THE JOURNAL OF FINANCE • VOL. LVII, NO. 1 • FEB. 2002

Nonlinear Pricing Kernels, Kurtosis Preference, and Evidence from the Cross Section of Equity Returns

ROBERT F. DITTMAR*

ABSTRACT

This paper investigates nonlinear pricing kernels in which the risk factor is endogenously determined and preferences restrict the definition of the pricing kernel. These kernels potentially generate the empirical performance of nonlinear and multifactor models, while maintaining empirical power and avoiding ad hoc specifications of factors or functional form. Our test results indicate that preference-restricted nonlinear pricing kernels are both admissible for the cross section of returns and are able to significantly improve upon linear single- and multifactor kernels. Further, the nonlinearities in the pricing kernel drive out the importance of the factors in the linear multi-factor model.

A PRINCIPAL IMPLICATION OF THE Capital Asset Pricing Model (CAPM) is that the pricing kernel is linear in a single factor, the portfolio of aggregate wealth. Numerous studies over the past two decades have documented violations of this restriction.[1] In response, researchers have examined the performance of alternative models of asset prices. These models have generally fallen into two classes: (1) multifactor models such as Ross' APT or Merton's ICAPM, in which factors in addition to the market return determine asset prices; or (2) nonparametric models, such as Bansal et al. (1993), Bansal and Viswanathan (1993), and Chapman (1997), in which the pricing kernel is not

* Dittmar is at the Kelley School of Business, Indiana University, Bloomington, Indiana. This paper is based on the author's dissertation at the University of North Carolina. Thanks to Dong-Hyun Ahn, Ravi Bansal, Utpal Bhattacharya, Mike Cliff, Jennifer Conrad, Amy Dittmar, Wayne Ferson, Ron Gallant, the journal editor, Richard Green, Mustafa Gültekin, Campbell Harvey, Karl Lins, Steve Monahan, Steve Slezak, George Tauchen, Marc Zenner, and an anonymous referee for valuable comments. Thanks also to seminar participants at Case Western Reserve University; Indiana University; the Office of the Comptroller of the Currency; Penn State University; the 1999 Western Finance Association meetings (Santa Monica); and the Universities of Cincinnati, Miami, Minnesota, North Carolina, Utah, Virginia, and Western Ontario for helpful comments and discussions. The author also thanks Eugene Fama and Kenneth French for making their data available. Any remaining errors are solely the author's responsibility.

[1] The literature documenting violations of this restriction is voluminous. A comprehensive set of references may be found in Campbell, Lo, and MacKinlay (1995).

linear in the market return. Empirical applications of these models suggest that they are much better at explaining cross-sectional variation in expected returns than the CAPM.[2]

Although these approaches perform well empirically, a number of limitations weaken their appeal. In particular, the models require ad hoc specifications of either the set of priced factors or the form of nonlinearity. Since the sets of potential factors and nonlinear functions are large, the researcher has considerable discretion over the form of the model to be investigated. Additionally, the form of the pricing kernel resulting from the nonparametric approaches does not derive from first principles. That is, given a set of assumptions on investors' preferences or return distributions, the nonlinear pricing kernels investigated in the nonparametric approaches do not follow endogenously. These limitations of the nonparametric and multifactor approaches are problematic in empirical applications because (1) tests based on ad hoc assumptions may lack power since they ignore the theoretical restrictions that might arise from a structural model and (2) the possibility exists for overfitting the data and factor dredging (Lo and MacKinlay (1990), Fama (1991)). In contrast, the set of factors in the CAPM (the market portfolio) and the form of the pricing kernel (linear) obtain as endogenous outcomes. Thus, the CAPM is free of the criticisms of arbitrary factor and functional form specification.

This paper investigates a pricing kernel that retains many of the attractive features of the pricing kernels investigated in nonparametric analyses while avoiding many of their limitations. The basis of our approach is to approximate an unknown marginal utility function in a static setting by a Taylor series expansion. The resulting pricing kernel is a polynomial function in aggregate wealth. The form of this Taylor series is restricted by imposing decreasing absolute prudence (Kimball (1993)) on investor's preferences. This restriction allows us to sign the first three polynomial terms in the expansion. The resulting pricing kernel is nonlinear, and therefore consistent with empirical evidence from nonparametric studies. Furthermore, it is a function of a risk factor that obtains endogenously and is restricted by preference assumptions, as in the CAPM. Consequently, the pricing kernel has the potential to explain some of the observed nonlinearities in the data. Concurrently, specification tests have improved power due to the preference restrictions imposed on the functional form of the pricing kernel.

As discussed above, our pricing kernel is a function only of the return on aggregate wealth. However, several recent papers have shown that the specification of aggregate wealth impacts the conclusions of empirical asset pricing studies. Consequently, we specify the priced factor as a function of both the return on equity and the return on human capital. We incorporate human capital, since recent evidence (Campbell (1996), Jagannathan and Wang

[2] Fama and French (1993, 1995, 1996) propose and investigate a multifactor alternative to the CAPM and find that it can capture more variation in expected returns than the CAPM. Bansal and Viswanathan (1993) and Bansal et al. (1993) explore various nonlinear pricing kernel specifications and find that these nonlinear specifications outperform linear specifications.

(1996)) suggests that the incorporation of human capital into the pricing kernel substantially improves the performance of the conditional CAPM. In contrast to this work, our pricing kernel allows human capital to impact asset prices nonlinearly through the polynomial pricing kernel. Moreover, we conjecture that mismeasurement of the market portfolio may have a particularly severe effect on the analysis of a nonlinear pricing kernel.

Our results indicate several interesting findings. First, we find that both a quadratic and a cubic pricing kernel are admissible for the cross section of industry portfolios, whereas the linear single-factor (CAPM) and linear multifactor (Fama-French) pricing kernels are not. Although the superior performance of nonlinear pricing kernels to linear pricing kernels has been documented in the literature (Bansal and Viswanathan (1993), Bansal et al. (1993), Chapman (1997)), to our knowledge the superiority of these kernels to a flexible multifactor model, such as the Fama–French model, has not. We find this result particularly interesting because the nonlinear pricing kernel that we investigate is subject to economic restrictions that do not affect the multifactor pricing kernel. In particular, the priced risk factor is obtained endogenously, and the signs of the coefficients of the pricing kernel are restricted by preference theory. In contrast, the priced risk factors in the multifactor model are specified exogenously, and the sign of the relationship between returns and these risk factors is unconstrained by economic theory. Furthermore, when the pricing kernel is specified as a cubic function of aggregate wealth augmented by the Fama–French factors, we find that these factors have no residual explanatory power for the cross section of returns. These results are important because they show that a pricing kernel grounded in preference theory can perform as well as, or better than, less restrictive factor models. Importantly, we find that human capital is critical to the improved performance of a nonlinear pricing kernel over linear single and multifactor pricing kernels. Moreover, it is incorporation of a *nonlinear* human capital measure that renders the pricing kernel admissible.

Although the pricing kernel that we investigate is restricted by preferences relative to multifactor or nonparametric pricing kernels, the kernel can be restricted further by preference theory. For example, specific preferences such as power utility are consistent with the decreasing absolute prudence restriction. We find that the nonlinear pricing kernel outperforms a pricing kernel implied by power utility. This evidence leads us to investigate the degree to which we can restrict the pricing kernel to be consistent with preferences and maintain improvement over the multifactor pricing kernels. In particular, we note that, under the assumption of decreasing absolute risk aversion, the pricing kernel itself should be decreasing. We impose this constraint in estimation and find that the resulting pricing kernel is no longer admissible for the cross section of returns. However, this pricing kernel continues to outperform the linear single and multifactor pricing kernels. This evidence suggests that nonlinearity can augment the performance of the pricing kernel framework. However, in order to describe the data, the pricing kernel must exhibit a fairly specific form of nonlinearity, which is captured by the cubic pricing kernel. Unfortunately, the cubic pricing kernel

cannot simultaneously deliver the nonlinearity necessary to price the assets under consideration and monotonically decrease. We conclude that a functional form that is able to maintain both of these properties is necessary to be both economically reasonable and admissible.

The remainder of the paper is organized as follows. In Section I, we discuss and motivate restrictions on agents' preferences that yield a specific nonlinear pricing kernel. The testing framework is discussed in Section II. Evidence on the performance of the model is provided in Section III. Section IV concludes.

I. Pricing Kernels and Moment Preference

To develop a specific nonlinear pricing kernel, we start with the intertemporal consumption and portfolio choice problem for a long-lived investor. As discussed in Hansen and Jagannathan (1991), the solution to an investor's portfolio choice problem can be expressed as the Euler equation

$$E[(1 + R_{i,t+1})m_{t+1}|\Omega_t] = 1, \tag{1}$$

where $(1 + R_{i,t+1})$ is the total return on asset i; m_{t+1} is the investor's intertemporal marginal rate of substitution, $U'(C_{t+1})/U'(C_t)$; and Ω_t is the information set available to the investor at time t. Harrison and Kreps (1979) show that m_{t+1} represents a pricing kernel that prices all risky payoffs under the law of one price and is nonnegative under the condition of no arbitrage. The assumption of the existence of a representative agent allows the pricing kernel to be expressed as a function of aggregate consumption. Although this specification is appealing from the standpoint of economic theory, considerable attention has been given to measurement and aggregation problems in available aggregate consumption proxies (e.g., Breeden, Gibbons, and Litzenberger (1989)). One method that is used to address this issue is to assume a static setting, and allow equation (1) to hold conditionally, as in Brown and Gibbons (1985). In this case, consumption and wealth are equivalent, and the intertemporal marginal rate of substitution can be expressed as a function of aggregate wealth, $U'(W_{t+1})/U'(W_t)$.

A further issue in this analysis is the form of the representative agent's utility function, $U(\cdot)$. A large body of literature investigates standard choices for $U(\cdot)$ and finds that the data imply unrealistic assumptions about investors' risk aversion or the riskless rate (e.g., Mehra and Prescott (1985), Weil (1989)). Thus, a suitable representation for the representative agent's utility function is unknown. To mitigate this problem, we express the pricing kernel generally as a nonlinear function of the return on aggregate wealth. Specifically, rather than take a stand on the exact form of the pricing kernel, we approximate it using a Taylor series expansion:

$$m_{t+1} = h_0 + h_1 \frac{U''}{U'} R_{W,t+1} + h_2 \frac{U'''}{U'} R^2_{W,t+1} + \ldots, \tag{2}$$

where $R_{W,t+1}$ represents the return on end-of-period aggregate wealth. As shown in equation (2), the marginal rate of substitution can be approximated as a polynomial in aggregate wealth in a static setting.

