WILEY-BLACKWELL

American Finance Association

http://www.jstor.org

Assessing Specification Errors in Stochastic Discount Factor Models

LARS PETER HANSEN and RAVI JAGANNATHAN*

ABSTRACT

In this article we develop alternative ways to compare asset pricing models when it is understood that their implied stochastic discount factors do not price all portfolios correctly. Unlike comparisons based on χ^2 statistics associated with null hypotheses that models are correct, our measures of model performance do not reward variability of discount factor proxies. One of our measures is designed to exploit fully the implications of arbitrage-free pricing of derivative claims. We demonstrate empirically the usefulness of our methods in assessing some alternative stochastic factor models that have been proposed in asset pricing literature.

IN THEORIES OF ASSET PRICING, portfolio payoffs are modeled as bundled contingent claims to a numeraire consumption good. When asset markets are frictionless, portfolio prices can be characterized as a linear valuation functional that assigns prices to the portfolio payoffs (e.g., see Ross (1978), Harrison and Kreps (1979), Kreps (1981), Chamberlain and Rothschild (1983), Hansen and Richard (1987), and Clark (1993)). These valuation functionals are typically represented as inner products of payoffs with *pricing kernels* or *stochastic discount factors*. As argued by Hansen and Richard (1987), observable implications of candidate models of asset markets are summarized conveniently in terms of their implied stochastic discount factors.[1]

While formal statistical testing of such asset pricing models can yield insights, it is also of interest to evaluate the performance of these models even when it is understood that the implied stochastic discount factors do not correctly price all portfolios. Pricing errors may occur either because the model

[1] We use the term *stochastic discount factor* as a label for a state-contingent discount factor. The market value today of an uncertain payoff tomorrow is represented by multiplying the payoff by the discount factor and adding across states of nature using the underlying probabilities. The discount factor is stochastic because it varies across states of nature. This variation captures corrections for risk as argued by Rubinstein (1976).

is viewed formally as an approximation, as in many linear factor pricing models, or because the empirical counterpart to the theoretical stochastic discount factor is error ridden (e.g., see Roll's (1977) critique of the single-period capital asset pricing model). In this article we associate a stochastic discount factor *proxy* with an asset pricing model and ask the question, *How large is the misspecification of the stochastic discount factor proxy?*

When there is only one asset, a natural measure of model misspecification is the pricing error associated with that asset—i.e., the difference between the price of the asset and the hypothetical price assigned by a candidate stochastic discount factor. In this case it is also straightforward to compare the relative performance of any two stochastic discount factor proxies. When there are a number of assets, there is a vector of pricing errors associated with each proxy. Comparing the specification error in two different proxies is not possible without taking a stand on the relative importance of the various assets. As we will verify, our measures of specification error focus on the most mispriced portfolios, while correcting for portfolio size in a particular way.

Conveniently, these maximum pricing errors are also the least squares distances between the stochastic discount factor proxy and families of stochastic discount factors that price correctly the vector of securities used in an econometric analysis. Except possibly when there are arbitrage opportunities present in the data set used in the empirical investigation, the set of correctly specified discount factors is nonempty and typically large (e.g., see Hansen and Jagannathan (1991)). However, as long as a proxy is misspecified, its least squares distance to the set of correctly specified discount factors will be strictly positive. In this article we consider two different distance measures corresponding to distances to two alternative families of stochastic discount factors. The second of these two families is smaller because the discount factors are restricted to be positive. Only such discount factors are consistent with the absence of arbitrage opportunities on the space of hypothetical derivative claims. The resulting measures of model misspecification are, by design, comparable across models and their associated proxies.

Pricing errors or closely related expected return errors are commonly used to assess asset pricing models. For instance, in linear factor models of returns the principle of no-arbitrage is used to characterize the sense in which security market prices can be *approximately* represented in terms of the prices of a small number of factors (e.g., see Ross (1976), Huberman (1982), Chamberlain and Rothschild (1983), and Shanken (1987)). Similarly, in present-value models with constant discount factors, Durlauf and Hall (1989) compute greatest lower bounds on the magnitude of pricing errors. As we will demonstrate, the least squares distance measures we propose also have a pricing-error interpretation. Like Shanken's (1987) analysis of linear factor models, one of our least squares distance measures can also be interpreted as the maximum pricing error for portfolios of payoffs with second moments equal to unity. One important way in which we extend the analysis of Shanken (1987) is that we also investigate implications for arbitrage-free pricing of derivative claims by restricting attention to positive stochastic discount factors.

This article is divided into five sections. In Section I we define two families of admissible stochastic discount factors, one associated with the pricing of the original collection of securities and the other associated with the assignment of arbitrage-free prices of derivative claims. In Section II we present our two alternative measures of model misspecification and deduce their pricing-error interpretations. We show that one of the specification-error measures depends *only* on the pricing errors of the collection of securities used in an econometric analysis, but not on pricing errors of derivative claims on those securities. We then demonstrate that the second measure takes account of potential pricing errors for derivative claims. In Section III we use duality theory to compute the specification-error measures in practice. In Section IV we establish the close connection between one of our measures of model misspecification and the expected return error measure of Shanken (1987) applied to linear factor pricing models. The duality results of Section III are then used in Section V to show that the econometric methods of Hansen, Heaton, and Luttmer (1995) are directly applicable. Finally, in Section VI we illustrate our apparatus by assessing several stochastic discount factor proxies that have been proposed in the literature.

I. Stochastic Discount Factors

In this section we construct the admissible set of stochastic discount factors. By a stochastic discount factor we mean a random variable that can be used to compute market prices today by discounting, state-by-state, the corresponding payoffs at a future date. As we will see, typically there is a large family of such random variables consistent with the asset market data used in an econometric analysis. Formally, we will use the apparatus of a Hilbert space to model the collection of portfolio payoffs, and we will use a continuous linear functional on that space as a way to represent the market values assigned to those payoffs. Given this setup, we will then construct two alternative families of stochastic discount factors. The smaller of the two families will be restricted so that the prices assigned to hypothetical derivative claims respect the Principle of No-Arbitrage.

For simplicity, we focus on asset market transactions that take place at two dates, t and $t + \tau$. At date t financial assets are purchased, while at date $t + \tau$ the payoffs are received. We let q_t denote the vector of prices used in an econometric analysis and let $x_{t+\tau}$ denote the corresponding vector of payoffs. In the first four sections we omit the date subscripts t and $t + \tau$ for notational convenience. As we see in Section V, to extract observable implications, we presume that these two periods are replicated over time in a manner that is stationary, at least asymptotically. The idea is to then apply a Law of Large Numbers to justify the approximation of population moments using time series averages.

A. *Payoff Space*

We follow Harrison and Kreps (1979), Chamberlain and Rothschild (1983), and others by modeling portfolio payoffs as elements of a Hilbert space. Such a space is convenient because it possesses all the nice properties of finite-dimensional vector spaces and accommodates some infinite-dimensional problems as well. In building a Hilbert space, we must specify the associated inner product (the counterpart to a dot product of two vectors). The corresponding norm of a payoff is then simply the square root of the inner product of that element with itself. We now specify formally the Hilbert space structure used in our article.

Let $\mathcal{F}$ be the conditioning information that is observed at the date of the asset payoffs. Associated with $\mathcal{F}$ is the space L^2 of all random variables with finite second moments that are in the information set $\mathcal{F}$. This space is used as the collection of hypothetical (and perhaps real) claims that could be traded. (For technical reasons, we restrict $\mathcal{F}$ and hence L^2 to be separable.) We endow L^2 with its usual inner product and norm:

$$\langle h_1 | h_2 \rangle \equiv E(h_1 h_2) \quad \text{and} \quad \|h\| \equiv \langle h | h \rangle^{1/2}, \quad h_1, h_2 \quad \text{in} \quad L^2. \tag{1}$$

Let P denote the space of portfolio payoffs used in an econometric analysis. In other words, the econometrician is presumed to have at his disposal historical data on some vector of basis payoffs (typically returns or excess returns). The space P then includes the basis payoffs along with synthetic portfolios constructed with these basis payoffs.

ASSUMPTION 1.1: *P is a closed linear subspace of L^2.*

The closure restriction is imposed for technical convenience.

Although in practice the payoff space P is built from a vector of basis payoffs, we do not assign any special role to a set of primitive payoffs beyond the fact that such payoffs are used to generate P. We view the basis payoffs as merely convenient building blocks for the space of portfolio payoffs used in assessing models. We do not require that the space P coincide with the entire collection of payoffs that can be traded by investors. Some asset payoffs available to investors may be precluded from an analysis for reasons of tractability. Alternatively, some of the conditioning information used by investors may not be fully reflected in the basis payoffs used in an empirical investigation.

One particularly convenient specification of the payoff space P is as follows. Let x denote an n-dimensional random vector with entries in L^2 and a nonsingular second-moment matrix. In this case, the entries of x are payoffs on n primitive securities and form a basis for the space of payoffs P:

$$P \equiv \{x \cdot \mathbf{c} : \mathbf{c} \in \mathbf{R}^n\}. \tag{2}$$

We require that Exx' be nonsingular, which implies that for each payoff p in P, there is a unique vector of portfolio weights $\mathbf{c}$ such that $p = \mathbf{c} \cdot x$. This payoff space structure will be used in our empirical investigation reported in Section

V. In using this space, we do not mean to require the portfolio weights used by the investors to be constant, for we would expect them to depend explicitly on conditioning information. Instead, the constant portfolio weight specification is adopted for econometric convenience.

When an econometrician models the evolution of conditioning information available to an investor, a richer specification of portfolio payoffs can be employed. For instance, suppose the payoff vector x is modeled as a factor general autoregressive conditional heteroskedasticity (GARCH) process, as in Bollerslev and Engle (1993) or King, Sentana, and Wadhwani (1994), or is modeled using the semi-nonparametric method of Gallant and Tauchen (1989). The information used in constructing conditional moments or distributions can also be exploited in forming hypothetical portfolios, as in Gallant, Hansen, and Tauchen (1990). In these cases we assume that $E(xx'|\mathscr{G})$ is nonsingular with probability one, where $\mathscr{G}$ contains information observed by economic agents at the trading date whose evolution is modeled by the econometrician to form portfolios. The payoff space P is then given by

$$P \equiv \{p \in L^2 : p = w \cdot x \text{ for some random vector } w \text{ of portfolio weights}$$
$$\text{that is in the conditioning information set } \mathscr{G}\}. \quad (3)$$

It is demonstrated in Hansen and Richard (1987) that P, as given by equation (3), satisfies Assumption 1.1.

