

Federal Reserve Bank of Minneapolis
Research Department Staff Report 167

March 1994

Assessing Specification Errors in Stochastic Discount Factor Models

Lars Peter Hansen*

University of Chicago, NORC,
and NBER

Ravi Jagannathan*

Federal Reserve Bank of Minneapolis
and University of Minnesota

ABSTRACT

In this paper 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 methods in assessing some alternative stochastic factor models that have been proposed in asset pricing literature.

*Comments by Ed Allen, John Cochrane, John Heaton, Peter Knez, Erzo Luttmer, Marc Roston, and S. Viswanathan are gratefully acknowledged. Expert research assistance was provided by Marc Roston and Amir Yaron. Hansen received funding from the National Science Foundation in support of this research.

The views expressed herein are those of the author(s) and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

Introduction

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 which 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 1990). 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.

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 is viewed formally as an approximation as, say, 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). The question addressed in this paper is *how large is the misspecification of the stochastic discount factor proxy?*

More formally, we associate a stochastic discount factor *proxy* with an asset pricing model, and then we measure the degree of model misspecification by the distances between the 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 paper 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 because 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) computed 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. Very similar to 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 paper 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 derive specification error bounds for a discount factor proxy that depend 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 go on and demonstrate the connection between our analysis and that of Shanken (1987) and Kandel and Stambaugh (1987). In Section III we extend the analysis by examining pricing errors for hypothetical derivative claims. Positivity of the stochastic discount factors is used as a device to restrict attention to only arbitrage-free assignments of prices to derivative claims. In Section IV we describe an associated set of econometric methods designed to account for the errors induced in our specification-error measures by using time-series data to approximate population moments of asset payoffs, prices, and proxies. Finally, in Section V we illustrate our apparatus by assessing several stochastic discount factor proxies that have been proposed in the literature.

I. Stochastic Discount Factors and Proxies

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 in this section, 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 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. We will then introduce a proxy used in an econometric analysis that assigns approximate prices to payoffs and suggest distance measures between the proxy and the two families of stochastic factors.

For simplicity, we focus on asset market transactions that take place at two dates, say date t and date $t + \tau$. At date t financial assets are purchased, while at date $t + \tau$ the payoffs are received. We will denote the price payoff pair by $(q_t, x_{t+\tau})$. In the first three sections we will omit the date subscripts t and $t + \tau$ for notational convenience. As we will see in Section IV, for extracting 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 approximating population moments using time series averages.

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 of 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 paper.

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:

$$(1.1) \quad \langle h_1 | h_2 \rangle \equiv E(h_1 h_2) \text{ and } \| h \| \equiv \langle h | h \rangle^{1/2} .$$

Let P denote the space of portfolio payoffs observed by an econometrician.

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

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 :

$$(1.2) \quad P \equiv \{x \cdot c : c \in \mathbf{R}^n\}.$$

Requiring that Exx' be nonsingular implies that for each payoff p in P , there is a unique vector of portfolio weights c such that $p = c \cdot x$.

In contrast to specification (1.2), sometimes it is of interest to look at infinite-dimensional specifications of P . For instance in linear factor models of asset payoffs, oftentimes an infinite-dimensional collection of asset payoffs is needed to generate a payoff space which contains the factors

Alternatively, even if there is a finite-dimensional set of *primitive* securities, the payoff space can be infinite dimensional. This occurs when an econometrician constructs "synthetic payoffs" using portfolio weights that depend on conditioning information as in Hansen and Richard (1987) and Gallant, Hansen, and Tauchen (1990), or managed portfolio payoffs as in Fama and Schwert (1977) and Breen, Glosten, and Jagannathan (1989). In this case we assume that $E(xx' | \mathcal{G})$ is nonsingular with probability one, where \mathcal{G} contains information observed by economic agents at the trading date and used by the econometrician to form portfolios. The payoff space P is then given by

$$(1.3) \quad P \equiv \{p \in L^2 : p = w \cdot x \text{ for some random vector } w \text{ of portfolio weights} \\ \text{that is in the conditioning information set } \mathcal{G}\}.$$

It is demonstrated in Hansen and Richard (1987) that P as given by (1.3) satisfies Assumption 1.1.

Our intention is to compare the hypothetical prices assigned by a given model to security market payoffs with the market prices 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. For payoffs such as returns or excess returns, the prices are, by construction, one or zero and hence do not vary over time. Consequently, the expected prices and actual prices coincide. However, we also wish to accommodate security market payoffs with prices that depend on conditioning information in ways that resulting time series of

prices fluctuate. For instance, suppose the econometrician has data on both a return, say r , and information available to investors at the purchase date, say z . Then for the purposes of the empirical analysis, the synthetic payoff zr can be constructed with expected price Ez . Given our intention to use time series averages, we are led to look at the alternative theories from the vantage point of their predicted 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 $\pi(p)$.

Assumption 1.2: π is a continuous linear functional 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 1990). 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 using an original finite-dimensional collection of distinct primitive payoffs. For instance, this is the case for both specifications (1.2) and (1.3) of P .

To illustrate the construction of the pricing functional, initially suppose P is given by (1.2). 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

$$(1.4) \quad \pi(x \cdot c) \equiv c \cdot Eq$$

as long as Eq is not a vector of zeros. The requirement that $|q|$ have a finite first moment can often be met by scaling the payoffs appropriately. One example of such a scaling is the conversion of asset payoffs into (gross) returns in which case q is simply a vector of ones. As we indicated previously, for some applications, it is of interest to allow q to be a vector of random variables in the conditioning information set of investors at the time asset trades are transacted. The restriction that q have a finite (absolute) first moment then limits to some extent the construction of synthetic payoffs using conditioning information of economic agents.

For an alternative illustration of π , suppose P is given by (1.3). This is the appropriate collection of payoffs when a researcher is willing to model formally the conditional second moment matrix $E(xx'|\mathcal{G})$ as in say, the work of Gallant, Hansen, and Tauchen (1990); the factor *GARCH* models of Bollerslev and Engle (1993); or King, Sentana, and Wadhvani

(1993). 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

$$(1.5) \quad \pi(w \cdot x) = E(w \cdot q).$$

It is shown in Appendix A that the resulting π is a bounded linear functional on P .

Returning 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 :

$$(1.6) \quad \pi(p) = E(pm) \text{ for all } p \in P.$$

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

$$(1.7) \quad Epm = EpEm + Cov(p, m).$$

Discounting the future is captured by the first term on the right side of (1.7) (so long as Em is less than one), and adjusting for risk 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 . 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 a random variable that is orthogonal to the space P of portfolio payoffs.

