THE JOURNAL OF FINANCE • VOL. XLVIII, NO. 4 • SEPTEMBER 1993

No Arbitrage and Arbitrage Pricing: A New Approach

RAVI BANSAL and S. VISWANATHAN*

ABSTRACT

We argue that arbitrage-pricing theories (APT) imply the existence of a low-dimensional nonnegative nonlinear pricing kernel. In contrast to standard constructs of the APT, we do not assume a linear factor structure on the payoffs. This allows us to price both primitive and derivative securities. Semi-nonparametric techniques are used to estimate the pricing kernel and test the theory. Empirical results using size-based portfolio returns and yields on bonds reject the nested capital asset-pricing model and linear APT and support the nonlinear APT. Diagnostics show that the nonlinear model is more capable of explaining variations in small firm returns.

THE IDEA THAT ONLY a few relevant state variables are needed to explain expected returns is the main driving force behind the seminal papers of Merton (1973) and Ross (1976). In these papers, there exist a small number of factor (state) portfolios that are functions only of the factor (state) information. Thus only risks related to these portfolios are priced. More importantly, Ross's approach implies that nonfactor risks are not priced and are diversified away.

Ross's (1976) unconditional version of the arbitrage-pricing theory (APT) assumes that payoffs are linear in the factors and the idiosyncratic noise. This linearity of payoffs and the no-arbitrage restriction leads to an approximate linear relationship between the unconditional mean returns and the factor loadings. Ross's ideas have been refined and extended by Chamberlain and Rothschild (1983), Huberman (1982), and Ingersoll (1984). In addition, Brock (1982), Connor (1984), Dybvig (1983), Grinblatt and Titman (1983), Chamberlain (1983), Milne (1988), Connor and Korajczyk (1989), and Bossaerts and Green (1989) provide conditions under which an equilibrium arbitrage-pricing relationship holds. In contrast to Ross's original linear arbitrage-pricing model (linear APT), these models imply that individual

* The Fuqua School of Business, Duke University. We owe a special debt to Ron Gallant and George Tauchen for the valuable time that they have spent in discussions with us. We thank two anonymous referees and the editor, René Stulz, for detailed comments that have improved the paper considerably. We thank Peter Bossaerts, Francisco Delgado, George Constantinides, Campbell Harvey, Burton Hollifield, David Hsieh, Ravi Jagannathan, Ellen McGratten, Victor Ng, Sanjay Srivastava, and seminar participants at Duke, Michigan, and Wisconsin, the WFA meetings in San Francisco, and the AFA meetings in Anaheim for their comments, and Campbell Harvey and Tom Smith for providing us with their data. The Business Associates Fund of the Fuqua School of Business provided financial support.

security prices are exactly linear in the factor prices. Again, to obtain this result, a linear payoff structure with fixed factor loadings is assumed.

We view the assumptions of the linear APT as being unnecessarily restrictive. In particular, the linear payoff assumption implies that only primitive securities having this linear structure can be priced. Derivative securities that are nonlinear functions of the linear payoffs cannot be priced by the linear APT model. Thus, for example, if total firm returns are linear in the factors, levered equity and holding period returns on discount bonds will be nonlinear functions of the factors. Consequently, one cannot legitimately use the APT for these payoffs. In addition, as Dybvig and Ingersoll (1982) show in the context of the single-factor market model (the CAPM), attempts to apply the linear pricing model to derivative securities lead to violations of the no-arbitrage condition. A similar criticism applies to the discrete time Merton-Constantinides intertemporal CAPM (ICAPM) (Constantinides (1989)) that is derived using distributional assumptions.

A different way to view the restrictions implied by such models is presented by Hansen and Jagannathan (1991). They show the existence of a unique (but possibly nonpositive) minimum variance pricing kernel (or stochastic discount factor) that is a payoff and characterize it. This pricing kernel is the projection of any valid pricing kernel on the space of payoffs and is the linear combinations of payoffs that is maximally correlated with any valid pricing kernel. This minimum variance pricing kernel is closely related to the maximally correlated portfolio considered in Breeden (1979) and Shanken (1987). In general, the minimum variance pricing kernel is a conditional linear combination of all payoffs. In the linear APT (or the Merton-Constantinides discrete time ICAPM), this pricing kernel is a conditional linear combination of only the factor payoffs and embodies the refutable restrictions of linearity and low dimensionality (see Hansen and Richard (1987) and Hansen and Jagannathan (1991)).

If the payoffs are nonlinear in the factors, these simple representations do not obtain. Nonlinearity in payoffs could arise because of derivative securities or because the primitive payoffs themselves are nonlinear functions of the factors. In this case, the minimum variance pricing kernel is not a conditionally linear combination of the factor (state) payoffs themselves and must include the securities that are nonlinear functions of the factor (state) payoffs.[1] Thus, the low dimensionality of the minimum variance pricing kernel does not obtain an environment that allows for nonlinear payoff structures.

To accommodate nonlinear payoff structures and retain the appealing restriction of low dimensionality on the pricing kernel, we take a different approach. Our approach exploits the theoretical underpinnings of equilibrium pricing models to obtain a low-dimensional nonnegative nonlin-

[1] Hansen and Jagannathan (1991) also discuss a nonnegative pricing kernel that has the minimum variance in the class of nonnegative pricing kernels and can be interpreted as an option on a portfolio constructed from the marketed payoffs. Since a similar criticism holds for this pricing kernel, we do not discuss it in detail.

ear pricing kernel which need not be a traded portfolio. Such a pricing kernel follows from the equilibrium restriction that the intertemporal marginal rate of substitution of some agent is a function of factor risk only. The nonnegativity of the pricing kernel is a necessary consequence of the absence of arbitrage opportunities (see Kreps (1981), Harrison and Kreps (1979) and Ross (1978)). Thus, we take the minimal set of assumptions that are required to have an asset-pricing theory with testable implications: no arbitrage and a dimensionality restriction. This approach to asset pricing is a parsimonious one and embeds versions of a number of asset-pricing models like the CAPM, the linear APT, discrete time ICAPM and the Rubinstein CAPM (R-CAPM).

In short, the main contribution of this paper is to identify a pricing kernel that prices all securities (including derivative securities), depends only on a few economy-wide factors, and satisfies the restriction of no arbitrage. Hence, this pricing kernel embodies in it all the restrictions of interest and provides a simple way to test the theory. Further, this pricing kernel is capable of pricing dynamic trading strategies.

In our approach, the only deep parameter of interest is the pricing kernel, restricted by no arbitrage and low dimensionality. Since the exact form of this nonlinear pricing kernel is unknown, we use the semi-nonparametric estimation (SNP) approach developed by Gallant and Tauchen (1989) to estimate the nonlinear pricing kernel. This approach to estimation allows us to embed a large class of parametric asset-pricing models including the linear APT, the CAPM, the discrete time ICAPM and the R-CAPM. The embedding of these asset-pricing models allows us to test them as nested models. In addition, The SNP approach allows the imposition of the restriction of nonnegativity implied by no arbitrage. In estimation, we use the Generalized Method of Moments (GMM) (Hansen (1982)) to avoid making distributional assumptions and identify the number of factors using a likelihood ratio type test.

We next discuss the link between our work and previous work by Kraus and Litzenberger (1976, 1983) and Glosten and Jagannathan (1992). Kraus and Litzenberger (1976, 1983) argue that the coskewness of an aggregate source of risk (like the market return) is priced. They use a second-order approximation of marginal utility to derive a CAPM type restriction where coskewness is priced. Their work is subsumed in our more general approach. Glosten and Jagannathan (1992) present a nonparametric approach to performance evaluation. They assume a pricing kernel and nonparametrically evaluate the conditional expectation of any payoff given the pricing kernel. This allows them to evaluate the performance of different mutual funds. However, neither of these approaches focus on evaluating the refutable implications of the no-arbitrage and dimensionality-restricted pricing kernel.

Our results are as follows. Tests that do not impose nonnegativity reject the nested parametric models and provide greater support for a nonlinear pricing kernel with two factors (the market return and the interest rate). The estimated pricing kernel is negatively correlated with the market. Hence the pricing kernel places higher value on claims that pay off in low realizations of the market return. The nonlinearities in the pricing kernel seem to be related to the variances of the market and interest rate, the covariance between the market and the interest rate and the skewness of the market return. Diagnostics using the excess returns of small firms indicate that the nonlinear model (as compared to the linear model) is more capable of capturing the variations in small-firm returns, including the January effect. Finally, the imposition of nonnegativity in estimation leads to lesser differences between the linear and nonlinear pricing kernels. These results support the approach to asset pricing that is introduced in this paper.

The outline of the paper is as follows. Section I discusses the restrictions that our approach places on asset prices. Section II provides an example to give some intuition regarding the assumed setup in this paper. Section III discusses the semi-nonparametric approach used in this paper and Section IV presents the details regarding the data set used. Section V discusses our estimation method and related details, while Section VI presents and interprets the results. Section VII concludes. All proofs are in the Appendix.

I. A General Representation

We present a general representation of asset returns and discuss the assumptions that are needed to derive the nonlinear state pricing theory or the nonlinear APT. In what follows, we use these two terms interchangeably. The key restriction we impose is that the intertemporal marginal utility of consumption of some individual is spanned by only M factors or state variables.[2]

Consider an agent who trades assets in securities markets at time t; the payoff to the assets are received at time $t + 1$. Let I_t represent the information set of all investors at time t and $\mathscr{P}$ represent the set of payoffs at time $t + 1$ that are traded at time t. The prices of assets are a function mapping $\mathscr{P}$ into R^+ and $\pi(x_{t+1})$ is the price at time t of the payoff x_{t+1}.

The first-order condition for the agent's investment decision (assuming no short sales constraints) is:

$$\mathbf{E}[mx_{t+1}|I_t] = \pi(x_{t+1}), \tag{1}$$

where m is the intertemporal marginal rate of substitution for the agent between time t and $t + 1$.

Let

$$\mathscr{U} = \left\{m|m \geq 0,\, E\left[m\frac{x_{t+1}}{\pi(x_{t+1})}|I_t\right] = 1.\right\} \tag{2}$$

Thus $\mathscr{U}$ is the set of candidate pricing kernels that satisfy nonnegativity and equation (1). With complete markets, the set $\mathscr{U}$ is a singleton that is traded. With incomplete markets, the marginal rates of substitution of different

[2] The factors or state variables are to be interpreted as the common sources of risk across all payoffs and do not contain idiosyncratic risk elements.

agents are not equated and the set $\mathscr{U}$ has multiple elements. Every element of this set prices all marketed payoffs in the set $\mathscr{P}$ and is nonnegative. However, some element of the set $\mathscr{U}$ must not be traded.[3] Thus the set $\mathscr{U}$ is not a subset of the set of marketed payoffs $\mathscr{P}$.

Ross (1978), Kreps (1981), and Harrison and Kreps (1979) show that the absence of arbitrage opportunities is equivalent to the existence of a nonnegative pricing kernel. Consequently, every element of the set $\mathscr{U}$ satisfies this nonnegativity restriction.

