of this nontrivial linear combination would have to be zero, violating Assumption 2.3. To accommodate securities whose purchase price differs from the sale price, we permit the second moment matrix of the composite vector x to be singular. Assumption 2.3 then requires that distinct portfolio weights used to construct the same payoff must have distinct expected prices.[1]

II.B: Minimum Distance Problems

There are two problems that underlie most of our analysis. Let $\mathcal{M}$ denote the set of all random variables with finite second moments that satisfy (2.2), and let $\mathcal{M}^+$ be the set of all nonnegative random variables in $\mathcal{M}$. In light of Assumption 2.2, both sets are nonempty. Let y denote some "proxy" variable for a stochastic discount factor that, strictly speaking, does not satisfy relations (2.2). Following Hansen and Jagannathan (1993),

we consider the following two *ad hoc* least squares measures of misspecification:

$$(2.5) \qquad \hat{\delta}^2 \;=\; \min_{m \in \mathcal{M}} E[(y - m)^2] \,,$$

and

$$(2.6) \qquad \hat{\gamma}^2 \;=\; \min_{m \in \mathcal{M}^+} E[(y - m)^2]$$

When the proxy y is set to zero, the minimization problems collapse to finding bounds on the second moment of stochastic discount factors as constructed by Hansen and Jagannathan (1991), He and Modest (1992) and Luttmer (1993). In particular, the bounds derived in Hansen and Jagannathan (1991) are obtained by setting y to zero and solving (2.5) and (2.6) when there are no short-sale constraints imposed (when C is set to R^n); the bound derived in He and Modest is obtained by solving (2.5) for y set to zero; and the bound derived by Luttmer (1993) is obtained by solving (2.6) for y set to zero. These second moment bounds will subsequently be used in deriving feasible regions for means and standard deviations. Clearly, the second moment bound implied by (2.6) is no smaller than that implied by (2.5) since it is obtained using a smaller constraint set.

Next consider the case in which the proxy y is not degenerate. Hansen and Jagannathan (1993) showed that the least squares distance between a proxy and the set $\mathcal{M}$ of (possibly negative) stochastic discount factors has an alternative interpretation of being the maximum pricing error per unit norm of payoffs in P, where the norm of a payoff is the square root of its

11

second moment. When the constraint set is shrunk to $\mathcal{M}^+$ as in problem (2.6), the dual interpretation takes account of potential pricing errors for hypothetical derivative claims. While Hansen and Jagannathan (1993) abstract from short-sale constraints in their analysis, pricing-error interpretations are applicable more generally.[2]

II.C: Conjugate Maximization Problems

In solving the least squares problems (2.5) and (2.6) and in our development of econometric methods associated with those problems, it is most convenient to study the conjugate maximization problems. They are given by

$$(2.7) \qquad \hat{\delta}^2 = \max_{\alpha \in C} \{Ey^2 - E[(y - x'\alpha)^2] - 2\alpha' Eq\} \quad ,$$

and

$$(2.8) \qquad \hat{\gamma}^2 = \max_{\alpha \in C} \{Ey^2 - E[(y - x'\alpha)^{+2}] - 2\alpha' Eq\}$$

where the notation h^+ denotes $\max\{h,0\}$. The conjugate problems are obtained by introducing Lagrange multipliers on the pricing constraints (2.2) and exploiting the familiar saddle point property of the Lagrangian. The α's then have interpretations as the multipliers on the pricing constraints.

The conjugate problems in (2.7) and (2.8) are convenient because the choice variables are finite-dimensional vectors whereas the choice variables in the original least squares problems are random variables that reside in possibly infinite-dimensional constraint sets. The specifications of the

conjugate problems are justified formally in Hansen and Jagannathan (1993) and Luttmer (1993). Of particular interest to us is that the criteria for the maximization problems are concave in α and that the first-order conditions for the solutions are given by:

$$(2.9) \quad Eq - E[(y - x'\hat{\alpha})x] \in C^*$$

in the case of problem (2.7) and

$$(2.10) \quad Eq - E[(y - x'\tilde{\alpha})^+x] \in C^*$$

along with the respective complementary slackness conditions

$$(2.11) \quad \hat{\alpha}'Eq - \hat{\alpha}'E[(y - x'\hat{\alpha})^+x] = 0,$$

and

$$(2.12) \quad \tilde{\alpha}'Eq - \tilde{\alpha}'E[(y - x'\tilde{\alpha})^+x] = 0.$$

In fact, optimization problem (2.7) is a standard quadratic programming problem. Interpreting the first-order conditions for these problems, observe that associated with a solution to problem (2.7) is a random variable $\hat{m} = (y - x'\hat{\alpha})$ in $\mathcal{M}$ and associated with a solution to problem (2.8) is a nonnegative random variable $\tilde{m} = (y - x'\tilde{\alpha})^+$ in $\mathcal{M}^+$. These random variables are the unique (up to the usual equivalence class of random variables that are equal with probability one) solutions to the original least squares problems.

Since Assumption 2.3 eliminates redundant securities and the random

13

variable $(y - x'\hat{\alpha})$ is uniquely determined, the solution $\hat{\alpha}$ to conjugate problem (2.7) is also unique. This follows because the value of the criterion must be the same for all solutions, implying that they all must have the same expected price $\hat{\alpha}'Eq$. The solution to conjugate problem (2.8) may not be unique, however. In this case the truncated random variable $(y - x'\tilde{\alpha})^+$ is uniquely determined, as is the expected price $\tilde{\alpha}'Eq$. On the other hand, the random variable $(y - x'\tilde{\alpha})$ is not necessarily unique, so we can not exploit Assumption 2.3 to verify that the solution $\tilde{\alpha}$ is unique. As we will now demonstrate, the set of solutions is convex and compact.

The convexity follows immediately from the concavity of the criterion function and the convexity of the constraint set. Similarly, the set of solutions must be closed because the constraint set is closed and the criterion function is continuous.

Boundedness of the set of solutions can be demonstrated by investigating the tail properties of the criterion functions. We consider two cases: directions θ for which $\theta'x$ is negative with positive probability and directions θ for which $\theta'x$ is nonnegative. To study the former case we take the criterion in (2.8) and divide it by $1 + |\alpha|^2$. For large values of $|\alpha|$ the scaled criterion is approximately:

$$(2.13) \qquad - E[(-x'\theta)^{+2}] \qquad \text{where } \theta = \alpha/[1 + |\alpha|^2]^{1/2} .$$

Hence $|\theta|$ is approximately one for large values of $|\alpha|$. Moreover, $\theta'x$ is a payoff in P. Consequently, the unscaled criterion will decrease (to $-\infty$) quadratically for large values $|\alpha|$.

Consider next directions θ for which $\theta'x$ is nonnegative. From Assumption 2.2 and relation (2.4) we have that

14

$$(2.14) \quad Em(\theta'x) \leq \theta'Eq$$

for some m that is strictly positive with probability one. Hence $\theta'Eq$ must be strictly positive unless $\theta'x$ is identically zero. However, when $\theta'x$ is identically zero, it follows from Assumption 2.3 and inequality (2.14) that $\theta'Eq$ is still strictly positive.

For directions θ for which the payoff $\theta'x$ is nonnegative, we study the tail behavior of the criterion after dividing by $(1 + |\alpha|^2)^{1/2}$, which yields approximately $-\theta'Eq$ for large values of $|\alpha|$. Hence in these directions the the unscaled criterion must diminish (to $-\infty$) at least linearly in $|\alpha|$. Thus in either case, we find that the set of solutions to conjugate problem (2.8) is bounded.

For some but not all of the results in the subsequent sections, we will need for there to exist a unique solution to conjugate problem (2.8). Since the set of solutions is convex, local uniqueness implies global uniqueness. To display a sufficient condition for local uniqueness, let x^* denote the component of the composite payoff vector x for which the pricing relation is satisfied with equality:

$$(2.15) \quad E\tilde{m}x^* = Eq^*$$

where q^* is the corresponding price vector. Also, let $1_{\{\tilde{m} > 0\}}$ be the the indicator function for the event $\{\tilde{m} > 0\}$. A sufficient condition for local uniqueness is that

Assumption 2.4: $Ex^*x^{*\prime}1_{\{\tilde{m} > 0\}}$ is nonsingular.

To see why this is a valid sufficient condition, observe that from the complementary slackness condition (2.12), $\tilde{m}$ is given by $(y - x^{*\prime}\hat{\beta})^+$ for some vector $\hat{\beta}$. Consequently,

$$(2.16) \qquad Eq^* = Ey1_{\{\tilde{m} > 0\}} - E(x^*x^{*\prime}1_{\{\tilde{m} > 0\}})\hat{\beta} \ .$$

When the matrix $E(x^*x^{*\prime}1_{\{\tilde{m} > 0\}})$ is nonsingular, we can solve (2.16) for $\hat{\beta}$.

II.D: Volatility Bounds and Restrictions on Means

The second moment bounds described in the previous subsection can be converted into standard deviation bounds via the formulas:

$$(2.17) \qquad \hat{\sigma} = [\hat{\delta}^2 - (Em)^2]^{1/2}$$
$$\tilde{\sigma} = [\tilde{\delta}^2 - (Em)^2]^{1/2}$$

where $\hat{\delta}^2$ and $\tilde{\delta}^2$ are constructed by setting the proxy to zero. When P contains a unit payoff, Em is also equal to the average price of that payoff and hence is restricted to be between the sale and purchase prices of the unit payoff. However, data on the price of a riskless payoff is often not available, so that it is difficult to determine Em. In these circumstances, bounds can be obtained for each choice of Em by adding a unit payoff to P (augmenting x with a 1) and assigning a price of v to that payoff (augmenting Eq with v). In forming the augmented cone, there should be no short sale constraints imposed on the additional security and hence no new price distortions should be introduced. The price assignment v is equivalent to a mean assignment for m. Mean-specific volatility bounds can then be obtained

16

using (2.7), (2.8) and (2.17).

The Principle of No-Arbitrage puts a limit on the admissible values of v: $v \in [\lambda_0, v_0]$ where λ_0 is the lower arbitrage bound and v_0 is the upper arbitrage bound. These bounds are computed using formulas familiar from derivative claims pricing:

(2.18) $\lambda_0 = - inf\{\alpha'Eq : \alpha \in C \text{ and } \alpha'x \geq -1\}$

and

(2.19) $v_0 = inf\{\alpha'Eq : \alpha \in C \text{ and } \alpha'x \geq 1\}$.

While λ_0 is always well defined via (2.18), v_0 may not be because there may not exist a payoff in P that dominates a unit payoff. In such circumstances, we define v_0 to be $+\infty$.