When the aim is to price derivative claims, we are interested in the extension of the pricing function π from P to L^2 using elements of \mathcal{M} . Recall that L^2 contains functions of the portfolio payoffs in P as long as the resulting random variables are finite and hence 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 linear functionals:

$$(1.8) \quad \pi_m(h) = Ehm \text{ for all } h \in L^2.$$

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_m 's of interest are those which satisfy:

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.

The set \mathcal{M}^{++} is convex but 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 only of nonnegative random variables. As long as (π, P) satisfies the no-arbitrage condition (Condition *N*), it follows from Kreps (1981) that

Assumption 1.3: \mathcal{M}^{++} is not empty.

Finally, 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 formula:

$$(1.9) \quad \hat{\pi}(h) \equiv E(yh)$$

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).

If the proxy y turns out to be an admissible stochastic discount factor, 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.¹ 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 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:

¹ 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 to an econometrician. Moreover, some derivations of the *APT* such as those in Huberman (1982) and Chamberlain and Rothschild (1983) obtain factor pricing only as an approximation.

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 used \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).²

Our aim is to use δ and δ^+ as *ad hoc* measures of model misspecification. By construction δ^+ is greater than or equal to δ . As we will see, both of these measures have alternative interpretations as bounds on the magnitude of pricing errors induced by using a proxy y . In the case of Problem 1, we are led to look only at pricing errors on the space of payoffs used in an econometric analysis, while in the case of Problem 2 pricing errors for hypothetical derivative claims on those payoffs (or more generally any hypothetical payoffs in L^2) are also incorporated.³

² 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 .

³ McCulloch and Rossi (1990) used exponential utility functions to measure the magnitude of expected return errors in factor models. There are two reasons why we adopt our more *ad hoc* measures of model misspecification. First, we wish to make comparisons across a wide class of models including ones in which investors have alternative preference orderings. Second, we study pricing errors instead of expected return errors, which makes it difficult to translate specification errors into optimization mistakes on the part of investors.

II. Least Squares Distance Between y and \mathcal{M}

In this section we solve and interpret Problem 1, which is to find the least squares distance between a proxy y and the set \mathcal{M} of (possibly negative) stochastic discount factors. Since our aim is to propose this distance as a measure of model misspecification, we establish the sense in which this measure is a bound on the pricing errors induced by using a proxy y in place of a stochastic discount factor in the set \mathcal{M} . We characterize this specification-error measure using ordinary least squares and two-stage least squares estimation theory, and we demonstrate the close connection between this measure and the proxy analysis in Kandel and Stambaugh (1987) and Shanken (1987).

Recall that the actual prices can be represented as $\pi(p) = Emp$ for a valid stochastic discount factor m , and the prices assigned by the proxy y are given by $\hat{\pi}(p) = Eyp$. Hence by the Cauchy-Schwarz Inequality we get:

$$(2.1) \quad |\pi(p) - \hat{\pi}(p)| \leq \|y - m\| \|p\|.$$

Since this inequality holds for *any* payoff in P and *any* stochastic discount factor m in \mathcal{M} ,

$$(2.2) \quad \sup\{|\pi(p) - \hat{\pi}(p)| : p \in P, \|p\| = 1\} \leq \delta$$

where δ is the least-squares distance between y and \mathcal{M} . As we will see shortly, the inequality in (2.2) can be replaced by an equality. Consequently, the maximum pricing-error problem on the left side of (2.2) provides us with an alternative interpretation of the least-squares distance between a proxy and the set of (possibly negative) stochastic discount factors. Just as there is a family of admissible stochastic discount factors, there is also a family $\hat{\mathcal{M}}$ of random variables that can be used to represent the proxy prices $\hat{\pi}$. Given the alternative pricing-error interpretation, it follows that the distance between the proxy y and \mathcal{M} is not sensitive to which random variable is selected from $\hat{\mathcal{M}}$ as the proxy.

We now show formally that the solution to the least squares problem has a simple and convenient characterization. It follows from the Riesz Representation Theorem that there is a unique element in the intersection of P and \mathcal{M} , which we denote by p^* . It turns out that p^* is the least squares projection of all random variables in \mathcal{M} onto P because $m - p^*$ is orthogonal to P for all m in \mathcal{M} and p^* is in P . This orthogonality also guarantees that the payoff p^* is the *minimum norm* random variable that assigns correct (expected) prices to portfolio payoffs in P .

Similarly, let \hat{p} denote the unique element of P that can be used to represent the *incorrect* $\hat{\pi}$ prices of portfolio payoffs in P :

$$(2.3) \quad \hat{\pi}(p) = E\hat{p}p \text{ for all } p \in P.$$

Although \hat{p} and y assign the same (approximate) prices to payoffs in P , they only coincide when y is in P . More generally, \hat{p} is the least squares projection of y onto P . We now verify that the solution \hat{m} to least squares Problem 1 is obtained by projecting $m - y$ onto P and adding y to the result:

$$(2.4) \quad \hat{m} = p^* - \hat{p} + y.$$

Notice that the right side of (2.4) is a valid stochastic discount factor. Moreover, the pricing error for the payoff $\tilde{p} \equiv (p^* - \hat{p}) / \|p^* - \hat{p}\|$ is

$$(2.5) \quad \pi(\tilde{p}) - \hat{\pi}(\tilde{p}) = E[(p^* - \hat{p})\tilde{p}] = \|\hat{m} - y\|.$$

It follows that (2.2) is satisfied with equality, and

$$(2.6) \quad \delta = \|\hat{m} - y\| = \|\hat{p} - p^*\|.$$

We now interpret these calculations. Since \hat{m} given in (2.4) is a valid stochastic discount factor, the payoff $p^* - \hat{p}$ captures what is missing in the discount factor proxy y in order for it to price correctly the portfolio payoffs in P . The mispricing of any given payoff p is due entirely to that part of p that is best approximated by $p^* - \hat{p}$ in a least squares sense. To see this, simply note that

$$(2.7) \quad \begin{aligned} \pi(p) - \hat{\pi}(p) &= E[(m - y)p] = E[(p^* - \hat{p})p] \\ &= E[(p^* - \hat{p})proj(p|p^* - \hat{p})] \\ &= E[(m - y)proj(p|p^* - \hat{p})] \end{aligned}$$

where $proj(p|h)$ denotes the population least squares regression of p onto h . Hence all of the pricing error in p arises from $proj(p|p^* - \hat{p})$. Moreover, the regression error $p - proj(p|p^* - \hat{p})$ contributes to the norm of p but not to the pricing error. Therefore, $p^* - \hat{p}$ is the payoff with the maximum pricing error per unit norm.