The intuitive content of the linear APT is the restriction that there is a pricing kernel (i.e., satisfying equation (1)) that is a linear, low-dimensional function of only a few factors. The presence of derivative securities and nonlinear payoffs implies that a low-dimensional linear pricing kernel need not exist. However, we can still obtain a nonlinear pricing kernel that is a function of a few factors. The key requirement for a nonlinear APT is the existence of an individual whose marginal rate of substitution is a function only of a few factors (Assumption 2 below). This intuitive claim is made clearer in the example that follows and is also evident in the work of Bossaerts and Green (1989) and Connor and Korajczyk (1989). The second assumption (Assumption 1 below) implies that factor information is sufficient to forecast any function of future factor realizations. As discussed below, this assumption ensures that the prices of discount bonds are only functions of the factor information.

Thus, we assume that:

ASSUMPTION 1: *There exists an M dimensional process f_t such that for all nonlinear functions $L(\cdot)$, $\mathbf{E}[L(f_\tau)|I_t, f_t] = \mathbf{E}[L(f_\tau)|f_t]$, $\tau > t$. Thus f_t is a sufficient statistic for predicting nonlinear functions of f_τ at time t.*

ASSUMPTION 2: *There exists an agent j whose intertemporal marginal rate of substitution, m^j, is a function, $H(f_{t+1})$, of the above M factors. Thus $H(f_{t+1}) \in \mathscr{U}$.*

In our approach, some agent's marginal rate of substitution is nonlinearly spanned by f_{t+1}, a vector process. Thus this process includes any history dependence that may be relevant in determining this marginal rate of substitution.[4]

Using Assumption 2, we rewrite equation (1) to obtain

$$\mathbf{E}\left[H(f_{t+1})\frac{x_{t+1}}{\pi(x_{t+1})}|I_t\right] = 1. \tag{3}$$

Assumption 2 embodies the intuitive content of Ross's ideas; the pricing kernel m is a nonlinear function of only the factors or state variables f_{t+1}

[3] The proof of this claim is as follows. If all $m \in \mathscr{U}$ are traded, then $E[m(\hat{m} - m)|I_t] = 0$ and $E[\hat{m}(\hat{m} - m)|I_t] = 0$ using equation (1). This leads to $E[(\hat{m} - m)^2|I_t] = 0$ or $\hat{m} = m$ almost surely.

[4] See also Constantinides (1989) and Chamberlain (1988) for related arguments.

and hence is low dimensional. In economic terms this assumption states that there is always a risk-averse investor who diversifies away nonfactor risk. In the context of models with linear payoff structures, Bossaerts and Green (1989), Chen and Ingersoll (1983), Connor and Korajczyk (1989), and Milne (1988) exploit Assumption 2 to derive linear arbitrage-pricing models. As Hansen and Jagannathan (1991, 1992) show, these linear models imply a pricing kernel that is a linear combination of M factor payoffs.

The standard approach to the linear APT model is to find the pricing kernel that is a linear combination of the factor payoffs. As we have previously argued, this approach is not feasible in the presence of nonlinear payoffs. Intuitively, nonlinear payoffs do not lie in the linear span of the factor payoffs and thus a M factor linear representation is not possible. In Proposition 1 (see Appendix), we formalize this by providing conditions under which nonlinear payoffs cannot be priced by a linear combination of the factors. In contrast, the pricing kernel $H(f_{t+1})$ (equation (3)) does not suffer from these shortcomings and can price any marketed security.[5]

To derive testable restrictions from the nonlinear APT we need to rewrite equation (3) and substitute for the unobservable state variables with observables known both to the econometrician and the agents in the economy. Thus, we need to find payoffs in $\mathscr{P}$ or prices of securities at time $t + 1$ that are free of nonfactor risk. Given Assumption 1, Assumption 2 implies that the prices of pure discount bonds in the next period are free of nonfactor risk. Also, any standard general equilibrium theory of arbitrage pricing must imply that a security written on aggregate consumption is free of nonfactor risk. However, the ex post payoff on individual securities will not satisfy this restriction and cannot be used for inverting the factors. Let $p_{t+1}^{b} \in B$ denote the $t + 1$ payoff or $t + 1$ prices of all securities that are free of nonfactor risk. To be able to write the pricing kernel in terms of observables we make an additional assumption.

ASSUMPTION 3: *There exists a continuous, one-to-one and onto (i.e., invertible) mapping from the set of factors f to some subset of the invariables in B.*

Assumption 3 has been made implicitly or explicitly in a large number of papers on the linear APT or discrete time ICAPM.[6] Explicit assumptions about the factors are made in papers by Bossaerts and Green (1989), Ferson

[5] An alternative approach is to assume that trading takes place in continuous time, price the primitive assets using Merton's intertemporal CAPM and then price derivative securities by no-arbitrage methods. However, the assumption of continuous trading is quite strong. Also, discrete time estimation requires integration of the continuous time pricing kernel leading to a nonlinear pricing kernel (see Longstaff (1989)).

[6] The equilibrium linear APT always implies invertibility. Since prices of individual assets are beta linear in the factor prices, a M factor model requires M securities whose prices are not perfectly correlated for inversion. Bossaerts and Green (1989) exploit this invertibility to replace the unobservable price of the consol.

and Harvey (1991), Hansen and Jagannathan (1992), Huberman and Kandel (1987) and others.[7]

In the next section we present an example where all yields to maturity on nominal risk-free bonds and the $t + 1$ payoff on the market portfolio always satisfy the restriction of being free of nonfactor risk and thus are functions of the M state variables. Hence these variables form a natural candidate for the set B. The inevitibility assumption holds in certain economic environments, examples of which are Cox, Ingersoll, and Ross (1985), Hansen and Singleton (1983), and Rubinstein (1976). We view the assumption of invertibility of being less restrictive than assuming linearity of payoffs or a specific parametric functional form for the pricing kernel.

Assumption 3 allows us to rewrite equation (3) as

$$\mathbf{E}\left[G(p_{t+1}^b)\frac{x_{t+1}}{\pi(x_{t+1})}|I_t\right] = 1, \tag{4}$$

where p_{t+1}^b is the low-dimensional ex post payoffs or prices at $t + 1$ that do not contain nonfactor risk. The low-dimensional function $G(p_{t+1}^b)$ is the relevant pricing kernel implied by the nonlinear APT. In contrast to most APT models, this pricing kernel can price dynamic trading strategies.

An alternative (and less-restrictive) derivation of equation (4) is as follows. We can rewrite equation (1) as:[8]

$$E\left[E[m|I_{t+1}]\frac{x_{t+1}}{\pi(x_{t+1})}|I_t\right] = 1. \tag{5}$$

Then the restriction that

$$E[m|I_{t+1}] = E\left[m|p_{t+1}^b\right] = G(p_{t+1}^b), \tag{6}$$

for some agent yields equation (4). Equation (6) states that p_{t+1}^b, a low-dimensional vector, is sufficient (relative to the information at time $t + 1$) for predicting some agent's marginal rate of substitution. Assumptions 2 and 3 imply equation (6).

Thus, the nonlinear APT theory implies the existence of a pricing kernel that satisfies the following restrictions.

[7] Papers by Connor and Korajczyk (1988, 1989) extract factor-mimicking portfolios using principal components methods. This approach implies the existence of traded portfolios (asymptotically in the cross section) that invert the factor payoffs. In the presence of nonlinear payoffs, standard approaches to testing the APT like principal components and factor analysis (see Chen, Roll, and Ross (1986) and Lehmann and Modest (1988)) are not feasible.

[8] This derivation is also followed in a companion paper on the pricing of international assets (Bansal, Hsieh, and Viswanathan (1993)).

Restriction 1: The pricing kernel satisfies the orthogonality conditions that

$$\mathbf{E}\left[\left(G(p_{t+1}^{b})\frac{x_{t+1}}{\pi(x_{t+1})} - 1\right)Z_t\right] = 0. \tag{7}$$

where Z_t is an instrument that belongs to the information set I_t.

Restriction 2: The pricing kernel is nonnegative; i.e., $G(\cdot) \geq 0$.

Restriction 3: The pricing kernel is low-dimensional.

Restriction 1 is the pricing restriction that any one period ahead payoff must be priced by the pricing kernel. Restriction 2 is the no-arbitrage restriction that the pricing operator is nonnegative and captures the idea that the positive payoffs must have positive prices. Restriction 3 captures the intuitive notion that there are few economy wide sources of risk that are priced and is the key insight of the arbitrage-pricing theory of Ross (1976) and the intertemporal asset-pricing theory of Merton (1973).

Since all empirical work on linear APT models has ignored the nonnegativity restriction, we follow this approach and first test Restrictions 1 and 3. Having done this, we proceed to impose nonnegativity (Restriction 2) in estimation and to test the model.[9]

II. A Nonlinear Arbitrage-Pricing Model: An Example

To illustrate the general theory in Section I, we present an explicit example. Suppose the dividend process in an endowment economy is given by

$$d_{it} = [g_i(f_t) + \epsilon_{it}]c_{t-1}, \tag{8}$$

where the M vector f_t (the factor process) is a stationary Markov process of the first order and $\mathbf{E}[\epsilon_{it}|f_t, \Phi_{t-1}] = 0$, where Φ_{t-1} represents the information available up until and including time $t-1$. The Markov assumption implies that Assumption 1 holds for the factor process.

In this endowment economy, aggregate consumption is just the sum of the dividends,

$$c_t = \sum_{i=1}^{N} d_{it} = c_{t-1}\sum_{i=1}^{N} g_i(f_t) = c_{t-1}F(f_t), \tag{9}$$

[9] Restrictions 1 and 2 are not completely independent. In a complete market, we need not impose Restriction 2 if the prices on all payoffs are positive. Then the unique pricing kernel cannot have any negative realizations on a set of positive measure. For if it did, we would consider the asset that pays one unit on this set of positive measure where the pricing kernel has a negative realization. The pricing kernel would imply a negative price for this asset, contradicting the observed positive price for all assets.

where $F(\cdot)$ is defined by equation (9) and where we assume that $\Sigma_{i=1}^{N} \epsilon_{it} = 0$. Thus the consumption process is diversified of the nonfactor risk.[10]

In this economy, we assume that there is a representative investor with time-separable utility. Additionally, this investor's marginal rate of substitution between t and $t + s$ is homogeneous of degree α in the growth rate of consumption c_{t+s}/c_t. This immediately implies that:

$$m_{t,t+s}(c_t, c_{t+s}) = H(\pi_{k=1}^{s} F(f_{t+k})) \tag{10}$$

which is consistent with Assumption 2 made previously.