III. Econometric Issues

In this section we develop consistency and asymptotic distribution results for the specification-error bounds presented in Section II. A key presumption underlying our analysis is that the data on asset payoffs and prices are replicated over time in some stationary fashion. That is, associated with the composite vector $(x',q',y)'$ is a stochastic process $\{(x_t',q_t',y_t)'\}$ whose sequence of empirical distributions approximate the joint distribution of $(x',q',y)'$. We denote integration with respect to the empirical distribution for sample size T as $\sum_T$. More precisely, for any z that is a (Borel measurable) function of (x',q',y) with a finite first moment, we will approximate Ez by $\sum_T z$ where

17

$$(3.1) \qquad \sum_T z \;=\; (1/T)\sum_{t=1}^{T} z_t \; .$$

Among other things, we require that this approximation becomes arbitrarily good as the sample size T gets large. That is we presume that $\{z_t\}$ obeys a Law of Large Numbers. A sufficient condition for this is:

Assumption 3.1: The composite process $\{(x_t',q_t',y_t)\}$ is stationary and ergodic.

Under this assumption, we can think of (x',q',y) as (x_0',q_0',y_0). Assumption 3.1 could be weakened in a variety of ways, but it is maintained for pedagogical simplicity. More generally, we might imagine that the process $\{(x_t',q_t',y_t)\}$ is asymptotically stationary, where the convergence to the stationary distribution is sufficiently fast to ensure that the Law of Large Numbers applies to averages of the form (3.1). In this case, the joint distribution of (x',q',y) is given by the stationary limit point of the process $\{(x_t',q_t',y_t)\}$.

To estimate the specification-error bounds, we suppose that a sample of size T is available and that the empirical distribution implied by this data is used in place of the population distribution. (Thus we are applying the *Analogy Principle* of Goldberger 1968 and Manski 1988). We introduce two random functions $\hat{\phi}$ and $\hat{\tilde{\phi}}$:

$$(3.2) \qquad \hat{\phi}(\alpha) \;=\; y^2 - (y - \alpha'x)^2 - 2\alpha'q,$$

and

18

$$(3.3) \qquad \tilde{\phi}(\alpha) = y^2 - (y - \alpha'x)^{+2} - 2\alpha'q .$$

The sample analog estimators of interest are given by

$$(3.4) \qquad (\hat{d}_T)^2 = \max_{\alpha \in C} \Sigma_T[\hat{\phi}(\alpha)]$$

and

$$(3.5) \qquad (\tilde{d}_T)^2 = \max_{\alpha \in C} \Sigma_T[\tilde{\phi}(\alpha)]$$

III.A: Consistent Estimation of the Specification-Error Bounds

We first establish the statistical consistency of the estimator sequences $\{\hat{d}_T\}$ and $\{\tilde{d}_T\}$:

Proposition 3.1: Under Assumptions 2.1-2.3 and 3.1, $\{\hat{d}_T\}$ and $\{\tilde{d}_T\}$ converge almost surely to $\hat{\delta}$ and $\tilde{\delta}$, respectively.

The proof of this proposition is given in Appendix A. The basic idea is that the population and sample criterion functions for the conjugate problems are concave and the sets of maximizers are convex. By Assumptions 2.1 and 3.1, the criterion functions converge pointwise (in α and β) almost surely to the population criterion functions introduced in Section II.C. In light of the concavity of the criterion functions, this convergence is uniform on compact sets almost surely (for example, see Rockafellar 1970). Finally, since the sets of maximizers of the limiting criterion functions

19

are compact, for sufficiently large T one can find a compact set such that the maximizers of the sample and population criteria are contained in that compact set (for example, see Hildenbrand 1974 and Haberman 1989). Hence the conclusion follows from the uniform convergence of the criteria on a compact set.

III.B: Asymptotic Distribution of the Estimators of the Bounds

We consider next the limiting distribution of the analog estimator sequences of the specification-error bounds. Our ability to express the objects of interest as solutions to the conjugate problems permits us to obtain results very similar to those in the literature on using likelihood ratios as devices for model selection in environments when models are possibly misspecified (for example, see Vuong 1989). We show that when the specification error bounds are positive, we obtain a limiting distribution that is equivalent to the one obtained by ignoring parameter estimation, and when the specification error bound is zero the limiting distribution is degenerate. (See Theorem 3.3 of Vuong 1989 page 307 for the corresponding result for likelihood ratios.)

Let $\hat{a}_T$ be a maximizer of $\sum_T \hat{\phi}$, $\hat{\alpha}$ a maximizer of $E\hat{\phi}$, $\tilde{a}_T$ a maximizer of $\sum_T \tilde{\phi}$, and $\tilde{\alpha}$ a maximizer of $E\tilde{\phi}$. To study the limiting behavior of the estimators, we use the decompositions:

$$(3.6) \qquad \sqrt{T}[(\hat{d}_T)^2 - \hat{\delta}^2] = \sqrt{T}\sum_T[\hat{\phi}(\hat{a}_T) - \hat{\phi}(\hat{\alpha})] + \sqrt{T}\sum_T[\hat{\phi}(\hat{\alpha}) - E\hat{\phi}(\hat{\alpha})],$$

and

$$(3.7) \qquad \sqrt{T}[(\tilde{d}_T)^2 - \tilde{\delta}^2] = \sqrt{T}\sum_T[\tilde{\phi}(\tilde{a}_T) - \tilde{\phi}(\tilde{\alpha})] + \sqrt{T}\sum_T[\tilde{\phi}(\tilde{\alpha}) - E\tilde{\phi}(\tilde{\alpha})].$$

20

As we will now demonstrate, the limiting distributions for the maximized values depend only on the second terms of these decompositions. In other words, the impact of replacing the unknown population maximizers by the sample maximizers in the sample criterion functions is negligible.

Take the case of the sequence $\{(\tilde{a}_T)^2\}$. Then by the concavity of $\tilde{\phi}$, we have the following gradient inequalities:

$$(3.8) \quad \tilde{\phi}(\tilde{a}_T) - \tilde{\phi}(\tilde{\alpha}) \leq (\tilde{m}x - q)\cdot(\tilde{a}_T - \tilde{\alpha})$$
$$= [(\tilde{m}x - q) - E(\tilde{m}x - q)]\cdot(\tilde{a}_T - \tilde{\alpha})$$
$$+ E(\tilde{m}x - q)\cdot(\tilde{a}_T - \tilde{\alpha}) .$$

However, it follows from the first-order conditions (including the complementary slackness conditions) for the population conjugate problem that

$$(3.9) \quad E(\tilde{m}x - q)\cdot(\tilde{a}_T - \tilde{\alpha}) = E(\tilde{m}x - q)\cdot\tilde{a}_T \leq 0 .$$

The inequality in (3.9) is obtained because $E(q - \tilde{m}x)$ is in the dual C^* while $\tilde{a}_T$ is constrained to be in C. Combining (3.8) and (3.9) we have that

$$(3.10) \quad 0 \leq \sqrt{T}\sum_T[\tilde{\phi}(\tilde{a}_T) - \tilde{\phi}(\tilde{\alpha})]$$
$$\leq \sqrt{T}\sum_T[(\tilde{m}x - q) - E(\tilde{m}x - q)]\cdot(\tilde{a}_T - \tilde{\alpha}) .$$

Therefore, $\{\sqrt{T}\sum_T[\tilde{\phi}(\tilde{a}_T) - \tilde{\phi}(\tilde{\alpha})]\}$ converges in probability to zero if the sample counterparts to the pricing errors obey a Central Limit Theorem and the maximizers can be chosen so that $\{(\tilde{a}_T - \tilde{\alpha})\}$ converges almost surely to

zero. This latter convergence can be demonstrated by exploiting the concavity of the population criterion function and the convexity of the constraint set (for example, see the discussion on page 1635 of Haberman 1989 and Appendix A).

Assumption 3.2: $\left\{ \sqrt{T} \Sigma_T \left[\begin{array}{c} \hat{\phi}(\hat{\alpha}) - E\hat{\phi}(\hat{\alpha}) \\ [(\hat{m}x - q) - E(\hat{m}x - q)] \end{array} \right] \right\}$ converges in distribution to a normally distributed random vector with mean zero and covariance matrix $\hat{V}$.

Assumption 3.3: $\left\{ \sqrt{T} \Sigma_T \left[\begin{array}{c} \tilde{\phi}(\tilde{\alpha}) - E\tilde{\phi}(\tilde{\alpha}) \\ [(\tilde{m}x - q) - E(\tilde{m}x - q)] \end{array} \right] \right\}$ converges in distribution to a normally distributed random vector with mean zero and covariance matrix $\tilde{V}$.

More primitive assumptions that imply the central limit approximations underlying Assumptions 3.2 and 3.3 are given by Gordin (1969) and Hall and Heyde (1980).

Let u denote a *selection vector* with a one in its first position followed by k + ℓ zeros. The limiting distributions for the specification-error bound estimators are:

Proposition 3.2: Suppose that $\hat{\delta} \neq 0$ and $\tilde{\delta} \neq 0$. Under Assumptions 2.1 - 2.3, 3.1 - 3.2, $\{\sqrt{T}[\hat{d}_T - \hat{\delta}]\}$ converges to a normally distributed random vector with mean zero and variance $u'\hat{V}u/(4\hat{\delta})$. Under Assumptions 2.1 - 2.4, 3.1 and 3.3, $\{\sqrt{T}[\tilde{d}_T - \tilde{\delta}]\}$ converges in distribution to a normally distributed random variable with mean zero and variance $u'\tilde{V}u/(4\tilde{\delta})$.

To use Proposition 3.2 in practice requires consistent estimation of $u'\hat{V}u$ or $u'\tilde{V}u$. Consider the case of $u'\tilde{V}u$. For each T form the scalar

22

sequence $\{\hat{\phi}_t(a_T): t=1,2,3, \ldots T\}$ and use one of the frequency zero spectral density estimators described by Newey and West (1987) or Andrews (1991), for example.

As is shown in Appendix A, when the price vector q is a vector of real numbers (degenerate random variables), the asymptotic distribution for $\{\sqrt{T}[\hat{\partial}_T - \partial]\}$ remains valid even when the population version of the conjugate maximum problem fails to have a unique solution (Assumption 2.4 is violated). In , this case, the lack of identification of the parameter vector $\tilde{\alpha}$ does not alter the distribution theory for the specification-error bound. While this special case is of considerable interest, it rules out the possibility of using conditioning information to form synthetic payoffs as described in Section II.