In summary, the maximum pricing error is the least squares distance between two projections: the projection of any stochastic discount factor m onto the space of portfolio payoffs P , and the projection of the proxy y onto P . Thus, we have established:

Proposition 2.1: Suppose Assumptions 1.1 - 1.2 are satisfied. Then

$$\begin{aligned}\delta &= \sup\{|\pi(p) - \hat{\pi}(p)| : p \in P, \|p\| = 1\} \\ &= \|\hat{p} - p^*\|.\end{aligned}$$

This close connection between pricing errors and the least squares distance between y and \mathcal{M} was to some extent anticipated from the continuity results in Green (1986), Kandel and Stambaugh (1987), and Glosten and Jagannathan (1993). As we will see subsequently, it is even more closely related to a result in Shanken (1987).

To illustrate the computation of δ , suppose that P is generated by a finite vector x of payoffs as in (1.2). In this case, it is straightforward to show that p^* is given by $\alpha^* \cdot x$ where

$$(2.8) \quad \alpha^* \equiv (Exx')^{-1}Eq$$

(e.g., see Hansen and Jagannathan 1991). Similarly, since \hat{p} is the least squares projection of y onto P , $\hat{p} = \hat{\alpha} \cdot x$ where $\hat{\alpha}$ is given by the population least squares formula:

$$(2.9) \quad \hat{\alpha} \equiv (Exx')^{-1}Exy.$$

Combining characterization (2.8) and (2.9), gives the least squares distance between p^* and \hat{p} as the square root of a quadratic form in the vector of pricing errors:

$$(2.10) \quad \delta = [(Eq - Exy)'(Exx')^{-1}(Eq - Exy)]^{1/2}.$$

Note that the distance matrix in the quadratic form on the right side of (2.10) is $(Exx')^{-1}$, which is different from what would be 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 spectral density matrix for time series process associated with the pricing-error vector $q - xy$. A drawback of the chi-square distance measure is its sensitivity to the choice of proxy y , whereas the distance measure in (2.10) depends only on the vector of payoffs used in the econometric analysis.⁴

⁴ Allen 1991 made essentially the same point in justifying an alternative distance measure to the ones developed here. Allen's focus was on the role of conditioning information in the pricing of single return whereas as ours looks across returns but can still accomodate conditioning information.

We now extend our calculations to asset pricing models with observable factors that are not necessarily in P . Such factor models have been studied empirically by Chen, Roll, and Ross (1986); Breeden, Gibbons, and Litzenberger (1989); and Cochrane (1992); among others. Suppose that f is k -dimensional vector of factors and that the candidate discount factor is of the form $\hat{\alpha} \cdot f$ for some k -dimensional vector $\hat{\alpha}$. Then the specification-error measure is given by

$$(2.11) \quad \delta = \| p^* - \text{proj}(\hat{\alpha} \cdot f \mid P) \| .$$

Since this criterion depends on the choice of $\hat{\alpha}$, a question is raised as to how $\hat{\alpha}$ is selected. One possibility is to use generalized method of moments estimation to construct an efficient estimator of $\hat{\alpha}$ using a time series on the payoff vector x , the corresponding price vector q and the proxy y (see Hansen 1982). Of course, such a procedure is justified formally only when the resulting pricing errors are zero, *i.e.*, $\delta = 0$. Alternatively, one might choose $\hat{\alpha}$ to minimize the maximum pricing error. This latter *ad hoc* approach turns out to have a simple connection to two-stage least squares as advocated by Basmann (1957) and Theil (1953), although the justification here is quite different. Also, it coincides with an approach that is often adopted in the factor pricing literature.

To see the two-stage least squares characterization, consider estimating $\hat{\alpha}$ in the relation:

$$(2.12) \quad m = \hat{\alpha}' f + u$$

where m is any member in \mathcal{M} . (Such a regression imitates equation (1) of Shanken 1987.) Let P be generated by the entries of x . To estimate $\hat{\alpha}$ in equation (2.12), first project both sides of (2.12) onto the space of payoffs generated by x :

$$(2.13) \quad \text{proj}(m|x) = \text{proj}(\hat{\alpha}' f|x) + \text{proj}(u|x).$$

Recall that the projection of m onto x is p^* . Write the regression of f onto x as:

$$(2.14) \quad \text{proj}(f|x) = \Phi x$$

where Φ is the matrix of population regression coefficients:

$$(2.15) \quad \Phi \equiv E f x' (E x x')^{-1}.$$

We choose $\hat{\alpha}$ to minimize the right side of (2.11). This is equivalent to selecting $\hat{\alpha}$ to be the vector of coefficients in the least squares regression of p^* and hence m onto Φx . Consequently, $\hat{\alpha}$ is the vector of population two-stage least squares coefficients where we first regress the vector f of factors onto x and then regress m onto Φx . Running the initial regression can equivalently be interpreted as forming the vector of factor mimicking payoffs as in Huberman, Kandel, and Stambaugh (1987).

The result of the second-stage regression is:

$$(2.16) \quad \begin{aligned} \hat{\alpha} &= [Efx'(Exx')^{-1}Exf']^{-1}Efx'(Exx')^{-1}Exm \\ &= [Efx'(Exx')^{-1}Exf']^{-1}Efx'(Exx')^{-1}Eq. \end{aligned}$$

The top line in (2.16) is the population version of the familiar two-stage least squares formula for $\hat{\alpha}$, and the second line uses the average price Eq in place of Exm . This latter expression shows that $\hat{\alpha}$ is not sensitive to the choice of stochastic discount factor m in \mathcal{M} . The corresponding formula for the specification error is

$$(2.17) \quad \delta =$$

$$\{Eq'\{(Exx')^{-1} - (Exx')^{-1}Exf'[Efx'(Exx')^{-1}Exf']^{-1}Efx'(Exx')^{-1}\}Eq\}^{1/2}.$$

Therefore, the least squares distance δ has the alternative interpretation of being the norm of the projection error of the minimum norm stochastic discount factor p^* onto the vector of factor mimicking payoffs Φx .

Typically econometricians use two-stage least squares when the variables on the right side of an equation such as (2.12) are correlated with the error term and the counterpart to x is not. If $\hat{\alpha}'f$ were itself in \mathcal{M} , then u would satisfy the orthogonality condition:

$$(2.18) \quad Eux = 0.$$

On the other hand, when factor pricing only holds as an approximation, u will be correlated with some of the entries of x .⁵ However, the two-stage least squares procedure will still consistently estimate $\hat{\alpha}$.⁶

⁵ Our use of the term *correlated* in this context is not standard. Unless P contains a unit payoff, we should use the more clumsy language *fails to be orthogonal*.