Given the assumptions on the marginal utilities and the payoffs, we can show that the following holds (see Appendix):

$$\begin{aligned} q_{it} &= c_t \overline{\phi}_i(f_t). \\ R_i(t+1) &= F(f_{t+1}) \left[\frac{\overline{\phi}_i(f_{t+1}) + \overline{g}_i(f_{t+1})}{\overline{\phi}_i(f_t)} \right] + \frac{\epsilon_{it+1}}{\overline{\phi}_i(f_t)} \\ R_M(t+1) &= F(f_{t+1}) \left[\frac{1 + \Sigma_{i=1}^{N} \overline{\phi}_i(f_{t+1})}{\Sigma_{i=1}^{N} \overline{\phi}_i(f_t)} \right] \\ y(t,s) &= 1/\{\mathbf{E}[H(\pi_{i=1}^{s} F(f_{t+r}))]\} = \zeta_{t+s}(f_t) \end{aligned} \tag{11}$$

where $\overline{\phi}(\cdot)$, $\overline{g}_i(\cdot)$ and $\zeta_{t+s}(\cdot)$ are functions only of the factor information. Equation (11) states that prices on securities, q_{it}, and yield to maturity on unit discount bonds, $y(t, s)$ (the bond payoffs s periods ahead), are low-dimensional functions of the factors, f_t. The return on the market portfolio, $R_M(t)$, is a function only of the factor process. However, returns on the individual securities, $R_i(t)$ are functions also of the noise process ϵ_{it+1}.

In this example, the marginal utility of consumption across time periods is a nonlinear nonnegative function only of the M factors or state variables f_t. The nonnegativity is guaranteed by the monotonicity restriction on preferences. However, the example does not lead to a pricing portfolio that is a convex combination of M factor portfolios or a linear beta relationship. Additional restrictions are required to obtain this linear beta relationship. Such restrictions include the linearity of the dividend process in the factors (made in Brock (1982), Connor and Korajczyk (1989), Bossaerts and Green (1989)) or the linearity of the marginal utility of consumption in the state variables (made in Engle, Rothschild, and Ng (1990)), or the joint normality of state variables and securities returns (see Constantinides (1989)). We do not make these additional assumptions. Consequently, the marginal

[10] Bossaerts and Green (1989), Milne (1988), and Hollifield (1990) make similar assumptions. Connor (1984) and Connor and Korajczyk (1989) provide assumptions for the infinite dimensional case. Similar assumptions need to be made to extend our model to an infinite number of assets. Brock (1982) presents a model with no idiosyncratic risks where consumption is a nonlinear function of the factors. Only under strong restrictions on technologies, tastes, and the exogenous processes does the usual linear pricing relationship emerge.

rate of substitution is a nonlinear function of the factors f_{t+1} and a linear M beta representation does not obtain.

III. The Semi-nonparametric Approach

The primary object of interest in our theory is the pricing kernel. In contrast to parametric models of asset pricing, we do not have a parametric pricing kernel that we can directly estimate using Restriction 1. To impose and test Restrictions 1 and 3, we use the semi-nonparametric approach discussed in Gallant (1987), Gallant and Nychka (1987), Gallant and White (1989), and especially Gallant and Tauchen (1989).

The idea of this approach is to construct a sequence of finite dimensional parametric models that can approximate the pricing kernel in the limit. In constructing the finite dimensional model, one uses a class of approximating functions that can approximate any function (in a given class of functions) as closely as we desire in some distance measure. In estimation, the size of the finite dimensional model (i.e., K, the length of the approximating functions to use) is determined by T, the sample size. As the sample size increases, the size of the finite dimensional model and the number of orthogonality conditions also increase. Papers by Andrews (1991), Gallant (1987), Gallant and Nychka (1987), and Gallant and White (1989) show how such procedures yield consistent estimates of the true unknown function.

We discuss the intuition behind the SNP approach to estimation that we employ (for more technical details see Bansal and Viswanathan (1992)). Define the set

$$\mathscr{M} = \left\{ G(\cdot) \middle| E\left[\left(G(p^b_{t+1})\frac{x_{t+1}}{\pi(x_{t+1})} - 1\right)Z_t\right] = 0 \right.$$
$$\left. \text{for all payoffs } x_{t+1} \in \mathscr{P} \text{ and } Z_i \text{ that is contained in } I_t \right\}. \quad (12)$$

The set $\mathscr{M}$ contains the set of pricing kernels that satisfy Restrictions 1 and 3 imposed by the nonlinear APT. The nonlinear APT model implies the null hypothesis that the set $\mathscr{M}$ is not empty.[11] With complete markets $\mathscr{M} = \mathscr{U}$ is a singleton set.[12]

We use neural networks to approximate the unknown pricing kernel (see Gallant and White (1989), Hornik, Stinchcombe, and White (1989), and White (1988) for a detailed discussion and other applications of neural networks).

[11] To be precise, the null hypothesis is tested given a fixed number of factors.

[12] When markets are complete, the pricing kernel estimated using the SNP approach will converge to the unique population-pricing kernel. With incomplete markets, there is more than one population-pricing kernel. However, the estimated pricing kernel will converge to one of these population-pricing kernels. For more details, see Bansal and Viswanathan (1992).

The method of neural networks (amended to include a linear term) approximates the unknown function $G(p_{t+1}^b)$ by

$$G(p_{t+1}^b) = \alpha_0 + \alpha \cdot p_{t+1}^b + \sum_{k=1}^{K} \beta_k g(\gamma_{0k} + \gamma_k \cdot p_{t+1}^b) \tag{13}$$

where the function $g(\cdot)$ is the logistic function $\exp(\cdot)/[1 + \exp(\cdot)]$ and is bounded. Our choice of neural nets as the approximating functions is motivated by the results in McCaffrey, Ellner, Gallant, and Nychka (1991) who show that the method of neural nets performs better than other methods like polynomial expansions or kernel estimation. Intuitively, the better performance of neural nets is attributable to the use of bounded functions in approximation.[13]

As formulated, equation (13) does not impose the restriction of nonnegativity on the pricing kernel (Restriction 2). The full imposition of Restrictions 1, 2, and 3 implies that the set $\mathscr{M}^+$, defined by,

$$\mathscr{M}^+ = \mathscr{M} \cap \{G(\cdot) | G(\cdot) \geq 0\} \tag{14}$$

is a nonempty set. We note that $\mathscr{M}^+ \subset \mathscr{M}$ and $\mathscr{M}^+ \subset \mathscr{U}$. Further, the three sets collapse to the same singleton set when markets are complete.

Our approach to nonnegativity is motivated by Hansen and Jagannathan (1991). They present a nonnegative pricing kernel that can be interpreted as an option on the set of marketed payoffs. We impose nonnegativity by truncating the neural net, $G(\cdot)$, as follows:

$$\begin{aligned} G^+(p_{t+1}^b) &= \mathrm{Max}\left[G(p_{t+1}^b), 0\right] \\ &= \mathrm{Max}\left[\alpha_0 + \alpha \cdot p_{t+1}^b + \sum_{k=1}^{K} \beta_k g(\gamma_{0k} + \gamma_k \cdot p_{t+1}^b), 0\right]. \end{aligned} \tag{15}$$

This method, when applied to the linear model, yields an option on the payoff space and thus embeds Hansen and Jagannathan's (1991) approach. Also, if the unrestricted neural net does not violate nonnegativity, the new method will yield the same estimated pricing kernel. When the unrestricted neural net violates nonnegativity, the imposition of Restriction 2 will nontrivially change the estimated pricing kernel.

The use of the maximum function can cause numerical problems in the optimization.[14] Hence, we use the function

$$G(p_{t+1}^b, \theta) = \frac{1}{2} G(p_{t+1}^b) + \frac{1}{2} \sqrt{\left(G(p_{t+1}^b)\right)^2 + \theta} \tag{16}$$

[13] In a slightly different context, Hollifield (1990) uses neural nets to test for neglected nonlinearities in return data.

[14] The gradient of the criterion function is everywhere zero in the region where function defined by equation (15) is zero. During numerical optimization, we may get stuck in such regions.

where θ is a small positive number and $G(p_{t+1}^b)$ is as defined in equation (13). When θ limits to 0, the function $G(p_{t+1}^b, \theta)$ is exactly equal to $G^+(p_{t+1}^b)$. However, for all $\theta > 0$, we have a smooth differentiable function. By smoothing the function in this manner, the numerical properties using $G(p_{t+1}^b, \theta)$ are better than those of the original function $G^+(p_{t+1}^b)$.

The length of the neural net, K, is determined by a deterministic rule that depends on the sample size T. Given a neural net of length K, we estimate the neural net by minimizing the GMM criterion function. In Bansal and Viswanathan (1992), we show that the estimated neural net is a strongly consistent estimator of the unknown pricing kernel that is the object of interest.

Asymptotic normality of SNP estimators for general environments has not been established. A recent paper by Andrews (1991) establishes the asymptotic normality of SNP estimators for i.i.d. environments. We follow Gallant and Tauchen (1989), Gallant, Rossi, and Tauchen (1992), and Bansal, Gallant, Hussey, and Tauchen (1992) and draw statistical inferences in the standard finite dimensional way. One justification for our statistical inference is to view the population-pricing kernel as belonging to a finite dimensional set of neural nets. Then the standard consistency arguments and asymptotics (for the GMM criterion function) as developed in Hansen (1982) hold and justify our statistical inference.

IV. Information on Data Sets

The data we use are the monthly size-based returns for firms listed on the New York Stock Exchange. The time period covered is from July 1964 to December 1989. Ten size-based value-weighted common stock returns for firms are used as the portfolios of interest. The number of monthly time observations is 306. The return on the New York Stock Exchange value-weighted index is used as the measure of return on the market. These data are taken from the Center for Research in Security Prices (CRSP) and are similar to those used in Ferson and Harvey (1991). Term structure data on the monthly nominal yield to maturity are taken from the Fama bond files on CRSP. The monthly nominal yields for one-month to nine-month maturities are taken from the yield file on CRSP. To incorporate information in the long end of the term structure, yields to maturity on one-year to five-year discount bonds are taken from the Fama and Bliss file.

V. Estimation Method

In this section we provide details regarding the estimation strategy. As shown in Section I, there are some ex post returns or yields to maturities of bonds (in the next period) that are free of nonfactor risk. These payoffs (or yields to maturities) form the basis variables that help in identifying the pricing kernel. Given the time interval we have in mind and availability of data, we confine our attention to the following variables as our candidate basis:

i. $1 + R_M(t + 1)$
ii. $1 + y(t + 1, 1)$
iii. $1 + (y(t + 1, 9) - y(t + 1, 3))$.

$R_M(t + 1)$ is the nominal return on the market, $y(t + 1, 1)$ is the nominal yield to maturity on the Treasury bill next period, and $y(t + 1, 9) - y(t + 1, 3)$ is the nominal yield spread (next period) on the nine-month Treasury bill relative to the three-month Treasury bill. We call p_{t+1}^b the realization of the M dimensional vector of basis variables at time $t + 1$ (see Table I for descriptive statistics on these variables).

Some observations regarding the choice of these returns as the basis variables are in order. The theory at hand implies that the market is an ex post payoff free of idiosyncratic risk if we view it as a return on a security whose dividend stream is aggregate consumption. Similarly, the prices of pure discount bonds and the yield spread do not involve any nonfactor risk

Table I

Descriptive Statistics

This table presents descriptive statistics for the potential factors and payoffs used in this paper. Panel A presents statistics for the one-month value-weighted market return ($R_M(t + 1)$), the one-month Treasury bill yield to maturity next period ($y(t + 1,\ t + 2)$), the three-month Treasury bill yield to maturity next period ($y(t + 1,\ t + 4)$), the nine-month Treasury bill yield to maturity next period ($y(t + 1,\ t + 10)$), and the five-year Treasury bond yield to maturity next period ($y(t + 1,\ t + 61)$). Panel B presents statistics for the monthly returns on the ten size-sorted portfolios ($R_i(t + 1)$, $i = 1, \ldots, 10$).