One difficulty with the polynomial expansion is the determination of the order at which the expansion should be truncated. Bansal et al. (1993) let the data determine the point of truncation. The difficulty with this approach is a loss of power; in allowing the data to guide the specification of the pricing kernel, the researcher risks overfitting the data. Furthermore, the economic interpretation of the resulting kernel is open to question. A more powerful alternative is to allow preference theory to guide the truncation. Thus, we rely on preference arguments to motivate the truncation of the polynomial. The standard arguments of positive marginal utility and risk aversion suggest that $U' > 0$ and $U'' < 0$. These restrictions yield a linear pricing kernel with a negative coefficient on the return on aggregate wealth, nesting the static CAPM. We further assume decreasing absolute risk aversion, which implies $U''' > 0$, as shown in Arditti (1967). This condition, coupled with truncating the series expansion after the quadratic term, yields a pricing kernel quadratic in the return on aggregate wealth, consistent with the three-moment CAPM.

We extend this progression of signing derivatives of utility functions by using the restriction of *decreasing absolute prudence* (Kimball (1993)). Kimball develops this restriction in response to Pratt and Zeckhauser (1987), who show that decreasing absolute risk aversion does not rule out certain counterintuitive risk-taking behavior. For example, any risk-averse agent should be unwilling to accept a bet with a negative expected payoff. Samuelson (1963) proves that if this agent had already accepted a bet with a negative expected payoff, that she should be unwilling to take another independent bet with a negative expected payoff. Pratt and Zeckhauser show that, if the agent's preferences are restricted only to exhibit decreasing absolute risk aversion, the agent may be willing to take this negative mean sequential gamble. Kimball shows that *standard risk aversion* rules out the aforementioned behavior. Sufficient conditions for standard risk aversion are decreasing absolute risk aversion *and* decreasing absolute prudence,

$$-\frac{d\dfrac{U'''}{U''}}{dW} = \frac{(U''')^2 - U''''U''}{(U'')^2} < 0. \tag{3}$$

Thus, assuming increasing marginal utility, risk aversion, and decreasing absolute risk aversion, equation (3) implies

$$U'''' < 0. \tag{4}$$

This condition shows that, by imposing standard risk aversion on agents' preferences, we are able to sign the coefficients of the first three polynomial terms in a Taylor series expansion.

Because preference theory does not guide us in determining the sign of additional polynomial terms, we assume that higher order polynomial terms are not important for pricing. More specifically, we implicitly assume that the covariance between returns and polynomial terms in aggregate wealth of order greater than three is zero.[3] Without this assumption, higher order terms, which we cannot definitively sign, enter the pricing kernel. Our view is that the power delivered by the sign restrictions outweigh the cost of omitting the higher order polynomial terms. Thus, with the assumption that the pricing kernel can be characterized by a low-order polynomial in aggregate wealth, imposing standard risk aversion on agents' preferences and truncating the expansion at the highest order term that can be signed together result in a pricing kernel that is cubic in the return on aggregate wealth.[4] The resulting pricing kernel is decreasing in the linear term of the pricing kernel, increasing in the quadratic term, and decreasing in the cubic term.

The pricing kernel that results from our analysis has several attractive features. First, the resulting pricing kernel does not take a strong stand regarding functional form. Additionally, the pricing kernel is nonlinear. Consequently, we conjecture that the polynomial pricing kernel will avoid problems associated with assuming a specific utility function and, instead, capture nonlinear features of the data, as do nonparametric pricing kernels. However, in contrast to the nonparametric kernels, the polynomial model is restricted by preference theory; preference assumptions drive the signs of the pricing kernel coefficients. These restrictions deliver greater economic and statistical power to tests of the model. In the subsequent sections, we conduct analyses of the performance of this kernel relative to alternative specifications of the pricing kernel.

As alluded to above, the polynomial expansion is also appealing in that it can be linked to preference for moments of the distribution of the return on wealth. Using the definition of covariance, equation (1) can be rewritten as

$$E[(1+R_{i,t+1})] = \frac{1}{E[m_{t+1}]} - \mathrm{Cov}[(1+R_{i,t+1}), m_{t+1}]\frac{1}{E[m_{t+1}]}. \tag{5}$$

Substituting equation (2) into equation (5) shows that expected returns are linked to covariances with the different orders of the polynomial in the return on aggregate wealth. Thus, a linear pricing kernel relates expected returns to covariance with the return on aggregate wealth, as in the CAPM. A quadratic pricing kernel relates expected returns to covariance with the return on aggregate wealth and the return on aggregate wealth squared. Since the coskewness of a random variable x with another random variable y can be represented as a function of $\mathrm{Cov}(x,y)$ and $\mathrm{Cov}(x,y^2)$, the quadratic pricing kernel is consistent with the three-moment CAPM. Similarly, a cubic

[3] This assumption may be justified if the joint distribution of returns and wealth is characterized by a four-moment density.

[4] This pricing kernel is consistent with a four-moment CAPM, as derived in Fang and Lai (1997).

pricing kernel is consistent with a model in the CAPM framework in which agents have preference over the first four moments of returns.

Analagous arguments can be made for higher moments; the pricing kernel in Bansal et al. (1993) incorporates a linear, quadratic, and quintic term, implying preference over variance, skewness, and the sixth moment. However, moments beyond the fourth are difficult to interpret intuitively and are not explicitly restricted by standard preference theory. In contrast, preference for the fourth moment, kurtosis, has both a utility-based and an intuitive rationale. Kurtosis can be described as the degree to which, for a given variance, a distribution is weighted toward its tails (Darlington (1970)). That is, kurtosis measures the bimodality of the distribution, or the probability mass in the tails of the distribution. Thus, kurtosis is distinguished from the variance, which measures the dispersion of observations from the mean, in that it captures the probability of outcomes that are highly divergent from the mean; that is, extreme outcomes. In a multivariate distribution, random variables may also exhibit cokurtosis. This measure captures the two variables' common sensitivity to extreme states.

Thus, a cubic pricing kernel can be justified under intuitive arguments, which suggests that investors are averse to extreme outcomes in a distribution, as well as utility-based arguments such as standard risk aversion. Consequently, we investigate a version of equation (2) that truncates the expansion at the return on aggregate wealth cubed

$$m_{t+1} = d_0 + d_1 R_{W,t+1} + d_2 R^2_{W,t+1} + d_3 R^3_{W,t+1}. \tag{6}$$

A pricing kernel specified in this way allows for an alternative functional form and potentially greater generality than that implied by the use of a specific utility function. However, since signs of the coefficients in the expansion are guided by theory, and we have limited the order of the expansion rather than allowing the data to determine the order of the expansion, we expect tests of the kernel's specification to be more powerful than a pure nonparametric approach.

II. Estimation Methods

As expressed in equation (6), the pricing kernel is a random variable with static coefficients. However, a large body of evidence suggests that return moments and prices of risk are time varying, and a wide array of studies have used this evidence as a basis for investigating static pricing models that hold conditionally (e.g., Harvey (1989), Ferson and Harvey (1991)). Although a static model will not hold conditionally in general, it may under certain conditions. For example, Campbell (1996) provides evidence that assets' intertemporal risks are proportional to their market risk. In this case, the asset pricing model can be expressed as a function only of market risk, allowing a static model to hold conditionally. Consequently, we analyze the model in conditional form by testing the implications of the Euler equation (1).

One potential implication of equation (1) holding conditionally is that the coefficients of the pricing kernel, d_n, are time varying. In a full-fledged pricing model, the conditional moments that drive these coefficients might be directly modeled (e.g., Harvey (1989)). Alternatively, in the more general situation described by (6), a functional form for the coefficients may be specified. Dumas and Solnik (1995) and Cochrane (1996) treat these coefficients as linear functions of time t information variables. The resulting pricing kernel is specified as

$$m_{t+1} = \delta_0' \mathbf{Z_t} + \delta_1' \mathbf{Z_t} R_{W,t+1} + \delta_2' \mathbf{Z_t} R_{W,t+1}^2 + \delta_3' \mathbf{Z_t} R_{W,t+1}^3. \tag{7}$$

This approach is advantageous in being a parsimonious approximation, but the functional form does not impose any restrictions on the signs of the coefficients. Consequently, we investigate a pricing kernel of the form

$$m_{t+1} = (\delta_0' \mathbf{Z_t})^2 - (\delta_1' \mathbf{Z_t})^2 R_{W,t+1} + (\delta_2' \mathbf{Z_t})^2 R_{W,t+1}^2 - (\delta_3' \mathbf{Z_t})^2 R_{W,t+1}^3. \tag{8}$$

As discussed in Section I, imposing decreasing absolute prudence implies that $U'''' < 0$, $U''' > 0$, and $U'' < 0$. Because the coefficients $(\delta_{n,t})^2$ are forced to be positive-valued in equation (8), this specification forces the preference restrictions implied by decreasing absolute prudence.

One more feature of the pricing kernel framework is exploited in estimation. Equation (1) implies that the mean of the pricing kernel should be equal to the inverse of the gross return on a riskless asset or, more generally, a zero-beta asset. That is, $E_t[m_{t+1}] = 1/E_t[R_{0,t+1}]$. This condition can be imposed by including a proxy for the riskless or zero-beta asset in the set of payoffs. Dahlquist and Söderlind (1999) and Farnsworth et al. (1999) find that imposing this restriction on the pricing kernel is important in the context of performance evaluation. Dahlquist and Söderlind also show that failure to impose this restriction can result in estimation of a valid pricing kernel that implies a mean-variance tangency portfolio that is not on the efficient frontier. To impose the mean restriction on the pricing kernel, we include a moment condition for the one-month T-bill in the estimation.

A. Estimating the Pricing Kernel

Using the Taylor series approximation with time-varying coefficients, equation (8), the Euler equation (1) can be expressed as

$$E[(1 + \mathbf{R}_{t+1})((\delta_0' \mathbf{Z}_t)^2 - (\delta_1' \mathbf{Z}_t)^2 R_{m,t+1} + (\delta_2' \mathbf{Z}_t)^2 R_{m,t+1}^2 - (\delta_3' \mathbf{Z}_t)^2 R_{m,t+1}^3) | \mathbf{Z}_t] = \mathbf{1}_N. \tag{9}$$

We collect the vector of errors

$$\mathbf{v}_{t+1} = (1 + \mathbf{R}_{t+1})((\mathbf{Z}_t \boldsymbol{\delta}_0)^2 - (\mathbf{Z}_t \boldsymbol{\delta}_1)^2 R_{m,t+1} + (\mathbf{Z}_t \boldsymbol{\delta}_2)^2 R_{m,t+1}^2 - (\mathbf{Z}_t \boldsymbol{\delta}_3)^2 R_{m,t+1}^3) - \mathbf{1}_N. \tag{10}$$

Equation (9) implies

$$E[\mathbf{v}_{t+1}|\mathbf{Z}_t] = 0, \tag{11}$$

which forms a set of moment conditions that can be utilized to test the asset pricing model via Hansen's (1982) generalized method of moments (GMM). Equation (11) implies the unconditional restriction $E[\mathbf{v}_{t+1} \otimes \mathbf{Z}_t] = 0$. The sample version of this condition is that

$$\mathbf{g}_T(\boldsymbol{\delta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{v}_{t+1} \otimes \mathbf{Z}_t' = \mathbf{0}_N. \tag{12}$$

T represents the number of time series observations and N the number of assets under consideration. Expression (12) is a system of $N \times K$ equations. The number of parameters in the model, p, is driven by the restrictions on equation (7). In the cubic case, $p = 4K$, whereas in the quadratic and linear cases, $p = 3K$ and $2K$, respectively.