B. Pricing

Our intention is to compare the hypothetical prices assigned by a given model to security market payoffs with market prices; we do this using time series averages. For this reason it is convenient to analyze the valuation implications of a model as reflected by the average or expected prices. We assume that portfolio payoffs in P obey the Law of One Price: To each portfolio payoff p in P there corresponds a unique *expected* price $\pi(p)$. We need the following assumption to facilitate the subsequent analysis.

ASSUMPTION 1.2: *The functional π is continuous and linear on P, and there exists a payoff $p \in P$ such that $\pi(p) = 1$.*

This assumption can often be derived from a more primitive no-arbitrage restriction when there is a security with limited liability (e.g., see Kreps (1981) or Clark (1993)). For many of the examples we consider, there are other, more mechanical, devices for verifying Assumption 1.2 because, in effect, the payoff spaces are constructed from a finite number of distinct primitive payoffs. This is the case for both specifications (2) and (3) of P.

To illustrate the construction of the pricing functional, initially suppose that P is given by equation (2). Corresponding to the payoff vector x, there is a price vector q of current period prices where $|q|$ has a finite first moment. Then Assumption 1.2 is satisfied for

$$\pi(x \cdot \mathbf{c}) \equiv \mathbf{c} \cdot Eq \quad (4)$$

as long as Eq is not a vector of zeros. The requirement that $|q|$ have a finite first moment is satisfied trivially when the basis payoffs are all returns. Such payoffs have unit prices by design and hence do not vary over time. Consequently, the expected prices and actual prices coincide, as do prices of portfolios with constant weights. In other words, the functional π assigns prices in this case because prices coincide with expected prices.

The question then emerges as to why we allow q to be random in the first place. We do so because we also wish to accommodate payoffs on portfolios of the primitive assets, where the amount invested depends on new information that becomes available over time. In such cases, the resulting time series of prices will fluctuate. For instance, suppose the econometrician has data on both a return, r, and information available to investors at the purchase date, z. Then, for the purposes of the empirical investigation, the synthetic payoff zr can be constructed with expected price Ez. Of course, the conditioning information could be captured in a more limited way by constructing a synthetic portfolio payoff that by design is still a return (it has unit price). For instance, if there are two initial returns, and if a portfolio is constructed with weights z, and $1 - z$ as in Breen, Glosten, and Jagannathan (1989), by construction the price will be unity (and hence will not fluctuate over time).

For an alternative illustration of π, suppose that P is given by equation (3). In this case, we restrict $q'E(xx'|\mathcal{G})^{-1}q$ to have a finite mean. The *expected pricing functional* is now given by

$$\pi(w \cdot x) = E(w \cdot q).\tag{5}$$

It is shown in Appendix A that the resulting π is a bounded linear functional on P. Random (fluctuating) prices can be avoided in this setting by shrinking the set P appropriately. For instance, P might be constructed to be the subset of payoffs given on the right side of equation (3) with constant prices.[2] This still permits the inclusion of conditioning information, but in a more limited way.

C. Stochastic Discount Factors

We now return to our general analysis. An admissible *stochastic discount factor* is a random variable m in L^2 such that the expected price of a payoff p can be represented as the inner product of the payoff and m:

$$\pi(p) = E(pm) \qquad \text{for all} \qquad p \in P.\tag{6}$$

Note that a stochastic discount factor m discounts payoffs state by state. It incorporates both a discount effect and an adjustment for risk. To see this, write

$$Epm = EpEm + \text{Cov}(p, m).\tag{7}$$

---

[2] It follows from the proof of Corollary 3.1 in Hansen and Richard (1987) that the resulting P satisfies Assumption 1.1 and that π satisfies Assumption 1.2.

Discounting of the future is captured by the first term on the right side of equation (7) (so long as Em is less than one), and adjustment for risk is captured by the second term. Let $\mathcal{M}$ denote the set of all admissible stochastic discount factors. It follows from the Riesz Representation Theorem that $\mathcal{M}$ is not empty and that there is a unique random variable in the intersection of $\mathcal{M}$ and P, i.e., there exists a unique stochastic discount factor that is also a portfolio payoff. This particular stochastic discount factor has the minimum norm among all of the elements of $\mathcal{M}$. All other stochastic discount factors can be represented as the sum of the minimum norm stochastic discount factor and of a random variable that is orthogonal to the space P of portfolio payoffs.

It is not tractable for an empirical researcher to study simultaneously the payoffs and prices on all traded assets. Moreover, it is sometimes of interest to identify a stochastic discount factor from some initial collection of portfolio payoffs and use it in pricing other assets (such as derivative claims on these payoffs). When this is the objective, it is necessary to restrict further the family of stochastic discount factors. This is because some of the stochastic discount factors in $\mathcal{M}$ will be negative with positive probability. Hence, while they assign the right prices to the collection of assets used in the empirical investigation, they will assign negative prices to some positive derivative claims on these payoffs. Such stochastic discount factors are unsuitable for pricing contingent claims.

We follow the usual approach in derivative claims pricing by considering arbitrage-free extensions of the pricing function π from the space P of portfolio payoffs used in an empirical investigation to the larger space L^2. Recall that L^2 contains functions of the portfolio payoffs in P as long as the resulting random variables are finite; hence, it contains a rich collection of potential derivative claims. Extensions of the pricing functional π are constructed by taking members m of $\mathcal{M}$ and forming the following linear functionals:

$$\pi_m(h) = Ehm \qquad \text{for all} \quad h \in L^2. \tag{8}$$

Then π_m agrees with π on P. Following Ross (1978) and Kreps (1981), some discount factors in $\mathcal{M}$ can be eliminated from consideration because the resulting extensions π_m of π introduce arbitrage opportunities on the space of potential derivative claims L^2. For pricing derivative claims, the π_ms of interest are those that satisfy the following condition.

Condition N: A pricing functional ρ does *not* induce arbitrage opportunities on a subspace H of L^2 if, for any $h \in H$ such that $h \geq 0$ and $\|h\| > 0$, $\rho(h) > 0$.

It is easy to show that (π_m, L^2) satisfies Condition N if, and only if, m is strictly positive with probability one (e.g., see Harrison and Kreps (1979) and Hansen and Richard (1987)). Hence when our concern is the pricing of derivative claims, we look at stochastic discount factors restricted to be in the subset $\mathcal{M}^{++}$ of $\mathcal{M}$ consisting of random variables m that are strictly positive with probability one.

As long as (π, P) satisfies the no-arbitrage condition (Condition N), then Assumption 1.3 follows from Kreps (1981).[3]

ASSUMPTION 1.3: $\mathcal{M}^{++}$ *is not empty.*

The set $\mathcal{M}^{++}$ is convex but is not necessarily closed. For much of our analysis, it will be convenient to include limit points by using the closure $\mathcal{M}^+$ in place of $\mathcal{M}^{++}$. This closure is just the subset of $\mathcal{M}$ consisting of only nonnegative random variables.

II. Least Squares Approximation of Proxies

In this section, we take as a starting point a *proxy* used in an empirical investigation in place of a correctly specified stochastic discount factor. This proxy assigns approximate prices to payoffs, and the question of interest is how to measure the magnitude of this approximation error. One example is the capital asset pricing model (CAPM) proxy $y = a + br^m$, where r^m is the return on the market. Another is the consumption CAPM proxy with power utility $y = \beta g^{-\gamma}$, where β is the subjective discount factor $\gamma \geq 0$ and where g is the consumption growth factor. We propose to measure the degree of model misspecification by the least squares distances between the proxy and the families $\mathcal{M}$ and $\mathcal{M}^{++}$. We then show how the resulting measures are directly tied to pricing errors, both in payoffs in P and of contingent claims.

Let y be a random variable in L^2 that is a *proxy* for a stochastic discount factor. This proxy is used to construct *approximate* prices for securities via the following formula:

$$\pi_a(h) \equiv E(yh) \tag{9}$$

for any payoff h in the space L^2 of potential derivative claims. For example, if the econometrician is studying the CAPM without using conditioning information, y will equal $a + br^m$ where r^m is the proxy for the return on the market and a and b are two unknown constants to be estimated in an empirical analysis (see Dybvig and Ingersoll (1982)). More will be said about this estimation subsequently.

If the proxy y turns out to be an admissible stochastic discount factor (that is, in $\mathcal{M}$), then the hypothetical expected prices assigned by the proxy will coincide with the actual expected prices. In general, the proxy will not be admissible either, because the model is, strictly speaking, misspecified or the observed stochastic discount factor is measured with error.[4] Consequently, pricing errors may be introduced on the original collection of payoffs, P, as well as on the collection of potential derivative claims, L^2. One measure of the

---

[3] Strictly speaking, this holds when L^2 is separable.

[4] See Roll (1977) for a discussion of why the CAPM is not testable, because the return on the market portfolio of all assets is not observable by an econometrician. Moreover, some derivations of the arbitrage pricing theory (APT), such as those in Huberman (1982) and Chamberlain and Rothschild (1983), obtain factor pricing only as an approximation.

magnitude of specification error induced by using the proxy as ̀a stochastic discount factor is its distance from the admissible set of stochastic discount factors (either $\mathcal{M}$ or $\mathcal{M}^+$). In view of this, we study the following least squares problems in the next two sections:

Problem 1:

$$\delta \equiv \min_{m \in \mathcal{M}} \|y - m\|.$$

Problem 2:

$$\delta^+ \equiv \min_{m \in \mathcal{M}^+} \|y - m\|.$$

In the case of Problem 2, we use $\mathcal{M}^+$ instead of $\mathcal{M}^{++}$ as a constraint set merely to guarantee that the infimum is attained. It is well known that both problems have unique solutions (e.g., see Luenberger (1969)).[5]

The least squares distances δ and δ^+ can be translated directly into pricing-error measures. We first show that δ is the (sharp) bound on the approximation-error functional:

$$\tilde{\pi}(p) \equiv \pi_a(p) - \pi(p). \tag{10}$$

By the Riesz Representation Theorem, there exists a unique payoff $\tilde{p}$ in P such that

$$\tilde{\pi}(p) = E(\tilde{p}p), \quad p \in P. \tag{11}$$

It also follows from this theorem that the bound on the approximation-error functional is given by $\|\tilde{p}\|$. In other words,

$$\|\tilde{p}\| = \max_{p \in P, \|p\|=1} |\tilde{\pi}(p)|; \tag{12}$$

hence $\|\tilde{p}\|$ is the maximum pricing error (per unit norm). The fact that $\|\tilde{p}\|$ gives a bound stems from the Cauchy–Schwarz Inequality:

$$|\tilde{\pi}(p)| \leq \|\tilde{p}\| \quad \|p\|. \tag{13}$$

The fact that the bound is sharp can be established by showing that the pricing error of the unit norm payoff $\tilde{p}/\|\tilde{p}\|$ is given by the left side of equation (13).