Variable	Mean	Standard Deviation	Maximum	Minimum
		Panel A		
$1 + R_M(t + 1)$	1.0093085	0.0455172	1.1650000	0.7816900
$1 + y(t + 1, t + 2)$	1.0056045	0.0022277	1.0136310	1.0024940
$1 + y(t + 1, t + 4)$	1.0059056	0.0022844	1.0133610	1.0027810
$1 + y(t + 1, t + 10)$	1.0062456	0.0022391	1.0136310	1.0030870
$1 + y(t + 1, t + 61)$	1.0065165	0.0021162	1.0129640	1.0032230
		Panel B		
$1 + R_1(t + 1)$	1.0185306	0.0881853	1.5525000	0.6930000
$1 + R_2(t + 1)$	1.0149160	0.0768238	1.4617500	0.6991400
$1 + R_3(t + 1)$	1.0141183	0.0707249	1.3994500	0.7021200
$1 + R_4(t + 1)$	1.0140217	0.0674318	1.3677700	0.7072200
$1 + R_5(t + 1)$	1.0124893	0.0655323	1.3386100	0.7158900
$1 + R_6(t + 1)$	1.0129478	0.0620391	1.3149600	0.7181400
$1 + R_7(t + 1)$	1.0127542	0.0575023	1.2500700	0.7274600
$1 + R_8(t + 1)$	1.0115448	0.0543981	1.2463000	0.7374110
$1 + R_9(t + 1)$	1.0111459	0.0514449	1.2157600	0.7496100
$1 + R_{10}(t + 1)$	1.008277	0.0435068	1.1765200	0.8007800

Table II

Correlation between Potential "Factors"

This table shows the correlation matrix of the monthly value-weighted market return ($R_M(t+1)$, the one-month Treasury bill yield to maturity next period ($y(t+1, t+2)$), the three-month Treasury bill yield to maturity next period ($y(t+1, t+4)$), the nine-month Treasury bill yield to maturity next period ($y(t+1, t+10)$), and the five-year Treasury bond yield to maturity next period ($y(t+1, t+61)$).

Correlation	$R_M(t+1)$	$y(t+1, t+2)$	$y(t+1, t+4)$	$y(t+1, t+10)$	$y(t+1, t+61)$
$R_M(t+1)$	1.0000	−0.0652	−0.07378	−0.0943	−0.0318
$y(t+1, t+2)$	−0.0652	1.0000	0.98880	0.9376	0.87344
$y(t+1, t+4)$	−0.07378	0.9888	1.0000	0.9923	0.8939
$y(t+1, t+10)$	−0.0943	0.9376	0.9923	1.00000	0.944016
$y(t+1, t+61)$	−0.0318	0.87344	0.8939	0.944016	1.0000

(see Sections I and II). Table II shows that there is strong cross-correlation between different yields to maturity. To allow for sharper identification of the underlying states in low-dimensional approximations, it is useful to use the yield spread. Furthermore, there is no loss of information from using the spread; information on the long-run expectations of the economy is still included.

Our choice of basis variables and the use of a linear leading term ($\alpha_0 + \alpha \cdot p_{t+1}^b$) in equation (13) allows us to embed versions of the CAPM, the discrete ICAPM and the linear APT. Hansen and Jagannathan (1992) assume similar forms for the pricing kernels based on the CAPM and the linear APT.[15] The other model that we embed is the Rubinstein (1976) logarithmic preferences model (the R-CAPM) that has been generalized by Epstein and Zin (1991) to state nonseparable preferences. These models imply that the reciprocal of the market return is the relevant pricing kernel. Since the market portfolio is one of our basis variables, this pricing kernel is embedded in the class of pricing kernels that we consider.

The primitive returns we are interested in pricing are the size-based returns, the market return, and the yield on the one-period bond. To keep the number of orthogonality conditions small, we confine our attention to the first, second, and third decile returns, the market return, and the yield of the one-period bond. Our focus on the small size deciles and the market return is based on the large number of papers that indicate the inability of linear pricing models (including the CAPM) to explain cross-sectional returns (see Shanken (1990) and Connor and Korajczyk (1988), for example). We include the yield on the one-period bond because the rejection of consumption-based models is partially due to their inability to simultaneously price this security and the market return.

[15] Conditional forms of the linear APT and discrete ICAPM imply coefficients that are history-dependent in an unknown way.

We have five primitive payoffs that we are interested in pricing. Call the $t+1$ realization of this vector of payoffs x_{t+1}. First, we estimate our pricing kernels without imposing the nonnegativity restriction (Restriction 2). Subsequently, we reestimate the pricing kernel imposing the nonnegativity restriction. Since the difference in the two estimations is in the approximating function and not in the estimation method, we discuss our estimation strategy using the unrestricted pricing kernel $G(p_{t+1}^{b})$ (see equation (13)) as the pricing kernel that is to be estimated. For payoff i, write

$$u_{it+1} = G(p_{t+1}^{b})x_{it+1} - 1. \tag{17}$$

From equation (7) u_{it+1} satisfies the restriction that for all instruments Z_{jt} that are contained in I_t,

$$E[u_{it+1}Z_{jt}] = 0. \tag{18}$$

We construct sample versions of this restriction and estimate $G(\cdot)$. A consistent estimator of $G(\cdot)$ is obtained by minimizing the GMM criterion function. Define the $I \times J$ vector $e_{t+1} = (u_{it+1}Z_{jt})_{i=1,\, j=1}^{I,\, J}$. With I elementary payoffs and J instruments we have $IJ = L$ orthogonality conditions which are used to estimate the pricing kernel. Call the sample mean of the vector e_{t+1}, se_{t+1}. $\hat{G}(p_{t+1}^{b})$, the estimator of $G(p_{t+1}^{b})$, is the function found by minimizing the GMM criterion function w.r.t. $G(\cdot)$. The GMM criterion function that is minimized is

$$N[(se(G))'W(se(G))] \tag{19}$$

where W is a $L \times L$ optimal weighting matrix as discussed in Hansen (1982) and Hansen and Singleton (1982).[16] Minimizing the GMM criterion function as given above w.r.t. $G(\cdot)$ is asymptotically equivalent to minimizing the GMM criterion with respect to the parameters of the finite length neural net (equation (13)) that approximates $G(\cdot)$. In what follows, let $J(G(p_{t+1}^{b}), M)$ be the value of the minimized criterion function. If the length of the net is K and the number of factors is M then the number of parameters to be estimated is given by $NP(M, K) = (M+1)K + (M+1) + K$.

The issue of how to decide on the number of factors remains. We make this decision by using a likelihood ratio type test for the GMM criterion function. We begin with M factors and estimate $G(\cdot)$. The estimation is repeated using $M+1$ factors while keeping the weighting matrix fixed. The specification

[16] The alternative is to follow Gallant and Tauchen (1989) and use semi-nonparametric maximum likelihood techniques. This technique requires SNP estimation of the joint conditional densities of all the primitive payoffs and conditioning instruments that are not functions of lagged primitive payoffs. Currently, such SNP estimation is computationally infeasible for more than two or three random variables.

with $M + 1$ factors nests the M factor specification. Hence, the difference in the minimized criterion function between the nested and the more general specification $J(G(p_{t+1}^b), M) - J(G(p_{t+1}^b), M+1)$ is distributed χ^2 with $NP(K, M+1) - NP(K, M)$ degrees of freedom. A similar approach is used to test for the nested models discussed earlier against the nonlinear arbitrage-pricing theory.

In Section III, we stated that the size of the finite dimensional model as captured by the integer K has to increase with T, the sample size. We follow Gallant and Tauchen (1989) and use saturation ratios as an intuitive way to restrict the number of parameters given T, the sample size. We view the specification of $K = 3$ as a conservative specification of the length of the neural net.[17] If L orthogonality conditions are used, let $U = (e_1, \ldots, e_T)$ be the $T \times L$ matrix of observations. Since we have panel data of length 306, U is a $306 \times L$ matrix and is the matrix defining the number of primitive observations available to the econometrician. GMM exploits the first two moments of this process. The $L \times L$ variance-covariance matrix $E[UU']$ has $UNP = (L \times (L+1))/2$ parameters to be estimated. Hence, in total we have $UNP + NP$ parameters to estimate. The saturation ratio is then defined by $SR = (L \times T)/(NP + UNP)$.[18] Tauchen (1986) shows that the overidentifying restrictions-based summary χ^2 statistic for model rejection is unaffected by the number of orthogonality conditions as long as the implied saturation ratio is greater than 20.

In our estimation, we set K equal to 3 and use a two-factor specification. In this case, the number of parameters, NP, is 15. Given the size of the sample, this is a fairly small number of parameters to be estimated in the neural net. Furthermore, with 25 orthogonality conditions, the SR is about 23. In Gallant and Tauchen (1989), the saturation ratio in their estimate of the conditional density is in the neighborhood of 11. We note that in the nonlinear APT, additional payoffs help restrict the behavior of the pricing kernel in a meaningful way. Adding new payoffs to estimate the pricing kernel may provide information about the pricing kernel on states that are not spanned by the previous payoffs.

We consider in our estimation four instrument sets that are labelled SET A, SET B, SET C, and SET D. Each set contains five instruments and with five payoffs, yields 25 orthogonality conditions and implies a saturation ratio of 23 observations per nuisance parameter. The set of returns that were priced are the return on the market, the yield on a one-period discount bond, and size-based returns from decile 1 to decile 3 (decile 1 is the smallest firm). The four sets of instruments used were:

$$\text{SET A} = \{1, 1 + y(t,1), 1 + R_M(t), 1 + y(t,9), 1 + 10*[y(t,9) - y(t,3)]\}$$

$$\text{SET B} = \{1, 1 + y(t,1), 1 + R_{\text{EWM}}(t), 1 + y(t,3), 1 + y(t-1,9)\}$$

[17] Some practical considerations regarding the maximum size of the weighting matrix that is invertible also restrict the dimension of K, the length of the net.

[18] Given T, as L increases, the saturation ratio, SR, goes to zero and the variance covariance matrix of U is not invertible.

$$\text{SET C} = \{1, 1 + y(t,1), 1 + R_{\text{EWM}}(t), 1 + y(t,7), 1 + |y(t,7) - y(t-1,7)|\}$$

$$\text{SET D} = \left\{1, 1 + y(t,7), 1 + y(t,5), 1 + |R_1(t)|, 1 + (R_2(t))^2\right\}$$

where $y(t, k)$ is the nominal yield to maturity on the k period ahead discount bond as of time t, $R_j(t)$ is the lagged return on the jth size portfolio, and $R_{\text{EWM}}(t)$ is the lagged equally weighted return. The choice of instruments was motivated by the empirical results of Ferson and Harvey (1991), Shanken (1990), and others that suggest that market returns and term structure variables help predict future returns.