Hansen (1982) shows that a test of model specification can be obtained by minimizing the quadratic form

$$J(\delta) = \mathbf{g}_T(\delta)'\mathbf{W}_T(\delta)\mathbf{g}_T(\delta), \tag{13}$$

where $\mathbf{W}_T$ is the GMM weighting matrix. Alternative approaches to GMM estimation are based on the specification of the weighting matrix. Hansen shows that the optimal weighting matrix is the covariance matrix of the moment conditions, $\mathbf{W}_T^* = [\mathbf{g}_T(\delta)\mathbf{g}_T'(\delta)]^{-1}$. Although the GMM estimates with respect to this matrix are efficient, several studies (e.g., Ferson and Foerster (1994)) suggest that the method may have poor finite sample properties. Furthermore, as pointed out in Chapman (1997), since the weighting matrix is the inverse of the second moment matrix of the pricing errors, a small J-statistic can be obtained through estimating a pricing kernel with highly volatile pricing errors. Thus, using the standard GMM estimator in an Euler equation test may result in acceptance of a pricing kernel due not to improved pricing ability, but instead due to the addition of noise to the pricing kernel.

Hansen and Jagannathan (1997) pursue a different approach. Rather than attempting to minimize the pricing errors weighted by their covariance matrix, the authors investigate the size of the correction to a model's pricing kernel that is necessary for it to be consistent with a pricing kernel that prices the assets. The solution to this problem uses the same criterion function as the standard GMM estimator, equation (13), but specifies the weighting matrix as the second moment of instrument-scaled returns:

$$W^{HJ} = E[(\mathbf{R}_{t+1} \otimes \mathbf{Z}_t)(\mathbf{R}_{t+1} \otimes \mathbf{Z}_t)'\,]. \tag{14}$$

We follow Jagannathan and Wang (1996) and Chapman (1997) in implementing this approach. The distribution of J^{HJ}, the resulting test statistic, is derived in Jagannathan and Wang and is used as a test of model specification.

There are several advantages to using the Hansen–Jagannathan estimator rather than the standard GMM estimator. First, the Hansen–Jagannathan approach provides a statistic that can be used to compare nonnested models. This statistic is termed the Hansen–Jagannathan distance measure and is given by the square root of the criterion function equation (13) using the Hansen–Jagannathan weighting matrix, equation (14). This distance measure is equivalent to $||\tilde{p}||$, where $\tilde{p}$ is the correction to the proxy stochastic discount factor necessary to make it consistent with the data. Since the distance measure is formed on a weighting matrix that is invariant across all models tested, it can be used to directly compare the performance not only of nested models, but nonnested models as well.

A second advantage to the Hansen–Jagannathan approach is that it largely avoids the pitfall of favoring pricing models that produce volatile pricing errors. The Hansen–Jagannathan criterion is a function of the inverse of the second moment matrix of returns rather than the inverse of the second moment matrix of pricing errors. Consequently, the Hansen–Jagannathan distance will fall only if the least-square distance to an admissible pricing kernel is reduced, and not if the proxy pricing kernel generates volatile pricing errors. Thus, the distance rewards models exclusively for improving pricing and not for adding noise.

One caveat is in order. The distribution of the Hansen–Jagannathan test statistic is a function of the optimal GMM weighting matrix. Consequently, when testing the *significance* of the Hansen–Jagannathan distance, one may find a high p-value because the parameters imply a "small" optimal GMM weighting matrix; that is, a weighting matrix characterized by highly volatile pricing errors. One potential safeguard against failing to reject a model due simply to noise in the pricing kernel is to analyze the significance of the parameter estimates. Whereas the distribution of the distance measure is rewarded for a small GMM weighting matrix, the distribution of the parameter estimates is *penalized* by a small GMM weighting matrix. That is, although a model may be accepted due to volatile pricing errors, the volatility will tend to reduce the significance of the parameter estimates. Consequently, we perform Wald tests to assess the significance of adding each marginal term in the pricing kernel. These tests provide some surety not only that a pricing kernel is not rewarded simply for being noisy, but also provides evidence as to the importance of adding polynomial terms, potentially alleviating concerns about overfitting.

A final advantage to the Hansen–Jagannathan distance measure is that the results may be more robust than in standard GMM estimates (Cochrane (2001)). Since the weighting matrix is not a function of the parameters, the results should be more stable. Despite this advantage, Ahn and Gadarowski (1999) suggest that the size of the test statistic is poor in finite samples; the distance measure rejects correctly specified models too often. These results suggest the possibility that using the Hansen–Jagannathan estimator rather

than the standard GMM estimator may trade size for power. To gauge the possible impact of this trade-off, we also estimate the models using the iterated GMM estimator of Hansen, Heaton, and Yaron (1996). Ferson and Foerster (1994) show that the iterated GMM estimator has superior finite sample properties relative to the standard GMM estimator.[5]

B. Measurement of the Market Portfolio

A principal difficulty in estimating asset pricing relationships based on the portfolio of aggregate wealth is mismeasurement of the market portfolio, as noted in Roll (1977). Stambaugh (1982) addresses this issue by examining many different market indices and finds that they produce similar inferences about the CAPM, even when common stocks represent only 10 percent of the index's value. However, Stambaugh does not investigate the impact of including a measure of human capital, as suggested in Mayers (1972). Recent studies, notably Jagannathan and Wang (1996) and Campbell (1996), suggest that human capital is an important determinant of the cross section of expected returns. Jagannathan and Wang note that dividend income represents only three percent of personal income in the United States over the period 1959 to 1992, whereas salary and wages represent 63 percent of personal income. Further, Diaz-Gimenez et al. (1992) show that approximately two-thirds of nongovernment tangible assets are owned by the household sector and only one-third of these assets is owned by the corporate sector. Of the corporate-owned assets, only one-third are financed by equity. This evidence suggests that equity may represent as little as one-ninth of aggregate wealth, a small proportion of total wealth relative to human capital.

There are complications in attempting to incorporate human capital in the wealth portfolio proxy. Mayers (1972) explicitly treats human capital as different from financial capital because it is not traded. However, Jagannathan and Wang (1996) argue that human capital can be more straightforwardly incorporated into aggregate wealth. The authors note that part of human capital is in fact traded or hedged in the form of home mortgages, consumer loans, life insurance, unemployment insurance, and medical insurance. Consequently, the authors suggest that the following representation is an appropriate first approximation to incorporating human capital into the portfolio of aggregate wealth:

$$R_{W,t+1} = \theta_0 + \theta_1 R_{m,t+1} + \theta_2 R_{l,t+1}, \tag{15}$$

where $R_{W,t+1}$ represents the return on aggregate wealth and $R_{l,t+1}$ represents the return on human capital.[6] It is important, however, to note that since only a portion of labor income is securitized that equation (15) represents an abstraction from the more explicit approach of Mayers.

[5] These results are untabulated, but are available from the author on request.

[6] See Jagannathan and Wang (1996) for a more complete discussion of the assumptions necessary for equation (15) to hold.

As in Jagannathan and Wang, we define the return on human capital as a two-month moving average of the growth rate in labor income:

$$R_{l,t+1} = \frac{L_t + L_{t-1}}{L_{t-1} + L_{t-2}}, \tag{16}$$

where L_t denotes the difference between total personal income and dividend income at time t. The return on human capital is a function of lagged labor income since the data become available with a one-month delay. Jagannathan and Wang use this two-month moving average in an attempt to minimize the impact of measurement errors.

To implement this method, we redefine the pricing kernel in equation (8) as follows:

$$m_{t+1} = \mathbf{Z}_t \delta_0 + \sum_{n=1}^{3} I_n [(\mathbf{Z}_t \delta_{n,vw})^2 R^n_{vw,t+1} + (\mathbf{Z}_t \delta_{n,lbr})^2 R^{n-1}_{l,t+1}], \tag{17}$$

where $R_{vw,t+1}$ represents the return on the value-weighted equity portfolio, $R_{l,t+1}$ represents the growth rate in labor income, and

$$I_n = \begin{cases} -1 & n = 1,3 \\ 1 & n = 2 \end{cases}. \tag{18}$$

We assume that the cross-products in higher order terms of the return on the wealth portfolio are zero. When cross-products are included in the estimation, the qualitative conclusions of the paper do not change and the performance of the nonlinear models improves.

C. Data and Estimation Details

Many sets of assets have been used in the empirical asset pricing literature for tests of candidate asset pricing models. In our main specification tests, we utilize the returns on 20 industry-sorted portfolios, where the industry definitions follow the two-digit SIC codes used in Moskowitz and Grinblatt (1999) and are described in Table I. As shown by King (1966), industry groupings proxy the investment opportunity set well; these groupings maximize intragroup and minimize intergroup correlations.

The choice of the instrument set Z_t is motivated by two considerations. First, the instruments should be a set of variables that are able to predict asset returns. Second, the choice of instruments should be parsimonious due to power considerations in GMM estimation (Tauchen (1986)). Consequently, we consider a set of instruments, $Z_t = \{1, r_{mt}, dy_t, ys_t, tb_t\}$, where 1 denotes a vector of ones, r_{mt} is the excess return on the CRSP value-weighted index at time t, dy_t is the dividend yield on the CRSP value-weighted index at time t, ys_t is the yield on the three-month Treasury bill in excess of the yield on

Table I

Summary Statistics: Industry Portfolios

Table I presents monthly means and standard deviations of the returns on 20 industry-sorted portfolios as in Moskowitz and Grinblatt (1999). Portfolios are equally weighted and formed on the basis of two-digit SIC codes. The data cover the period July 31, 1963, through December 31, 1995.

Panel A: Mean Returns

Industry	Mean Return	Industry	Mean Return
Mining	0.0128	Electrical Equipment	0.0148
Food & Beverage	0.0137	Transport Equipment	0.0138
Textile & Apparel	0.0112	Manufacturing	0.0138
Paper Products	0.0132	Railroads	0.0140
Chemical	0.0139	Other Transportation	0.0132
Petroleum	0.0132	Utilities	0.0097
Construction	0.0125	Department Stores	0.0113
Primary Metals	0.0112	Other Retail	0.0133
Fabricated Metals	0.0144	Finance, Real Estate	0.0110
Machinery	0.0130	Other	0.0131

Panel B: Standard Deviations

Industry	Standard Dev.	Industry	Standard Dev.
Mining	0.0674	Electrical Equipment	0.0740
Food & Beverage	0.0495	Transport Equipment	0.0665
Textile & Apparel	0.0693	Manufacturing	0.0681
Paper Products	0.0575	Railroads	0.0571
Chemical	0.0525	Other Transportation	0.0671
Petroleum	0.0560	Utilities	0.0368
Construction	0.0619	Department Stores	0.0674
Primary Metals	0.0593	Other Retail	0.0626
Fabricated Metals	0.0623	Finance, Real Estate	0.0575
Machinery	0.0647	Other	0.0656

the one-month Treasury bill at time t, and tb_t is the return on a Treasury bill closest to one month to maturity at time t. These variables have been shown to be predictors of future returns in various studies. The value-weighted CRSP index is examined in Harvey (1989) and Ferson and Harvey (1991). Fama and French (1988, 1989) investigate the predictive power of the dividend yield. Campbell (1987) shows that term premia in Treasury bill returns can predict stock returns. Finally, Fama and Schwert (1977), Ferson (1989), and Shanken (1990) examine the T-bill return.