Notice that if $\hat{\delta} = 0$ or $\partial = 0$, Proposition 3.2 breaks down. This occurs if y is a valid stochastic discount factor in which case the solutions to the population conjugate problems are $\hat{\alpha} = \tilde{\alpha} = 0$. As a consequence, $\hat{\phi}_t(\hat{\alpha})$ and $\hat{\phi}_t(\tilde{\alpha})$ are both identically zero giving rise to a degenerate limiting distribution for $\{\sqrt{T}(\hat{d}_T)^2\}$ and $\{\sqrt{T}(\partial_T)^2\}$. Our subsequent results on the convergence of the parameters can be used to establish that the rate of convergence of $\{(\hat{d}_T)^2\}$ and $\{(\partial_T)^2\}$ is T, and is given by a weighted sum of chi-squared distributions (see Vuong 1989). As a result $\{\hat{d}_T\}$ and $\{\partial_T\}$ converge at the rate $\sqrt{T}$, although the limiting distribution is not normal.

III.C: Asymptotic Distribution of the Parameter Estimators

In several situations it is useful to examine the solutions to the conjugate problems used in constructing the bounds. For example, it may be of interest to examine whether a particular asset or group of assets are

important in determining the bound or it may of interest to determine whether the coefficient vector is zero, in which case the bound is degenerate. In developing a central limit approximation to do this type of statistical inference, we initially consider the case where there are no assets that are subject to short sales constraints. In other words, we assume that the cone C is R^n. Since, in the absence of market frictions, the estimation problem is posed as an unconstrained maximization problem, the limiting covariance matrices for the asymptotic distribution of the coefficient estimators have a form that is familiar both from M estimation (*e.g.*, see Huber 1981) and from *GMM* estimation (*e.g.*, see Hansen 1982). Our formal derivation of this distribution theory is given in Appendix C and uses a result from Pakes and Pollard (1989). A byproduct from our analysis in the appendix is a (modest) weakening of the assumptions imposed in Hansen (1982) to accommodate kinks in the moment conditions used in estimation.

The population moment conditions of interest are:

(3.11) $E[x(y-x'\hat{\alpha}) - q] = 0,$

for the specification-error bound in which the no-arbitrage restriction is not fully exploited, and

(3.12) $E[x(y-x'\tilde{\alpha})^+ - q] = 0,$

when the no-arbitrage is exploited. Equalities (3.11) and (3.12) are simply the first-order conditions (2.9) and (2.10) for the conjugate maximum problems when short-sale constraints are not imposed. The sample analog estimators satisfy:

$$(3.13) \quad \sum_T [x(y\text{-}x'\hat{a}_T) \text{ - } q] = 0$$

and

$$(3.14) \quad \sum_T [x(y\text{-}x'\tilde{a}_T)^+ \text{ - } q] = 0 \;.$$

While the equations for $\hat{\alpha}$ are linear, those for $\tilde{\alpha}$ are nonlinear. In the latter case, we use a linear approximation to the moment conditions in deriving the central limit approximation for the parameters:

$$(3.15) \quad x(y\text{-}x'\alpha)^+ \text{ - } q \;\approx\; x(y\text{-}x'\tilde{\alpha})^+ \text{ - } q \text{ - } xx'\mathbf{1}_{\{y\text{-}x'\tilde{\alpha}\,\geq 0\}}(\alpha\text{-}\tilde{\alpha}) \;.$$

$$= x(y\text{-}x'\alpha)\mathbf{1}_{\{y\text{-}x'\tilde{\alpha}\,\geq 0\}} \text{ - } q \;.$$

Notice that the function of α on the left side of (3.15) is differentiable except at values of $\tilde{\alpha}$ such that $y\text{-}x'\tilde{\alpha} = 0$. We assume that such sample points are "unusual":

Assumption 3.4: $Pr\{y\text{-}x'\tilde{\alpha} = 0\} = 0.$

To evaluate further the quality of the approximation in (3.15), let $r(\alpha)$ denote the random approximation error:

$$(3.16) \quad r(\alpha) \;\approx\; | x(y\text{-}x'\alpha)(\mathbf{1}_{\{y\text{-}x'\alpha\geq 0\}}\text{-}\mathbf{1}_{\{y\text{-}x'\tilde{\alpha}\,\geq 0\}}) | \;.$$

It follows from the Cauchy-Schwarz Inequality that

$$(3.17) \qquad r(\alpha) \;\; \leq \;\; |x(y\text{-}x'\alpha)| \; | \, (1_{\{y\text{-}x'\alpha \geq 0\}} - 1_{\{y\text{-}x'\tilde{\alpha} \geq 0\}}) \, |$$

$$\leq \;\; |xx'\alpha - xx'\tilde{\alpha}| \; | \, (1_{\{y\text{-}x'\alpha \geq 0\}} - 1_{\{y\text{-}x'\tilde{\alpha} \geq 0\}}) \, |$$

$$\leq \;\; |x|^2 |\alpha\text{-}\tilde{\alpha}|$$

where the second inequality follows because $|x'\alpha - x'\tilde{\alpha}|$ dominates $|x(y\text{-}x'\alpha)|$ whenever $y\text{-}x'\alpha$ and $y\text{-}x'\tilde{\alpha}$ have opposite signs. Therefore, the random approximation error satisfies:

$$(3.18) \qquad r(\alpha)/|\alpha\text{-}\tilde{\alpha}| \;\; \leq \;\; |x|^2$$

for $\alpha \neq \tilde{\alpha}$ implying that the *modulus of differentiability*

$$(3.19) \qquad dmod(\varepsilon) \;\; = \;\; sup\{r(\alpha)/|\alpha\text{-}\tilde{\alpha}| \;\; : \;\; |\alpha\text{-}\tilde{\alpha}| < \varepsilon, \text{ for } \alpha \neq \tilde{\alpha}\}$$

is dominated by $|x|^2$. Combined with Assumption 2.1 this implies that for any positive value of ε, $E[dmod(\varepsilon)]$ is finite. As $\varepsilon \rightarrow 0$, $dmod(\varepsilon)$ goes to zero except when $1_{\{y\text{-}x'\tilde{\alpha}=0\}} = 1$. In this case it is possible to choose α such that $|\alpha\text{-}\tilde{\alpha}| < \varepsilon$ and $1_{\{y\text{-}x'\alpha<0\}} = 1$ so that $r(\alpha) = |xx'|$. However Assumption 3.4 implies that this occurs with probability zero so that as $\varepsilon \rightarrow 0$, $dmod(\varepsilon)$ converges almost surely to zero. As is shown in Appendix C, these restrictions are sufficient for us to study the asymptotic behavior of the estimator $\{a_T\}$ using the linearization on the right side of (3.15):

$$(3.20) \qquad E[x(y\text{-}x'\alpha)1_{\{y\text{-}x'\tilde{\alpha} \geq 0\}} - q] \;\; = \;\; 0.$$

26

To use linear equation system (3.20) to identify $\tilde{\alpha}$, we need the matrix $E(xx'1_{\{y-x'\tilde{\alpha}\geq 0\}})$ to be nonsingular. Given Assumption 3.4, this rank condition is equivalent to one in Assumption 2.4 because x and x^* must coincide when no short-sale constraints are imposed. The counterpart to this rank condition for $\hat{\alpha}$ is that the second moment matrix $E(xx')$ be nonsingular as required by Assumption 2.3.

Working with the two linear moment conditions, we obtain the approximations:

$$(3.21) \qquad \sqrt{T}(\tilde{a}_T - \tilde{\alpha}) \approx -[E(xx'1_{\{y-x'\tilde{\alpha}\geq 0\}})]^{-1}\sqrt{T}\Sigma_T[x(y-x'\tilde{\alpha})^+ - q]$$

$$\sqrt{T}(\hat{a}_T - \hat{\alpha}) \approx -[E(xx')]^{-1}\sqrt{T}\Sigma_T[x(y-x'\hat{\alpha}) - q]$$

where the notation $\approx$ is used to denote the fact that the differences between the left and right sides of (3.21) converge in probability to zero. These approximations are justified formally in Appendix B. Let $w = [0 \ I_n]$. Combining approximations (3.21) with Assumptions 3.2 and 3.3 gives us the asymptotic distribution of the analog estimators.

Proposition 3.3: Suppose Assumptions 2.1-2.3, 3.1 and 3.2 are satisfied. Then $\{\sqrt{T}(\hat{a}_T-\hat{\alpha})\}$ converges in distribution to a normally distributed random vector with mean zero and covariance matrix: $[E(xx')]^{-1}w\hat{V}w'[E(xx')]^{-1}$. Suppose Assumptions 2.1-2.4, 3.1 and 3.3-3.4 are satisfied. Then $\{\sqrt{T}(\tilde{a}_T-\tilde{\alpha})\}$ converges in distribution to a normally distributed random vector with mean zero and covariance matrix: $[E(xx'1_{\{y-x'\tilde{\alpha}\geq 0\}})]^{-1}w\hat{V}w'[E(xx'1_{\{y-x'\tilde{\alpha}\geq 0\}})]^{-1}$.

To apply these limiting distributions in practice requires consistent estimators of the asymptotic covariance matrices. The terms $w\hat{V}w'$ and $w\hat{V}w'$

can be estimated using one of the spectral methods referenced previously. Under assumptions maintained in Proposition 3.3, the matrices $E(xx')$ and $E(xx' 1_{\{y-x' \tilde{\alpha} \geq 0\}})$ can be estimated consistently by their sample analogs, where the estimator $\tilde{a}_T$ is used in place of $\tilde{\alpha}$ in estimating the second of these matrices.

We now briefly describe how the distribution theory is modified when some short-sale constraints are imposed (C is a proper subset of R^n). We will focus on the limiting behavior of $\sqrt{T}(\tilde{a}_T - \tilde{\alpha})$, but the results for $\sqrt{T}(\hat{a}_T - \hat{\alpha})$ are very similar. As in Section II, we partition x by whether or not $\tilde{m}$ prices the payoffs with equality or not, that is by whether

$$(3.22) \qquad E\tilde{m}x^i \;=\; Eq^i, \; \text{or} \; E\tilde{m}x^i \;<\; Eq^i \quad .$$

The component coefficient estimators that multiply x^i's for which there is strict inequality will equal zero with arbitrarily high probability as the sample size gets large. Hence the limiting distribution is degenerate for these component estimators.