Because we use conditioning information we are creating dynamic trading strategies that are nonlinear functions of the returns. The pricing kernel is forced to price such nonlinear trading strategies. By using nonlinear functions as instruments we also test the theory against information not contained in the first moments of the payoffs.[19]

We use our factors, in sequence, the one-period-ahead market return, the next period yield on the one-period bond, and the yield spread of the next period. Thus the one-factor specification is the market factor, the two-factor specification is the market, and the one-period yield next period and the three-factor specification is the two-factor specification plus the yield spread next period. While we did try other possibilities for the third factor based on the yield curve, we do not report them here as the results were similar.

In addition, for each estimated pricing kernel, we test the performance of the pricing kernel using pricing restrictions not used in estimation. Following Hansen (1982), these Q additional orthogonality conditions yield a quadratic form of the GMM type discussed above that is distributed χ^2 with Q degrees of freedom. This $\chi^2(Q)$ statistic tests whether the estimated pricing kernel prices additional trading strategies not used in estimation. We emphasize that, strictly speaking, this test is not a conventional overidentifying restrictions test as the additional dynamic trading strategies may nontrivially restrict the behavior of potential pricing kernels.

VI. Results

The simplest linear specification (that embeds versions of the linear APT) was tested first (see equation (13)). Results in Table III, row 6 show that the linear model can be rejected very strongly for three sets of instruments (p-value is 0.001), while the fourth instrument set (SET D) yields a p-value of 0.0967. Thus the linear pricing kernel seems to be unable to explain the cross-sectional variations in returns.

The χ^2 test for the one-factor model (the CAPM) yields strong rejections (see Table III, row 14). Tests for the two- versus three-factor specification

[19] Bollersley, Engle, and Woolridge (1988) and a large number of other studies show strong ARCH and GARCH type nonlinearity in the returns data.

Table III

Tests of the Linear APT

This table shows results for the linear APT model with the one-month value-weighted return and the one-month yield to maturity as factors. With two factors (M = 2) and a linear model (K = 0), the number of parameters (NP) to be estimated is 3 and the saturation ratio (SR) is 23. The five payoffs used in estimation are the first- second-, and third-decile returns, the market return, and the one-month Treasury bill yield to maturity. For each of the four instrument sets used in estimation, we have five instruments and five payoffs and thus 25 orthogonality conditions (OC). For each instrument set, we first report the GMM criterion value (χ^2 Value for model), the degrees of freedom (DoF), and the associated p-value (Tail probability). Second, we report the mean and standard deviation of the estimated pricing kernel. Third, we report the χ^2 value and associated p-value for the GMM tests based on moment restrictions not used in estimating the pricing kernel. This test uses twelve payoffs and three instruments (36 DoF). Fourth, we report the χ^2 value, DoF, and associated p-value for the specification test for the nested one-factor model (only the market return and a constant) versus the two-factor model. Fifth, we report the χ^2 value, DoF, and associated p-value for the test whether an additional third factor (the yield spread) is required.

1.	Specification test (K, M, NP, SR)	0, 2, 3, 23	0, 2, 3, 23	0, 2, 3, 23	0, 2, 3, 23
2.	No. of returns, No. of inst., No. of OC	5, 5, 25	5, 5, 25	5, 5, 25	5, 5, 25
3.	Instrument set	Set A	Set B	Set C	Set D
4.	χ^2 Value for model	51.39	52.16	46.60	30.97
5.	DoF	22	22	22	22
6.	Tail probability	0.0004	0.0003	0.0016	0.0967
7.	PK mean	0.968	0.987	0.982	0.996
8.	PK standard deviation	0.192	0.169	0.167	0.1892
Out-of-sample tests					
9.	No. of returns, No. of inst., DoF	12, 3, 36	12, 3, 36	12, 3, 36	12, 3, 36
10.	χ^2 Value	80.86	73.22	76.66	75.90
11.	Tail probability	0.0000	0.0002	0.0001	0.0001
Nested one-factor model test					
12.	K, M, NP	0, 1, 2	0, 1, 2	0, 1, 2	0, 1, 2
13.	χ^2 Value, DoF	61.20, 23	57.84, 23	54.36, 23	37.21, 23
14.	Tail probability	0.0000	0.0001	0.0002	0.0309
15.	Test for additional factor (K, M, M + 1)	0, 2, 3	0, 2, 3	0, 2, 3	0, 2, 3
16.	Difference, DoF	0.0498, 1	0.3800, 1	0.0002, 1	0.0000, 1
17.	Tail probability	0.8234	0.5376	0.9887	1.0000

strongly support the two-factor case (Table III, row 17).[20] Finally, a test using additional dynamic trading strategies was conducted. This additional trading strategies test used twelve payoffs (the ten decile returns plus the market return and the yield on the one-period bond) and three instruments not used in estimation yielding 36 orthogonality conditions.[21] This test shows that

[20] Even though we have a two-factor specification, the market return seems to be the dominant factor. The pricing kernel covaries to a much greater degree with the market than with the short interest rate. This finding is consistent with the empirical work of Ferson and Harvey (1991).

[21] The three new instruments were the yield spread on the five-year bond, the lagged excess market return, and the cube of the lagged first-decile return.

the estimated pricing kernel cannot price trading strategies not used in estimation (Table III, row 11).

Given the poor performance of the linear specification, we test the nonlinear specification discussed above. The one-factor (market) nonlinear specification (which subsumes the R-CAPM) is rejected strongly (see Table IV, row 14). Hence, we investigate the two-factor nonlinear model. Table IV, row 6 shows that the χ^2 specification test yields much higher p-values for the

Table IV

Tests of the Nonlinear APT

This table shows results for the nonlinear APT model with the one-month value-weighted return and the one-month yield to maturity as factors. With two factors ($M = 2$) and three nets ($K = 3$), the number of parameters (NP) to be estimated is 15 and the saturation ratio (SR) is 22. The five payoffs used in estimation are the first-, second-, and third-decile returns, the market return, and the one-month Treasury bill yield to maturity. For each of the four instrument sets used in estimation, we have five instruments and five payoffs and thus 25 orthogonality conditions (OC). For each instrument set, we first report the GMM criterion value (χ^2 value for model), the degrees of freedom (DoF), and the associated p-value (Tail probability). Second, we report the mean and standard deviation of the estimated pricing kernel. Third, we report the χ^2 value and associated p-value for the GMM tests based on moment restrictions not used in estimating the pricing kernel. This test uses twelve payoffs and three instruments (36 DoF). Fourth, we report the χ^2 value, DoF, and associated p-value for the specification test for the nested one-factor model (only the market return and a constant) versus the two-factor model. Fifth, we report the χ^2 value, DoF, and associated p-value for the test whether an additional third factor (the yield spread) is required. Sixth, we report the χ^2 value, DoF, and associated p-value for the test for the nested linear APT model versus the nonlinear APT model.

1. Specification test (K, M, NP, SR)	3, 2, 15, 22	3, 2, 15, 22	3, 2, 15, 22	3, 2, 15, 22
2. No. of returns, No. of inst., No. of OC	5, 5, 25	5, 5, 25	5, 5, 25	5, 5, 25
3. Instrument set	Set A	Set B	Set C	Set D
4. χ^2 Value for model	13.38	18.96	18.83	11.31
5. DoF	10	10	10	10
6. Tail probability	0.2032	0.0408	0.0426	0.3338
7. PK mean	0.997	0.995	0.983	0.983
8. PK standard deviation	0.824	0.762	1.054	0.492
Out-of-sample tests				
9. No. of returns, No. of inst., DoF	10, 2, 36	10, 2, 36	10, 2, 36	10, 2, 36
10. χ^2 Value	52.78	57.95	63.79	74.368
11. Tail probability	0.0352	0.0116	0.0029	0.0002
Nested one-factor model test				
12. K, M, NP	0, 1, 2	0, 1, 2	0, 1, 2	0, 1, 2
13. χ^2 Value, DoF	42.75, 14	31.58, 14	32.19, 14	20.67, 14
14. Tail probability	0.0001	0.0046	0.0038	0.1104
15. Test for additional factor (K, M, M + 1)	0, 2, 3	0, 2, 3	0, 2, 3	0, 2, 3
16. Difference, DoF	0.812, 4	0.155, 4	0.072, 4	0.000, 4
17. Tail probability	0.9368	0.9972	0.9994	1.0000
Nested linear model test				
18. Difference in function value, DoF	31.59, 12,	44.48, 12	18.26, 12	8.30, 12
19. Tail probability	0.0016	0.0000	0.1080	0.7613

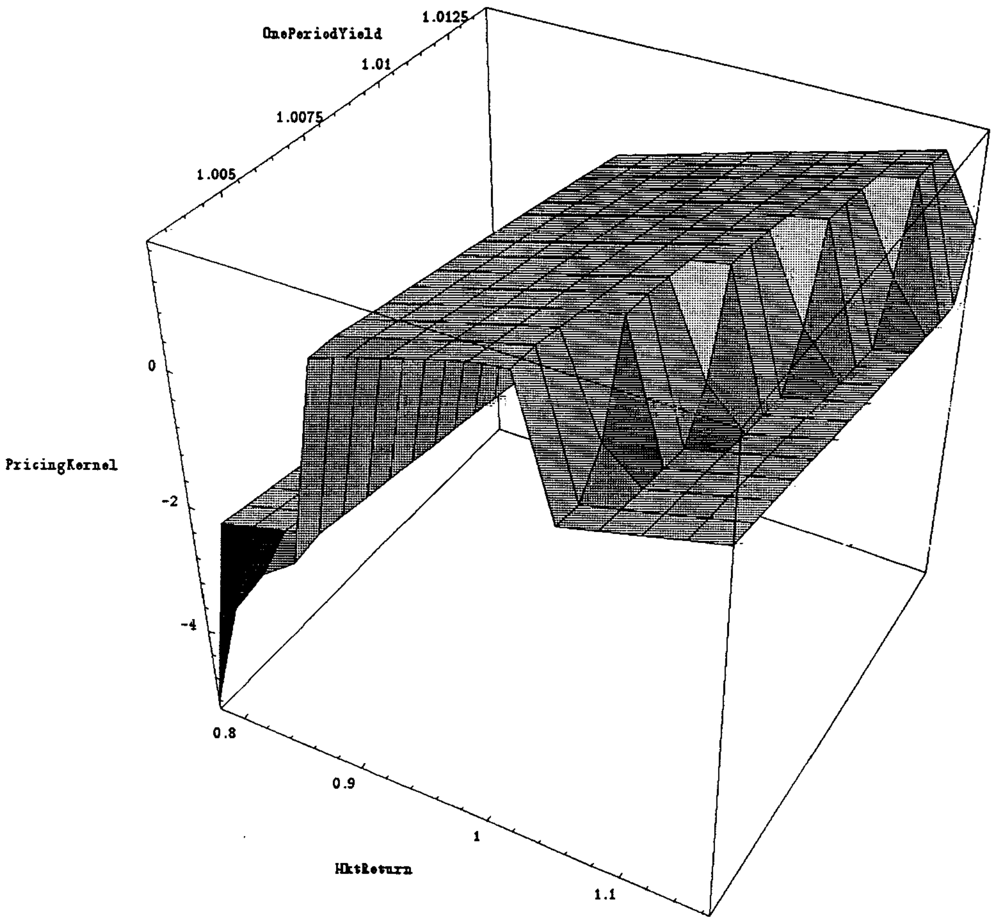

Figure 1. Pricing kernel using estimates from instrument set A. This figure uses the estimated parameters of the nonlinear pricing kernel for instrument set A and plots the pricing kernel over the range of the observed values for the one-month market return and the one-month Treasury bill yield to maturity.

nonlinear model as compared to the linear model. For instrument sets A and D, the p-values are around 20 and 30 percent respectively. For instrument sets B and C, the p-values are around 4 percent. Thus there seems to be substantially more support for a nonlinear model than for a linear model. The nonlinear pricing kernel shows substantial nonlinearities in the region where market returns are high (a region with about 15 percent of the market return observations). This is illustrated by Figures 1 and 2. In fact, all the four estimated pricing kernels are negatively correlated with the market return (see Table VIII). This negative correlation is consistent with the general implications of the R-CAPM that payoffs which hedge against movements in the market are valuable. Again, likelihood type tests strongly prefer the two-factor model to the three-factor model (Table IV, row 17).