The data used to compute the industry portfolio returns, value-weighted index return, dividend yield, yield spread, and risk-free return are obtained from CRSP. The data used to compute the labor return series is obtained from the NIPA data available on DataStream. Labor income at time t is

Table II

Summary Statistics: Instruments

Table II displays a summary of the predictive power of the instrumental variables used in the paper, $\mathbf{Z_t} = \{r_{m,t}, dy_t, ys_t, tb_t\}$, where $r_{m,t}$ represents the return on the value-weighted CRSP index, dy_t is the dividend yield on the value-weighted CRSP index, ys_t is the excess yield on the Treasury bill closest to three months to maturity over the Treasury bill closest to one month to maturity, and tb_t is the return on the Treasury bill closest to one month to maturity. The data cover the period July 30, 1963, through December 31, 1995. The predictive power of the instruments is assessed by the linear projection

$$R_{i,t+1} = d_0 + \mathbf{d}\mathbf{Z}_t + u_{t+1}.$$

The column labeled χ_4^2 presents Newey and West (1987a) Wald tests of the hypothesis

$$H_0 : \mathbf{d} = \mathbf{0}$$

with p-values in parentheses. The statistics are computed using the Newey and West (1987b) heteroskedasticity and autocorrelation-consistent covariance matrix.

Industry	χ_4^2	Industry	χ_4^2
Mining	13.132	Electrical Equipment	39.929
	(0.011)		(0.000)
Food & Beverage	24.334	Transport Equipment	39.333
	(0.000)		(0.000)
Textile & Apparel	32.750	Manufacturing	44.992
	(0.000)		(0.000)
Paper Products	22.826	Railroads	11.156
	(0.000)		(0.025)
Chemical	24.012	Other Transportation	19.671
	(0.000)		(0.001)
Petroleum	1.510	Utilities	8.522
	(0.825)		(0.074)
Construction	25.663	Department Stores	19.031
	(0.000)		(0.001)
Primary Metals	22.071	Other Retail	32.012
	(0.000)		(0.000)
Fabricated Metals	39.437	Finance, Real Estate	28.469
	(0.000)		(0.000)
Machinery	33.585	Other	37.818
	(0.000)		(0.000)

computed as the per capita difference between total personal income and dividend income. The data cover the period July 31, 1963, through December 31, 1995, totalling 390 observations.

Sample statistics for the returns on the 20 industry portfolios and the components of the market proxy are presented in Table I. The average returns over the sample period for the payoffs range from 101 basis points per month for the utilities industry to 151 basis points per month for the fabricated metals industry. In Table II, we present a summary of the predictive

power of the instrumental variables for the payoffs. We project the payoffs onto the instruments:

$$R_{i,t+1} = \mathbf{b}'\mathbf{Z}_t + u_{t+1}.$$

The table contains statistics for a Wald test of the null hypothesis that the instruments have no predictive power for the payoffs. Consistent with the results of previous studies, the table shows that the information variables serve as good instruments for the payoffs.

III. Results

A. Model Specification Tests

In this section, we discuss tests of the Euler equation (1) when the pricing kernel is expressed with quadratic time-varying coefficients, as in equation (8). We analyze the cubic pricing kernel and also the linear and quadratic pricing kernels that are nested in the cubic case. Results are presented with and without human capital as a component of the return on aggregate wealth.

Table III presents results of specification tests when the measure of aggregate wealth does not include human capital. The table presents average values of the coefficients $d_{n,t}$, $n = 1,2,3$ corresponding to the nth order of the return on the market portfolio. The table also presents the Hansen–Jagannathan distance measure and p-values for the Hansen–Jagannathan test of model specification. The first row of each panel, labeled "Coefficients," presents the value of the estimated coefficient evaluated at the mean of the instruments. As shown in the table, with this specification, the linear, quadratic, and cubic pricing kernels are all rejected at the five percent significance level for this data set. The distance measures and p-values for the tests of significance of the coefficients suggest marginal improvement from moving from a linear specification to a nonlinear specification. The quadratic pricing kernel reduces the distance measure from 0.735 to 0.709, a drop of 3.5 percent relative to the linear pricing kernel. The test of the significance of the d_2 terms suggest that this improvement is marginally significant (p-value 0.027), indicating that incorporation of the quadratic term in the pricing kernel improves the fit of the model. These results are consistent with the findings of Harvey and Siddique (2000). However, the addition of a cubic term does not materially improve the performance of the pricing kernel.

We next analyze the impact of incorporating a measure of human capital in the return on aggregate wealth. These results are displayed in Table IV. The outcome of the specification tests are markedly different from those in Table III. All three of the pricing kernels improve substantially relative to the case in which human capital is not included in the measure of aggregate wealth. The distance measure implied by the linear pricing kernel falls to

Table III

Specification Tests: Polynomial Pricing Kernels with Human Capital Excluded

Table III presents results of GMM tests of the Euler equation condition,

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0}$$

using the polynomial pricing kernels, m_{t+1} nested in equation (7). The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix $E[(R_{t+1} \otimes \mathbf{Z}_t)(R_{t+1} \otimes \mathbf{Z}_t)']$. The columns present the coefficients of the pricing kernel evaluated at the means of the instruments. The coefficients are modeled as

$$d_n = I_n(\delta_n'\mathbf{Z}_t)^2 \qquad I_n = \begin{cases} -1 & n = 2,4 \\ 1 & n = 3. \end{cases}$$

P-values for Wald tests of the joint significance of the coefficients are presented in parentheses. The final column presents the Hansen–Jagannathan distance measure with *p*-values for the test of model specification in parentheses. The set of returns used in estimation are those of 20 industry-sorted portfolios augmented by the return on a one-month Treasury bill, covering the period July 31, 1963, through December 31, 1995, and the measure of aggregate wealth does not include human capital.

	$d(\overline{Z})_{0t}$	$d(\overline{Z})_{1t}$	$d(\overline{Z})_{2t}$	$d(\overline{Z})_{3t}$	Dist
		Panel A: Linear			
Coefficient	1.088	−3.936			0.735
P-value	(0.000)	(0.000)			(0.000)
		Panel B: Quadratic			
Coefficient	0.950	−4.702	77.633		0.709
P-value	(0.000)	(0.000)	(0.027)		(0.001)
		Panel C: Cubic			
Coefficient	0.941	−4.545	77.655	−4.647	0.703
P-value	(0.000)	(0.000)	(0.053)	(0.654)	(0.000)

0.719, a decline of 2.2 percent relative to the linear kernel omitting human capital. This result is consistent with the findings of Jagannathan and Wang (1996), who find that incorporating human capital improves the performance of the conditional CAPM. However, the linear pricing kernel is rejected at the five percent significance level (p-value 0.019).

Considerable further improvement is observed by moving from a linear to a nonlinear specification. The results in Panel B of Table IV indicate that a quadratic specification of the pricing kernel results in an additional decrease in the distance measure of 12.5 percent relative to the linear kernel with human capital. This pricing kernel cannot be rejected at the 10 percent

Table IV

Specification Tests: Polynomial Pricing Kernels with Human Capital Included

Table IV presents results of GMM tests of the Euler equation condition,

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0}$$

using the polynomial pricing kernels, m_{t+1} nested in equation (7). The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix $E[(R_{t+1} \otimes \mathbf{Z}_t)(R_{t+1} \otimes \mathbf{Z}_t)']$. The columns present the coefficients of the pricing kernel evaluated at the means of the instruments. The coefficients are modeled as

$$d_n = I_n(\delta_n'\mathbf{Z}_t)^2 \qquad I_n = \begin{cases} -1 & n = 2,4 \\ 1 & n = 3 \end{cases}.$$

P-values for Wald tests of the joint significance of the coefficients are presented in parentheses. The final column presents the Hansen–Jagannathan distance measure with *p*-values for the test of model specification in parentheses. The set of returns used in estimation are those of 20 industry-sorted portfolios covering the period July 31, 1963, through December 31, 1995, augmented by the return on a 30-day Treasury bill. The measure of aggregate wealth includes human capital.

	$d(\overline{Z})_{0t}$	$d(\overline{Z})_{1vw}$	$d(\overline{Z})_{1l}$	$d(\overline{Z})_{2vw}$	$d(\overline{Z})_{2l}$	$d(\overline{Z})_{3vw}$	$d(\overline{Z})_{3l}$	Dist
				Panel A: Linear				
Coefficient	1.197	−3.674	−1.020					0.719
P-value	(0.000)	(0.000)	(0.000)					(0.000)
				Panel B: Quadratic				
Coefficient	0.725	−3.447	−12.326	46.043	9,839.447			0.629
P-value	(0.000)	(0.000)	(0.000)	(0.008)	(0.000)			(0.100)
				Panel C: Cubic				
Coefficient	0.392	−2.079	−0.013	74.010	17,374.873	−0.084	−72,319.964	0.578
P-value	(0.000)	(0.003)	(0.001)	(0.002)	(0.000)	(0.945)	(0.000)	(0.229)

significance level (*p*-value 0.100). Incorporating the quadratic return on wealth term contributes significantly to the fit of the pricing kernel, as indicated by the test of the significance of the d_2 terms (*p*-values 0.002 and 0.000). Thus, incorporating a nonlinear function of the return on human capital appears to have a dramatic impact on the fit of the pricing kernel.

The performance of the pricing kernel is further enhanced by incorporating the cubic return on wealth, as shown in Panel C. The distance measure falls to 0.578, a decline of 8.1 percent relative to the quadratic pricing kernel, and a decrease of 21.4 percent relative to the conditional CAPM estimated in Panel A of Table III. Moreover, the specification test cannot reject the cubic pricing kernel at the 10 percent significance level (*p*-value 0.229),

and the d_{3l} term contributes significantly to the improvement in the distance measure (p-value 0.000).[7] These results suggest that, by allowing for preference restrictions implied by decreasing absolute risk aversion and decreasing absolute prudence, that the performance of a pricing kernel grounded in preference theory can capture cross-sectional variation in returns. The results of Tables III and IV suggest that incorporating only nonlinear functions of the return on the value-weighted index or a linear function of the return on labor is insufficient to generate an admissible pricing kernel. However, by utilizing both the return on labor and the nonlinearities implied by the series expansion, we are able to generate an admissible pricing kernel.[8]

B. Multifactor Alternatives

As noted earlier in the paper, multifactor models of asset prices have been more successful in pricing the cross section of equities than have single-factor models. However, multifactor models provide the researcher with considerable freedom since the models give little guidance for the choice of factors. In contrast, the pricing kernel in this paper explicitly defines the relevant factor for pricing, the portfolio of aggregate wealth. Further, preference theory imposes restrictions on the signs of the coefficients on each term in the pricing kernel. In this section, we gauge the ability of the polynomial pricing kernel to price the cross section of industry portfolios relative to a popular multifactor model, the Fama and French (1993) three-factor model. This model is not nested in the polynomial pricing kernel, but the performance of all of the models can be compared using their Hansen–Jagannathan distance measures, as discussed previously.