Now that we have obtained a characterization of a sharp bound on the pricing-error functional, it remains to demonstrate that this bound is indeed δ. Note that the random variable $y - m$ can be used to represent the approximation-error functional $\tilde{\pi}$ for any admissible stochastic discount factor m. Since $\tilde{p}$ can be used to represent the same approximation-error functional on

---

[5] By uniqueness in this setting, we follow the usual convention of treating the equivalence class of all random variables that are equal almost surely as a single element of L^2.

P, $y - m - \tilde{p}$ is orthogonal to P. Moreover, because $\tilde{p}$ is in P, $\tilde{p}$ is the least square projection of $y - m$ onto P for any m in $\mathcal{M}$. Therefore,

$$\|\tilde{p}\| \leq \delta. \tag{14}$$

Also, since $\tilde{p}$ can be used to represent the approximation-error functional and since y can be used to represent the approximate pricing functional, the random variable $y - \tilde{p}$ can be used to represent π. Consequently, this random variable is in $\mathcal{M}$, which implies that

$$\|\tilde{p}\| \geq \delta. \tag{15}$$

Taken together, inequalities (14) and (15) give us the following alternative interpretation of δ:

PROPOSITION 2.1: *Suppose Assumptions 1.1 and 1.2 are satisfied. Then*

$$\delta = \max_{p \in P, \|p\|=1} |\pi^a(p) - \pi(p)|.$$

The close connection between pricing errors and the least squares distance between y and $\mathcal{M}$ is to be expected from the continuity results of Green (1986), Kandel and Stambaugh (1987), and Glosten and Jagannathan (1993). As we see in Section IV, it is even more closely connected to a result in Shanken (1987).

As a byproduct of the proof of this proposition, it follows that one characterization of the solution to the least squares Problem 1 is $y - \tilde{p}$, where $\tilde{p}$ is the unique random variable in P that can be used to represent the approximation-error functional. Hence the solution entails finding an additional pricing factor among the portfolio payoffs to add to the proxy. This pricing factor $\tilde{p}$ is the smallest adjustment, in a least squares sense, required to make $y - \tilde{p}$ an admissible stochastic discount factor; the magnitude of this adjustment is δ. In the next section we give a more algorithmic characterization of the solution to both least squares problems.

Next we deduce the pricing-error characterization of δ^+. This characterization entails looking at pricing-errors for payoffs in the span of contingent claims on events in $\mathcal{F}$, i.e., on payoffs in L^2. Although we are interested in cases in which P is not sufficiently rich to permit us to "price by arbitrage" the payoffs in L^2, there still exist nontrivial arbitrage bounds on the possible price assignments. These arbitrage bounds are fully captured by the range of price assignments implied by the stochastic discount factors in $\mathcal{M}^{++}$.

For the moment, imagine forming pricing errors using a single stochastic discount factor in $\mathcal{M}^{++}$, for example, m. Let π_m denote the corresponding extension of π to all of L^2. For any h in L^2, the pricing error for this claim satisfies the Cauchy–Schwarz bound:

$$|Emh - Eyh| \leq \|m - y\| \|h\|. \tag{16}$$

Define the following:

$$\delta_m^+ \equiv \sup\{|\pi_m(h) - \pi_a(h)| : h \in L^2, \|h\| = 1\}. \qquad (17)$$

Then

$$\delta_m^+ = \|m - y\|, \qquad (18)$$

which can be verified by letting $h = (m - y)/\|m - y\|$. Clearly, the pricing error bound δ_m^+ is sensitive to the choice of the stochastic discount factor m in $\mathcal{M}^{++}$. We eliminate this sensitivity by computing the *mini-max* bound:

$$\begin{aligned} \inf\{\delta_m^+ : m \in \mathcal{M}^{++}\} &= \inf_{m \in \mathcal{M}^{++}} \|y - m\| \\ &= \min_{m \in \mathcal{M}^+} \|y - m\| \\ &\equiv \delta^+. \end{aligned} \qquad (19)$$

Therefore, the least squares distance between y and the convex set $\mathcal{M}^{++}$ gives the following *mini-max* bound on the pricing errors for all hypothetical derivative claims:

PROPOSITION 2.2: *Suppose that Assumptions 1.1–1.3 are satisfied. Then*

$$\delta^+ = \min_{m \in \mathcal{M}^+} \max_{h \in L^2, \|h\|=1} |\pi_m(h) - \pi_a(h)|.$$

As empirical researchers, we are interested in the solutions to least squares Problems 1 and 2 for three reasons. First, we will use δ and δ^+ as measures of model misspecification. By construction, δ^+ is greater than or equal to δ. As we have seen, both of these measures have alternative interpretations as bounds on the magnitude of pricing errors induced by using a proxy y. In particular, optimization Problem 2 is of interest when it is envisioned that the identified stochastic discount factor will also be used to price securities not used in the empirical investigation. Second, we will use the solutions to these problems as devices for identifying or selecting among the multitude of stochastic discount factors that correctly price assets. The identified discount factors are the closest ones (in a least squares sense) to the proxy. Third, the solutions to these optimization problems help to diagnose the deficiencies of a given model or family of models of a stochastic discount factor.

At this juncture, some queries might emerge concerning our use (or abuse) of these optimization problems. For instance, why measure model misspecification by these least squares criteria, and why use these criteria as devices for identifying alternative stochastic discount factors? Taking a more precise stand on the nature of the model misspecification or stipulating in more detail how the competing models will be used may result in better justified criteria for measuring the extent of the misspecification. However, such an exercise will necessarily restrict the scope of such comparisons to a narrower range of

models.[6] Further, using the least squares criteria in conjunction with a para-metric model has the virtue that the parametric model is employed in a less rigid way than usual without discarding it completely.

III. Using Duality to Solve the Least Squares Problems

To solve the least squares problems in practice, it is most convenient to use appropriately formulated dual or conjugate problems. The conjugate problems are also of value in interpreting the solution. We treat formally the case in which P is generated by a finite-dimensional random vector x with a nonsingular second-moment matrix, although comparable results could be obtained for the other specifications of P presented in Section I. We initially focus on optimization Problem 2, in which we minimize the distance between the proxy and the family of nonnegative stochastic discount factors. We do so only because this problem is slightly more difficult to solve than is optimization Problem 1. After solving Problem 2, we will return briefly to the solution to Problem 1.

We rewrite the *primal* Problem 2 as[7]

$$(\delta^+)^2 = \min_{m \in L^2, m \geq 0} \|y - m\|^2 \quad \text{subject to} \quad E(mx) = Eq. \quad (20)$$

To justify formally the use of Lagrange multipliers in solving this problem, in the Appendix A we verify that, given the no-arbitrage restriction (Assumption 1.3), the expected price vector Eq is an interior point in the set $\mathscr{C} \equiv \{Emx : m \in L^2, m \geq 0\}$ (e.g., see Luenberger (1969), p. 236). Consequently, we are led to investigate the saddle-point problem:

$$(\delta^+)^2 = \min_{m \in L^2, m \geq 0} \sup_{\lambda \in \mathbf{R}^n} \{E[(y - m)^2] + 2\lambda' E(xm) - 2\lambda' Eq\}$$
$$= \max_{\lambda \in \mathbf{R}^n} \min_{m \in L^2, m \geq 0} \{E[(y - m)^2] + 2\lambda' E(xm) - 2\lambda' Eq\}. \quad (21)$$

To construct the criterion of the conjugate problem, we must solve the inner minimization problem in the second line of equation (21). This turns out to be

---

[6] A nice example of a more structured approach to assessing model misspecification is found in McCulloch and Rossi (1990). These authors use exponential utility functions to measure the magnitude of expected return errors in factor models. Unfortunately, their approach is not conducive to making comparisons across a wide class of models, including models in which investors have alternative preference orderings to those implied by the exponential functions of the next period's market wealth.

[7] This construction parallels a construction in Luttmer (1994). Luttmer considers a slightly different problem. In line with his interest in computing volatility bounds on stochastic discount factors when transaction costs are present, y is 0 in his setup, but the payoff space P is constructed more generally to include short-sale constraints.

an easy problem to solve, as we will now demonstrate. Rewrite the portion of the saddle-point criterion containing m as

$$E[(y - m)^2] + 2\lambda'E(xm) = E[(y - \lambda'x - m)^2] - \lambda'E(xx')\lambda + 2\lambda'E(yx).$$
$$(22)$$

Since the second and third terms on the right side of equation (22) do not involve m, to construct the criterion function for the conjugate problem we solve the problem

$$\min_{m\in L^2, m\geq 0} E[(y - \lambda'x - m)^2].$$
$$(23)$$

Optimization problem (23) is a least squares problem whereby nonnegative random variables m are used to approximate the $y - \lambda'x$. We use the notation $(h)^+$ to denote an option payoff equal to max $\{h, 0\}$. The solution to problem (23) is simply the option payoff $(y - \lambda'x)^+$, as we now demonstrate. Construct an orthogonal decomposition of the target payoff $y - \lambda'x$ in the least squares criterion:

$$y - \lambda'x = (y - \lambda'x)^+ + [-(-y + \lambda'x)^+].$$
$$(24)$$

This decomposition is orthogonal because the second term is zero whenever the first term is not. Moreover, since the first term is itself nonnegative, it can be approximated perfectly by a nonnegative random variable. However, the second term is nonpositive, and the closest (in the least squares sense) nonnegative random variable to it is the degenerate random variable that is equal to zero with probability one. Therefore, the solution to equation (23) is given by the first term in the decomposition. To verify this conclusion formally, rewrite the criterion in equation (23) as

$$E[(y - \lambda'x - m)^2] = E\{[(y - \lambda'x)^+ - m]^2\} + 2E[(-y + \lambda'x)^+m]$$
$$+ E\{[(-y + \lambda'x)^+]^2\}.$$
$$(25)$$

The first two terms are nonnegative and are equal to zero only if $m = (y - \lambda'x)^+$, and the third term does not depend on m.