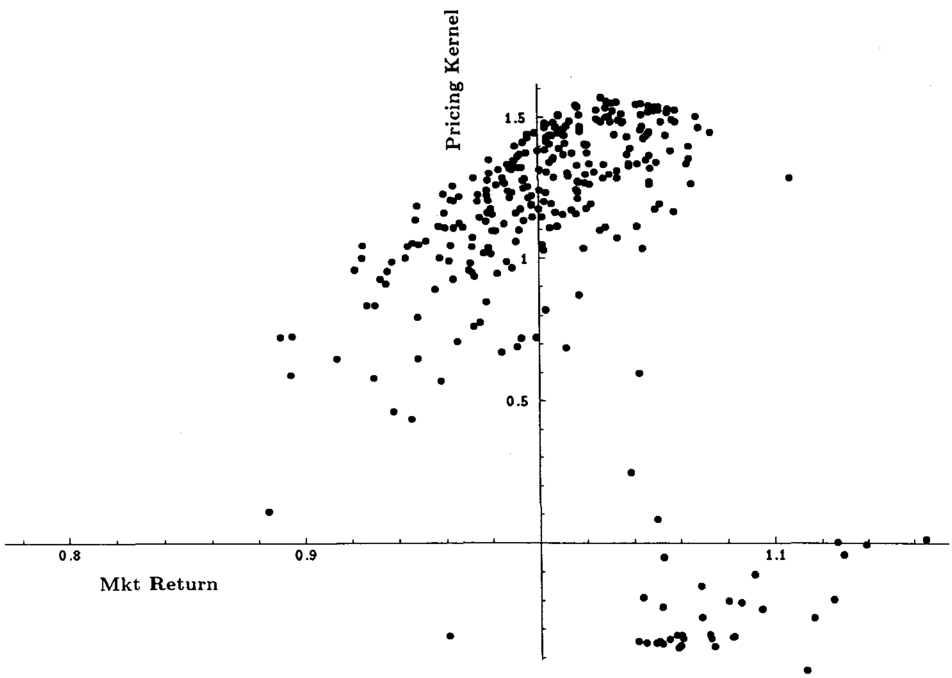

Figure 2. Estimated pricing kernel versus actual market returns: Instrument set A. This figure shows the scatterplot of the estimated pricing kernel (using instrument set A) against the one-month market return.

An additional trading strategies test similar to that done for the linear model was conducted. The test shows that it is hard to jointly price the new dynamic trading strategies (Table IV, row 11). However, the performance of the nonlinear pricing kernel is considerably better than that of the linear pricing kernel. This evidence seems to provide further support for a nonlinear specification.

Specification tests for the embedded linear model were conducted fixing the weighting matrix estimated using the larger nonlinear model. These results are reported in Table IV, row 19. For instruments sets A and B, this test rejects the linear model very strongly (p-values are 0.0016 and 0.0000). For instrument set C, the rejection is marginal (p-value 0.1080), while for instrument set D, the linear model (in comparison to the nonlinear model) cannot be rejected. This result for instrument D is not surprising and is consistent with the result that the linear model was rejected with this instrument set only at 9.67 percent (see Table III, row 6). These specification tests and the previously discussed results provide strong evidence that the nonlinear model is the preferred model.

To understand the better performance of the nonlinear specification, diagnostics using the excess return on the small firm were conducted. In particular, the fact that the errors of the orthogonality conditions form a martingale difference sequence is used. Hence the error $RES_{t+1} = \hat{G}(p_{t+1}^{b})(R_1(t+1) -$

$y(t, 1)$) is regressed against measurable functions of the lagged errors. Given the martingale difference property this regression should have a R^2 of zero and insignificant coefficients. We run the following two regressions for the linear and the nonlinear specifications.

Diagnostic 1:

$$RES_t = \delta_0 + \delta_{JAN} 1_{\{t=JAN\}} + \eta_1 RES_{t-1} + \eta_2 RES_{t-2} + \zeta t$$

Diagnostic 2:

$$RES_t = \delta_0 + \delta_{\{JAN\}} 1_{t=JAN} + \eta_1 |RES|_{t-1} + \eta_2 |RES|_{t-2} + \theta_1 RES_{t-1} + \zeta t$$

In the first diagnostic, we run the error against two lags and a January dummy. The January dummy is motivated by the considerable empirical evidence that small firm excess returns have significantly different January means. The second diagnostic uses lags of the absolute values of the lagged errors along with the lagged residual itself. We do this to allow for nonlinear dependence.

Tables V, VI, and VII shows the results for the two diagnostics. The first diagnostic regression without the January dummy shows the hypothesis that

Table V

First Diagnostic for the Linear and Nonlinear Models with January Dummy

This table provides diagnostics for the linear and nonlinear APT models by regressing the error from the pricing restriction associated with the excess return on the small-firm portfolio against a constant, a January dummy, and two lags of the errors themselves. The R^2 and Durbin Watson statistic for the regression, the estimated coefficients, and their associated t-values are shown in the table.

Panel A: Linear Model				
	$R^2 = 0.209$	Adjusted $R^2 = 0.201$	DW = 1.983	
Variable	Coefficient	Standard Error	t-statistic	p-value
Constant (δ_0)	−0.00727	0.0046	−1.573	0.1169
January dummy (δ_{JAN})	0.1310	0.0160	8.154	0.0000
One-lag residual (η_1)	0.1895	0.0521	3.631	0.0003
Two-lag residual (η_2)	−0.0570	0.0522	−1.093	0.2754
Panel B: Nonlinear Model				
	$R^2 = 0.006$	Adjusted $R^2 = -0.004$	DW = 2.01	
Variable	Coefficient	Standard Error	t-statistic	p-value
Constant (δ_0)	0.0059	0.0095	0.632	0.5282
January dummy (δ_{JAN})	−0.0112	0.0330	−0.341	0.7331
One-lag residual (η_1)	0.0140	0.0576	0.243	0.8083
Two-lag residual (η_2)	0.0754	0.0577	1.308	0.1918

Table VI

First Diagnostic for the Linear and Nonlinear Models without January Dummy

This table provides diagnostics for the linear and nonlinear APT models by regressing the error from the pricing restriction associated with the excess return on the small-firm portfolio against a constant and two lags of the errors themselves. The R^2 and Durbin Watson statistic for the regression, the estimated coefficients, and their associated t-values are shown in the table.

Panel A: Linear Model

Variable	$R^2 = 0.032$ Coefficient	Adjusted $R^2 = 0.026$ Standard Error	DW = 2.0041 t-statistic	p-value
Constant (δ_0)	0.00366	0.0048	0.750	0.4536
One-lag residual (η_1)	0.1783	0.0575	3.097	0.0021
Two-lag residual (η_2)	−0.0753	0.0575	−1.308	0.1918

Panel B: Nonlinear Model

Variable	$R^2 = 0.006$ Coefficient	Adjusted $R^2 = -0.001$ Standard Error	DW = 2.011 t-statistic	p-value
Constant (δ_0)	0.00505	0.0090	0.558	0.5775
One-lag residual (η_1)	0.0146	0.0576	0.254	0.7994
Two-lag residual (η_2)	0.7478	0.0575	1.299	0.1950

Table VII

Second Diagnostic for the Linear and Nonlinear Models

This table provides diagnostics for the linear and nonlinear APT models by regressing the error from the pricing restriction associated with the excess return on the small-firm portfolio against a constant, a January dummy, and one lag of the GMM estimation error and two lags of the absolute value of the GMM estimation error. The R^2 and Durbin Watson statistic for the regression, the estimated coefficients, and their associated t-values are shown in the table.

Panel A: Linear Model

Variable	$R^2 = 0.2114$ Coefficient	Adjusted $R^2 = 0.200$ Standard Error	DW = 1.944 t-statistic	p-value
Constant (δ_0)	−0.0140	0.0074	−1.888	0.0600
January dummy (δ_{JAN})	0.1327	0.0160	8.258	0.0000
Absolute value of one-lag residual (η_1)	0.1043	0.0715	1.458	0.1459
Absolute value of two-lag residual (η_2)	0.0042	0.0714	0.059	0.9524
One-lag residual (θ_1)	0.1850	0.0516	3.581	0.0004

Panel B: Nonlinear Model

Variable	$R^2 = 0.0014$ Coefficient	Adjusted $R^3 = -0.011$ Standard Error	DW = 1.997 t-statistic	p-value
Constant (δ_0)	0.0096	0.0118	0.816	0.4149
January dummy (δ_{JAN})	−0.0103	0.0331	−0.313	0.7542
Absolute value of one-lag residual (η_1)	−0.0352	0.0707	−0.497	0.6195
Absolute value of two-lag residual (η_2)	−0.0113	0.0647	−0.175	0.8608
One-lag residual (θ_1)	0.0023	0.0633	0.036	0.9711

Table VIII
Correlation between Estimated Pricing Kernels

This table provides descriptive statistics for the four pricing kernels estimated using the four instrument sets. Panel A of the table shows the mean, standard deviation, minimum, and maximum value taken in sample and the number of negative value (in sample) for each of the four estimated pricing kernels. Panel B shows the correlation between each estimated pricing kernel and the value-weighted market return and the one-month Treasury bill yield to maturity.

Panel A

Pricing Kernel	Mean	Standard Deviation	Maximum	Minimum	No. of Negative Values
PK25A	0.99685	0.82546	1.56795	−8.81818	38
PK25B	0.99542	0.76841	1.56795	−8.81818	12
PK25C	0.98272	1.05454	17.11337	−0.68537	13
PK25D	0.98352	0.49233	1.16983	−3.12667	3

Panel B

Correlation	PK25A	PK25B	PK25C	PK25D
$R_M(t+1)$	−0.04423	−0.09156	−0.08168	−0.19543
$y(t+1, t+2)$	0.01479	0.04164	−0.00093	0.02284

the moment has been set equal to zero is accepted strongly. For the linear model, both diagnostics (with a January dummy) show R^2 and adjusted R^2 around 20 percent. Furthermore, the coefficient on the January dummy is positive and significantly different from zero and the error structure shows statistically significant persistence. In contrast, the nonlinear model shows virtually zero R^2 for both diagnostics and there is no significant coefficient on either the January dummy or lagged errors. This implies that the nonlinear model is better able to explain the variation in small firm returns and the January effect. Thus, the January effect seems to be a peculiar kind of nonlinearity that linear models cannot capture and nonlinear models can. Similar results for the linear and nonlinear model were obtained with returns in deciles 2 and 3 (small firms, but not in the smallest decile).[22] However, the January effect was not as marked as it was for the smallest decile in the linear model. In part, the rejections of the linear model can be attributed to their inability to explain small firm returns.