Fama and French (1992) provide evidence that firms' market capitalization and market-to-book ratios appear to outperform the CAPM beta in capturing cross-sectional variation in returns. Fama and French (1993), noting this evidence, propose the following model for returns

$$E[r_{i,t+1}] = \beta_{MRP}E[r_{MRP,t+1}] + \beta_{SMB}E[r_{SMB,t+1}] + \beta_{HML}E[r_{HML,t+1}]. \quad (19)$$

In this model, $r_{MRP,t+1}$ represents the excess return on the market portfolio over the risk free rate, $r_{SMB,t+1}$ represents the excess return on a portfolio of small capitalization stocks over large capitalization stocks, and $r_{HML,t+1}$

[7] The magnitude of the average coefficient is quite large (-7.232×10^4). This magnitude is driven by the size of the higher orders of the return on labor income. The mean of the monthly return on labor income is 0.0057, whereas the mean of the monthly return on labor income cubed is 4.233×10^{-7}. Thus, the coefficient on the cubic term is quite large to reflect the scaling of the return on labor income cubed.

[8] In untabulated results, we repeat the estimation of the pricing kernels using the iterated GMM estimator in Hansen et al. (1996). The results of this estimation mirror the Hansen–Jagannathan distance estimates. Consequently, both sets of tests suggest that nonlinear pricing kernels with reasonable economic restrictions perform well in pricing the cross section of industry-sorted returns.

represents the excess return on a portfolio of high market-to-book stocks over low market-to-book stocks.[9] The authors suggest that the returns to the portfolios *SMB* and *HML* represent hedge portfolios in the sense of Merton (1973). In later work (Fama and French (1995, 1996)), the authors suggest that the size and book-to-market factors may capture some systematic distress factor. The model in expression (19) can be expressed in stochastic discount factor form. As in Jagannathan and Wang (1996), note that equation (19) implies

$$m_{t+1}^{FF} = \delta_0 + \delta_{MRP} r_{MRP,t+1} + \delta_{SMB} r_{SMB,t+1} + \delta_{HML} r_{HML,t+1}. \tag{20}$$

In this setting, the coefficients δ_n capture the prices of factor n risk. We allow for time variation in these coefficients by assuming a linear specification in the instruments.[10]

Results for the estimation of the Fama–French model are presented in Panel A of Table V. The results suggest that the pricing kernel implied by the model fares poorly in describing the cross section of industry returns. The distance measure for the model of 0.714 (p-value 0.000) is comparable to the distance measure for the linear pricing kernel incorporating human capital. Further, the distance measure of the Fama–French model is substantially higher than that of either the quadratic or the cubic pricing kernel with human capital. Thus, the results suggest that, although preference restrictions are imposed on the nonlinear pricing kernels and the kernels are specified as functions of the return on aggregate wealth, the nonlinear kernels outperform the Fama–French model in pricing the cross section of industry returns.

To further investigate the ability of the Fama–French factors to price the cross section of equity returns compared to the polynomial pricing kernels, we estimate the polynomial models augmented by the *SMB* and *HML* factors of the Fama–French model. Results of these tests are also presented in Table V. In the case of the quadratic pricing kernel, the distance measure falls from 0.629 to 0.588 with the Fama–French factors included. The p-value of the specification test for the quadratic kernel augmented by the Fama–French factors falls to 0.040, indicating that the loss of degrees of freedom resulting from the incorporation of the Fama–French factors more than offsets any improvement in the fit of the pricing kernel. However, the *SMB* factor continues to be marginally significant, with a p-value of 0.005. In contrast, when the Fama–French factors are included in the cubic pricing kernel, the model cannot be rejected (p-value 0.140), and neither the *SMB* nor the *HML* coefficients are significantly different than zero. These results suggest that,

[9] We would like to thank Eugene Fama for providing these data.

[10] We do not investigate a specification for the factor coefficients that is quadratic in the instruments as in equation (8) because doing so imposes restrictions on the signs of the coefficients. The coefficients of the Fama–French model are not restricted in sign; consequently, imposing sign restrictions would unfairly penalize the model.

Table V

Specification Tests: Fama–French Pricing Kernel

Table V presents results of GMM estimation of the Euler equation restriction

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0}$$

using the pricing kernel, m_{t+1} implied by the Fama and French (1993) three-factor model, as in equation (20). The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix. *P*-values for Wald tests of the joint significance of the coefficients are presented in parentheses. The final column presents the Hansen–Jagannathan distance measure, with *p*-values for the test of model specification in parentheses. In Panel B, the Fama–French pricing kernel is augmented by a quadratic function of the return on wealth, and in Panel B, the pricing kernel is augmented by both a quadratic and a cubic function of the return on wealth. The set of returns used in estimation are those of 20 industry-sorted portfolios covering the period July 31, 1963, through December 31, 1995, augmented by the return on a one-month Treasury bill.

	$d(\overline{Z})_{0t}$	$d(\overline{Z})_{mrp,t}$	$d(\overline{Z})_{smb,t}$	$d(\overline{Z})_{hml,t}$	$d(\overline{Z})_{1vw}$	$d(\overline{Z})_{1l}$	$d(\overline{Z})_{2vw}$	$d(\overline{Z})_{2l}$	$d(\overline{Z})_{3vw}$	$d(\overline{Z})_{3l}$	Dist
					Panel A: Fama–French Factors Only						
Coefficient	1.147	−5.299	−5.929	−1.977							0.632
P-value	(0.000)	(0.045)	(0.000)	(0.069)							(0.008)
					Panel B: Quadratic Augmented by Fama–French Factors						
Coefficient	1.252		1.044	−0.970	−3.298	−5.665	57.803	205.206			0.588
P-value	(0.000)		(0.005)	(0.525)	(0.000)	(0.000)	(0.000)	(0.000)			(0.040)
					Panel C: Cubic Augmented by Fama–French Factors						
Coefficient	0.404		0.735	1.766	−2.167	−1.146	87.818	16,510.940	−0.486	−62,684.668	0.555
P-value	(0.000)		(0.621)	(0.941)	(0.005)	(0.000)	(0.020)	(0.000)	(0.999)	(0.000)	(0.140)

in the cross section of industry-sorted portfolios, the cubic pricing kernel captures much of the variation in returns that is explained by the Fama–French factors. This result is particularly interesting since the signs of the coefficients are restricted by preference theory, and the factor obtains from first principles.

C. Comparison with Power Utility

In the previous sections, we investigate models that are restricted to be consistent with established assumptions governing agents' preferences. However, the polynomial pricing kernels that are investigated in this paper are divorced from the more rigorous restrictions imposed by assuming a specific utility function. Caballé and Pomansky (1996) show that all HARA utility functions display standard risk aversion, consistent with the cubic pricing kernel. In this section, we assume that agents' preferences are characterized by power utility, and investigate the ability of the resulting pricing kernel to explain cross-sectional variation in returns, as in Brown and Gibbons (1985). In doing so, we examine the trade-offs between this parsimonious specification of the pricing kernel and the more general specification implied by the Taylor series expansion.

Brown and Gibbons (1985) investigate a static setting in which a representative agent exhibits power utility. In this case, the pricing kernel can be expressed as

$$m_{t+1} = a_0(1 + R_{W,t+1})^{-a_1}, \tag{21}$$

where a_1 is the representative agent's relative risk aversion. One issue in the implementation of equation (21) is the incorporation of human capital. In estimation, the optimization of the Euler equation is ill behaved when we allow $R_{W,t+1}$ to be an unrestricted linear function of the return on labor and the return on the value-weighted index, as in equation (15). Consequently, similar to Campbell (1996), we assume that the return on wealth can be expressed as

$$R_{W,t+1} = a_2 R_{m,t+1} + (1 - a_2) R_{l,t+1}. \tag{22}$$

Although this formulation imposes additional restrictions on the relationship between returns, the value-weighted portfolio, and the labor return, it offers a straightforward way to incorporate human capital in the pricing kernel expression (21).

Results of this estimation are presented in Table VI. As shown in the table, the pricing kernel implied by power utility is rejected via the Hansen–Jagannathan distance measure both with and without human capital. These results are qualitatively similar to those of Hansen and Jagannathan (1997), who use the distance measure to evaluate the power utility pricing kernel defined over aggregate consumption. Both forms of the pricing kernel per-

Table VI

Specification Tests: Power Utility Pricing Kernel

Table VI presents results of GMM estimation of the Euler equation restriction

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0}$$

using the pricing kernel, m_{t+1} implied by power utility, as in equation (21). The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix. *P*-values for Wald tests of the significance of the coefficients are presented in parentheses. The final column presents the Hansen–Jagannathan distance measure, with *p*-values for the test of model specification in parentheses. The set of returns used in estimation are those of 20 industry-sorted portfolios covering the period July 31, 1963, through December 31, 1995, augmented by the return on a one-month Treasury bill.

Coefficient:	a_0	a_1	a_2	Dist
	Panel A: Human Capital Excluded			
	1.014	3.963		0.740
P-value	(0.000)	(0.000)		(0.000)
	Panel B: Human Capital Included			
	1.036	3.671	0.558	0.740
P-value	(0.000)	(0.743)	(0.733)	(0.000)

form worse than the linear pricing kernel. Further, the incorporation of human capital in the pricing kernel does not appear to materially improve the performance of the power utility pricing kernel and seems to contribute noise to the parameter estimates. The results suggest that although power utility is consistent with the preference restrictions imposed on the cubic pricing kernel, the parsimony provided by a specific utility function comes at a large cost in terms of the fit of the model.[11]

Some intuition for the source of improvement in the performance of the polynomial pricing kernel relative to the power utility specification is provided by decomposing the distance measure as discussed in Hansen and Jagannathan (1997). Recall from Section II.A that the distance measure can be expressed as $||\tilde{p}||$, where $\tilde{p}$ is the adjustment to the model pricing kernel necessary to reduce the distance to an admissible pricing kernel to zero. Using the definition of the norm of p, Hansen and Jagannathan note that

$$||\tilde{p}|| = \sqrt{E[\tilde{p}]^2 + \mathrm{Var}[\tilde{p}]}. \tag{23}$$

[11] The comparison between the power utility model and the polynomial models is not entirely fair because the coefficients of the polynomial models are allowed to vary over time. In contrast, the time preference and risk aversion parameters of the power utility model are fixed. The focus of this paper is on conditional pricing models; however, in order to provide a fair comparison, we estimate the polynomial models with fixed coefficients as well. The resulting distance measure for the cubic pricing kernel is 0.634, compared to 0.740 for the power utility kernel, a difference of approximately 14 percent.