⁶ A similar connection to two-stage least squares was obtained by Durlauf and Hall (1989) in their investigation of specification errors in present-value models of equity prices.

Pursuing further this two-stage least squares link, suppose that all of the entries of the factor vector f are in P . In this case Φx is equal to f and hence the two-stage least squares estimator of $\hat{\alpha}$ collapses to the ordinary least squares estimator:

$$(2.19) \quad \begin{aligned} \hat{\alpha} &= (E f f')^{-1} E f m \\ &= (E f f')^{-1} E q_f \end{aligned}$$

where q_f is the vector of factor prices. Notice that in this case $\hat{\alpha}' f$ prices f correctly on average by construction. Hence minimizing the specification error leads one to select $\hat{\alpha}$ to price correctly the factor vector f .

To compute the corresponding specification-error measure, we partition $x' \equiv [f', x_2']$ and $q' \equiv [q_1', q_2']$. Project x_2 onto f and let e_2 denote the corresponding error. Then by the partitioned inverse formula

$$(2.20) \quad \delta = [(E q_2 - E x_2 f' \hat{\alpha})' (E e_2 e_2')^{-1} (E q_2 - E x_2 f' \hat{\alpha})]^{1/2},$$

which is the pricing-error counterpart to equation (5) in Shanken (1987).⁷

The corresponding analysis for Ross's (1976) linear factor model of asset prices is essentially the same except that the space P is now generated by some sequence of payoffs $\{p_j\}$. Suppose that each element of this sequence has a factor decomposition:

$$(2.21) \quad p_j = \alpha_j \cdot f + u_j$$

where f is a k -dimensional vector of *factors* in P , α_j is vector of *factor loadings* in \mathbf{R}^k , and u_j an *idiosyncratic* payoff in P that is orthogonal to the vector of factors. One of the factors can be taken to be unity. Let U be the space generated by the idiosyncratic components $\{u_j\}$.

In general, arbitrage-pricing arguments only imply that factor pricing holds as an approximation (see Ross 1976 and Huberman 1982).⁸ The corresponding approximate prices, $\hat{\pi}(p)$, are constructed by assigning zero prices to the idiosyncratic components. In other words, for any p in P

$$(2.22) \quad \hat{\pi}(p) \equiv \pi(\alpha \cdot f)$$

⁷ The pricing-error counterpart to Corollary 1 in Shanken (1987) and its analog in Kandel and Stambaugh (1987) is obtained by verifying that p^* is a scaled translation of the return on the tangency portfolio.

⁸ See Connor (1984) for some sufficient conditions for *exact factor pricing*.

where $\alpha \cdot f$ is the least square projection of p onto the vector f of factors. Then $\hat{\pi}$ is a bounded linear functional on P . By the Riesz Representation Theorem, the functional $\hat{\pi}$ can always be represented using a payoff $\hat{p} \equiv \hat{\alpha} \cdot f$ for some $\hat{\alpha} \in \mathbf{R}^k$. Since the entries of f are assumed to be in P and are priced correctly by construction, it follows that \hat{p} is the least squares projection of *any* stochastic discount factor m in \mathcal{M} onto the vector of factors. The maximum pricing error is again just the norm of the projection error of p^* onto the factor space. Since the factors f are priced correctly by $\hat{\pi}$, the maximum pricing error (per unit norm) taken over all of P is equal to the maximum pricing error (per unit norm) taken over the idiosyncratic space U .

III. Least-Squares Distance Between y and \mathcal{M}^+

In posing least squares Problem 1, we included stochastic discount factors that are negative with positive probability. As was noted in Section I, such stochastic discount factors may induce arbitrage opportunities on a *complete* set of asset payoffs such as L^2 , but not necessarily on an *incomplete* set such as P . Ignoring positivity is consistent with the goals of Shanken (1987) and Kandel and Stambaugh (1987) to investigate the use of discount factor proxies for the linear factor models of asset pricing including the single-period capital asset pricing model. As emphasized by Dybvig and Ingersoll (1982), the single-period capital asset pricing model is typically not designed for pricing derivative securities. On the other hand, many intertemporal asset pricing theories do imply predictions for pricing derivative claims as well. For this reason, we now explore connection between least-squares Problem 2, which restricts stochastic discount factors to be nonnegative, and the pricing-error implications for pricing derivative claims. In addition, we show how to compute the solution to the least squares problem by solving the corresponding conjugate or dual problem.

The error in pricing a derivative claim induced by using the proxy y is not well defined because we cannot necessarily infer a unique price for the derivative claim from the prices of the original collection of securities. Instead there is a range of possible prices consistent with absence of arbitrage. Alternative stochastic discount factors in \mathcal{M}^{++} may assign distinct prices to a derivative claim, which complicates the construction of pricing errors.

For the moment, imagine forming pricing errors using a single stochastic discount factor in \mathcal{M}^{++} , say 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:

$$(3.1) \quad |Emh - Eyh| \leq \|m - y\| \|h\|.$$

Define:

$$(3.2) \quad \delta_m^+ \equiv \sup\{|\pi_m(h) - \hat{\pi}(h)| : h \in L^2, \|h\|=1\}.$$

Then

$$(3.3) \quad \delta_m^+ = \|m - y\|,$$

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

$$(3.4) \quad \begin{aligned} \inf\{\delta_m^+ : m \in \mathcal{M}^{++}\} &= \inf_{m \in \mathcal{M}^{++}} \|y - m\| \\ &= \min_{m \in \mathcal{M}^+} \|y - m\| \\ &\equiv \delta^+. \end{aligned}$$

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

Proposition 3.1: Suppose that Assumptions 1.1 - 1.3 are satisfied. Then

$$\delta^+ = \inf_{m \in \mathcal{M}^{++}} \sup\{|\pi_m(h) - \hat{\pi}(h)| : h \in L^2, \|h\| = 1\}.$$

To compute the bound, it is most convenient to use the dual or conjugate problem associated with Problem 2. 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 other specifications of P . We write the primal problem as⁹

$$(3.5) \quad (\delta^+)^2 = \min_{m \in L^2, m \geq 0} \|y - m\|^2 \text{ subject to } E(mx) = Eq.$$