As mentioned, the nonlinear pricing kernel exhibits significant nonlinearities when market returns are higher. To provide some intuition for the nonlinear kernel, the pricing kernel estimated in instrument set A was regressed against polynomial functions of the market returns and the one-period interest rate. These results, presented in Table IX indicate that the

[22] These results are not reported.

Table IX

Understanding the Nonlinearities in the Pricing Kernel

This table provides intuition for the nature of nonlinearities in the pricing kernel estimated in instrument set A by regressing the pricing kernel against a constant, the market return, the square of the market return, the cube of the market return, the product of the market return, and one-month Treasury bill yield to maturity and the square of the one-month Treasury bill yield to maturity.

Variable	Coefficient	Standard Error	t-statistic	p-value
Constant	715.586	201.55	2.55	0.0004
$R_M(t+1)$	−1324.557	426.35	−3.10	0.0021
$(R_M(t+1))^2$	−409.55	236.02	−1.74	0.0837
$R_M(t+1) \times y(t+1, t+2)$	1806.65	398.74	4.53	0.0000
$(R_M(t+1))^3$	108.27	79.86	1.36	0.1761
$(y(t+1, t+2))^2$	−895.28	198.91	−4.50	0.0000

$R^2 = 0.2466$, Adjusted $R^2 = 0.2340$.

nonlinearities seem to be related to the variances of the market and the interest rate, the covariance between the market return and interest rate, and the skewness of the market return.

Tests that do not impose nonnegativity are supportive of a nonlinear model. However, the nonlinear model has considerably more negative values than the linear pricing kernel (see Table VIII for the number of negative values). Thus we redo our estimation for instrument sets A and B with the imposition of nonnegativity in estimation. We note that imposing nonnegativity in small samples may actually deteriorate the quality of the pricing kernel. An unrestricted neural net may better approximate a positive pricing kernel than a neural net that is constrained to be positive.

Since the linear pricing kernel has virtually no negative values, the imposition of nonnegativity does not make a difference in the estimates. As before, we reject the linear model very strongly for the instrument sets A and B (see Table X, row 6). Similarly, tests for the number of factors support a two-factor model. Thus the behavior of the linear pricing kernel does not change when nonnegativity is imposed in estimation.

Not surprisingly, the imposition of nonnegativity changes our results on the nonlinear pricing kernel. In particular, the nonlinear model is rejected strongly for both instrument sets A and B (Table XI, row 6). Thus the imposition of nonnegativity results in greater rejections for the nonlinear model. Finally, the nested test of the linear specification using the weighting matrix estimated for the larger nonlinear specification yields p-values of 0.1191 and 0.4053 for the two instrument sets. Thus the performance of the linear and nonlinear models is closer when nonnegativity is imposed.[23]

[23] Diagnostics (not reported) based on the excess return errors on the small firm decile were undertaken (as in the case where nonnegativity was not imposed). These diagnostics showed persistence in the errors for the nonlinear model when nonnegativity was imposed as in the linear case.

Table X

Tests of the Linear APT (with Nonnegativity)

This table shows results for the linear APT model with the one-month value-weighted return and the one-month yield to maturity as factors and where nonnegativity is imposed in estimation. With two factors (M = 2) and a linear model (K = 0), the number of parameters (NP) to be estimated is 3 and the saturation ratio (SR) is 23. The five payoffs used in estimation are the first-, second-, and third-decile returns, the market return, and the one-month Treasury bill yield to maturity. For each of the four instrument sets used in estimation, we have five instruments and five payoffs and thus 25 orthogonality conditions (OC). For each instrument set, we first report the GMM criterion value (χ^2 Value for model), the degrees of freedom (DoF), and the associated p-value (Tail probability). Second, we report the mean and standard deviation of the estimated pricing kernel. Third, we report the χ^2 value and associated p-value for the GMM tests based on moment restrictions not used in estimating the pricing kernel. This test uses twelve payoffs and three instruments (36 DoF). Fourth, we report the χ^2 value, DoF, and associated p-value for the specification test for the nested one-factor model (only the market return and a constant) versus the two-factor model. Fifth, we report the χ^2 value, DoF, and associated p-value for the test whether an additional third factor (the yield spread) is required.

1. Specification test (K, M, NP, SR)	0, 2, 3, 23	0, 2, 3, 23
2. No. of returns, No. of inst., No. of OC	5, 5, 25	5, 5, 25
3. Instrument set	Set A	Set B
4. χ^2 Value for model	51.63	52.22
5. DoF	22	22
6. Tail probability	0.0003	0.0003
7. PK mean	0.968	0.986
8. PK standard deviation	0.188	0.169
Out-of-sample tests		
9. No. of returns, No. of inst., DoF	12, 3, 36	12, 3, 36
10. χ^2 Value	81.86	73.23
11. Tail probability	0.0000	0.0002
Nested one-factor model test		
12. K, M, NP	0, 1, 2	0, 1, 2
13. χ^2 Value, DoF	61.21, 23	57.85, 23
14. Tail probability	0.0000	0.0001
15. Test for additional factor (K, M, M + 1)	0, 2, 3	0, 2, 3
16. Difference in function value, DoF	0.0039, 1	0.0026, 1
17. Tail probability	0.9502	0.9593

In the estimation θ is set equal to 0.01.

We reiterate that the imposition of nonnegativity may deteriorate the approximation quality and hence the estimates of the nonlinear pricing kernel in small samples.

VII. Conclusions

In this paper, we take the view that the theoretical and empirical content of the APT is the existence of a low-dimensional, nonnegative pricing kernel. The no-arbitrage restriction implies the existence of a nonnegative pricing

Table XI

Tests of the Nonlinear APT (with Nonnegativity)

This table shows results for the nonlinear APT model with the one-month value-weighted return and the one-month yield to maturity as factors and where nonnegativity is imposed in estimation. With two factors (M = 2) and three nets (K = 3), the number of parameters (NP) to be estimated is 15 and the saturation ratio (SR) is 22. The five payoffs used in estimation are the first-, second-, and third-decile returns, the market return, and the one-month Treasury bill yield to maturity. For each of the four instrument sets used in estimation, we have five instruments and five payoffs and thus 25 orthogonality conditions (OC). For each instrument set, we first report the GMM criterion value χ^2 Value for model), the degrees of freedom (DoF), and the associated p-value (Tail probability). Second, we report the mean and standard deviation of the estimated pricing kernel. Third, we report the χ^2 value and associated p-value for the GMM tests based on moment restrictions not used in estimating the pricing kernel. This test uses twelve payoffs and three instruments (36 DoF). Fourth, we report the χ^2 value, DoF, and associated p-value for the specification test for the nested one-factor model (only the market return and a constant) versus the two-factor model. Fifth, we report the χ^2 value, DoF, and associated p-value for the test whether an additional third factor (the yield spread) is required. Sixth, we report the χ^2 value, DoF, and associated p-value for the test for the nested linear APT model versus the nonlinear APT model.

1. Specification test (K, M, NP, SR)	2, 2, 15, 22	3, 2, 15, 22
2. No. of returns, No. of inst., No. of OC	5, 5, 25	5, 5, 25
3. Instrument set	Set A	Set B
4. χ^2 Value for model	41.99	32.29
5. DoF	10	10
6. Tail probability	0.0000	0.0003
7. PK mean	0.989	1.005
8. PK standard deviation	0.276	0.712
Out-of-sample tests		
9. No. of returns, No. of inst., DoF	10, 2, 36	10, 2, 36
10. χ^2 Value	65.28	62.65
11. Tail probability	0.0020	0.0039
Nested one-factor model test		
12. K, M, NP	0, 1, 2	0, 1, 2
13. χ^2 Value, DoF	47.69, 14	36.72, 14
14. Tail probability	0.0000	0.0008
15. Test for additional factor (K, M, M + 1)	0, 2, 3	0, 2, 3
16. Difference in function value, DoF	2.73, 4	0.020, 4
17. Tail probability	0.6040	0.9999
Nested linear model test		
18. Difference in function value, DoF	18.25, 12	14.61, 12
19. Tail probability	0.1083	0.4053

In the estimation θ is set to equal 0.01.

kernel and the APT restricts the kernel to be a function of a few factors or state variables. In particular, we argue that the usual linearity assumptions on the payoff structure that are used to derive linear APT models are unnecessary and provide conditions under which a low-dimensional nonnegative nonlinear pricing kernel exists. All the testable implications of the APT are then embedded in the restriction that there exists a nonnegative pricing kernel that prices all payoffs and is a function of a few factors or state variables.

In contrast to parametric models of asset pricing, the exact functional form of the pricing kernel is unknown to the econometrician. Hence, we proceed to estimate the pricing kernel and test the theory of using semi-nonparametric techniques. We use a GMM-based approach where we approximate the unknown nonlinear pricing kernel by neural networks. To embed versions of the linear APT, the CAPM, and the discrete time ICAPM we augment the neural net with a leading linear term. Furthermore, using likelihood ratio type tests, this approach allows us to test for the parametric nested models against the general nonlinear APT discussed in the paper.

In our first set of results, we do not impose nonnegativity in estimation. Nested parametric models are strongly rejected and the results provide greater support for a nonlinear pricing kernel with two factors: the market return and the one-period yield in the next period. The nonlinearities in the kernel seem to be related to the variances of the market and the interest rate, the covariance between the market and the interest rate, and the skewness of the market return. In addition, diagnostics using the excess returns of the small firms indicate that the nonlinear model is more capable of capturing the variation in these returns. With the imposition of nonnegativity in estimation, the performance of the linear and nonlinear pricing kernels is closer.

Our approach to arbitrage-pricing theory emphasizes the no-arbitrage and dimensionality restrictions and does not require that payoffs are linear in the factors as is typically assumed in the literature. This generalization of arbitrage pricing yields nonlinear pricing kernels on which the theory places restrictions. The econometric methodology that we present allows us to test the theory. We believe that our empirical results provide some support for a nonlinear arbitrage-pricing theory as compared to typical parametric models considered in the literature. Moreover, results from a related paper on high-frequency international data (see Bansal, Hsieh, and Viswanathan (1993)) corroborate our belief that this approach is promising.

Appendix

PROPOSITION 1: *Suppose there are K factor payoffs that are traded. Then, generically, the addition of a derivative security (even one that only involves factor risk but is not linearly spanned by the K factors) implies that a pricing kernel that is a linear combination of the factor payoffs cannot price all traded securities.*

Proof of Proposition 1: From Hansen and Jagannathan (1991), the linear APT implies that a linear combination of the K traded factor returns is a pricing kernel for returns (call this m_1). However, the minimum variance pricing kernel is also a linear combination of payoffs (call this m_2). Then for any payoff, x, with a unit price (i.e., a return) we must have

$$E[(m_1 - m_2)x|Z] = 0 \tag{A1}$$

Both m_1 and m_2 are linear combination of the payoffs. Thus we obtain that:

$$E[(m_1 - m_2)(m_1 - m_2)|Z] = E\left[(m_1 - m_2)^2|Z\right] = 0 \tag{A2}$$

From this it follows that $m_1 = m_2$ almost surely.