Table VII

Decomposition of Distance Measure

Table VII presents a decomposition of the Hansen–Jagannathan distance measure, $||\tilde{p}|| = \sqrt{E[\tilde{p}]^2 + \mathrm{Var}[\tilde{p}]}$. The variable p represents the adjustment to the model pricing kernel needed to make it admissible. The column labeled "Mean ($\tilde{p}$)" represents the average of the estimated $\tilde{p}$, the column labeled "Std. ($\tilde{p}$)" represents its standard deviation, and the column labeled "Distance" represents the Hansen–Jagannathan distance. The rows labeled "Linear: No HC," "Quadratic: No HC," and "Cubic: No HC" represent the decomposition for the polynomial pricing kernels, omitting human capital. The rows labeled "Linear," "Quadratic," and "Cubic" represent the decomposition for the polynomial pricing kernels, including human capital. In both sets of decompositions, the coefficients of the pricing kernel are modeled as quadratic in the instruments and sign restrictions are imposed. The final row, labeled "Power," represents the results for the power utility pricing kernel without human capital.

Model	Mean ($\tilde{p}$)	Std. ($\tilde{p}$)	Distance
Linear: No HC	0.0003	0.7346	0.7346
Quadratic: No HC	0.0004	0.7093	0.7093
Cubic: No HC	0.0005	0.7035	0.7035
Linear	0.0003	0.7191	0.7191
Quadratic	0.0021	0.6292	0.6292
Cubic	0.0017	0.5779	0.5779
Power	0.0002	0.7836	0.7836

Thus, the distance measure has two components; it is a function of the expected deviation from some admissible pricing kernel and the variance of that deviation. In this sense, the Hansen–Jagannathan distance captures a sense of both the average and the variability of a proxy pricing kernel's pricing errors.

Table VII presents estimates of $E[\tilde{p}]$ and $Std[\tilde{p}]$ for the linear, quadratic, and cubic pricing kernels, with and without human capital. The table also presents these estimates for the power utility pricing kernel. As in Hansen and Jagannathan (1997), most of the distance measure results from $Std[\tilde{p}]$. That is, a proxy pricing kernel with a small distance measure tends to reduce the volatility of the adjustment necessary to make the proxy admissible. The power utility pricing kernel has the lowest average value for $\tilde{p}$, suggesting that, on average, it is the pricing kernel that requires the least adjustment to be admissible. The linear pricing kernels require the next lowest mean adjustment and the quadratic and cubic pricing kernels with human capital require mean adjustments that are considerably larger than those of the remaining pricing kernels.

However, the variability of the adjustment required to make the linear or power pricing kernel valid dwarfs the mean term, rendering the pricing kernels inadmissible. As shown in the table, virtually all of the distance comes from this variability. In contrast to the linear and power pricing kernels, the quadratic and cubic pricing kernels with human capital require much smaller standard deviation adjustments to render the kernels admissible. This is the source of the improvement in the nonlinear pricing kernels compared to standard parametric pricing kernels represented by the linear case and the

power utility case. All of the pricing kernels fare relatively well in capturing the mean of the pricing kernel, but, although the mean difference is close to zero, the linear and power kernels deviate highly from the set of admissible pricing kernels. In contrast, the quadratic and cubic pricing kernels are close not only on average to the set of admissible kernels, but the variability of their deviations is much lower.

D. Properties of the Estimated Pricing Kernels

Thus far, we have imposed conditions on the coefficients of the pricing kernel that guarantee that the functions behave *locally* in a manner consistent with preference theory. We may also impose stronger conditions on preferences that restrict the *global* behavior of the pricing kernel. In a setting with standard preferences and static prices of risk, the pricing kernel can be interpreted as a scaled marginal utility. Consequently, under these assumptions, the pricing kernel should be positive in order to be consistent with positive marginal utility (and the no arbitrage condition), and decreasing in order to be consistent with decreasing absolute risk aversion. Neither of these conditions have been imposed on the pricing kernels that we have estimated thus far.[12]

Figure 1 depicts plots of the polynomial pricing kernels that we estimate in Section III.A. The plot depicts the functional form of the pricing kernels when the coefficients of the pricing kernel are evaluated at the means of the instrumental variables. Figure 1a depicts the linear pricing kernel; the plot shows that this kernel is decreasing in both the return on the index and the return on labor, which is guaranteed by the restrictions on the signs of its coefficients. In contrast, the quadratic and cubic pricing kernels are not globally decreasing. Figures 1b and 1c show that, when the coefficients of the pricing kernels are fixed at the mean of the instruments, these pricing kernels may be increasing in both the return on labor and return on wealth. This plot suggests that the pricing kernels that we have estimated are not likely to be globally consistent with standard preferences.

How important are the restrictions imposed by standard preferences in terms of model fit? We address this question by imposing functional forms on the coefficients of the pricing kernel that guarantee nonnegativity and a nonpositive first derivative of the estimated pricing kernel. To ensure that the first condition holds, we estimate the models with the following restriction:

$$m_{t+1} \geq 0.$$

As noted in Hansen and Jagannathan (1991), this condition can easily be imposed on the pricing kernel in estimation. We follow Chen and Knez (1996) in our implementation of this restriction in GMM estimation. The second

[12] We would like to thank the referee for suggesting that we investigate this issue.

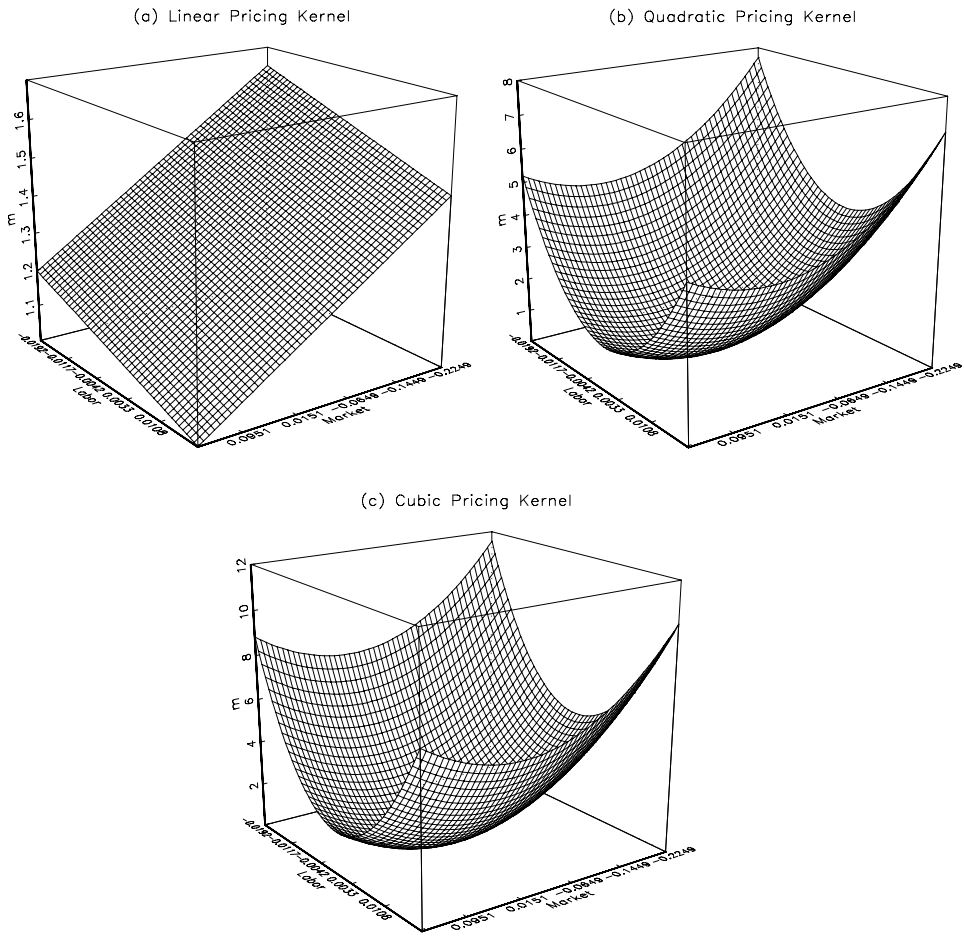

Figure 1. Estimated pricing kernels. Figure 1 depicts point estimates of the pricing kernels estimated without global restrictions. The point estimates are calculated at the mean of the instrumental variables and the support for the graphs is the observed range of the return on labor and the value-weighted index. The coefficients of the pricing kernels are estimated via GMM utilizing the Euler equation condition,

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0},$$

where m_{t+1} represents a polynomial pricing kernel. The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix $E[(R_{t+1} \otimes \mathbf{Z}_t)(R_{t+1} \otimes \mathbf{Z}_t)']$. The coefficients are modeled as

$$d_n = I_n(\delta_n'\mathbf{Z}_t)^2 \qquad I_n = \begin{cases} -1 & n = 2,4 \\ 1 & n = 3. \end{cases}$$

The set of returns used in estimation are those of 20 industry-sorted portfolios covering the period July 31, 1963, through December 31, 1995, augmented by the return on a 30-day Treasury bill. The measure of aggregate wealth includes human capital.

condition, $m'_{t+1} \leq 0$, is more difficult to implement. We enforce it by imposing the following functional form on the pricing kernel:

$$d_{3t} = \min\left[d_{3t}, -\frac{d_{1t} + 2d_{2t}R_{w,t+1}}{3R_{w,t+1}^2}\right]. \tag{24}$$

If this constraint binds, then, to ensure that the derivative is negative,

$$d_{1t} = \min[d_{1t}, -d_{2t}R_{w,t+1}]. \tag{25}$$

In the case of the quadratic pricing kernel, we impose the constraint

$$d_{1t} = \min[d_{1t}, -2d_{2t}R_{w,t+1}]. \tag{26}$$

The combination of these constraints ensures the negativity of the first derivative of the pricing kernel.

Results of this estimation are presented in Table VIII and suggest two conclusions. First, imposing restrictions that are consistent with standard preferences has a large cost in terms of fit. The Hansen–Jagannathan distance for the quadratic kernel rises to 0.668 and that of the cubic kernel to 0.645, and both models are rejected by the specification test.[13] However, the second conclusion implied by the table is that these nonlinear pricing kernels continue to outperform the linear single-factor pricing kernel and the Fama–French linear multifactor model. The restricted quadratic kernel reduces the distance measure by 7 percent relative to the linear model and by 6 percent relative to the Fama–French model. The restricted cubic kernel reduces the Hansen–Jagannathan distance by an additional 3.5 percent. Furthermore, the nonlinear terms continue to contribute significantly to the improvements in model fit.