Consider next the estimator of the remaining subvector of $\tilde{\alpha}$, which we denote $\tilde{\beta}$. Because of the degeneracy just described, we can, in effect, treat the limiting distribution of the estimator of $\tilde{\beta}$ separately. Let $\tilde{C}$ be the lower-dimensional cone associated with estimating $\tilde{\beta}$. If $\tilde{\beta}$ is an interior point of $\tilde{C}$, then the argument leading up to Proposition 3.3 can be imitated to deduce a limiting normal distribution for the parameter estimator. However if $\tilde{\beta}$ is at the boundary of the cone $\tilde{C}$, the limiting distribution may be a nonlinear function of a normally distributed random vector (see Haberman 1989, page 1645).[3]

As in any econometric estimation problem with inequality constraints,

the problematic feature of this limiting distribution theory is the manner in which it depends on the true parameter vector β including the associated discontinuities. This feature makes the distribution theory harder to use in practice and, in other settings, has led researchers to compute approximate bounds on probabilities of test statistics (see, for example Wolak 1991 and Boudoukh, Richardson and Smith 1992). Recall, however, that in our derivation of the distribution theory for the specification-error bounds, we were able to circumvent the need for a distribution theory for the parameter estimators. Thus even though the distribution theory for the parameter estimators becomes more complicated in the presence of market frictions, the distribution theory for the specification-error bounds remains simple.

III.D: Consistent Estimation of the Arbitrage Bounds

As we discussed in Section II the second moments bounds can be converted into standard deviation bounds if the mean of m is known or if it can be estimated using the price of a risk-free asset. When Em is not known it must be prespecified. Let v be the hypothesized mean of m when a risk free asset is not available. Proposition 3.1 can be applied to establish the consistency of the second moment bound estimators for each *admissible* price assignment v. In the case of δ^2, for the price assignment to be admissible, it must not induce arbitrage opportunities onto the augmented collection of asset payoffs and prices. Any price (mean) assignment in the open interval (λ_0, v_0) is admissible in this sense.

The final question we explore in this section is whether the arbitrage bounds, λ_0 and v_0 given in (2.18) and (2.19), can be consistently estimated using the sample analogs:

(3.23) $\ell_T = -inf\{\alpha'\Sigma_T q : \alpha \in C$ and

$$\alpha'x_t \geq -1 \text{ for all } t=1,2,...,T\}$$

and

(3.24) $u_T = inf\{\alpha'\Sigma_T q : \alpha \in C$ and

$$\alpha'x_t \geq 1 \text{ for all } t=1,2,...,T\} \ .$$

The estimated upper arbitrage bound u_T is always finite when there is a payoff on a limited liability security that is never observed to be zero in the sample. Our estimated range of the admissible values for the (average) price of a unit payoff and hence mean of m is $[\ell_T, u_T]$. Notice that these bounds can be computed by solving simple linear programming problems. In Appendix A we prove:

Proposition 3.4: Under Assumptions 2.1-2.3 and 3.1, $\{\ell_T\}$ converges to λ_0 almost surely. If υ_0 is finite, then $\{u_T\}$ converges to υ_0 almost surely; and if $\upsilon_0 = +\infty$, then $\{u_T\}$ diverges to $+\infty$ almost surely.

IV. Applications and Extensions

In this section we discuss several applications and extensions of the analysis of Section III. First we discuss consistent estimation of the set of feasible means and standard deviations of stochastic discount factors. Previously we showed that for a given mean of the stochastic discount factor, the standard deviation bound can be consistently estimated. However, the mean of the stochastic discount factor typically is not known. As a result it is important to understand the sense in which the entire feasible region

30

can be approximated.

We next examine extensions of the distribution theory of Section III that are useful in answering several questions about the bounds and asset pricing models. First we extend the analysis of Section III to examine whether, given an initial set of asset returns, additional asset returns result in a change in the volatility bound. We call this set of tests *region subset tests*. Snow (1991) used this type of test to examine whether returns on a portfolio of stocks of firms with small capitalization contained additional information about the volatility of stochastic discount factors over and above that found in the return on the market portfolio; Cochrane and Hansen (1992) used it to determine whether conditioning information is important; Knez (1993) used it in his investigation of the links between the markets for Treasury bills, certificates of deposit and commercial paper; and De Santis (1993) used it to study the significance of returns on foreign securities *vis-a-vis* domestic securities in the construction of the volatility bounds.[4]

A particular example of the region subset test occurs when checking whether a constant discount factor would correctly price the assets under consideration. This is a test of whether the volatility bounds have content and is an important initial hypothesis to examine since, if true, the bounds do not preclude constant discount factors (risk-neutral pricing).

We then show how the feasible regions for the means and standard-deviations can be used to test a specific model of the discount factor. Burnside (1992) and Cecchetti, Lam and Mark (1992) have developed a version of this test when there are no assets subject to short-sale constraints or transactions costs. We show how this test can be implemented in a relatively simple manner by exploiting the results of Section III.

31

Further we formulate the test so that it is also applicable when there are assets subject to short-sale constraints. As a result this provides (large sample) statistical foundation to the tests of asset pricing models suggested by He and Modest (1992), and Luttmer (1993).

Finally we outline an extension of the specification-error bound analysis that is useful when the discount factor proxy under consideration depends upon a vector of unknown parameters. We consider an estimator of the parameter vector that minimizes the specification-error bound and briefly describe how to develop an asymptotic distribution for this estimator and for the implied bound.

Some of the formal discussion in this section focuses on the case when positivity is imposed in the construction of the volatility and specification-error bounds. Moreover, when considering volatility bounds, we study the more usual case in which data on the prices of a unit payoff are not used in the econometric analysis. Comparable results without positivity or with a unit payoff require the obvious modifications and are sometimes computationally simpler.

IV.A: Consistent Estimation of the Feasible Region of Means and Standard Deviations

As discussed in Section I and II, it is often of interest to construct approximations to the feasible region of ordered pairs of means and standard deviations of stochastic discount factors implied by security market data. Such a region can be computed with or without imposing the no-arbitrage restriction that the stochastic discount factors be positive. Let S_0 denote the region without positivity and S_0^+ the (closure) of the region with positivity. Similarly, let S_T and the S_T^+ denote the sample counterparts.

The question we now turn to is in what sense are S_T and S_T^+ good approximations to S_0 and S_0^+?

From our results in Section III we know that when there is a unit payoff, all four regions are vertical rays because the (average) price of this payoff is the mean discount factor. In this case the points of origin of the rays S_0 and S_0^+ can be estimated consistently by the points of origin of the corresponding rays S_T and S_T^+.

In the more usual case when data on a unit payoff and its price is not available, matters are a little more complicated. The feasible regions are no longer vertical rays but instead are unions of such rays resulting in convex sets with nonempty interiors. The boundaries of these sets can be represented as (possibly extended) real-valued functions of the ordinate (hypothetical mean), and our previous analysis implies pointwise (in the mean) convergence of the sample analog functions to their population counterparts. This result implies *uniform convergence* of the sample analog functions in following sense.

Since the lower and upper arbitrage bounds can be consistently estimated, for large enough T, the sample analog functions under positivity are finite on any compact subset of (λ_0, v_0). When positivity is ignored the functions are finite on any compact subset of R. Further these functions are convex functions of the hypothetical mean of the discount factor. As a result (see Theorem 10.8 of Rockafellar 1970) the sample analog functions converge uniformly, almost surely, on any compact subset of (λ_0, v_0) in the case of positivity and on any compact set when positivity is ignored. One difficulty is that the approximations deteriorate as the mean assignment, v, approaches the arbitrage bounds in the case of positivity, or when v gets large when positivity is ignored.

33

The deterioration of the sample analog to S_0^+ in the vicinity of a finite arbitrage bound turns out not to be problematic. To see this, instead of viewing the boundaries of the regions as functions of the ordinate, we explore the approximation error from a set-theoretic vantage point in R^2. Consider first the case in which $v_0 < +\infty$. Associated with a sample of size T is an approximation error as measured by the Hausdorff metric:[5]

$$(4.1) \qquad \eta_T = max\{\gamma(S_T^+, S_0^+), \gamma(S_T^+, S_0^+)\}$$

where:

$$(4.2) \qquad \gamma(K_1, K_2) = \sup_{(v_1, w_1) \in K_1} \inf_{(v_2, w_2) \in K_2} |(v_1, w_1) - (v_2, w_2)| \quad .$$

Since the arbitrage bounds can be consistently estimated and the lower boundaries of $\{S_T^+\}$ approach the lower boundary of S_0^+ uniformly on any compact interval within the arbitrage bound, the approximation error sequence $\{\eta_T\}$ converges to zero almost surely.[6]

Measuring the approximation error via the Hausdorff metric allows ordered pairs to get close without restricting them to have the same ordinate. In other words, we no longer confine our attention to "vertical" measures of distance, as is the case when we view the boundaries of the regions as functions of the hypothetical (expected) prices of a unit payoff. The added flexibility in the Hausdorff metric permits us to exploit better the consistent estimation of the upper and lower arbitrage bounds (Proposition 3.5).

When v_0 is infinite, the approximation error η_T defined by (4.1) will be infinite. As a remedy, we replace γ by

$$(4.3) \qquad \gamma_\rho(C_1, C_2) = \sup_{\substack{(v_1, w_1) \in K_1 \\ 0 \le v_1 \le \rho}} \inf_{\substack{(v_2, w_2) \in K_2 \\ 0 \le v_2 \le \rho}} |(v_1, w_1) - (v_2, w_2)|$$

where ρ is any arbitrary positive number greater than the lower arbitrage bound λ_0. Then the modified approximation error will be well defined and finite for sufficiently large T and will converge almost surely to zero. Thus we still get uniform convergence as long as the ordinate is restricted to a finite interval.

IV.B: Region Subset Tests

The first set of tests we consider are whether the volatility bounds can be constructed using a smaller vector of security payoffs. As in section III.C, we initially consider the case where there are no assets that are subject to short-sale constraints, and we assume that the parameters are uniquely identified. Let z denote an $(n-1)$-dimensional vector of assets under consideration with price vector s, and let f be the k-dimensional vector including the k-1 ass payoffs that are to be used to construct the bound augmented by a unit payoff. Formally, the hypothesis of interest can be represented as:

$$(4.4) \qquad E[z(f'\vartheta)^+ - s] = 0,$$
$$E[(f'\vartheta)^+ - v_0] = 0.$$

One possibility is to test this hypothesis for a prespecified v_0, and the

other is to test whether it is satisfied for some $v_0 > 0$.