We now obtain the conjugate of Problem 2 by substituting the solution to equation (25) into the criterion of the saddle-point problem (21):

Problem 2':

$$(\delta^+)^2 = \max_{\lambda\in\mathbf{R}^n} E\{y^2 - [(y - \lambda'x)^+]^2 - 2\lambda'q\}.$$

Notice that the criterion for the optimization Problem 2' is concave in the multiplier vector λ, and the constraint set is finite dimensional. This finite-dimensional character of the conjugate problem makes it much more tractable to solve than the original primal problem. Although the optimal choice of λ and hence the square root of the optimized value of the criterion δ^+ cannot typically

be expressed in terms of simple matrix manipulations, both can be computed easily using standard numerical methods.

We can obtain the conjugate to Problem 1 using an analogous (but actually simpler) argument. Recall that optimization Problem 1 finds the minimum distance between the proxy y and the family of all discount factors (including negative ones). By mimicking our previous logic, it is easy to verify that the conjugate to optimization Problem 1 is given by

Problem 1′:

$$\delta^2 = \max_{\lambda \in \mathbf{R}^n} E[y^2 - (y - \lambda'x)^2 - 2\lambda'q].$$

The first-order conditions for this problem are

$$E[x(y - \tilde{\lambda}'x) - q] = 0, \tag{26}$$

which can be interpreted as finding the vector $\tilde{\lambda}$ such that $y - \tilde{\lambda}'x$ is an admissible stochastic discount factor. The solution for λ is given by

$$\tilde{\lambda} = (Exx')^{-1}E(xy - q), \tag{27}$$

as long as there are no redundant payoffs in the basis vector x. In relating this to the analysis leading up to Proposition 2.1, the random variable $\tilde{p}$ used to represent the approximation-error functional is given by

$$\tilde{p} = \tilde{\lambda}'x. \tag{28}$$

As we remarked in Section II, the "pricing factor" $\tilde{\lambda}'x$ is the smallest adjustment in a least squares sense required to make $y - \tilde{\lambda}'x$ an admissible stochastic discount factor. The magnitude of this adjustment is given by

$$\delta = [(Exy - Eq)'(Exx')^{-1}(Exy - Eq)]^{1/2}. \tag{29}$$

Since we are proposing to use δ as a measure of model misspecification, it is of interest to compare it to the chi-square test of a stochastic discount factor model suggested by Hansen and Singleton (1982) and Brown and Gibbons (1985). Notice from equation (29) that the chi-square statistic δ^2 is a quadratic form in the pricing-error vector $Exy - Eq$. The distance or weighting matrices in these two quadratic forms are different, however. The distance matrix in the quadratic form on the right side of (29), $(Exx')^{-1}$, is invariant to the choice of the proxy and is different from the one used in computing a large sample chi-square test of the null hypothesis that the pricing-error vector $(Eq - Exy)$ is zero. In the latter case, the distance matrix is proportional to the inverse of the asymptotic covariance matrix for a central limit approximation. Equivalently, it is the inverse of the spectral density matrix for the time series process associated with the pricing-error vector $q - xy$. A drawback of the chi-square statistic as a measure of model misspecification is its sensitivity to the choice

of proxy y and its reward for sampling error associated with the sample mean of $q - xy$. The reward built into the chi-square statistic is reflected in the use of the *inverse* of the asymptotic covariance matrix.[8]

Now consider the conjugate Problem 2'. The solution $\hat{\lambda}$ is not necessarily unique, but the resulting random variable $(y - \hat{\lambda}'x)^+$ is unique.[9] As with Problem 1', the first-order conditions

$$E[(y - \hat{\lambda}'x)^+x - q] = 0 \qquad (30)$$

for Problem 2' provide us with a simple interpretation of this random variable. It is the stochastic discount factor in $\mathcal{M}^+$ that is closest to the proxy y.[10] In other words, the smallest alteration in the proxy required to price correctly the payoffs in P is $-\hat{\lambda}'x$ in states for which $y - \hat{\lambda}'x$ is nonnegative and is $-y$ otherwise. The quantity δ^+ measures the magnitude of this correction. Therefore, we are again led to find a payoff in P to use in correcting the discount factor proxy. In this case a possibly nonlinear transformation is required to ensure that the resulting random variable is both an admissible stochastic discount factor and is nonnegative. Of course, if the solution to Problem 1' turns out to be nonnegative, then it will also be the solution to Problem 2'.

To summarize, we have derived the conjugates to the least squares problems posed in Section II. These conjugate problems are easy to solve in practice, and they give us an operational way to make comparisons of competing misspecified models of stochastic discount factors. As a side benefit, the solutions identify minimal perturbations as explicit functions of portfolio payoffs that correct the proxies (misspecified discount factors) for their inherent pricing errors. Also, the magnitude of the multiplier vectors can be used to assess the importance of particular basis payoffs for correcting the proxy.

IV. Linear Factor Pricing Models

Although our interest is in a more general class of asset pricing models, in this section we derive some results that are special to linear factor pricing models. This discussion is designed to show the connection between our work and some previous contributions to the empirical asset pricing literature. In particular, we address two issues. We first look at the relationship between pricing errors such as we have described here and expected return errors commonly employed in the linear factor pricing literature. We then provide a

---

[8] Allen (1991) makes essentially the same point in justifying an alternative distance measure to the ones developed here. Allen's focus is on the role of conditioning information in the pricing of a single return, whereas ours looks across returns but can still accommodate conditioning information.

[9] The solution $\hat{\lambda}$ is unique when the second-moment matrix of the random vector $x\mathbf{1}_{\{y-\hat{\lambda}'x>0\}}$ is nonsingular where $\mathbf{1}_{\{y-\hat{\lambda}'x>0\}}$ is an indicator function equal to 1 when $y - \hat{\lambda}'x$ is strictly positive and 0 otherwise.

[10] Hansen and Jagannathan (1991) attempt to prove an infinite dimensional version of this result when $y = 0$. However, there is a mistake in their Lemma A5.

simple characterization of the solution to least squares Problem 1 for linear factor pricing models.

A. Expected Payoff Errors

An alternative to using the proxy to deduce prices is to use it to infer expected payoffs and, in particular, expected returns. We now explore the connection between the expected return error bound deduced by Shanken (1987) and least squares Problem 1. The connection turns out to be close for factor models in which factor prices are chosen according to the least squares criterion of Problem 1.

Consider a stochastic discount factor m with a positive mean. When we use the covariance decomposition (7), we have the familiar characterization of a risk return tradeoff:

$$E(p) = \frac{\pi(p)}{E(m)} - \frac{\text{cov}(m, p)}{E(m)}, \tag{31}$$

where $1/E(m)$ plays the role of a riskless return. (More generally, $E(m)$ is the average price of a unit payoff.) We interpret $-\text{cov}(m, p)/E(m)$ as a measure of the compensation for holding a risky portfolio. When $m = a + br^m$, where r^m is the market return, it is straightforward to deduce the familiar single-beta representation of the risk-return tradeoff in terms of the market return.

Now suppose we use proxy y for m in formula (31) to deduce the approximate expected payoffs $E^a(p)$ to an investment of $\pi(p)$. When m and y have the same mean (that is, when they assign the same price to a unit payoff), the expected payoff error is

$$E^a(p) - E(p) = \frac{\text{cov}(m - y, p)}{E(y)}. \tag{32}$$

By means of the Cauchy–Schwarz Inequality, we bound the absolute expected payoff error as follows:

$$|E^a p - Ep| \leq \frac{\|m - y\|\, \text{std}(p)}{Ey}. \tag{33}$$

To sharpen this bound, we are led to solve Problem 3.

Problem 3:

$$\hat{\delta} \equiv \min_{m \in \mathcal{M}, Em = Ey} \|y - m\|.$$

Then

$$|E^a(p) - E(p)| \leq \frac{\hat{\delta}\, \text{std}(p)}{E(y)}. \tag{34}$$

Optimization Problem 3 can be converted into a special case of optimization Problem 1 and solved accordingly. The conversion is done by replacing P with the larger space $\hat{P}$, which is formed by the span of a unit payoff and P, and by assigning a price of Ey to the unit payoff. Let $\hat{\mathcal{M}}$ denote the smaller collection of admissible stochastic discount factors for this larger space. By construction, all of the discount factors will have mean Ey, since that is the (average) price assigned to a unit payoff. In other words,

$$\hat{\mathcal{M}} = \{m \in \mathcal{M} : Em = Ey\}, \tag{35}$$

which gives us an equivalent way to represent the constraint set of Problem 3. By simply mimicking the proof of Proposition 2.1, it can be shown that the expected payoff error bound given in equation (34) is sharp.

PROPOSITION 4.1: *Suppose that P does not contain a unit payoff, that Assumptions 1.1 and 1.2 are satisfied, and that $Ey > 0$. Then*

$$\frac{\hat{\delta}}{Ey} = \max_{p \in P, std(p)=1} |E^a(p) - E(p)|.$$

This proposition shows the connection between one of our measures of model misspecification and the expected return error bound of Shanken (1987). An unpleasant aspect of this result is that it is predicated on the proxy's price assignment for a unit payoff. That is, in forming approximate expected payoffs, we take as given the valuation assignment of Ey to a unit payoff. Although we have deliberately chosen to exclude a unit payoff from P in our formal statement of Proposition 4.1, in fact the result still holds if the unit payoff is included in P and if the proxy assigned the correct price to it. From this vantage point, Proposition 4.1 seems to be of limited interest, because the proxy is required to price correctly a unit payoff. However, as we will now see, for a linear factor model this concern is mitigated by the fact that the implicit proxy is not known ex ante but instead depends on an unknown parameter vector.

For Ross's (1976) linear factor model, the stochastic discount factor proxy is of the form

$$y = f' \theta_0 + \gamma_0, \tag{36}$$

where f is a random vector of factors and θ_0 and γ_0 are components of a parameter vector that is to be estimated. For instance, as we noted in Section II, in the absence of conditioning information, the CAPM implies such a discount factor proxy for f equal to the market return (possibly scaled and translated). When the factor components have a mean of zero, are mutually uncorrelated, and have unit standard deviations, the elements of θ_0 can be interpreted as factor prices and γ_0 is the price assignment to a unit payoff. Since we are confronted by the unknown parameter vector θ, we study the following modification of optimization Problem 1.