To better gauge the impact of imposing these restrictions on the polynomial pricing kernels, we again plot the functional form of the pricing kernels in Figure 2. As in Figure 1, the coefficients are fixed at the means of the instrumental variables. The pricing kernels plotted in Figure 2 are markedly different from those plotted in Figure 1. In particular, both pricing kernels are very nearly linear over the range of the labor and value-weighted index return series. The quadratic pricing kernel exhibits mild curvature in the value-weighted index, but is linear and very nearly flat in the labor return. The cubic pricing kernel displays much more marked departures from nonlinearity when the labor return is extremely low but, like

[13] This result occurs primarily due to the imposition of the second condition, nonpositivity of the derivative of the pricing kernel. When the pricing kernel is restricted only to be positive, but not necessarily decreasing, the performance of the models is similar to that exhibited in Section III.A.

Table VIII

Specification Tests: Polynomial Pricing Kernels with Global Restrictions and Human Capital Included

Table VIII presents results of GMM tests of the Euler equation condition,

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0}$$

using the polynomial pricing kernels, m_{t+1} nested in equation (7). The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix $E[(R_{t+1} \otimes \mathbf{Z}_t)(R_{t+1} \otimes \mathbf{Z}_t)']$. The columns present the coefficients of the pricing kernel evaluated at the means of the instruments. The coefficients are modeled as

$$d_n = I_n(\delta_n'\mathbf{Z}_t)^2 \qquad I_n = \begin{cases} -1 & n = 2,4 \\ 1 & n = 3. \end{cases}$$

In addition to constraining the signs of the coefficients, the following constraints are placed on the pricing kernel:

$$m_{t+1} \geq 0 \qquad m'_{t+1} \leq 0.$$

P-values for Wald tests of the joint significance of the coefficients are presented in parentheses. The final column presents the Hansen–Jagannathan distance measure with *p*-values for the test of model specification in parentheses. The set of returns used in estimation are those of 20 industry-sorted portfolios covering the period July 31, 1963, through December 31, 1995, augmented by the return on a 30-day Treasury bill. The measure of aggregate wealth includes human capital.

	$d(\overline{Z})_{0t}$	$d(\overline{Z})_{1vw}$	$d(\overline{Z})_{1l}$	$d(\overline{Z})_{2vw}$	$d(\overline{Z})_{2l}$	$d(\overline{Z})_{3vw}$	$d(\overline{Z})_{3l}$	Dist
				Panel A: Linear				
Coefficient	1.343	−2.688	−0.001					0.698
P-value	(0.000)	(0.000)	(0.000)					(0.000)
				Panel B: Quadratic				
Coefficient	1.246	−2.411	−0.624	14.157	454.223			0.668
P-value	(0.000)	(0.000)	(0.000)	(0.016)	(0.000)			(0.000)
				Panel C: Cubic				
Coefficient	1.103	−3.024	−4.723	8.785	28,600.468	−31.276	−1.082*	0.645
P-value	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.044)	(0.000)	(0.000)

*($\times$ 10^6).

the quadratic pricing kernel, is close to linear over most of the labor return support. These plots suggest that by imposing conditions on the derivative of the pricing kernel, we suppress much of the nonlinearity that appears to be important in fitting the cross section of returns.

These results suggest that we are left with a trade-off. The standard economic paradigm suggests that the pricing kernel should be decreasing in its argument. Our results suggest that a nonlinear pricing kernel can be consistent with this restriction and outperform linear single- and multiple-

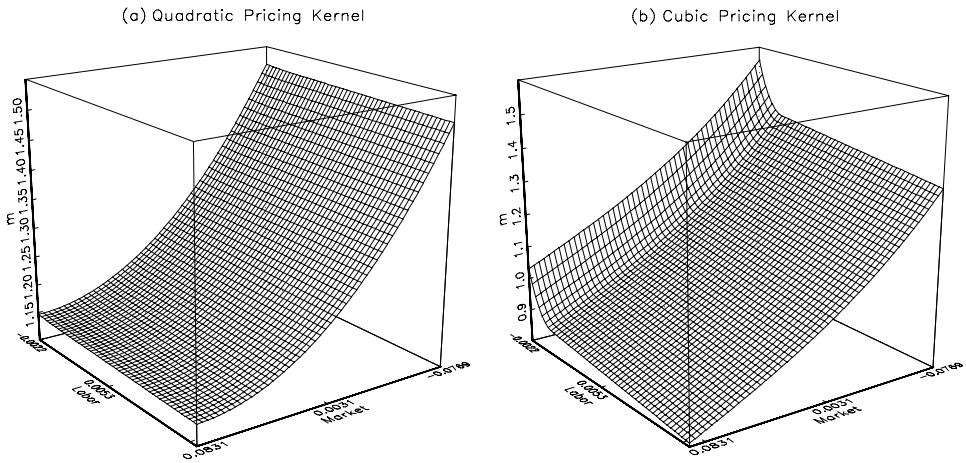

Figure 2. Estimated pricing kernels with global restrictions. Figure 2 depicts point estimates of the pricing kernels estimated with global restrictions. The point estimates are calculated at the mean of the instrumental variables and the support for the graphs is the observed range of the return on labor and the value-weighted index. The coefficients of the pricing kernels are estimated via GMM utilizing the Euler equation condition,

$$E[(1 + \mathbf{R}_{t+1})m_{t+1}|\mathbf{Z}_t] - \mathbf{1}_N = \mathbf{0}$$

where m_{t+1} represents a polynomial pricing kernel. The coefficients are estimated using the Hansen and Jagannathan (1997) weighting matrix $E[(R_{t+1} \otimes \mathbf{Z}_t)(R_{t+1} \otimes \mathbf{Z}_t)']$. The coefficients are modeled as

$$d_n = I_n(\delta_n'\mathbf{Z}_t)^2 \qquad I_n = \begin{cases} -1 & n = 2,4 \\ 1 & n = 3. \end{cases}$$

In addition to constraining the signs of the coefficients, the following constraints are placed on the pricing kernel:

$$m_{t+1} \geq 0 \qquad m'_{t+1} \leq 0.$$

The set of returns used in estimation are those of 20 industry-sorted portfolios covering the period July 31, 1963, through December 31, 1995, augmented by the return on a 30-day Treasury bill. The measure of aggregate wealth includes human capital.

factor pricing kernels. However, imposing this restriction significantly reduces the nonlinearity in the pricing kernel, and consequently significantly impacts the kernel's ability to fit the data. Failing to impose the restriction on the derivative of the pricing kernel results in a substantial improvement in fit, as shown in Section III.A, but produces a pricing kernel that is at odds with standard economic models.

E. Discussion and Interpretation of the Results

Several noteworthy results emerge from the tests conducted in this paper. First, the pricing kernels implied by both a linear single- and a linear multifactor model appear unable to explain the cross-sectional variation in portfolio returns. However, if we allow for nonlinearity in the pricing kernel, either quadratic or cubic in aggregate wealth, and impose restrictions on agents' preferences, we are able to describe cross-sectional variation in returns. One noteworthy feature of the nonlinear pricing kernels is their incorporation of a measure of the return on human capital. The importance of human capital in explaining the cross section of returns has been documented in Campbell (1996) and Jagannathan and Wang (1996). However, in both of these studies, the return on human capital impacts the cross section of returns linearly. The evidence in this paper suggests that this linear impact is not sufficient to explain cross-sectional variation in returns. Rather, it is a nonlinear function of the return on human capital that improves the performance of the model.

To gather some further insight into the sources of improvement in the kernels, we examine the relation of the estimated pricing kernels to the volatility bounds of Hansen and Jagannathan (1996). The bounds represent the minimum volatility that a pricing kernel must exhibit, given its mean, to be admissible. In this respect, the bounds depict the set of admissible pricing kernels in mean–standard deviation space. Since the pricing kernel approach relates the first moment of returns to the second moment of the discount factor, this provides further insight into the specification of the model. The analysis differs from the specification test of the Hansen–Jagannathan distance measure, which asks whether there is some *specific* admissible pricing kernel that is statistically indistinguishable from that of the model.

The Hansen–Jagannathan bounds for the industry portfolios augmented by the one-month T-bill return are presented in Figures 3 and 4. As suggested by the decomposition in Table VII, the pricing kernels perform fairly well in terms of mean deviation from the set of admissible pricing kernels. Further, the graph in Figure 3 suggests that there is not much distinction between the cubic and quadratic pricing kernels when human capital is omitted. In contrast, the quadratic and cubic pricing kernels with human capital, depicted in Figure 4, match the mean of the pricing kernel fairly well, but also come much closer to matching the volatility of the set of pricing kernels. The cubic pricing kernel is actually able to generate sufficient volatility to be inside the Hansen–Jagannathan bounds, but its mean is slightly too high for the pricing kernel to actually lie within the bounds. Again, this result is consistent with the decomposition results, which suggest that the quadratic and cubic pricing kernels require larger *mean* adjustments than the remaining pricing kernels in order to render them admissible.

The Hansen–Jagannathan plots, together with the decomposition of the distance measure, indicate that the incorporation of human capital substan-

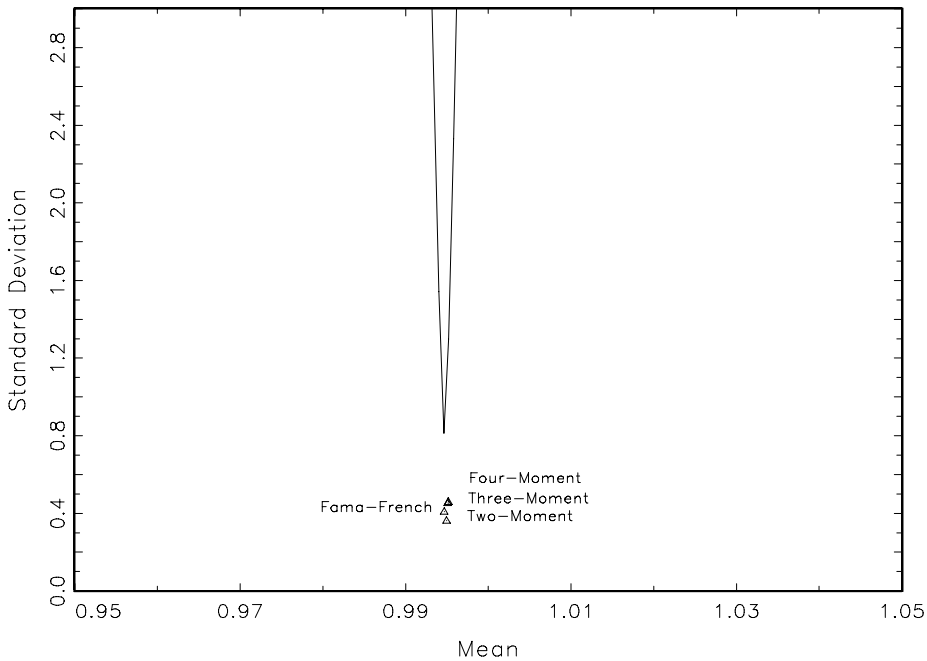

Figure 3. Hansen–Jagannathan bounds, pricing kernels omitting human capital. Figure 3 depicts the Hansen–Jagannathan (1991) bounds on the mean and standard deviation of admissible pricing kernels for the industry-sorted portfolio returns augmented by the one-month T-bill return. Mean–standard deviation pairs for the polynomial pricing kernels as well as the pricing kernels of the Fama–French model are shown as small triangles. The human capital measure is omitted in the estimation of the pricing kernels, and the coefficients of the pricing kernel are modeled as quadratic in the instruments.