Rescale the linear APT pricing kernel so that this portfolio with weights α^*. The portfolio weights on m_2 (again we rescale to obtain a portfolio), α are given by (see Hansen and Jagannathan (1991)):

$$\alpha = cV^{-1}\mu \tag{A3}$$

where V is the $(K+1 \times K+1)$ variance covariance matrix and μ is the $K+1$ dimension mean vector. The first K elements of the mean vector correspond to the traded factors and the $K+1$st element corresponds to the added security. Finally, c is a constant of proportionality to ensure that the portfolio weights sum to one. Given that $m_1 = m_2$ a.s. it follows that $(\alpha_1, \dots, \alpha_K)' = \alpha^*$ and $\alpha_{K+1} = 0$. Rewriting equation (A3) as $c\mu = V\alpha$ and considering the $K+1$st equation in this set of equations, we obtain that

$$\frac{1}{c} \times h'_{K+1}\alpha^* = \mu_{K+1} \tag{A4}$$

where $h_{K+1} = [\sigma_{1,K+1}, \dots, \sigma_{K,K+1}]'$ is a $K \times 1$ vector of the covariances of the $K+1$st security with the K factor portfolios. Since equation (A4) is a linear manifold of dimension less than $\mathscr{R}^{K+1}$, the set of points $(\sigma_{1,K+1}, \dots, \sigma_{K,K+1}, \mu_{K+1})$ that satisfies equation (A4) is a set of Lebesgue measure zero (this is in the Lebesgue measure in $\mathscr{R}^{K+1}$). Consequently, generically, any new security (in particular, a derivative security not spanned by the factors) will not satisfy the above condition and will have nonzero weight in the minimum variance pricing kernel. Thus a linear factor pricing kernel with only K factor portfolios does not obtain. □

Proofs Relating to Example: First we note that

$$m_{t,t+s}d_{it+s} = H(\pi_{r=1}^{s}F(f_{t+r}))c_{t+s-1}[g_i(f_{t+s}) + \epsilon_{it+s}]. \tag{A5}$$

Since

$$\mathbf{E}[\epsilon_{it+s}H(\pi_{r=1}^{s}F(f_{t+r}))c_{t+s-1}|f_{t+s}, \Phi_{t+s-1}] = 0, \tag{A6}$$

the price of a security simplifies to

$$q_{it} = c_{t-1}\sum_{s=1}^{\infty}\mathbf{E}\left[H(\pi_{r=1}^{s}F(f_{t+r}))\frac{c_{t+s-1}}{c_{t-1}}|\Phi_t\right]. \tag{A7}$$

Recursive use of the equation (9) implies that c_{t+s-1}/c_{t-1} is a function of $f_t, \dots, f_{t+s-1}$ alone; in particular

$$\frac{c_{t+s-1}}{c_{t-1}} = \pi_{r=0}^{s-1}F(f_{t+r}). \tag{A8}$$

This allows us to write

$$q_{it} = c_{t-1} \sum_{s=1}^{\infty} \mathbf{E}\left[H(\pi_{r=1}^{s} F(f_{t+r})) \cdot \pi_{r=0}^{s-1} F(f_{t+r}))|\Phi_t\right]. \tag{A9}$$

Since the information in f_t is sufficient for predicting nonlinear functions of f_{t+r}, $r \geq 1$, it follows that the right-hand side is a function $\phi_i(f_t)$ of f_t only. Using $c_{t-1} = c_t/F(f_t)$ and defining $\bar{\phi}_i(f_t) = \phi_i(f_t)/F(f_t)$ we obtain that $q_{it} = c_t \bar{\phi}_i(f_t)$.

To determine returns, define $\bar{g}_i(f_t) = g_i(f_t)/F(f_t)$. Then $\bar{g}_i(f_t) = 1$ by equation (9). The equation for $R_i(t+1)$ comes by direct substitution in the definition, $R_i(t+1) = (q_{it+1} + d_{it+1})/q_{it}$ and by using equation (9) and the definition of $g_i(f_{t+1})$. The market return, $R_M(t+1)$ is given by using the weights $q_{it}/\Sigma_{i=1}^{N} q_{it} = \phi_i(f_t)/\Sigma_{i=1}^{N} \phi_i(f_t)$ on the individual firm returns and simplifying. The equation for $y(t, t+s)$ follows from Assumption 1 and the pricing equation for the discount bond. □

REFERENCES

Andrews, D. W. K., 1991, Asymptotic normality of series estimators for nonparametric and semi-parametric regression, *Econometrica* 59, 307–346.

Bansal, R., R. Gallant, B. Hussey, and G. Tauchen, 1992, Nonparametric estimation of structural models for high-frequency currency market data, Working paper, Duke University.

Bansal, R., D. Hsieh, and S. Viswanathan, 1993, A new approach to international arbitrage pricing, Working paper, Duke University, Fuqua School of Business.

Bansal, R., and S. Viswanathan, 1992, No arbitrage and arbitrage pricing: A new approach, Working paper, Duke University, Fuqua School of Business.

Bollerslev, T., R. F. Engle, and J. M. Woolridge, 1988, A capital asset pricing model with time-varying covariances, *Journal of Political Economy* 96, 116–131.

Bossaerts, P., and R. C. Green, 1989, A general equilibrium model of changing risk premia: Theory and tests, *Review of Financial Studies* 2, 467–493.

Breeden, D., 1979, An intertemporal asset-pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.

Brock, W. A., 1982, Asset prices in a production economy, in John J. McCall, ed.: *The Economics of Information and Uncertainty* (University of Chicago Press, Chicago).

Chamberlain, G., 1983, Funds, factors and diversification in arbitrage pricing theory, *Econometrica* 51, 1305–1323.

———, 1988, Asset pricing in multiperiod securities markets, *Econometrica* 56, 1283–1300.

———, and M. Rothschild, 1983, Arbitrage and mean variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Chen, N. F., and J. Ingersoll, 1983, Exact pricing in linear models with finitely many assets: A note, *Journal of Finance* 38, 985–988.

Chen, N. F., R. Roll, and S. A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Connor, G., 1984, A unified beta pricing theory, *Journal of Economic Theory* 34, 13–31.

———, and R. A. Korajczyk, 1988, Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics* 21, 255–289.

———, 1989, An intertemporal beta pricing model, *Review of Financial Studies* 2, 373–392.

Constantinides, G., 1989, Theory of valuation: Overview and recent developments, in S. Bhattacharya and G. Constantinides, eds.: *Theory of Valuation: Frontiers of Modern Financial Theory* (Rowman and Littlefield, Totawa, N.J.).

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, An intertemporal general equilibrium model of asset prices, *Econometrica* 51, 363–384.

Dybvig, P. H., 1983, An explicit bound on deviations from APT pricing in a finite economy, *Journal of Financial Economics* 12, 483–496.

———, and J. E. Ingersoll, 1982, Mean variance theory in complete markets, *Journal of Business* 55, 233–251.

Engle, R., V. Ng, and M. Rothschild, 1990, Asset pricing with a factor-ARCH covariance structure: Empirical estimates for Treasury bills, *Journal of Econometrics* 45, 213–237.

Epstein, L. G., and S. E. Zin, 1991, Substitution, risk aversion, and temporal behavior of consumption and asset returns, *Journal of Political Economy* 99, 263–286.

Ferson, W. E., and C. R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

Gallant, R., 1987, Identification and consistency in semi-nonparametric regression, in Truman F. Bewley, ed.: *Advances in Econometrics Fifth World Congress*, Vol. 1 (Cambridge University Press, New York).

———, and D. W. Nychka, 1987, Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 363–390.

Gallant, R., S. Rossi, and G. Tauchen, 1992, Stock returns and volumes, *Review of Financial Studies* 5, 199–242.

Gallant, R., and G. Tauchen, 1989, Semi-nonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing implications, *Econometrica* 57, 1091–1120.

Gallant, R., and H. White, 1989, On learning the derivatives of an unknown mapping with multilayer feedforward networks, Discussion paper, University of California at San Diego, Department of Economics.

Glosten, L., and R. Jagannathan, 1992, A contingent claim approach to performance evaluation, Working paper, Federal Reserve Bank of Minneapolis Research Department.

Grinblatt, M., and S. Titman, 1983, Factor pricing in a finite economy, *Journal of Financial Economics* 12, 495–507.

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

———, and R. Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

———, 1992, Assessing specification errors in stochastic discount factor models, Working paper, University of Minnesota.

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

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

———, 1983, Stochastic consumption, risk aversion and the temporal behavior of asset returns, *Journal of Political Economy* 91, 245–265.

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

Hollifield, B., 1990, Linear asset pricing with predictable time varying betas and risk premia: A general equilibrium approach, Working paper, Carnegie Mellon University.

Hornik, K., M. Stinchcombe, and H. White, 1989, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, Discussion paper, University of California, San Diego, Department of Economics.

Huberman, G., 1982, A simple approach to arbitrage pricing theory, *Journal of Economic Theory* 28, 182–191.

———, and S. Kandel, 1987, Mean variance spanning, *Journal of Finance* 42, 873–888.

Ingersoll, J., 1984, Some results in the theory of arbitrage pricing, *Journal of Finance* 39, 1021–1039.

Kraus, A., and R. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *Journal of Finance* 31, 1085–1100.

———, 1983, On the distributional conditions for a consumption-oriented three moment CAPM, *Journal of Finance* 38, 1381–1391.

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

Lehmann, B. N., and D. M. Modest, 1988, The empirical foundations of the arbitrage pricing theory, *Journal of Financial Economics* 21, 213–254.

Longstaff, F., 1989, Temporal aggregation and the continuous time asset pricing model. *Journal of Finance* 44, 871–887.

McCaffery, D. F., S. Ellner, R. Gallant, and D. W. Nychka, 1991, Estimating the Lyapunov exponent of a chaotic system with nonparametric regression, Working paper, North Carolina State University, Department of Statistics.

Merton, R., 1973, An intertemporal asset pricing model, *Econometrica* 41, 867–880.

Milne, F., 1988, Arbitrage and diversification in a general equilibrium asset economy, *Econometrica* 56, 815–840.

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

———, 1978, A simple approach to the valuation of risky streams, *Journal of Business* 51, 453–475.

Rubinstein, 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.

———, 1990, Intertemporal asset pricing: An empirical investigation, *Journal of Econometrics* 45, 99–120.

Tauchen, G., 1986, Statistical properties of generalized method of moments estimators obtained from financial market data, *Journal of Business and Economic Statistics* 4, 397–416.

White, H., 1988, Multilayer feedforward networks can learn arbitrary mappings: Connectionist nonparametric regression with automatic and semi-automatic determination of network complexity, Discussion paper, University of California, San Diego, Department of Economics.