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

$$(3.6) \quad \begin{aligned} (\delta^+)^2 &= \min_{m \in L^2, m \geq 0} \sup_{\lambda \in \mathbb{R}^n} \{E[(y - m)^2] + 2\lambda' E(xm) - 2\lambda' Eq\} \\ &= \max_{\lambda \in \mathbb{R}^n} \min_{m \in L^2, m \geq 0} \{E[(y - m)^2] + 2\lambda' E(xm) - 2\lambda' Eq\}. \end{aligned}$$

To construct the criterion of the conjugate problem, we must solve the inner minimization problem in the second line of (3.6). This turns out to be an easy problem to solve, as

⁹ This construction parallels closely a construction in Luttmer (1993). Luttmer considered a little different problem. In line with his interest in computing volatility bounds on stochastic discount factors when transactions costs are present, y is zero in his setup but the payoff space P is constructed more generally to include short-sale constraints.

we will now demonstrate. Rewrite the portion of the the saddle-point criterion containing m as:

$$(3.7) \quad E[(y - m)^2] + 2\lambda' E(xm) = E[(y - \lambda'x - m)^2] - \lambda' E(xx')\lambda + 2\lambda' E(yx).$$

Since the second and third terms on the right-hand side of (3.7) do not involve m , to construct the criterion function for the conjugate problem we solve:

$$(3.8) \quad \min_{m \in L^2, m \geq 0} E[(y - \lambda'x - m)^2].$$

Optimization problem (3.8) 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 (3.8) 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:

$$(3.9) \quad y - \lambda'x = (y - \lambda'x)^+ + [-(-y + \lambda'x)^+].$$

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. On the other hand, the second term is nonpositive and the closest (in the least squares sense) nonnegative random variable to it is the degenerate random variable equal to zero with probability one. Therefore, the solution to (3.8) is given by the first term in the decomposition. To verify this conclusion formally, rewrite the criterion in (3.8) as:

$$(3.10) \quad E[(y - \lambda'x - m)^2] = E\{[(y - \lambda'x)^+ - m]^2\} + 2E[(-y + \lambda'x)^+ m] + E\{[(-y + \lambda'x)^+]^2\}.$$

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

To construct the conjugate problem, we substitute the solution to (3.8) into the criterion of the saddle-point problem (3.6), resulting in:

$$(3.11) \quad (\delta^+)^2 = \max_{\lambda \in \mathbb{R}^n} E(y^2) - E[(y - \lambda'x)^+]^2 - 2\lambda' E q.$$

Notice that the criterion for the optimization problem on the right-hand side of (3.11) is concave in the multiplier vector λ , and constraint set is finite dimensional. Although $(\delta^+)^2$

cannot typically be expressed in terms of simple matrix manipulations, it can be computed numerically by solving the conjugate problem. Without further restrictions, the multiplier vector λ that solves this conjugate problem is not necessarily unique, but the resulting random variable $(y - \lambda'x)^+$ is unique. The first-order conditions for optimization Problem 3.11 are:

$$(3.12) \quad E[(y - \lambda'x)^+ x - q] = 0,$$

i.e., corresponding to an optimal choice $\tilde{\lambda}$ is a unique stochastic discount factor $(y - \tilde{\lambda}'x)^+$ in \mathcal{M}^+ . Finally, the solution will be unique when the second moment matrix of the random vector $x\mathbf{1}_{\{y - \tilde{\lambda}'x > 0\}}$ is nonsingular where $\mathbf{1}_{\{y - \tilde{\lambda}'x > 0\}}$ is an indicator function equal to one when $y - \tilde{\lambda}'x$ is strictly positive and zero otherwise.

The same approach can also be applied to obtain an alternative representation of δ^2 from that described in Section II. In this case the conjugate maximization problem is quadratic and given by:

$$(3.13) \quad \delta^2 = \max_{\lambda \in \mathbb{R}^n} E(y^2) - E[(y - \lambda'x)^2] - 2\lambda'Eq.$$

The solution, $\hat{\lambda}$, to this maximization problem satisfies:

$$(3.14) \quad \hat{\lambda}'x = \hat{p} - p^*$$

where \hat{p} and p^* are the random variables in Section II used to represent the approximate and actual prices, respectively. One way to obtain starting values for λ when solving Problem 3.11 numerically is to use the solution $\hat{\lambda}$ to the quadratic problem (3.13). If the resulting random variable $\hat{\lambda}'x$ is nonnegative, then $\hat{\lambda}$ is also a solution to Problem 3.11.

IV. 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 $\{(q_t, x_{t+\tau}, y_{t+\tau}) : t = 1, 2, \dots, T\}$ for sample size T . The stochastic process generating this 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 set of discount factors is given by:

$$(4.1) \quad d_T = \left\{ \max_{\lambda \in \mathbb{R}^n} (1/T) \sum_{t=1}^T [(y_{t+\tau})^2 - (y_{t+\tau} - \lambda' x_{t+\tau})^2 - 2\lambda' q_t] \right\}^{1/2};$$

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

$$(4.2) \quad d_T^+ = \left\{ \max_{\lambda \in \mathbb{R}^n} (1/T) \sum_{t=1}^T [(y_{t+\tau})^2 - (y_{t+\tau} - \lambda' x_{t+\tau})^2 - 2\lambda' q_t] \right\}^{1/2}.$$

While finite sample results would be preferable, they appear to be hard to obtain at the level of generality of this paper. For this reason, Hansen, Heaton, and Luttmer (1994) derived the limiting (or asymptotic) distribution of these estimators (see Proposition 3.2 in Hansen, Heaton, and Luttmer, 1994). From their results it follows that:

$$(4.3) \quad T^{1/2}(d_T - \delta) \xrightarrow{\mathcal{D}} N[0, \hat{\sigma}^2 / (4\delta^2)]$$

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

$$(4.4) \quad T^{1/2} \sum_{t=1}^T [(y_{t+\tau})^2 - (y_{t+\tau} - \hat{\lambda}' x_{t+\tau})^2 - 2\hat{\lambda}' q_t - \delta^2] \xrightarrow{\mathcal{D}} N(0, \hat{\sigma}^2).$$

Recall that $\hat{\lambda}$ maximizes the population conjugate problem. Hence the numerator term $\hat{\sigma}^2$ in the asymptotic variance in (4.3) 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 (4.4). This can be accomplished by forming a scalar time series sequence $\{u_{t,T} : t = 1, 2, \dots, T\}$ where:

$$(4.5) \quad u_{t,T} \equiv (y_{t+\tau})^2 - (y_{t+\tau} - \hat{\ell}'_T x'_{t+\tau})^2 - 2q'_t \hat{\ell}_T$$

and $\hat{\ell}_T$ is the value of λ that maximizes (4.1). 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 $\{u_{t,T} - (d_T)^2 : t = 1, 2, \dots, T\}$. Let s_T denote the resulting estimator. Then statistical inference can be based on the approximation:

$$(4.6) \quad \frac{T^{1/2} d_T}{2s_T} (d_T - \delta) \xrightarrow{\mathcal{D}} N(0, 1).$$