Problem 4:

$$\delta = \min_{(\theta,\gamma)} \min_{m \in \mathcal{M}} \|f'\theta + \gamma - m\|.$$

Note that the first-order condition for the choice of γ is given by

$$Ey - \tilde{m} = 0, \tag{37}$$

where $\tilde{m}$ is the optimal choice of m in $\mathcal{M}$. In other words, the parameters for the discount factor proxy are chosen so that the fitted proxy has the same mean as the admissible stochastic discount factor that is closest to it. In light of Propositions 2.1 and 4.1, the least squares distance δ from the proxy y to the space $\mathcal{M}$ of stochastic discount factors gives both the maximum pricing error and the maximum expected payoff error (times Ey) for linear factor models.

The key ingredient in this argument is that there is an unknown constant term in the stochastic discount factor that gets chosen to minimize the least squares criterion. The presumed linearity in the other factors is not essential to the argument. Of course, not all stochastic discount factor parameterizations have an unknown constant term. (See Hansen and Singleton (1982) and Brown and Gibbons (1985) for examples based on consumption CAPMs with parameterized utility functions.) Once it is conceded that the proxy might misprice a unit payoff, least squares Problem 3 seems less appealing to us. Why should we constrain the family of discount factors to price the unit payoff in precisely the same way the proxy does?

B. Factor Mimicking Payoffs

We complete our discussion by providing a characterization of the solution to Problem 4. Because of the solution characterization of Problem 1 provided in Section II, we know that for each θ,

$$\min_{m \in \mathcal{M}} \|f'\theta + \gamma - m\| = \|\mathrm{Proj}(f|P)'\theta + \mathrm{Proj}(1|P)\gamma - \mathrm{Proj}(m|P)\|, \tag{38}$$

where $\mathrm{Proj}(\cdot|P)$ denotes the least squares projection onto the closed linear space P. The operations $\mathrm{Proj}(f|P)$ and $\mathrm{Proj}(1|P)$ have the simple interpretations of forming factor mimicking payoffs as analyzed by Huberman, Kandel, and Stambaugh (1987). Also, since all admissible stochastic discount factors imply the same pricing functional on P, $\mathrm{Proj}(m|P)$ does not vary with m. We let p^* denote this projection, which is itself an admissible stochastic discount factor. As a consequence of equation (38), we can reduce Problem 4 to

$$\delta = \min_{(\theta,\gamma)\in P} \|\mathrm{Proj}(f|P)'\theta + \mathrm{Proj}(1|P)\gamma - p^*\|. \tag{39}$$

The optimization problem in equation (39) is quadratic in the parameter vector and its solution is easy to interpret: choose θ and γ to price correctly the factor

mimicking payoffs. This interpretation follows directly from the first-order conditions (remembering that p^* prices correctly all payoffs in P).

V. Implementation and Approximation

The analysis so far has been conducted using population expectations, which must be approximated in practice. As we indicated in Section I, we presume the existence of time series data of the form $\{(q_t, x_{t+\tau}, y_{t+\tau}): t = 1, 2, \ldots, T\}$ for sample size T. The stochastic process generating these data is modeled as being stationary, at least asymptotically. Our approach is to solve sample counterparts to the conjugate maximization problems discussed at the end of Section III. Consequently, our estimator (d_T) of the distance between the proxy and the set of discount factors is given by

$$d_T = \left\{ \max_{\lambda \in \mathbf{R}^n} (1/T) \sum_{t=1}^{T} \left[(y_{t+\tau})^2 - (y_{t+\tau} - \lambda' x_{t+\tau})^2 - 2\lambda' q_t \right] \right\}^{1/2}, \quad (40)$$

and our estimator (d_T^+) of the distance between the proxy and the set of positive stochastic discount factors is

$$d_T^+ = \left\{ \max_{\lambda \in \mathbf{R}^n} (1/T) \sum_{t=1}^{T} \left[(y_{t+\tau})^2 - (y_{t+\tau} - \lambda' x_{t+\tau})^{+2} - 2\lambda' q_t \right] \right\}^{1/2}. \quad (41)$$

While finite sample results would be preferable, they appear to be hard to obtain at the level of generality of this article. For this reason, Hansen, Heaton, and Luttmer (1995) derive the limiting (or asymptotic) distribution of these estimators (see Proposition 3.2 in Hansen, Heaton, and Luttmer (1995)). From their results it follows that

$$T^{1/2}(d_T - \delta) \xrightarrow{\mathscr{D}} N[0, \, \hat{\sigma}^2/(4\delta^2)], \quad (42)$$

where the scalar $\hat{\sigma}^2$ is the variance in the following central limit approximation:

$$T^{1/2} \sum_{t=1}^{T} \left[(y_{t+\tau})^2 - (y_{t+\tau} - \hat{\lambda}' x_{t+\tau})^2 - 2\hat{\lambda}' q_t - \delta^2 \right] \xrightarrow{\mathscr{D}} N(0, \hat{\sigma}^2). \quad (43)$$

Recall that $\hat{\lambda}$ maximizes the population conjugate problem. Hence the numerator term $\hat{\sigma}^2$ in the asymptotic variance in equation (42) comes from the central limit approximation for the criterion of the conjugate maximization problem evaluated at the solution to the population problem and centered appropriately. The denominator term $4\delta^2$ is present because of the mean-value approximation in the transformation of δ^2 to δ. However, the limiting distribution for d_T does not include an adjustment for the fact that $\hat{\lambda}$ is approximated by the

sample solution to the conjugate maximization problem. Even though this latter approximation always results in a larger estimate of δ, it turns out that the impact of estimating $\hat{\lambda}$ does not alter the limiting distribution for δ.

To use this limiting distribution in practice requires that we obtain an estimate of the scalar asymptotic variance $\hat{\sigma}^2$ in equation (43). This can be accomplished by forming a scalar time series sequence $\{u_{t,T}: t = 1, 2, \ldots, T\}$, where

$$u_{t,T} \equiv (y_{t+\tau})^2 - (y_{t+\tau} - \hat{\ell}'_T x_{t+\tau})^2 - 2qt'\hat{\ell}_T \qquad (44)$$

and where $\hat{\ell}_T$ is the value of λ that maximizes equation (40). Notice that the sample mean of $\{u_{t,T}\}$ is $(d_T)^2$. Then $\hat{\sigma}^2$ can be estimated by using one of the frequency zero spectral density estimators described by Newey and West (1987) or Andrews (1991) applied to the time series sequence $\{u_{t,T} - (d_T)^2: t = 1, 2, \ldots, T\}$. Let s_T denote the resulting estimator. Then statistical inference can be based on the approximation

$$\frac{T^{1/2}d_T}{2s_T}(d_T - \delta) \xrightarrow{\mathcal{D}} N(0,1). \qquad (45)$$

The validity of this approximation requires that δ be strictly positive.

Although the limiting distribution of d_T is altered when δ is zero, this phenomenon occurs only when the proxy is a valid discount factor, i.e., when

$$E(yx - q) = 0. \qquad (46)$$

This restriction is what underlies the generalized-method-of-moments estimation and the inference methods described in Hansen and Singleton (1982). In particular, for a given proxy, this moment condition can be checked using an asymptotic chi-square test.

The same distribution theory given in equation (42) also applies when the proxy y depends on unknown parameters, as in observable factor models or utility-based models with unknown preference parameters. For instance, consider observable factor models for which the time series of proxies satisfy

$$y_{t+\tau} = \hat{\theta}' f_{t+\tau} + \hat{\gamma}, \qquad (47)$$

where $\hat{\theta}$ is unknown. The sample estimator of the specification-error bound can now be represented as

$$d_T = \left\{ \min_{(\theta,\gamma)} \max_{\lambda \in \mathbf{R}^n} (1/T) \sum_{t=1}^{T} [(\theta'f_{t+\gamma} + \gamma)^2 - (\alpha'f_{t+\tau} - \lambda'x_{t+\tau})^2 - 2\lambda'q_t] \right\}^{1/2}. \qquad (48)$$

The limiting distribution in equation (42) still remains valid as long as the population least-squares distance, δ, between the family of proxies and the

family of valid stochastic discount factors is strictly positive. In using formula (48) we are implicitly employing the sample two-stage least squares estimator for $\hat{\lambda}$ described in Section II.

A parallel set of results applies to the estimator d_T^+ of δ^+. For instance,

$$T^{1/2}(d_T^+ - \delta^+) \xrightarrow{\mathscr{D}} N[0, \, \tilde{\sigma}^2/(4\delta^{+2})], \tag{49}$$

where the scalar $\tilde{\sigma}^2$ is the variance in the following central limit approximation:

$$T^{1/2} \sum_{t=1}^{T} [(y_{t+\tau})^2 - (y_{t+\tau} - x_{t+\tau}'\tilde{\lambda})^{+2} - 2qt'\tilde{\lambda} - \delta^{+2}] \xrightarrow{\mathscr{D}} N(0, \, \tilde{\sigma}^2). \tag{50}$$

VI. Applications

In this section we apply the least squares measures of specification errors developed in the previous sections to a variety of alternative models of stochastic discount factors that have appeared in the literature. We use a common vector of six asset returns in making comparisons among models.

A. Data Description

Time series of six monthly returns for the period 1959:1–1990:12 were constructed as follows. (In this application, τ is one.) The first return is the equally weighted portfolio of New York Stock Exchange (NYSE) stocks in the largest size decile; the second is an equally weighted portfolio of NYSE stocks in the smallest size decile; the third is a portfolio of long-term government bonds. These three assets are a subset of the assets used by Ferson and Constantinides (1991) and Ferson and Harvey (1992). The remaining three returns were constructed as "managed portfolios" in which $1 - z$ units were invested in the one-month Treasury bill and z units were invested in the largest size decile portfolio. For the fourth return, the portfolio weight z is the annual yield difference between Aaa and Baa bonds; for the fifth return, z is the annual yield difference between Baa bonds and the one-month Treasury bill; and for the last return, z is the annual yield difference between one-year and one-month Treasury bills. In all cases the yield differences were selected from the previous time periods so that they were in the conditioning information sets of investors at the time of the investment. More details of the data construction and sources are given in Appendix B.

B. Constant Discount Factors

We begin by evaluating constant discount factor models, i.e.,

$$y = \beta \tag{51}$$

for some β. We are mainly interested in these models as points of reference, because constant discount factor models imply that the asset valuation is risk neutral. We perform two calculations. First, we compute the constant discount factors that minimize the two specification error measures. We also find which value of β is most plausible from the vantage point of the chi-square statistic.