Consider first the case in which v_0 is prespecified. To map this into the setup of Section III, form the n-dimensional vector x by augmenting z with a unit payoff and form q by augmenting s with the "price" v_0. Then hypothesis (4.4) can be interpreted as a zero restriction on the coefficient vector $\tilde{\alpha}$ employed in Sections II and III. The components of coefficient vector $\tilde{\alpha}$ set to zero are just the entries of z that are omitted from f. A large sample Wald test of this zero restriction can be formed by applying the limiting distribution in Proposition 3.3. Alternatively, we could construct a test by solving the *GMM* optimization problem:

$$(4.5) \quad \min_{\theta} \sqrt{T}\Sigma_T\left[\begin{matrix}z(f'\theta)^+ - s \\ (f'\theta)^+ - v_0\end{matrix}\right]' [(w\nabla w')_T]^{-1} \sqrt{T}\Sigma_T\left[\begin{matrix}z(f'\theta)^+ - s \\ (f'\theta)^+ - v_0\end{matrix}\right] \quad .$$

where $(w\nabla w')_T$ is one of the spectral estimators referenced in Section III. Since this test is embedded within a *GMM* estimation problem, the analysis in Section III.C. can be easily modified to show that the minimized value of the criterion function is distributed as a chi-square random variable with $n - k$ degrees of freedom (see Hansen 1982). When the hypothesis of interest is modified to be for some $v_0 > 0$, this *GMM* approach is modified by minimizing the criterion in (4.5) by choice of θ and v_0 with a corresponding loss in the degrees of freedom.

One special case of this setup is a test for risk-neutral pricing. In this case k is one and f contains only a unit payoff. In other words, a constant discount factor prices the securities correctly on average and the volatility bound is zero. Furthermore, the central limit approximations for the volatility bound estimator (without imposing risk-neutral pricing) is degenerate since the second moment of the discount factor is a constant. For

36

both of these reasons, a test for risk-neutral pricing is a useful starting point in an empirical investigation.

Conducting such tests can proceed as described with one modification. There is no sampling error associated with the second moment condition in (4.4) implying that $w \mathbb{V} w'$ is singular. Instead this second condition should be imposed ($\theta = v_0$) and testing should be based only on the initial set of moment conditions. When v_0 is not known, this second condition should be omitted and θ should be restricted to be nonnegative when solving minimization problem (4.5).

Region subset tests without positivity turn out to be closely connected to tests of factor structure and mean-standard deviation boundary intersection tests (*e.g.*, see Braun 1992 and Knez 1993 for an elaboration). This connection follows from the duality of the mean-standard deviation frontier for stochastic discount factors and the comparable frontier for returns (see Hansen and Jagannathan 1991 for an elaboration). Also, the Wald and *GMM* test statistics coincide because the moment conditions are linear in the parameter vector θ.

The tests using the criterion (4.5) rely upon the distribution theory of the estimator of $\tilde{\alpha}$. To use the theory of Section III.C requires that there are no securities subject to short-sale constraints. As we discussed in that subsection, the presence of the inequality restriction on the parameter vector $\tilde{\alpha}$ can complicate the distribution theory of the parameter estimators. As a result, testing zero restrictions on a subvector of $\tilde{\alpha}$ when some of the remaining coefficients are against a nonnegativity constraint can be problematic. On the other hand, the results of Haberman (1989) could be used to develop such a test when the zero restrictions are imposed on at least all of the securities to which the short-sale constraints apply.

37

IV.C: Testing a Specific Model of the Discount Factor using Volatility Bounds

Suppose that in addition to asset market data, a model of the discount factor is posited and a time series of observations of the discount factor is available: $\{m_t: \quad t=1, ..., T\}$. One way to test the model is to examine whether it satisfies the volatility bounds discussed in Sections II and III. Since observations of the discount factor are available, the average price of a unit payoff can be estimated by the mean of m. Specifically, form x by augmenting the original vector of payoffs with a unit payoff; form q by augmenting the original vector of prices with the random variable m; and form C by constructing the the Cartesian product of the original cone with R. In effect, we had added an unit payoff with an average price m that is not subject to a short-sale constraint. In forming a test, we can apply the results of Section II and III.B with one minor modification. The random functions $\hat{\phi}$ and $\overset{\gamma}{\phi}$ are now constructed by setting the proxy y to zero and subtracting m^2:

$$(4.6) \qquad \hat{\phi}(\alpha) \ = \ -(-\alpha'x)^2 - 2\alpha'q - m^2 ,$$

and

$$(4.7) \qquad \overset{\gamma}{\phi}(\alpha) \ = \ -(-\alpha'x)^{+2} - 2\alpha'q - m^2 .$$

Subtracting m^2 does not alter the solutions to either the sample or population maximization problems. It does, however, change the maximized values of the criteria functions. The volatility bounds for Em will be

satisfied, if, and only if

$$(4.8) \qquad \hat{\zeta} = \max_{\alpha \in C} E\hat{\phi}(\alpha) \leq 0,$$

when positivity is ignored, or

$$(4.9) \qquad \zeta = \max_{\alpha \in C} E\tilde{\phi}(\alpha) \leq 0,$$

when positivity is imposed. The limiting distribution reported in Proposition 3.2 (appropriately modified) can be applied to construct a test of these hypotheses using sample analog estimators of $\hat{\zeta}$ and ζ.[7] Again, we have formulated the problem so that approximation error due to parameter estimation plays no role in the limiting distributions for these sample analogs.

In practice we find the solutions for the sample maximization problems, estimate the asymptotic standard errors, and form one-sided tests. In particular, let $\tilde{c}_T$ be the maximized value of $\Sigma_T(\tilde{\phi})$ over the constraint set C. Then $\{\sqrt{T}[\tilde{c}_T - \zeta]\}$ converges in distribution to a normal random variable with mean zero and variance $u'\mathcal{V}u$. This variance can be estimated in the manner described in Section III.B. Since ζ is not specified under the null hypothesis (4.9), the "conservative" choice of $\zeta = 0$ is used in constructing the test statistic.

Finally when there are no transactions costs, short sales-constraints or other constraints to be considered, the asymptotic distribution of the estimators can be used to construct a different test, (analogous to a Likelihood Ratio test) that exploits the inequality restriction in the

first-stage estimation of the bound. Considering the case when positivity is imposed, the parameter vector $\tilde{\alpha}$ that solves problem (4.9) satisfies the moment conditions:

$$(4.10) \qquad E[x(x'\tilde{\alpha})^+ - q] = 0 .$$

The inequality restriction (4.9) implies the further moment condition:

$$(4.11) \qquad E\{[\tilde{\phi}(\tilde{\alpha}) - \gamma]\} = 0,$$

for some $\gamma \leq 0$.

Without a constraint on the parameter γ, moment conditions (4.10) and (4.11) exactly identify $\tilde{\alpha}$ and γ. However, with the restriction that γ be nonpositive, we can set up a *GMM* criterion in the parameter vector (α, ζ) using the moment conditions (4.10) and (4.11) and minimize this function subject to the constraint that $\zeta \leq 0$. If the population moments of *m* are on the boundary of the feasible region S^+ (that is, when $\gamma = 0$), then the appropriately scaled minimized *GMM* criterion function has a limiting distribution that has probability one-half of being zero and the remaining half of the probability is allocated according to a chi-square one distribution with one degree of freedom. This chi-square distribution then bounds the distribution of the *GMM* test statistics (under the null hypothesis) for other negative values of γ. When positivity of the stochastic discount factor is ignored and the $\hat{\phi}$ random function is used in place of $\tilde{\phi}$, it can be shown that the resulting test statistic coincides with the test statistic based solely on (4.6).[8]

Similar approaches to testing a model of the discount factor can be

applied when a time series for m can be constructed from simulated data instead of actual data. In this case the randomness of $\hat{\phi}(\tilde{\alpha})$ can be decomposed additively into two components, one due to the randomness of the security market payoffs and prices and the other due to the simulation of m. As in the work of McFadden (1989), Pakes and Pollard (1989), Lee and Ingram (1991) and Duffie and Singleton (1993), the asymptotic variance in the limiting distribution will now have an extra component due to the sampling error induced by simulation.

When the first two moments of m can be computed numerically with an arbitrarily high degree of accuracy, we can proceed as follows. Augment the price vector with Em instead of m and subtract Em^2 from the criteria instead of m^2 as in (4.6) and (4.7). This same strategy can be employed to assess the accuracy of the estimated feasible region for means and standard deviations of stochastic discount factors. For any hypothetical mean-standard deviation pair for m, one can compute the corresponding test statistic and probability value.

IV.D: Minimizing the Specification-Error Bound for Parameterized Families of Models

Recall that the specification-error bounds provide a way to assess the usefulness of an asset pricing model even when it is technically misspecified. In many situations the discount factor proxy depends on unknown parameters. For example, in a representative consumer model with constant relative risk aversion preferences, the pure rate of time preference and the coefficient of relative risk aversion are typically unknown. In this case one way to estimate the parameters of the model is to minimize the specification error. Alternatively, in an observable factor model, the

41

discount factor proxy depends on a linear combination of the factors with unknown coefficients. As in the work of Shanken (1987) one could imagine selecting factor coefficients to minimize the specification error. We now sketch how the results of Section III.B extend in a straightforward manner to obtain a distribution theory for the minimized value of specification-error bound.

Suppose that the discount factor proxy y depends on the parameter vector $\beta \in B$ where B is a compact set. The population optimization problems of interest are now:

$$(4.12) \quad \hat{\delta}^2 = \min_{\beta \in B} \ \max_{\alpha \in C} \ \left[Ey(\beta)^2 - E\{(y(\beta) - x'\alpha)^2\} - 2\alpha'Eq \right],$$

and

$$(4.13) \quad \hat{\gamma}^2 = \min_{\beta \in B} \ \max_{\alpha \in C} \ \left[Ey(\beta)^2 - E\{[(y(\beta) - x'\alpha)^+]^2\} - 2\alpha'Eq \right] \ .$$

When $\hat{\delta}$ and $\hat{\gamma}$ are strictly positive and the parameterized family of stochastic discount factors satisfies the appropriate smoothness and moment restrictions, an extended version of Theorem 3.2 can be obtained for the sample analog estimators of $\hat{\delta}$ and $\hat{\gamma}$. Again the limiting distribution will be the same as if the solutions to the population optimization problems were known *a priori*.

The approach can be extended to compare the smallest specification-errors for two nonnested families of models. Such a comparison potentially can be used as a device for selecting between the two families of models. Vuong (1989) examined a very similar problem by using the large

sample behavior of likelihood ratios for two nonnested families of misspecified models (in particular, see the discussion in Section 5 of Vuong 1989); and we can imitate and adapt his analysis to our problem. More precisely, let δ_1 and δ_2 denote the specification-error bounds associated with two such families. Take the null hypothesis to be:

$$(4.14) \qquad \delta_1 = \delta_2 .$$

Under the null hypothesis, the smallest specification-error associated with each parameterized family is the same. As a consequence the performance of the two parameterized families can not be ranked once sampling error is accounted for. This hypothesis can be tested by using the corresponding distribution theory for the difference between the analog estimators of δ_1 and δ_2 scaled by the square root of the sample size.