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 only occurs when the proxy is a valid discount factor, *i.e.*, when

$$(4.7) \quad E(yx - q) = 0.$$

This restriction is what underlies the generalized-method-of-moments estimation and 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 (4.3) 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:

$$(4.8) \quad y_{t+\tau} = \hat{\alpha}' f_{t+\tau}$$

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

$$(4.9) \quad d_T = \left\{ \min_{\alpha \in \mathbb{R}^k} \max_{\lambda \in \mathbb{R}^n} (1/T) \sum_{t=1}^T [(\alpha' f_{t+\tau})^2 - (\alpha' f_{t+\tau} - \lambda' x_{t+\tau})^2 - 2\lambda' q_t] \right\}^{1/2}.$$

The limiting distribution in (4.3) 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 (4.9) 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,

$$(4.10) \quad T^{1/2}(d_T^+ - \delta^+) \xrightarrow{\mathcal{D}} N[0, \bar{\sigma}^2/(4\delta^{+2})].$$

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

$$(4.11) \quad T^{1/2} \sum_{t=1}^T [(y_{t+\tau})^2 - (y_{t+\tau} - x'_{t+\tau} \tilde{\lambda})^2 - 2q'_t \tilde{\lambda} - \delta^{+2}] \xrightarrow{\mathcal{D}} N(0, \bar{\sigma}^2).$$

V. 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.

V.A. Data Description

Time series of six monthly returns for the period 1959:1-1990:12 were constructed as follows. (Hence τ is one in this application.) The first return is the equally-weighted portfolio of 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 Constantinides and Ferson (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 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.

V.B. Constant Discount Factors

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

$$(5.1) \quad y = \beta$$

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 performed two calculations. First, we computed the constant discount factors that minimize the two specification error measures. We also found which value of β is most plausible from the vantage point of the chi-square statistic.

Consider the least squares problem:

$$(5.2) \quad \begin{aligned} \delta &= \min_{\beta} \min_{m \in \mathcal{M}} E[(\beta - m)^2]^{1/2} \\ &= \min_{m \in \mathcal{M}} \min_{\beta} E[(\beta - m)^2]^{1/2} \\ &= \min_{m \in \mathcal{M}} \text{std}(m) \end{aligned}$$

where $std(m)$ 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 .998 with a standard deviation of .329. As just argued, this standard deviation is also our estimate of δ with a standard error of .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^+ out to three decimal points. The minimum chi-square value for the family of constant discount factors is 41.6. Since we used six security returns, and there is one free parameter (β), the degrees of freedom for the associated chi-square test is five and the probability value is essentially zero. Thus, there is considerable statistical evidence against risk-neutral pricing.

V.C. Power Utility

We now extend the collection of models 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 γ positive. The implied stochastic discount factor is the intertemporal marginal rate of substitution between time $t + 1$ and time t :

$$(5.3) \quad y_{t+1} = \beta(c_{t+1}/c_t)^{-\gamma}$$

where β is the subjective discount factor. We used consumption of nondurables as our proxy for c_t leaving us with a two parameter family of stochastic discount factors. In Table 5.1 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 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 only from .328 to .335, which is certainly small range considering the size of the estimated standard errors. The minimum specification errors for the entire family of stochastic discount factors is .327, only expanding the range slightly. Thus, the estimated maximum pricing errors always exceed 30 percent of the norm. Moreover, the power utility model barely reduces the estimated pricing error relative to the constant discount factor model described previously.

TABLE 5.1: Power Utility

$\beta = .95$			
γ	δ	δ^+	χ^2
0	.332 (.054)	.333 (.054)	2,824.9
1	.333 (.053)	.333 (.054)	2,850.9
5	.333 (.053)	.333 (.054)	1,557.7
10	.333 (.053)	.333 (.053)	705.9
15	.334 (.052)	.334 (.053)	419.2
30	.335 (.052)	.336 (.052)	184.1
$\beta = 1.00$			
γ	δ	δ^+	χ^2
0	.329 (.054)	.329 (.055)	46.6
1	.329 (.054)	.329 (.054)	42.6
5	.329 (.054)	.329 (.054)	47.8
10	.328 (.053)	.328 (.054)	55.7
15	.328 (.053)	.328 (.054)	58.1
30	.328 (.053)	.328 (.053)	57.4
minimized	.327 (.053)	.327 (.053)	35.2

The stochastic discount factor is $\beta(c_{t+1}/c_t)^\gamma$. Numbers in parentheses are estimated standard errors. The standard errors were computed using the method described in Newey and West (1987) with $m = 15$. (Very similar results were obtained using $m = 9$ and 12.) The entries in the row labelled "minimized" were computed by selecting parameter values for β and γ that minimized the corresponding criteria. In the column labelled 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 $\gamma = 249.5$.

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:

$$(5.4) \quad \| p \| = [E(p)]^2 + [std(p)^2]^{1/2}.$$

We applied this decomposition to $\hat{p} - p^*$ recalling that δ is the norm of this random variable, we computed the means and standard deviations of $\hat{p} - p^*$ for the different specifications of y . We found that the mean was typically very small relative to the standard deviation, indicating that most of the norm of $\hat{p} - p^*$ is attributable to its standard deviation. As emphasized by Cochrane and Hansen (1992), the standard deviations of \hat{p} tend to be small for the class of power utility models, so that most of the variation in $\hat{p} - p^*$ comes from variation in p^* , which is invariant to the choice of proxy.

Finally, even though the range of the chi-square statistics is extremely large for the parameter values reported in Table 5.1, they all have extremely small probability values for the null hypothesis that the model is correctly specified. Minimizing the value of the chi-square statistic by choice of β and γ , we obtain coefficient estimates that are perverse with a very large (absolute) value of γ and chi-square statistic that remains quite large.¹⁰

V.D. Consumption Externality

We now alter the preferences of consumers by introducing an externality in preferences as in Abel (1990). An extreme version of this preference specification has the time t fiduciary utility function be 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 for consumption is modified to be:

$$(5.5) \quad mu_t = (c_t)^{-\gamma}(c_{t-1})^{\gamma-1}$$

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

$$(5.6) \quad y_{t+1} = \beta(c_{t+1}/c_t)^{-\gamma}(c_t/c_{t-1})^{\gamma-1}$$

¹⁰ While the chi-square value is reduced to 35.2, this remains far in the right tail of a chi-square our 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.