Consider the least squares problem:

$$
\begin{aligned}
\delta &= \min_{\beta} \min_{m \in \mathcal{M}} E[(\beta - m)^2]^{1/2} \\
&= \min_{m \in \mathcal{M}} \min_{\beta} E[(\beta - m)^2]^{1/2} \\
&= \min_{m \in \mathcal{M}} \text{std}(m),
\end{aligned} \tag{52}
$$

where $\text{std}(m)$ is the standard deviation of the stochastic discount factor m. Hence, in this special case, the discount factor closest to the family of constant random variables is just the stochastic discount factor that is least variable and the distance is given by the standard deviation of that least variable stochastic discount factor.

For the data set used in our application, the estimated mean of the least volatile stochastic discount factor is 0.998, with a standard deviation of 0.329. As we have argued, this standard deviation is also our estimate of δ, with a standard error of 0.054. The least variable stochastic discount factor is positive for almost all of the sample points, so there is no difference between the estimates of δ and δ^+ in the first three decimal points.

For comparison, we also report the corresponding results obtained by applying generalized method of moments (GMM), as in Hansen (1982). When the model is correctly specified and β is known, the GMM criterion is constructed to have a chi-square limiting distribution, with degrees of freedom equal to the number of assets (six). When β is unknown, a GMM estimate is obtained by minimizing the criterion function. The limiting distribution of the minimized value of the GMM criterion remains chi-square, but with one less degree of freedom. The minimum chi-square value for the family of constant discount factors is 41.6. Since the degree of freedom for the associated chi-square test is five, the probability value is essentially zero. Thus there is considerable statistical evidence against risk-neutral pricing.

C. Power Utility

We now extend the collection of models that we explore using the familiar class of time-separable power utility models. We parameterize the time t marginal utility to be $(c_t)^{-\gamma}$ for positive γ. The implied stochastic discount factor is the intertemporal marginal rate of substitution between time $t + 1$ and time t:

$$
y_{t+1} = \beta(c_{t+1}/c_t)^{-\gamma}, \tag{53}
$$

<div align="center">

Table I

Specification Errors for Power Utility

</div>

The stochastic discount factor proxy is $\beta(c_{t+1}/c_t)^{-\gamma}$, where β is the rate of time preference, c_t is the per capita nondurables consumption for month t, and γ is the coefficient of relative risk aversion of the representative consumer. Details of the data construction and sources are given in Section VI.A and Appendix B. The column labeled δ gives the least squares distances between the proxies and the family of stochastic discount factors, and the column labeled δ^+ gives the distances between the proxies and the family of nonnegative stochastic discount factors. The column labeled χ^2 gives the chi-square test statistics for the null hypothesis that all pricing errors are zero. Numbers in parentheses are estimated standard errors. The standard errors are computed using the method described in Newey and West (1987) with $m = 15$. (Very similar results are obtained using $m = 9$ and 12.) The entries in the row labeled "minimized" are computed by selecting parameter values for β and γ that minimized the corresponding criteria. In the column labeled δ the minimizers are $\beta = 1.03$ and $\gamma = 33.3$; in the column δ^+ the minimizers are $\beta = 1.03$ and $\gamma = 28.6$; and in the column labelled χ^2 the minimizers are $\beta = 1.608$ and $\gamma = 249.5$.

γ	δ	δ^+	χ^2
	$\beta = 0.95$		
0	0.332	0.333	2,824.9
	(0.054)	(0.054)	
1	0.333	0.333	2,850.9
	(0.053)	(0.054)	
5	0.333	0.333	1,557.7
	(0.053)	(0.054)	
10	0.333	0.333	705.9
	(0.053)	(0.053)	
15	0.334	0.334	419.2
	(0.052)	(0.053)	
	$\beta = 1.00$		
0	0.329	0.329	46.6
	(0.054)	(0.055)	
1	0.329	0.329	42.6
	(0.054)	(0.054)	
5	0.329	0.329	47.8
	(0.054)	(0.054)	
10	0.328	0.328	55.7
	(0.053)	(0.054)	
15	0.328	0.328	58.1
	(0.053)	(0.054)	
Minimized	0.327	0.327	35.2
	(0.053)	(0.053)	

where β is the subjective discount factor. We use consumption of nondurables as our proxy for c_t, which leaves us with a two-parameter family of stochastic discount factors. In Table I we report estimates of both least squares distance measures for several alternative choices of β and γ.

Several important conclusions emerge from this table. First, the positivity restriction on the admissible stochastic discount factors has very little impact for these data, as is shown by the slight difference between the distance measures δ and δ^+. Second, there is very little distance variation across the

discount factor proxies. For the (β, γ) pairs reported, the distances range from only 0.328 to 0.335, which is certainly a small range considering the size of the estimated standard errors. The minimum specification errors for the entire family of stochastic discount factors is 0.327, only expanding the range slightly. Thus the estimated maximum pricing errors always exceed thirty percent of the norm. Moreover, the power utility model barely reduces the estimated pricing error relative to the constant discount factor model described previously.

To understand better why the specification-error criterion is so large, recall that the norm of a random variable can be decomposed into a mean component and a standard deviation component via the formula

$$\|p\| = [E(p)]^2 + [\text{std}(p)^2]^{1/2}. \tag{54}$$

We applied this decomposition to $\tilde{p}$ while recalling that δ is the norm of this random variable. More precisely, we computed the means and standard deviations of $\tilde{p}$ for the different specifications of y. We find that the mean is typically very small relative to the standard deviation, indicating that most of the norm of $\tilde{p}$ is attributable to its standard deviation. The random variable $\tilde{p}$ can be decomposed as $\tilde{p} = \hat{p} - p^*$, where $\hat{p}$ is the least square projection of the proxy y on the payoff space P, and where p^* is the unique stochastic discount factor in P. As is emphasized by Cochrane and Hansen (1992), the standard deviations of the $\hat{p}$s are small for the class of power utility models, so most of the variation in the $\tilde{p}$s comes from variation in p^*, which is invariant to the choice of proxy. This is confirmed by the fact that the correlation among the $\tilde{p}$s is extremely high across the power utility models (in excess of 0.999).

Even though the range of the chi-square statistics is extremely large for the parameter values reported in Table I, they all have extremely small probability values for the null hypothesis that the model is correctly specified. By minimizing the value of the chi-square statistic by our choice of β and γ, we obtain coefficient estimates that are perverse, with a very large (absolute) value of γ and with a chi-square statistic that remains quite large.[11]

Recall from formula (29) that the specification-error δ is similar to a scaled version of the chi-square statistic. Both objects are square roots of quadratic forms in the pricing error vector. As is illustrated in Table I, the relative magnitudes (across proxies) of these two objects can be quite different in practice. The critical distinction is that a chi-square statistic uses a distance measure in the quadratic form, which changes with the proxy and rewards variation. This feature of the chi-square statistic shows up because, from the vantage point of classical statistics, it is harder to reject models (at a fixed significance level) with highly variable discount factor proxies. By design, our

---

[11] While the chi-square value is reduced to 35.2, this remains far to the right tail of a chi-square distribution. However, the accuracy of the asymptotic approximations with large (in absolute value) powers are likely to be extremely poor, because the sample moment calculations are dominated by a small number of recessionary (negative growth rate) data points.

<div align="center">

Table II

Lagrange Multipliers for Power Utility
</div>

The stochastic discount factor proxy is $\beta(c_{t+1}/c_t)^{-\gamma}$. The numbers reported are the Lagrange multipliers for the pricing constraints (denoted by the vector λ in equation (40) in the text) for the least squares optimization problem. A distinct multiplier is reported for each of the six securities listed in the first column. Numbers in parentheses are estimated standard errors. The standard errors are computed using the method described in Newey and West (1987) with $m = 15$. Descriptions of the securities and details of the data construction and sources are given in Section VI.A and Appendix B.

	$\beta = 1.00$		
Security	$\gamma = 0$	$\gamma = 5$	$\gamma = 15$
1	8.60	8.61	8.63
	(3.89)	(3.88)	(3.88)
2	1.81	1.80	1.77
	(0.96)	(0.97)	(0.98)
3	4.74	4.79	4.89
	(1.80)	(1.81)	(1.82)
4	−13.46	−13.47	−13.50
	(3.71)	(3.72)	(3.72)
5	3.76	3.75	3.73
	(0.89)	(0.88)	(0.88)
6	−5.56	−5.58	−5.64
	(1.77)	(1.77)	(1.78)

least squares specification-error measures do not reward stochastic discount factor variability.

In Table II we report the estimates of the Lagrange multipliers that are associated with the pricing constraints. According to the duality approach of Section III, these multipliers are also the portfolio weights for the factor payoff $\tilde{p}$ that we use to correct the proxy. Given the high correlation among the $\tilde{p}$s, it is not surprising that the Lagrange multipliers are very similar in magnitude as the power parameter γ is altered. Notice in particular that the coefficients on the three managed portfolios (payoffs 4, 5, and 6) are all large relative to their standard errors.

D. Consumption Externality

We now alter the preferences of consumers by introducing an externality in preferences as is done in Abel (1990). An extreme version of this preference specification has the time t felicity function as a power of the ratio c_t/c_{t-1}^*, where c_{t-1}^* is time $t - 1$ community-wide consumption. The corresponding marginal utility of consumption is modified to be

$$mu_t = (c_t)^{-\gamma}(c_{t-1})^{\gamma-1}, \tag{55}$$

<div align="center">

Table III

Specification Errors for Power Utility With Consumption Externality

</div>

The stochastic discount factor proxy is $\beta(c_{t+1}/c_t)^{-\gamma}(c_t/c_{t-1})^{\gamma-1}$, where β is the rate of time preference, c_t is the per capita nondurables consumption for month t, and γ is the coefficient of relative risk aversion of the representative consumer. Details of the data construction and sources are given in Section VI.A and Appendix B. The column labeled δ gives the least squares distances between the proxies and the family of stochastic discount factors, and the column labeled δ^+ gives the distances between the proxies and the family of nonnegative stochastic discount factors. The column labeled χ^2 gives the chi-square test statistics for the null hypothesis that all pricing errors are zero. Numbers in parentheses are estimated standard errors. The standard errors are computed using the method described in Newey and West (1987) with $m = 15$. The entries in the row labeled "minimized" are computed by selecting parameter values for β and γ that minimize the corresponding criteria. In the column labeled δ the minimizers are $\beta = 0.998$ and $\gamma = 7.3$; in the column δ^+ the minimizers are $\beta = 0.999$ and $\gamma = 4.8$; and in the column labelled χ^2 the minimizers are $\beta = 21.1$ and $\gamma = 237.6$.