Finally, we sketch the distribution theory for the coefficient estimators when there is a parameterized family of discount factor proxies. Suppose that the unique solution β of (4.13) is contained in the interior of B, the parameterization family satisfies the appropriate smoothness and moment restrictions, and no short-sale constraints are imposed. The population moment conditions are given by:

$$(4.15) \qquad E\{x[y(\beta)-x'\tilde{\alpha}]^+ - q\} = 0;$$

and

$$(4.16) \qquad E\left[\frac{\partial y}{\partial \beta}(\beta)\{y(\beta) - [y(\beta) - \tilde{\alpha}'x]^+\}\right] = 0.$$

The distribution theory for the analog estimators $\{\tilde{b}_T\}$ and $\{\tilde{a}_T\}$ of β and $\tilde{\alpha}$ respectively can be deduced by taking linear approximations to the sample moment conditions (4.15) and (4.16) and appealing to the results of Appendix C.

V. Concluding Remarks

In this paper we provided statistical methods for assessing asset pricing models using specification-error and volatility bounds. In developing these procedures, it was advantageous to exploit duality theory and represent the measurements of interest as solutions to unconstrained conjugate maximization problems. This duality approach simplifies both computations and statistical inferences. The resulting statistical procedures can account for market frictions due to transactions costs or short-sale constraints, and are often easier to interpret than standard tests of asset pricing models. For the most part these methods are quite easy to implement, even when market frictions are considered. They are designed to provide a better understanding of the statistical failures of some popular asset pricing models and to offer guidance in improving these models.

Among other things, the results in this paper allow one to do the following: (i) to test whether a specific model of the stochastic discount factor satisfies the volatility bounds implicit in asset market returns; (ii) to compare the information about the means and standard deviations of discount factors contained in different sets of asset returns; and (iii) to test hypotheses about the size of possible pricing errors of misspecified asset pricing models. An advantage of (i) is that the resulting test is robust to misspecification of the *joint* distribution of asset returns and the

stochastic discount factor. In regards to (ii), our results permit these comparisons among data sets to be made independent of a specific stochastic discount factor model. Our motivation for (iii) is to shift the focus of statistical analyses of asset pricing models away from whether the models are correctly specified and towards measuring the extent to which they are misspecified.

Appendix A: Consistency

In this appendix we demonstrate formally the results of Section III.A, and III.D. We maintain Assumptions 2.1-2.3, and 3.1 throughout.

Let $\mathcal{U}$ denote a compact set in R^n. For any subset h of $\mathcal{U}$, we let $cl(h)$ denote the closure of h. Let $\mathcal{K}$ denote the collection of all nonempty closed subsets of $\mathcal{U}$. We use the Hausdorff metric η on $\mathcal{K}$ given by

$$(A.1) \qquad \eta(h_1, h_2) = max\{ \sup_{\alpha_1 \in h_1} \inf_{\alpha_2 \in h_2} |\alpha_1 - \alpha_2| , \sup_{\alpha_2 \in h_2} \inf_{\alpha_1 \in h_1} |\alpha_1 - \alpha_2| \} .$$

to define notions of convergence of compact sets. For some of our results we will use the construct of a *lim sup* of a sequence in $\mathcal{K}$. We follow Hildenbrand (1974) and define:

Definition A.1: For a sequence $\{h_j\}$ in $\mathcal{U}$, *lim sup* h_j $=$ $\bigcap_m cl (\bigcup_{j \geq m} h_j)$.

Since the *lim sup* is the intersection of a decreasing sequence of closed sets, it is closed and not empty. An alternative way to characterize the *lim sup* is to imagine forming sequences of points by selecting a point from each h_j. All of the limit points of convergent subsequences are in the *lim sup*, and, in fact, all of elements of the *lim sup* can be represented in this manner.

We shall make reference to an implication of a Corollary on page 30 of Hildenbrand (1974) that characterizes the set of minimizers of an "approximating" function over an "approximating set."

46

Lemma A.1: Suppose

(i) $\{\psi_j\}$ is a sequence of continuous functions mapping $\mathcal{U}$ into R that converges uniformly to ψ_∞;

and

(ii) $\{h_j\}$ converges to h_∞.

Then *lim sup* $g_j \subset g_\infty$ where $g_j \equiv \{u \in h_j : \psi_j(u) \leq \psi_j(u')$ for all $u' \in h_j\}$, and *lim min* $\underset{h_j}{\psi_j} = \underset{h_\infty}{min} \ \psi_\infty$.

Proof: To verify that this follows from the Corollary in Hildenbrand, let $\mathcal{J}$ denote the set of positive integers augmented by $+\infty$, and endow $\mathcal{J}$ with the usual metric for a one-point compactification. Then in light of (i), the sequence $\{\psi_j\}$ in conjunction with ψ_∞ defines a continuous function on $\mathcal{J} \times \mathcal{U}$; and in light of (ii), the sequence $\{h_j\}$ in conjunction with h_∞ defines a continuous compact correspondence mapping $\mathcal{J}$ into $\mathcal{U}$. The conclusion of the Lemma A1 then follows from the Corollary together with part (ii) of Proposition 1 of page 22 in Hildenbrand. *Q.E.D.*

Turning to the result in Section III.A, we formally establish Proposition 3.1:

Proof of Proposition 3.1:

We treat only the consistency of $\{\hat{a}_T\}$ because the corresponding argument

47

for $\{\hat{d}_T\}$ is very similar. Assumptions 2.1 and 3.1 imply that $\{\sum_T \hat{\phi}(\alpha)\}$ converges almost surely to $E\hat{\phi}(\alpha)$ for each $\alpha \in C$. Since for each T, $\sum_T \hat{\phi}$ is concave as is $E\hat{\phi}$, Theorem 10.8 of Rockafellar implies that $\{\sum_T \hat{\phi}\}$ converges uniformly on any compact set in R^n. Further as argued in Section II, the set of maximizers of $E\hat{\phi}$ is bounded. For a positive number N, define $C_N = \{\alpha \in C : |\alpha| \le N\}$ and $D_N = \{\alpha \in C : |\alpha| = N\}$. Then C_N and D_N are compact. By choosing N to be sufficiently large we can ensure that C_N contains all of the maximizers of $E\hat{\phi}$ over the constraint set C and that none of the maximizers are in D_N. Let δ_N be the maximized value of $E\hat{\phi}$ over D_N. Then by choice of N we have that $\delta_N < \hat{\vartheta}$. Since $\{\sum_T \hat{\phi}\}$ converges uniformly to $E\hat{\phi}$ on C_N almost surely, for sufficiently large T, the maximizers of $\sum_T \hat{\phi}$ over C_N are also not in D_N. By the convexity of C and concavity of $\sum_T \hat{\phi}$, it follows that for sufficiently large T, the maximizers of $\sum_T \hat{\phi}$ over C_N coincide with those over C. Consequently, the almost sure convergence of $\{\hat{\vartheta}_T\}$ to $\hat{\vartheta}$ follows from the almost sure uniform convergence of $\{\sum_T \hat{\phi}\}$ on C_N. Q.E.D.

We now turn to the results in Section III.D and investigate the statistical consistency of sample analog estimators $\{\ell_T\}$ and $\{u_T\}$ for the arbitrage bounds λ_0 and υ_0. Recall that the arbitrage bounds are representable as solutions to linear programming problems. Since there is no natural compact set for the choice variables in these problems, we must explore "directions to infinity." We study these "directions" using a compactification of the parameter space.

First consider any $\alpha \in C$ such that $\alpha'x \ge -1$ with probability one. Then with probability one $\alpha'x_t \ge -1$ for all t with probability one and $\{\alpha'\sum_T q\}$ converges almost surely to $\alpha'Eq$. Define $\ell_T^- = -\ell_T$ and $\lambda_0^- = -\lambda_0$. Since $\ell_T^- \le \alpha'\sum_T q$, it follows that $\lim \sup \ell_T^- \le \lambda_0^-$ with probability one.

Similarly, if $v_0 < \infty$, $\lim \sup u_T \leq v_0$. Hence our interest is in the *lim inf* ℓ_T and *lim inf* u_T.

To construct a compact parameter space, we map the original parameter space for each problem into the closed unit ball in R^s which we denote as $\mathcal{U}$. We consider explicitly the case of u_T. The proofs for the case of ℓ_T are completely analogous to the case for u_T and are omitted.

Notice that the constraint set used in defining v_0 can be represented as the set of all $\alpha \in C$ satisfying the equation:

(A.2) $E[(1 - \alpha'x)^+] = 0$.

Consider now a transformation of the parameter space by mapping the parameter space into the unit ball. The mapping $\zeta = \alpha/(1 + |\alpha|)$ maps R^s into the open unit ball. To compactify the transformed parameter space, we consider adding the boundary points of the unit ball. Notice that we can recover the original parameterization by considering the inverse mapping:

(A.3) $\alpha = \zeta/(1 - |\zeta|)$

for $|\zeta| < 1$. Using the transformation in (A.3), instead of considering those α's that satisfy (A.2) we consider:

(A.4) $D^* = \{ \zeta \in \mathcal{U} \cap C \mid E\{[(1 - |\zeta|) - x'\zeta]^+\} = 0 \}$

This transformation potentially adds solutions to (A.2) by including the boundary of the unit ball. The potentially problematic values of ζ are those

49

for which $x'\zeta \geq 0$, $\zeta \in C$ and $|\zeta| = 1$. We rule this out by limiting attention to values of ζ in

$$(A.5) \qquad \hat{D} = \{ \zeta \in \mathcal{U} \cap C \mid \zeta'Eq \leq (1-|\zeta|)(v_0+1) \}$$

Notice that any ζ in $\hat{D}$ for which $|\zeta| \neq 1$ satisfies $\zeta'Eq/(1-|\zeta|) \leq (v_0+1)$. In effect by focusing on ζ's in $\hat{D}$ we are eliminating ζ's corresponding to payoffs with "high" prices. This does not cause us problems because we are concerned with estimated upper arbitrage bounds that are too low, not too high. Also, any ζ in $\hat{D}$ for which $|\zeta| = 1$ must have an (average) price that is nonpositive. This eliminates the troublesome points (directions) from D^*. Let D_T^* be the sample analog of D^* and $\hat{D}_T$ be the sample analog of $\hat{D}$. We first consider the limiting behavior of $D_T^* \cap \hat{D}_T$:

Lemma A.2: Suppose that $v_0 < \infty$. Then $\lim \sup D_T^* \cap \hat{D}_T \subset D^* \cap \hat{D}$.