In Table 5.2 we report results for the same parameter configurations as in Table 5.1. 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.

V.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 assuming that consumers preferences are time and state separable with a logarithmic fiduciary utility function (power utility with $\gamma = -1$). More generally, Epstein and Zin (1991) rationalized this discount factor by relaxing state-separability but 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 and equal to .311 with a standard error of .055. This is a small improvement over the best specification error from the power utility model.¹¹

V.F. Observable Factor Models

The final stochastic discount factor proxies we consider are those implied by three alternative linear factor models. For all three models there are two factors: a constant 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 used both the equally-weighted and the valued-weighted returns as alternative measures of the market return giving rise to two alternative linear factor models. The third model imitates Breeden, Gibbons, and Litzenburger (1989) by using the consumption growth rate as an observable factor.

¹¹ Interestingly, the chi-square statistic is 128, so naive use of classical chi-square statistics might lead one to argue in favor of the power utility model.

TABLE 5.2: Power Utility With Consumption Externality

$\beta = .95$			
γ	δ	δ^+	χ^2
0	.332 (.054)	.332 (.054)	2,762.1
1	.332 (.054)	.332 (.054)	2,837.7
5	.332 (.053)	.332 (.054)	807.6
10	.332 (.053)	.332 (.053)	244.0
15	.332 (.052)	.332 (.053)	123.9
30	.331 (.052)	.332 (.053)	52.9
$\beta = 1.00$			
γ	δ	δ^+	χ^2
0	.328 (.054)	.339 (.055)	42.2
1	.328 (.054)	.328 (.055)	42.3
5	.328 (.054)	.328 (.055)	42.1
10	.328 (.053)	.329 (.054)	42.8
15	.328 (.053)	.329 (.053)	42.8
30	.330 (.052)	.332 (.053)	44.2
minimized	.328 (.053)	.328 (.054)	27.5

The stochastic discount factor is $\beta(c_{t+1}/c_t)^\gamma(c_t/c_{t-1})^{\gamma-1}$. Numbers in parentheses are estimated standard errors. The standard errors were computed using the method described in Newey and West (1987) with $m = 15$. The entries in the row labelled "minimized" were computed by selecting parameter values for β and γ that minimized the corresponding criteria. In the column labelled δ , the minimizers are $\beta = .998$ and $\gamma = 7.3$; in the column δ^+ , the minimizers are $\beta = .999$ and $\gamma = 4.8$; and in the column labelled χ^2 , the minimizers are $\beta = 21.1$ and $\gamma = 237.6$.

The results are reported in Table 5.3. 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, .286, and value of the chi-square statistic, 30.2, were found when the variable factor was 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 consumption factor model, there was 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 30 times more variable, illustrating 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 5.3 might erroneously conclude that the consumption factor model ‘does better’ than the other two factor models. As a further indication of how 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.

Using the duality approach described in Section III, we can interpret the portfolio weights as Lagrange multipliers on the pricing constraints. In all cases we look at proxies that are chosen to minimize the specification errors. The estimated multipliers and their standard errors are reported in Table 5.4. Notice that 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 by using variables in the conditioning information sets of investors. The role of securities such as these in statistical inferences lend support to the empirical factor models of Cochrane (1992) and Jagannathan and Wang (1993), whereby the coefficients on the factors are also permitted to depend on variables in the conditioning information sets of investors.

TABLE 5.3: Linear Factor Models

Coefficients estimated by minimizing the specification error		
Variable Factor	δ	χ^2
Equally-Weighted Return	.286 (.054)	30.1
Value-Weighted Return	.289 (.052)	33.5
Consumption Growth	.325 (.052)	38.9
Coefficients estimated by minimizing the value of χ^2 statistic		
Variable Factor	δ	χ^2
Equally-Weighted Return	.286	29.3
Value-Weighted Return	.290	31.5
Consumption Growth	1.57	12.0

The stochastic discount factor is $\hat{\alpha} \cdot f_{t+1}$ where f_{t+1} contains a constant and one of the variable factors listed in the table. Estimated standard errors for δ are given in parentheses. The standard errors were computed using the method described in Newey and West (1987) with $m = 15$.

TABLE 5.4: Estimates of Lagrange Multipliers on the Pricing Constraints

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

Numbers in parentheses are estimated standard errors. The standard errors were computed using the method described in Newey and West (1987) with $m = 15$.

TABLE 5.5: Correlations of $\hat{p} - p^*$ Across Proxies

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

As noted previously, the payoffs that are orthogonal to $\hat{p} - p^*$ are correctly priced by the discount factor proxy y , *i.e.*, all of the mispricing is confined to that part of the payoff which is in the linear span of $\hat{p} - p^*$. In Table 5.5 we report the correlations of $\hat{p} - p^*$ across the factor models in Table 5.5. Not surprisingly, the correlations among the payoffs with the maximum pricing error (per unit norm) across factor models are quite high (they all exceed .87). Abstracting from sampling error, this high correlation is indicative that whatever is missing in one proxy is also missing in the other. However, the following caution is in order. If our asymptotic measures of sampling error understate the underlying sampling variability, then a portfolio that performs well in sample might appear to be similarly mispriced by correctly specified stochastic discount factors (random variables in \mathcal{M}). Thus even if all of the factor models were correctly specified the estimated $\hat{p} - p^*$ could well be highly correlated across models.

VI. Concluding Remarks

In this paper we have explored alternative ways to compare stochastic discount factor proxies when it is understood that the models under consideration are misspecified. In contrast to comparisons made 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 showed 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 25 percent of the norm for some payoffs. Bansal, Hsieh, and Viswanathan (1993) have employed our apparatus to document the advantages of nonlinear factor models over linear ones in pricing international securities.

We proposed and justified two alternative models of model misspecification, the second of which exploits the positivity of arbitrage-free stochastic discount factors for pricing derivative claims. For the data we used in this paper, the differences between the two measures was negligible. This is unlikely to be true for all data sets. For instance, in light of the findings in Hansen and Jagannathan (1991), it seems likely that holding-period returns for Treasury bills would lead to a much more substantial discrepancy between specification error measures. Moreover, from the work of Luttmer (1993) it is likely to be important to take explicit account of transactions costs. Hansen, Heaton, and Luttmer (1994) show how to incorporate such market frictions into the measures described here.

Appendix A: Mathematical Results

The first proposition establishes a result stated in Section I.