γ	δ	δ^+	χ^2
	$\beta = 0.95$		
0	0.332	0.332	2,762.1
	(0.054)	(0.054)	
1	0.332	0.332	2,837.7
	(0.054)	(0.054)	
5	0.332	0.332	807.6
	(0.053)	(0.054)	
10	0.332	0.332	244.0
	(0.053)	(0.053)	
15	0.332	0.332	123.9
	(0.052)	(0.053)	
	$\beta = 1.00$		
0	0.328	0.339	42.2
	(0.054)	(0.055)	
1	0.328	0.328	42.3
	(0.054)	(0.055)	
5	0.328	0.328	42.1
	(0.054)	(0.055)	
10	0.328	0.329	42.8
	(0.053)	(0.054)	
15	0.328	0.329	42.8
	(0.053)	(0.053)	
Minimized	0.328	0.328	27.5
	(0.053)	(0.054)	

where we have imposed the equilibrium condition that $c^*_{t-1} = c_{t-1}$. The stochastic discount factor proxy is now

$$y_{t+1} = \beta(c_{t+1}/c_t)^{-\gamma}(c_t/c_{t-1})^{\gamma-1}. \tag{56}$$

In Table III we report results for the same parameter configurations as in Table I. The specification-error results are very similar. The chi-square statistics are reduced a little, but not sufficiently to make the null hypothesis,

that the proxies are in $\mathcal{M}^+$, appear plausible. Therefore, the introduction of a consumption externality has very little impact on either the specification-error measures or the statistical inferences.

E. Reciprocal of the Market Return

We next consider the inverse market portfolio as a discount factor proxy. This random variable can be justified as a discount factor by following Rubinstein (1976) and by assuming that consumers' preferences are time and state separable and have a logarithmic felicity function (a power utility with $\gamma = -1$). More generally, Epstein and Zin (1991) rationalize this discount factor by relaxing state-separability and imposing a logarithmic risk correction in a recursive utility formulation. We treat this as an error-ridden proxy for the familiar reasons delineated by Roll (1977). The specification-error measures, with and without positivity imposed, are again essentially the same; they are equal to 0.311, with a standard error of 0.055. This is a small improvement over the best specification error from the power utility model.[12]

F. Observable Factor Models

The final stochastic discount factor proxies that we consider are those implied by three alternative linear factor models. For each of the three models there are two factors: a constant factor and a single variable factor. The first two of these models employ the implicit stochastic discount factor for the one-period CAPM: a constant plus a scale multiple of the market return. We use both the equally-weighted and the value-weighted returns as alternative measures of the market return; this gives rise to two alternative linear factor models. The third model imitates Breeden, Gibbons, and Litzenberger (1989) by using the consumption growth rate as an observable factor.

The results are reported in Table IV. Since linear factors models are not typically designed to price derivative claims, we report only the δ measure of model misspecification. In generating these results, the unknown coefficients were estimated both by minimizing the specification error and by minimizing the chi-square statistic. Again, according to the chi-square statistic there is substantial statistical evidence against all three models. The smallest specification error, 0.286, and the value of the chi-square statistic, 30.2, are found when the variable factor is the equally-weighted return, although the specification error is almost the same when the variable factor is the value-weighted return.

Particularly in the case of the consumption factor model, there is a substantial difference in the selected discount proxy, depending on whether the specification-error measure is minimized or the chi-square statistic is minimized. In the latter case the sample standard deviation of the implied approximating stochastic discount factor is more than thirty times more variable, illustrating

---

[12] Interestingly, the chi-square statistic is 128, so a naive use of classical chi-square statistics might lead one to argue in favor of the power utility model.

<div align="center">

Table IV

Specification Errors for Linear Factor Models
</div>

The stochastic discount factor is $\hat{\alpha}'f_{t+1}$ where f_{t+1} contains a constant and one of the variable factors listed in the first column of the table, and $\hat{\alpha}'$ is the estimated parameter vector. Details of the data construction and sources are given in Section VI.A and Appendix B. The column labeled δ gives the least squares distances between the family of proxies and the family of stochastic discount factors. The column χ^2 gives the chi-square test statistic for the null hypothesis that all pricing errors are zero. Estimated standard errors for δ are given in parentheses. The standard errors are computed using the method described in Newey and West (1987) with $m = 15$.

Variable Factor	δ	χ^2
Coefficients Estimated by Minimizing the Specification Error		
Equally-weighted return	0.286	30.1
	(0.054)	
Value-weighted return	0.289	33.5
	(0.052)	
Consumption growth	0.325	38.9
	(0.052)	
Coefficients Estimated by Minimizing the Value of the χ^2		
Equally-weighted return	0.286	29.3
Value-weighted return	0.290	31.5
Consumption growth	1.57	12.0

that the chi-square criterion rewards variability in the discount factor proxy. A naive econometrician using only the chi-square values reported in the last column of the second panel in Table IV might erroneously conclude that the consumption factor model does better than the other two factor models. This conclusion is problematic, because the chi-square statistics are uniformly large relative to their degrees of freedom. While these statistics show that particular classes of models are misspecified, they are not designed for making comparisons among misspecified models. As we emphasized in Section III, our least squares measures are better suited for this task. To see that minimizing the chi-square statistic gives a different picture than minimizing the least squares specification error, notice that the least squares specification error exceeds one for the consumption factor model when the parameter estimates are obtained by minimizing the chi-square statistic.

The estimated multipliers on the pricing constraints and their standard errors are reported in Table V. Again, the coefficients on the three managed portfolios (payoffs 4, 5, and 6) are all large relative to their standard errors. Recall that these portfolio payoffs are constructed using variables in the conditioning information sets of investors. The role of securities such as these in constructing the "price factor" p given in equation (28) lends support to the unconditional version of the CAPM derived by Jagannathan and Wang (1993, 1996) and the empirical factor models of Cochrane (1992) and Bansal, Hsieh, and Viswanathan (1993).

In Table VI we report the correlations in the additional factor $\tilde{p}$ used to correct the candidate discount factors. Not surprisingly, the correlations among the payoffs with the maximum pricing error (per unit norm) across

Table V

Lagrange Multipliers for Factor Models

The stochastic discount factor is $\hat{\alpha}'f_{t+1}$ where f_{t+1} contains a constant and one of the variable factors listed in the first column of the table, and $\hat{\alpha}'$ is the estimated parameter vector. The numbers reported are the Lagrange multipliers for the pricing constraints (denoted by the vector λ in equation (40) in the text) for the least squares optimization problem. A distinct multiplier is reported for each of the six securities listed in the first column. Numbers in parentheses are estimated standard errors. The standard errors are computed using the method described in Newey and West (1987) with $m = 15$. Details of the data construction and sources are given in Section VI.A and Appendix B.

Security	Variable Factor			
	None	Consumption Growth	Equally-Weighted Return	Value-Weighted Return
1	8.80	8.93	8.64	6.90
	(4.63)	(4.60)	(4.71)	(4.71)
2	1.71	1.50	−0.05	1.41
	(1.10)	(1.03)	(0.80)	(1.01)
3	4.80	5.42	6.10	6.30
	(2.20)	(2.32)	(2.10)	(2.04)
4	−13.52	−13.70	−13.00	−12.80
	(3.94)	(4.10)	(3.74)	(3.70)
5	3.80	3.64	3.10	2.90
	(1.0)	(1.09)	(0.90)	(0.80)
6	−5.60	−5.90	4.90	−4.70
	(1.50)	(1.50)	(1.50)	(1.40)

Table VI

Correlations of $\tilde{p}$ Across Proxies

This table reports the correlations among the estimated $\tilde{p}$s for the proxies implied by four different models with alternative variable factors, where $\tilde{p}$ [given in equation (28)] is the smallest adjustment in the least squares sense required to make the stochastic discount factor price the given set of assets right. The model with the variable factor labeled "None" has a discount factor that is constant. Details of the data construction and sources are given in Section VI.A and Appendix B.

Variable Factors	Consumption Growth	Equally-Weighted Return	Value-Weighted Return
None	0.99	0.87	0.88
Consumption growth		0.92	0.94
Equally-weighted return			0.96

factor models are quite high. (They all exceed 0.87). Abstracting from any sampling error, this high correlation indicates that whatever is missing in one proxy is also missing in the other. However, the following caution is in order. Suppose that the mispricing is an artifact of poor asymptotic approximations. For instance, suppose that our asymptotic measures of sampling error under-

state the true sampling variability by making a portfolio that performs well in ·a sample appear to be mispriced. In this case, even if all of the factor models are correctly specified, the estimated $\bar{p}$s could well be highly correlated across models.

VII. Concluding Remarks

In this article we explore alternative ways to compare stochastic discount factor proxies when it is understood that the models under consideration are misspecified. In contrast to comparisons based on chi-square statistics, our measures do not reward variability of the proxies. Instead, they focus on measuring maximum pricing errors (per unit norm). Using our alternative least squares measures, we show the sense in which power utility models do not perform much better than constant discount factor models. While the estimated specification errors are a little smaller for CAPM-style models in which the stochastic discount factor is a linear combination of a constant plus a scale factor times a measure of the market return, even these models imply specification errors that exceed twenty-five percent of the norm for some payoffs. Bansal, Hsieh, and Viswanathan (1993) use our apparatus to document the advantages of nonlinear factor models over linear ones in pricing international securities; Bakshi and Chen (1995) use it to investigate the extent to which investors directly incorporate wealth into their preference-orderings when making portfolio decisions. Finally, Chen and Knez (1995) extend our setup to measure the degree to which markets are integrated.

We propose and justify two alternative measures of model misspecification. The first measure relies on linearity of the pricing functional. The second measure exploits the positivity of arbitrage-free stochastic discount factors for pricing derivative claims. The second measure is conceptually more appealing when evaluating candidate discount factors for the pricing of derivative claims. Associated with this measure is a nonnegative stochastic discount factor that prices the original collection of payoffs correctly. This discount factor is the closest (in a least squares sense) to the prespecified proxy, and it assigns prices to derivative claims within (or at the boundary of) the range of arbitrage-free price assignments to derivative claims. Thus the misspecified parametric model can be used to select among the family of nonnegative stochastic discount factors in assigning prices to derivative claims.