Proof: First notice that since $\Sigma_T q$ converges to Eq almost surely, then $\eta(\hat{D}_T, \hat{D})$ converges almost sure to 0. We next establish that $\lim D_T^* = D^*$. To do this we first show that $\Sigma_T[(1 - |\zeta|) - x'\zeta]^+$ converges uniformly to $E[(1 - |\zeta|) - x'\zeta]^+$ on $\mathcal{U}$. Note that $\mathcal{U} \cap C$ is compact and that:

$$(A.6) \qquad E\left\{ |[(1 - |\zeta_1|) - x'\zeta_1]^+ - [(1 - |\zeta_2|) - x'\zeta_2]^+| \right\}$$
$$\leq (1 + (E|x|^2)^{1/2})|\zeta_1 - \zeta_2|.$$

This is sufficient for the Uniform Law of Large Numbers of Hansen (1982) to apply. Hence from Lemma A.1, the *lim sup* of the sequence of minimizers of

50

$\sum_T [(1 - |\zeta|) - x'\zeta]^+$ over $\mathcal{U} \cap C$ is contained in the set of minimizers of $E[(1 - |\zeta|) - x'\zeta]^+$. Since $v_0 < \infty$, the set $\overset{*}{D}$ is not empty and $\overset{*}{D}$ is the set of minimizers of $E[(1 - |\zeta|) - x'\zeta]^+$. With probability one, any point in $\mathcal{C}$ must also be in D^*_T for all $T \geq 1$. Since D^* is separable, a common probability measure one set of sample points can be selected so that $D^* \subseteq D^*_T$ for all $T \geq 1$. As a result $\lim D^*_T = D^*$. The conclusion follows. *Q.E.D.*

Lemma A.3: Suppose that $v_0 < \infty$. Then $\lim \inf u_T \geq v_0$.

Proof: First note that

(A.7) $u_T = \min\{ \zeta' \sum_T q/(1-|\zeta|) \mid \zeta \in D^*_T \cap \hat{D}_T \}$ for sufficiently large T,

  and

$$v_0 = \min\{ \zeta' Eq/(1-|\zeta|) \mid \zeta \in D^* \cap \hat{D} \}.$$

Hypothetical expansions of the constraint set $D^*_T \cap \hat{D}_T$ for u_T can only result in smaller values of the maximized criterion. For instance, suppose the constraint set is augmented to include all of the points in $D^* \cap \hat{D}$. Then Lemma A.2 implies that this sequence of augmented constraint sets converges to $D^* \cap \hat{D}$. The conclusion then follows from Lemma A.1. *Q.E.D.*

Finally, we consider the case in $v_0 = \infty$.

Lemma A.4: Suppose that $v_0 = \infty$. Then $\{u_T\}$ diverges with probability one.

Proof: Since $v_0 = \infty$, there are no values of $\alpha \in C$ such that $\alpha'x \geq 1$ with

51

probability one. Consequently, the only values of ζ in D^* are ones for which $|\zeta| = 1$. We consider two cases. First suppose that $D^* = \emptyset$. The uniform convergence of $\Sigma_T[(1 - |\zeta|) - x'\zeta]^+$ to $E[(1 - |\zeta|) - x'\zeta]^+$ implies that for sufficiently large T, $D_T^* = \emptyset$ and $u_T = \infty$. Next suppose that $D \neq \emptyset$. Since there are no arbitrage opportunities (Assumption 2.2), $\zeta'Eq > 0$ for any ζ in D^* such that $\|\zeta'x\| > 0$. Also, Assumption 2.2 together with the no-redundancy Assumption 2.3 imply that $\zeta'Eq > 0$ for any ζ in D^* such that $\|\zeta'x\| = 0$. Furthermore, D^* is closed implying that

(A.8) $\varepsilon = \inf\{\zeta'Eq : \zeta \in D^*\} > 0$.

Since $\{\Sigma_T q\}$ converges to Eq almost surely and D_T^* converges almost surely to D^*, it follows from Lemma A.1 that with probability one for sufficiently large T, $\zeta'\Sigma_T q > \varepsilon/2$ for all $\zeta \in D_T^*$. The convergence of $\{D_T^*\}$ to D^* coupled with the fact that all elements of D^* have norm one then implies that $\{u_T\}$ diverges almost surely. $Q.E.D.$

Taken together, Lemmas A.2, A.3 and A.4 imply Proposition 3.4.

Appendix B: Asymptotic Distribution of Bounds Estimators

In this appendix we show that in the case in which the prices of the payoffs are constant, the asymptotic distribution of the estimated bounds can be demonstrated even when the parameter vector is not uniquely identified (even when Assumption 2.4 is not satisfied).

Proof of Proposition 3.2:

We consider the case of $\tilde{a}_T$. The case of $\hat{a}_T$ is similar. Let h_T be the set of maximizers of $\sum_T \hat{\phi}$ and let h_∞ be the set of maximizers of $E\hat{\phi}$. For each T, let $\tilde{a}_T$ be a measurable selection from h_T (see Theorem 1 of Hildenbrand (1974), page 54). Since *lim sup* $h_T = h_\infty$ almost surely and h_∞ is compact, there is a sequence $\{\alpha_T\}$ in h_∞ such that *lim* $|\tilde{a}_T - \alpha_T| = 0$ almost surely (see Appendix A). Further an implication of Lemma A.1 of Hansen and Jagannathan (1991) is that all $\alpha \in h_\infty$ result in the same random variable $\tilde{m} = (y-\alpha'x)^+$. Also (2.12) implies that for $\alpha \in h_\infty$, $\alpha'q = E\{y(y-\alpha'x)^+ - (y-\alpha'x)^{+2}\}$, so that $\alpha'q$ is the same for all $\alpha \in h_\infty$. As a result the random variable $\hat{\phi}(\alpha)$ is the same for all $\alpha \in h_\infty$. Now consider the decomposition of $\sqrt{T}\sum_T [(d_T^\sim)^2 - \hat{\phi}^2]$ as in (3.7):

$$(B.1) \qquad \sqrt{T}[(\tilde{a}_T)^2 - \hat{\phi}^2] = \sqrt{T}\sum_T [\hat{\phi}(\tilde{a}_T) - \hat{\phi}(\tilde{\alpha}_T)] + \sqrt{T}\sum_T [\hat{\phi}(\tilde{\alpha}_T) - E\hat{\phi}(\tilde{\alpha}_T)] \quad .$$

As in relation (3.10), we have:

$$(B.2) \qquad 0 \leq \sqrt{T}\sum_T [\hat{\phi}(\tilde{a}_T) - \hat{\phi}(\tilde{\alpha}_T)]$$
$$\leq \sqrt{T}\sum_T [(\tilde{m}x - q) - E(\tilde{m}x - q)]\cdot(\tilde{a}_T - \tilde{\alpha}_T) \quad .$$

Since $|\tilde{a}_T - \alpha_T|$ converges almost surely to 0, the result follows. *Q.E.D.*

Appendix C: Asymptotic Distribution

In this appendix we consider the asymptotic distribution of our parameter estimator. We begin by demonstrating that restrictions used in Hansen (1982) can be extended along the lines of Pollard (1985) and Pakes and

Pollard (1989) to accommodate "kinks" in the functions used to represent the moment conditions. Then from the discussion in Section III.C, it is straightforward to show that Proposition 3.3 follows from the main result in this appendix.

The notation used in this appendix conflicts with some of the notation used elsewhere in the paper. We let β_0 denote the parameter vector of interest and β any hypothetical point in the underlying parameter space $\mathcal{P}$. The parameter space is restricted to satisfy:

Assumption C.1: $\mathcal{P}$ contains an open ball in R^k about β_0.

We will use the construct of a random function. A random function ψ maps the set of sample points into the space of vector-valued continuous functions on $\mathcal{P}$. We require that $\psi(\beta)$ be an n-dimensional random vector for each β in $\mathcal{P}$.

We also consider an approximating function

$$\psi_t^\bullet(\beta) = \psi_t(\beta_o) + \Delta_t(\beta-\beta_o)$$

that is linear β. The composite random function satisfies:

Assumption C.2: $\{(\psi_t{}',\psi_t^\bullet{}')'\}$ is stationary and ergodic and has finite first moments.

We now specify the sense in which $\psi_t^\bullet$ is required to approximate ψ_t. The approximation error induced by using is $\psi_t^\bullet$ in place of ψ_t is

$$r_t(\beta) = |\psi_t(\beta) - \psi_t^\bullet(\beta)| .$$

54

Define:

$$dmod_t(\delta) = sup\{r_t(\beta)/ \mid \beta\text{-}\beta_o \mid : \mid \beta\text{-}\beta_o \mid < \delta, \beta \neq \beta_o\} .$$

Note that $dmod_t(\cdot)$ is monotone in δ. Therefore, we can take almost sure limits as δ declines to zero. We impose the following restrictions on mod_t.

Assumption C.3: $\lim\limits_{\delta \to 0} dmod_t(\delta) = 0$ almost surely.

Assumption C.4: $E[dmod_t(\delta)] < \infty$ for some $\delta > 0$.

To satisfy Assumptions C.3 and C.4, Δ_t is typically taken to be the matrix of partial derivatives of ψ_t at β_0 when ψ_t is differentiable at β_0 and be *well behaved* for the other sample points. The random variable $mod_t(\delta)$ is interpreted as the *modulus of differentiability* for ψ_t at β_o.

The approach adopted in Hansen (1982) is to restrict the *modulus of continuity* of the derivative of ψ_t to converge almost surely to zero and to have a finite expectation for some neighborhood of the parameter. It follows from the Mean-Value Theorem that restrictions imposed in Hansen (1982) on the local behavior of ψ_t imply Assumptions C.3 and C.4.