Proposition A.1: Suppose P is given by (1.3). Then π given by (1.5) 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

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

Furthermore,

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

Finally, note that

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

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

The next proposition imitates a result in Luttmer (1993). 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, the corresponding price vector q has a finite first moment and 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 continuous function of the price assignment γ . Consequently, there is an open ball about 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: Data

The large and small capitalization stock portfolio returns were constructed using data from the CRSP tapes. We sorted firms in the NYSE into ten size 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 months in the subsequent calendar year. We repeat this procedure for each calendar year and then splice the monthly series for different calendar years together.

Monthly returns on long term government bonds and one-month Treasury bills were taken 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 the Federal Reserve Bulletins. Nominal returns were deflated using the price deflator series for consumption of non-durables from CITIBASE.

Monthly returns on the equally-weighted as well as the value-weighted index of stocks in the NYSE and AMEX used to construct candidate marginal rates of substitution are taken 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

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

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

References

- Abel, A.B. (1990), "Asset Prices Under Habit Formation and Catching Up With the Joneses," *American Economic Review* 80, 38-42.
- Allen, E.R., III (1991), "Evaluating Consumption-Based Models of Asset Pricing," University of Chicago Ph.D. Dissertation.
- Bansal, R., D.A. Hsieh, and S. Viswanathan (1993), "A New Approach to International Arbitrage Pricing," *The Journal of Finance* 48:5, 1719-1747.
- Basmann, R.L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica* 25, 77-83.
- Bollerslev, T. and R.F. Engle (1993), "Common Persistence in Conditional Variances," *Econometrica* 61:1, 167-186.
- Breeden, D., M. Gibbons, and R. Litzenberger (1989), "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance* 44, 231-262.
- Breen, W., L.R. Glosten, and R. Jagannathan (1989), "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance* 44:4.
- Chamberlain, G. and M. Rothschild (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica* 51, 1281-1304.
- Chen, N.F., R. Roll, and S.A. Ross (1986), "Economic Forces and the Stock Market," *Journal of Business* 59:3, 383-403.
- Clark, S.A. (1990), "The Valuation Problem in Arbitrage Price Theory," *Journal of Mathematical Economics* 19.
- Cochrane, J.H. (1992), "A Cross-Sectional Test of a Production-Based Asset Pricing Model," NBER Working Paper 4025.
- Cochrane, J.H. and L.P. Hansen (1992), "Asset Pricing Explorations for Macroeconomics," *NBER Macroeconomics Annual 1992*, Cambridge, MA: MIT Press, 115-165.
- Connor, G. (1984), "A Unified Beta Pricing Theory," *Journal of Economic Theory* 34, 13-31.
- Durlauf, S.N. and R.E. Hall (1989), "Bounds on the Variances of Specification Errors in Models With Expectations," NBER Working Paper 2936.

- Dybvig, P.H. and J.E. Ingersoll, Jr. (1982), "Mean-Variance Theory in Complete Markets," *Journal of Business* 55, 233-251.
- Epstein, L. and S. Zin (1991), "Substitution, Risk Aversion, and Temporal Behavior of Consumption and Asset Returns II: An Empirical Analysis," *Journal of Political Economy* 99, 263-286.
- Fama, E.F. and G.W. Schwert (1977), "Asset Returns and Inflation," *Journal of Financial Economics* 11, 115-146.
- Ferson, W.E. and G.M. Constantinides (1991), "Habit Persistence and Durability in Aggregate Consumption: Empirical Tests," *Journal of Financial Economics* 29, 199-240.
- Ferson, W.E. and C.R. Harvey (1992), "Seasonality and Consumption-Based Asset Pricing," *Journal of Finance* 47:2, 511-552.
- Gallant, R., L.P. Hansen, and G. Tauchen (1990), "Using Conditional Moments of Asset Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution," *Journal of Econometrics* 45, 141-179.
- Glosten, L.R. and R. Jagannathan (1993), "A Contingent Claim Approach to Performance Evaluation," forthcoming in *Journal of Empirical Finance*.
- Green, R. (1986), "Benchmark Portfolio Inefficiency and Deviations from the Security Market Line," *Journal of Finance* 41:2, 295-312.
- Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.
- Hansen, L.P., J. Heaton, and E. Luttmer (1994), "Econometric Evaluation of Asset Pricing Models," manuscript.
- Hansen, L.P. and R. Jagannathan (1991), "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99, 225-262.
- Hansen, L.P. and S.F. Richard (1987), "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica* 55, 587-613.
- Hansen, L.P. and K.J. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica* 50, 1269-1286.

- Harrison, M.J. and D.M. Kreps (1979), "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory* 20, 381-408.
- Huberman, G. (1982), "A Simple Approach to Arbitrage Pricing Theory," *Journal of Economic Theory* 28, 183-191.
- Huberman, G., S. Kandel, and R.F. Stambaugh (1987), "Mimicking Portfolios and Exact Arbitrage Pricing," *The Journal of Finance* XLII:1, 1-9.
- Jagannathan, R. and Z. Wang (1993), "The CAPM Is Alive and Well," University of Minnesota manuscript.
- Kandel, S. and R.F. Stambaugh (1987), "On Correlations and Inferences About Mean-Variance Efficiency," *Journal of Financial Economics* 18, 61-90.
- King, M., E. Sentana, and S. Wadhvani (1993), "Volatility and Links Between National Stock Markets," forthcoming in *Econometrica*.
- Kreps, D.M. (1981), "Arbitrage and Equilibrium in Economies With Infinitely Many Commodities," *Journal of Mathematical Economics*, 15-35.
- Luenberger, D.G. (1969), *Optimization by Vector Space Methods*, NY: John Wiley & Sons, Inc.
- Luttmer, E. G. J. (1993), "Asset Pricing in Economies With Frictions," University of Chicago Ph.D. Dissertation.
- McCulloch, R.E. and P.E. Rossi (1990), "Posterior, Predictive, and Utility-Based Approaches to Testing the Arbitrage Pricing Theory," *Journal of Financial Economics* 28, 7-38.
- Newey, W.K. and K. West (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.
- Roll, R. (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, S.A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory* 13, 341-360.
- Ross, S.A. (1978), "A Simple Approach to the Valuation of Risky Streams," *Journal of Business* 3, 453-476.

Rubenstein, M. (1976), "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics* 7, 407-425.

Shanken, J. (1987), "Multivariate Proxies and Asset Pricing Relations," *Journal of Financial Economics* 18, 91-110.

Theil, H. (1953), "Repeated Least-Squares Applied to Complete Equation Systems," The Hague: Central Planning Bureau (mimeographed).