For the data used in this article, the differences between the two measures of model misspecification are negligible. This, by itself, is an interesting insight for stochastic discount factor models designed for use in pricing derivative securities. However, this finding is unlikely to be true for all data sets. For instance, the findings in Hansen and Jagannathan (1991) suggest that using holding-period returns for Treasury bills might lead to a more substantial discrepancy between specification-error measures. Moreover, from the work of Luttmer (1994) it is important to take explicit account of transaction costs. Hansen, Heaton, and Luttmer (1995) show how to incorporate such market frictions into the measures described here.

Appendix A

The first proposition establishes a result stated in Section I.

PROPOSITION A.1: *Suppose that P is given by equation (3). Then* π*, given by equation (5), is a bounded linear functional.*

Proof: To see this, consider any $w \cdot x$ with a norm less than or equal to one. Let Λ denote any factorization of the conditional second-moment matrix $E(xx'|\mathcal{G})$, and note that

$$
\begin{aligned}
\|\Lambda w\|^2 &= E[w'E(xx'|\mathcal{G})w)] \\
&= E(w'xx'w) \\
&= \|w \cdot x\|^2.
\end{aligned}
\tag{A.1}
$$

Furthermore,

$$
\begin{aligned}
E|w \cdot q| &= E|w'\Lambda'(\Lambda')^{-1}q| \\
&\leq \|\Lambda w\|\|(\Lambda')^{-1}q\|.
\end{aligned}
\tag{A.2}
$$

Finally, note that

$$
\begin{aligned}
\|(\Lambda')^{-1}q\|^2 &= E[q'E(xx'|\mathcal{G})^{-1}q] \\
&< \infty.
\end{aligned}
\tag{A.3}
$$

It follows that π is well defined on P and bounded.   Q.E.D.

The next proposition imitates a result in Luttmer (1994). It is used to justify the duality approach applied in Section III.

PROPOSITION A.2: *Suppose that P is generated by a random vector x with a nonsingular second-moment matrix, that the corresponding price vector q has a finite first moment, and that Assumption 1.3 is satisfied. Then Eq is an interior point in the set* $\{Emx: m \in L^2, m \geq 0\}$*.*

Proof: Let $\mathcal{C}^1(\gamma) \equiv \{\alpha: |\alpha| = 1 \text{ and } \alpha'\gamma \geq 0\}$ and $\mathcal{C}^2 \equiv \{\alpha: \| (\alpha'x)^+ \| = 0 \text{ and } |\alpha| = 1\}$. For any $\gamma \in \mathbf{R}^n$, let $\pi_\gamma(\alpha'x) = \alpha'\gamma$. Then $\mathcal{C}^1(\gamma) \cap \mathcal{C}^2$ is empty if and only if (P, π_γ) satisfies Condition N. (There are no-arbitrage opportunities on (P, π_γ).) Therefore, $\mathcal{C}^1 (Eq) \cap \mathcal{C}^2$ is empty. Since $\mathcal{C}^1 (Eq)$ and $\mathcal{C}^2$ are both compact sets, there is a positive distance between them (where distance is measured using the Hausdorff metric). Moreover, $\mathcal{C}^1$ is a continuous function of the price assignment γ. Consequently, there is an open ball around Eq such that, for all γ in this ball, $\mathcal{C}^1 (\gamma) \cap \mathcal{C}^2$ is empty. For any such γ, the linear functional π_γ can be represented as $\pi(p) = E(pm)$ for some nonnegative $m \in L^2$. (In fact, m can be chosen to be strictly positive with probability one.)   Q.E.D.

Appendix B

The large and small capitalization stock portfolio returns are constructed using data from the Center for Research in Security Prices (CRSP) tapes. We sorted firms on the NYSE into decile classes based on the market value of common stocks outstanding in June of each year. We then constructed the time series of monthly returns for an equally-weighted portfolio of stocks in the largest and smallest size deciles for the twelve v months in the subsequent calendar year. We repeated this procedure for each calendar year and then spliced together the monthly series for different calendar years.

Monthly returns on long-term government bonds and one-month Treasury bills were from *Stocks, Bonds, Bills and Inflation – 1991 Year Book* of Ibbotson Associates. Yields on one-month and one-year Treasury bills, as well as Aaa and Baa bonds, were taken from *Federal Reserve Bulletins.* Nominal returns were deflated using the price deflator series for consumption of nondurables from CITIBASE.

Monthly returns on the equally-weighted as well as value-weighted index of stocks in the NYSE and American Stock Exchange (AMEX), which were used to construct candidate marginal rates of substitution, are from the CRSP tapes. Monthly per capita consumption data used in constructing candidate marginal rates of substitution were constructed using data from CITIBASE.

Table B.1

Summary Statistics of Return Data

The six securities are described in Section VI.A.

Security	Mean	Standard Deviation	Skewness	Kurtosis
1	1.008	0.044	−0.122	4.975
2	1.013	0.071	0.899	10.402
3	1.005	0.029	0.789	6.216
4	1.006	0.056	1.066	9.124
5	1.021	0.167	0.697	10.824
6	1.001	0.056	1.043	13.598

Table B.2

Correlations Among Security Returns

The six securities are described in Section VI.A.

Security	2	3	4	5	6
1	0.70	0.33	0.92	0.90	0.72
2		0.14	0.63	0.60	0.42
3			0.40	0.31	0.30
4				0.93	0.80
5					0.81

Table B.1 gives the summary statistics for the return data. Notice that there is substantial variation in the characteristics of the six assets. For example, asset 4 has about the same average return as asset 3 but is almost twice as volatile.

Table B.2 gives the correlation among the six asset returns.

REFERENCES

Abel, Andrew B., 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review* 80, 38–42.

Allen, Edward R., III, 1991, Evaluating consumption-based models of asset pricing, Ph.D. dissertation, University of Chicago.

Andrews, Donald W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 307–346.

Bakshi, Gurdip S., and Zhiwu Chen, 1997, The spirit of capitalism and stock market prices, *The American Economic Review*, Forthcoming.

Bansal, Ravi, David A. Hsieh, and S. Viswanathan, 1993, A new approach to international arbitrage pricing, *Journal of Finance* 48, 1719–1747.

Bollerslev, Tim, and Robert F. Engle, 1993, Common persistence in conditional variances, *Econometrica* 61, 167–186.

Breeden, Douglas, Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989, Economic significance of predictable variations in stock index returns, *Journal of Finance* 44, 1177–1189.

Brown, David P., and Michael R. Gibbons, 1985, A single econometric approach for utility-based asset pricing models, *Journal of Finance* 40, 359–381.

Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Chen, Zhiwu, and Peter J. Knez, 1995, Measurement of market integration and arbitrage, *The Review of Financial Studies* 8, 287–325.

Clark, Stephen A., 1993, The valuation problem in arbitrage price theory, *Journal of Mathematical Economics* 22, 463–478.

Cochrane, John H., 1992, A cross-sectional test of a production-based asset pricing model, Working paper, NBER.

Cochrane, John H., and Lars P. Hansen, 1992, Asset pricing explorations for macroeconomics, in *NBER Macroeconomics Annual 1992* (MIT Press, Cambridge, Mass.), 115–165.

Durlauf, Steven N., and Robert E. Hall, 1989, Bounds on the variances of specification errors in models with expectations, Working paper, NBER.

Dybvig, Philip H., and Jonathan E. Ingersoll, Jr., 1982, Mean-variance theory in complete markets, *Journal of Business* 55, 233–251.

Epstein, Larry G., and Stanley E. Zin, 1991, Substitution, risk aversion, and temporal behavior of consumption and asset returns II: An empirical analysis, *Journal of Political Economy* 99, 263–286.

Ferson, Wayne E., and George M. Constantinides, 1991, Habit persistence and durability in aggregate consumption: Empirical tests, *Journal of Financial Economics* 29, 199–240.

Ferson, Wayne E., and Campbell R. Harvey, 1992, Seasonality and consumption-based asset pricing, *Journal of Finance* 47, 511–552.

Gallant, A. Ronald, and George Tauchen, 1989, Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications, *Econometrica* 57, 1091–1120.

Gallant, A. Ronald, Lars P. Hansen, and George Tauchen, 1990, Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution, *Journal of Econometrics* 45, 141–179.

Glosten, Lawrence R., and Ravi Jagannathan, 1993, A contingent claim approach to performance evaluation, *Journal of Empirical Finance*, Forthcoming.

Green, Richard C., 1986, Benchmark portfolio inefficiency and deviations from the security market line, *Journal of Finance* 41, 295–312.

Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hansen, Lars P., John Heaton, and Erzo G. J. Luttmer, 1995, Econometric evaluation of asset pricing models, *The Review of Financial Studies* 8, 237–274.

Hansen, Lars P., and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hansen, Lars P., and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, Lars P., and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1286.

Harrison, J. Michael, and David M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Huberman, Gur, 1982, A simple approach to arbitrage pricing theory, *Journal of Economic Theory* 28, 183–191.

Huberman, Gur, Shmuel Kandel, and Robert F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *Journal of Finance* 43, 1–9.

Jagannathan, Ravi, and Zhenyu Wang, 1993, The CAPM is alive and well, Research Department Staff Report 165, Federal Reserve Bank of Minneapolis.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Kandel, Shmuel, and Robert F. Stambaugh, 1987, On correlations and inferences about mean-variance efficiency, *Journal of Financial Economics* 18, 61–90.

King, Mervyn, Enrique Sentana, and Sushil Wadhwani, 1994, Volatility and links between national stock markets, *Econometrica* 62, 901–938.

Kreps, David M., 1981, Arbitrage and equilibrium in economies with infinitely many commodities, *Journal of Mathematical Economics* 8, 15–35.

Luenberger, David G., 1969. *Optimization by Vector Space Methods* (John Wiley & Sons, Inc., New York).

Luttmer, Erzo G. J., 1994, Asset pricing in economies with frictions, Working paper, Northwestern University.

McCulloch, Robert E., and Peter E. Rossi, 1990, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, *Journal of Financial Economics* 28, 7–38.

Newey, Whitney K., and Kenneth D. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Roll, Richard, 1977, A critique of the asset pricing theory's tests, part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129–176.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Ross, Stephen A., 1978, A simple approach to the valuation of risky streams, *Journal of Business* 3, 453–476.

Rubinstein, Mark, 1976, The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics* 7, 407–425.

Shanken, Jay, 1987, Multivariate proxies and asset pricing relations, *Journal of Financial Economics* 18, 91–110.