We use Assumptions C.3 - C.4 to study the sense in which $\sum_T \psi$ is stochastically differentiable. Hence look at the approximation error

$$\varepsilon_T(\delta) = sup \{ \mid \sum_T \psi(\beta) - \sum_T \psi^*(\beta) \mid / \mid \beta\text{-}\beta_o \mid : \mid \beta\text{-}\beta_o \mid < \delta, \beta \neq \beta_o\} .$$

By the Triangle Inequality we have that

$$\varepsilon_T(\delta) \leq \sum_T dmod(\delta) \quad .$$

Thus by Assumptions C.1-C.2, we have that

(C.1) $\lim\limits_{\delta \to 0} \lim\limits_{T \to \infty} \sup \varepsilon_T(\delta) \leq \lim\limits_{\delta \to 0} Edmod(\delta)$

$= 0.$

This in turn implies the stochastic differentiability condition in Pollard (1985) because the counterpart to $\varepsilon_T(\delta)$ in Pollard's condition is scaled by $\sqrt{T} |\beta-\beta_o| /(1 + \sqrt{T} |\beta-\beta_o|)$, which is less than one. Also, the iterated limit in (C.1) implies the limit taken in Pollard's condition because ε_T is monotone in δ. The differentiability of limiting moment function $E\psi$ follows directly from Assumption C.4. Therefore, $\sum_T \psi - E\psi$ satisfies the stochastic differentiability condition with derivative at β_o given by $\sum_T \Delta - E\Delta$. Since $\{\psi_t^*\}$ is stationary and ergodic, $\{\sum_T \Delta - E\Delta\}$ converges almost surely to zero hence the derivative is asymptotically negligible.

Next we impose a *global identification* condition on the approximating function ψ_t^*. Since the approximation of ψ_t by ψ_t^* is *local*, this condition can also be viewed as a *local identification* condition on the original function ψ_t.

Assumption C.5: $E |\Delta_t| < \infty$ and $E\Delta$ has full rank k.

This rank condition on the derivative together with the stochastic differentiability conditions already established imply the equicontinuity

condition (iii) in Theorem 3.3 of Pakes and Pollard (1989) (see the discussion on page 1043 of Pakes and Pollard).

We study the behavior of an estimator b_T that solves the equations:

$$a_T \Sigma_T \psi(b_T) = 0$$

for sufficiently large T. The (k x n) random matrix a_T selects the linear combination of moment conditions to be used in estimation.

Assumption C.6: $\{b_T\}$ converges in probability to β_0.

Assumption C.7: $\{a_T\}$ converges in probability to a nonrandom matrix a_0 where $a_0 E \Delta$ is nonsingular.

Finally, to obtain a limiting distribution for $\{b_T\}$ we assume:

Assumption C.8: $\{\sqrt{T} \Sigma_T \psi(\beta_0)\}$ converges in distribution to a normally distributed random vector with mean zero and nonsingular covariance matrix V_0.

Sufficient conditions for Assumption C.8 can be obtained using martingale approximations as described by Gordin (1969), Hall and Heyde (1980) and Hansen (1985). This condition implies that $E\psi(\beta_0)$ is equal to zero.

The following extension of Theorem 3.1 in Hansen (1982) is now a direct consequence of Theorem 3.3 and Lemma 3.5 in Pakes and Pollard (1989).

57

Theorem C.1:   Suppose that Assumptions C.1-C.8 are satisfied.   Then $\{\sqrt{T}(b_T-\beta_0)\}$ converges in distribution to a normally distributed random vector with mean zero and covariance matrix $[a_0 E(\Delta_t)]^{-1} a_0 V_0 a_0{}' [E(\Delta_t{}')a_0]^{-1}$.

Estimation of $E\Delta$ follows as in Hansen (1982) as long as Δ can be expressed in terms of a random matrix function D that satisfies $\Delta = D(\beta_0)$ where D is continuous at β_0 with probability one and has a modulus of continuity with a finite first moment for some $\delta > 0$.   In this case, $\{\Sigma_T D(b_T)\}$ converges in probability to $E\Delta$.

58

Footnotes

[1]A weaker version of this restriction would replace Eq by q. In effect, Assumption 2.3 does more than eliminate redundant securities. It also precludes cases in which distinct portfolio weights give rise to the same payoff, possibly different prices but the same expected prices.

[2]Formally, the pricing-error interpretation for least squares problem (2.6) is

$$\hat{\delta} = \inf_{m \in \mathcal{M}} \sup_{\substack{p \in P \\ Ep^2 = 1}} |Emp - Eyp| \quad ,$$

and for (2.7) is

$$\tilde{\delta} = \inf_{m \in \mathcal{M}^+} \sup_{\substack{p \in H \\ Ep^2 = 1}} |Emp - Eyp|$$

where H is a complete set of derivative claims on the payoffs in P.

[3]Haberman characterized this nonlinear function as a particular projection onto a closed convex set formed by translating $\mathcal{C}$ by $-\beta$. Although Haberman (1989) only considers the case in which the data are *iid*, his characterization of the limiting distribution applies more generally with a covariance matrix replaced by a spectral density matrix at frequency zero.

[4]The impetus for this work was the econometric discussion in an unpublished precursor to this paper: Hansen and Jagannathan (1988).

[5]The Hausdorff metric is usually employed for compact sets to ensure that the resulting distance is finite. Because of the vertical character of the regions and the existence of finite arbitrage bounds, the Hausdorff distance will be finite even though the sets are not bounded.

[6]The Euclidean distance in (4.2) could be replaced by the square root of a quadratic form in the differences between two points as long as a positive weight is given to both dimensions.

[7]Even if hypothesis (4.9) is satisfied, the sample analog may be infinite, making implementation problematic. This happens when the sample mean is outside the estimated arbitrage bounds. This phenomenon does not arise for hypothesis (4.8).

[8]Burnside (1992) and Cecchetti, Lam and Mark (1992) developed and studied alternative versions of the volatility bounds tests when no transactions costs are introduced. The test used by Cochrane and Hansen (1992) abstracted from positivity and can be formulated equivalently using $\hat{\phi}$ in (4.6). See Burnside (1992) for a Monte Carlo comparison of various volatility tests including the ones proposed here.

References

Andrews, D.W.K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59:3, 817-858.

Bansal, R., D.A. Hsieh, and S. Viswanathan, 1992, "A New Approach to International Arbitrage Pricing," Duke University, Fuqua School of Business manuscript.

Boudoukh, J., M. Richardson, and T. Smith, 1992, "Testing Inequality Restrictions Implied from Conditional Asset Pricing Models." New York University manuscript.

Braun, P.A., 1992, "Tests of Consistent Pricing and No Arbitrage," Northwestern University, unpublished manuscript.

Brown, D.P. and M.R. Gibbons, 1985, "A Simple Econometric Approach for Utility-based Asset Pricing Models," *Journal of Finance* 40, 359-381.

Burnside, C., 1992, "Hansen-Jagannathan Bounds as Classical Tests of Asset Pricing Models," NBER Workshop manuscript.

Cecchetti, S.G., P. Lam, and N.C. Mark, 1992, "Testing Volatility Restrictions on Intertemporal Marginal Rates of Substitution Implied by Euler Equations and Asset Returns," Ohio State University manuscript.

Cochrane J.H., and L.P. Hansen, 1992, "Asset Pricing Explorations for Macroeconomics," *NBER Macroeconomics Annual 1992*, Cambridge, MA: MIT Press, 115-165.

De Santis, G., 1993, "Volatility Bounds for Stochastic Discount Factors: Tests and Implications from International Stock Returns," University of Chicago Ph.D. Dissertation.

Duffie, D., and K.J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica* 61:4, 929-952.

Epstein, L.G., and S.E. Zin, 1991, "Substitution, Risk Aversion, and Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy* 99:2, 263-286.

Foley, D.K., 1970, "Economic Equilibrium with Costly Marketing," *Journal of Economic Theory* 2:3, 276-291.

Goldberger, A., 1968, *Topics in Regression Analysis*, New York: Macmillan.

Gordin, M.I., 1969, "The Central Limit Theorem for Stationary Processes," *Soviet Mathematics Doklady* 10, 1174-1176.

Haberman, S.J., 1989, "Concavity and Estimation," *The Annals of Statistics* 17:4, 1631-1661.

Hall, P., and C.C. Heyde, 1980, *Martingale Limit Theory and Its Application*, NY: Academic Press.

Hansen, L.P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Hansen, L.P., 1985, "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics* 30, 203-238.

Hansen, L.P., and R. Jagannathan, 1988, "Restrictions on Intertemporal Marginal Rates of Substitution Implied by Asset Prices," unpublished manuscript.

Hansen, L. P., and R. Jagannathan, 1991, ,"Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99:2, 225-262.

Hansen, L.P., and R. Jagannathan, 1993, "Assessing Specification Errors in Stochastic Discount Factor Models," manuscript.

Hansen, L.P., and K.J. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50:5, 1269-1286.

He, H., and D.M. Modest, 1992, "Market Frictions and Consumption-Based Asset Pricing," manuscript.

Heaton, J., 1993, "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specification," MIT Sloan School Working Paper.

Hildenbrand, W., 1974, *Core and Equilibria of a Large Economy*, Princeton NJ: Princeton University Press.

Huber, P., 1981, *Robust Statistics*, Wiley, New York.

Jouni, E., and H. Kallal, 1992, "Martingales, Arbitrage and Equilibrium in Securities Markets with Transaction Costs," University of Chicago manuscript.

Knez, P.J., 1993, "Pricing Money Market Securities with Stochastic Discount Factors," University of Wisconsin manuscript.

Kreps, D.M., 1981, "Arbitrage and Equilibrium in Economies with Infinitely Many Commodities," *Journal of Mathematical Economics*, 15-35.

Lee, B., and B. Ingram, 1991, "Simulation Estimation of Time Series Models," *Journal of Econometrics*, 47, 197-205.

Luttmer, E., 1993, "Asset Pricing in Economies with Frictions," Northwestern University, Department of Finance Working Paper #151.

MacKinlay, A.C., and M.P. Richardson, 1991, "Using Generalized Method of Moments to Test Mean-Variance Efficiency," *The Journal of Finance* 46, 511-527.

Manski, C., 1988, *Analog Estimation Methods in Econometrics*, NY: Chapman and Hall.

McFadden, D., 1989, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995-1026.

Newey, W.K., and K. West, 1987, "A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703-708.

Pakes, A., and D. Pollard, 1989, "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57:5, 1027-1057.

Pollard, D., 1985, "New Ways to Prove Central Limit Theorems," *Economic Theory* 1, 295-314.

Prisman, E.Z., 1986, "Valuation of Risky Assets in Arbitrage Free Economies with Frictions," *The Journal of Finance*, 41, 545-559.

Rockafellar, R.T., 1970, *Convex Analysis*, Princeton: Princeton University Press.

Shanken, J., 1987, "Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique," *Journal of Financial Economics* 18, 91-110.

Snow, K.N., 1991, "Diagnosing Asset Pricing Models Using the Distribution of Asset Returns," *Journal of Finance* 46, 955-984.

Vuong, Q.H., 1989, "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica* 57:2, 307-333.

Wolak, F. A., 1991, "The Local Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica* 59:4, 981-995.