tially improves the nonlinear pricing kernels' ability to match the volatility of the set of pricing kernels that are admissible for the industry portfolios. That is, incorporation of human capital substantially lowers the standard deviation of the adjustment necessary to make the nonlinear pricing kernels admissible. This result is initially surprising, since the labor return series is relatively smooth; the monthly standard deviation of the equity index is 4.3 percent compared to 0.4 percent for the return on labor series. However, the labor return is much more leptokurtic than the index return; the excess kurtosis of the labor return is 5.87 compared to 2.70 for the index return. These moments suggest that accounting for human capital through the labor return does not contribute substantially to the improvement of *linear* measures of risk (i.e., variance).[14] Rather, the high kurtosis of the labor return

[14] This conclusion is also reached in Fama and Schwert (1977), who find that betas implied by a market index are not materially different from those implied by correcting for nontraded human capital as in Mayers (1972).

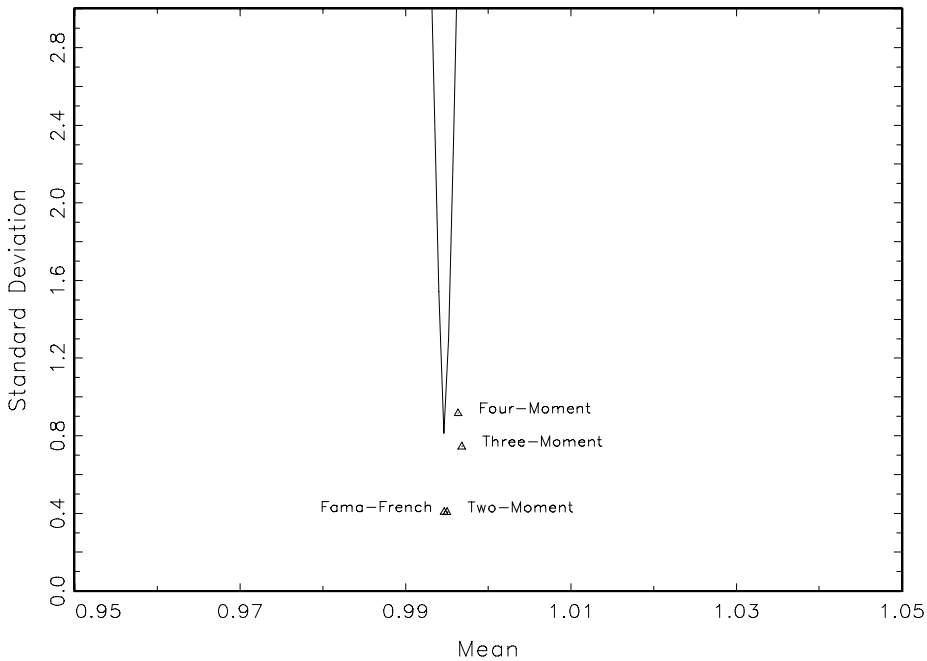

Figure 4. Hansen–Jagannathan bounds, pricing kernels with human capital. Figure 4 depicts the Hansen–Jagannathan (1991) bounds on the mean and standard deviation of admissible pricing kernels for the industry-sorted portfolio returns augmented by the one-month T-bill return. Mean–standard deviation pairs for the polynomial pricing kernels as well as the pricing kernels of the Fama–French model are shown as small triangles. The human capital measure is included in the estimation of the pricing kernels, and the coefficients of the pricing kernel are modeled as quadratic in the instruments.

series compared to the index return series suggests that the labor return better captures *nonlinear* risk.

Another intriguing result is presented in Section III.D. The results of this section indicate that imposing global restrictions on the nonlinear pricing kernels does not invalidate their improvement upon the linear pricing kernels. This result is encouraging because it suggests that we can go a long way in describing asset prices using fundamental preference restrictions. However, the resulting pricing kernels are no longer admissible. Thus, the results suggest that the data require a highly nonlinear pricing kernel, and logic indicates that this kernel should be decreasing. Unfortunately, the polynomial framework does not allow both of these features to be simultaneously present in the pricing kernel. When the kernel is not restricted to be decreasing, the estimates sacrifice this restriction in favor of high nonlinearity. When the restriction of a decreasing pricing kernel is imposed, the polynomial is not able to generate sufficient nonlinearity to be consistent with the data. Thus, the results suggest the need for a functional form of the

pricing kernel that is decreasing and simultaneously exhibits a significant degree of nonlinearity.

One further point deserves attention. In general, it is difficult to determine whether apparent nonlinearities in the data arise due to missing risk factors or a nonlinear relationship between returns and proposed risk factors. For example, an omitted stochastic volatility factor might give rise to an apparent nonlinear relationship between a posited risk factor such as the market portfolio and returns. However, this paper is explicit both about the priced risk factor, the market portfolio, and the form of nonlinearity that arises through agents' preference restrictions. Consequently, the failure to reject the model specification despite the discipline imposed by a model's specific factor and functional form suggest that the nonlinear relationship between returns and the market portfolio is robust. Furthermore, not only does the model survive the specification tests, it does so in a setting in which a highly successful linear multifactor model cannot.

IV. Conclusion

This paper investigates nonlinear pricing kernels that represent a link between nonparametric and parametric approaches to describing cross-sectional variation in equity returns. The common element in this paper's pricing kernels and those of nonparametric models is nonlinearity in priced risk factors. In contrast to these nonparametric approaches, and in common with parametric approaches, the pricing kernels are defined over an endogenous risk factor, and preference restrictions govern the sign of the relationship between returns and the terms in the pricing kernel. The risk factor is the return on aggregate wealth, and the nonlinearity arises from an expansion of a representative investor's Euler equations for portfolio and consumption choice. Adding the additional assumption that the agent's preferences exhibit decreasing absolute prudence allows us to restrict the sign of the first three terms of this expansion. We show that this framework is consistent with a setting in which agents are averse to kurtosis, and consequently asset returns are affected by covariance, coskewness, and cokurtosis with the return on aggregate wealth.

Tests of the model show that incorporating nonlinearity substantially improves upon the pricing kernel's ability to describe the cross section of returns. In particular, when human capital is incorporated into the measure of aggregate wealth, a quadratic and cubic pricing kernel are able to fit the cross section of industry-sorted portfolio returns, whereas a linear pricing kernel and a pricing kernel implied by power utility cannot. Moreover, the marginal contribution of each nonlinear term is statistically important for improving the fit of the pricing kernel. Further, we find that the nonlinear pricing kernels are able to price the cross section of returns substantially better than the Fama and French (1993) three-factor model; the quadratic and cubic models are not rejected whereas the Fama–French model is, and the polynomial pricing kernels produce smaller pricing errors. Additionally,

we find that incorporating the cubic term in the pricing kernel drives out the significance of both the size and book-to-market factor in the Fama–French model. Furthermore, the nonlinear pricing kernel implied by power utility is not admissible for the cross section of industry portfolios, despite the fact that power utility is consistent with decreasing absolute prudence. This result suggests that a specific form of nonlinearity, rather than generic nonlinearity, is important for pricing.

A particularly important source of improvement in the pricing kernel is its incorporation of human capital. However, the results suggest that a *linear* measure of human capital is insufficient to render the pricing kernel admissible. Instead, it is *nonlinear* measures of human capital that improve the performance of the pricing kernel. The results show that, when human capital is incorporated into aggregate wealth, a pricing kernel restricted by preferences and first principles can fit the cross section of returns well, in a setting in which a successful multifactor model cannot. Further, the nonlinear pricing kernel continues to outperform linear single- and multifactor pricing kernels when additional global restrictions are imposed on its functional form. In particular, restricting the pricing kernel to be decreasing over its support generates a pricing kernel that, while inadmissible, dominates the linear pricing kernels in describing the cross section of returns.

This last result provokes an interesting question. Why does the admissible pricing kernel have the wrong shape? That is, what features of the data or the functional form of the polynomial pricing kernel render the kernel inadmissible when monotonicity is imposed? The results suggest the possibility that fitting the data necessitates a highly nonlinear pricing kernel. However, a polynomial cannot simultaneously provide this high degree of nonlinearity and a globally decreasing functional form. What functional relationship between aggregate wealth and returns can provide both of these conditions? What features of the data necessitate the high degree of nonlinearity? These questions remain important issues to be addressed in future research.

REFERENCES

Ahn, Seung, and Christopher Gadarowski, 1999, Small sample properties of the model specification test based on the Hansen–Jagannathan distance, Working paper, Arizona State University.

Arditti, Fred D., 1967, Risk and the required return on equity, *Journal of Finance* 22, 19–36.

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

Bansal, Ravi, and S. Viswanathan, 1993, No-arbitrage and arbitrage pricing: A new approach, *Journal of Finance* 48, 1231–1262.

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

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

Caballé, Jordi, and Alexey Pomansky, 1996, Mixed risk aversion, *Journal of Economic Theory* 71, 485–513.

Campbell, John, 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Campbell, John, 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.

Campbell, John, Andrew Lo, and A. Craig MacKinlay, 1995, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).

Chapman, David, 1997, Approximating the asset pricing kernel, *Journal of Finance* 52, 1383–1410.

Chen, Zhiwu, and Peter J. Knez, 1996, Portfolio performance measurement: Theory and applications, *Review of Financial Studies* 9, 511–555.

Cochrane, John, 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.

Cochrane, John, 2001, *Asset Pricing* (Princeton University Press, Princeton, NJ).

Dahlquist, Magnus, and Paul Söderlind, 1999, Evaluating portfolio performance with stochastic discount factors, *Journal of Business* 72, 347–383.

Darlington, Richard B., 1970, Is kurtosis really "peakedness?" *The American Statistician* 24, 19–22.

Diaz-Gimenez, Javier, Edward C. Prescott, Terry Fitzgerald, and Fernando Alvarez, 1992, Banking in computable general equilibrium economies, *Journal of Economic Dynamics and Control* 16, 533–559.

Dumas, Bernard, and Bruno Solnik, 1995, The world price of foreign exchange risk, *Journal of Finance* 50, 445–479.

Fama, Eugene, 1991, Efficient capital markets: II, *Journal of Finance* 46, 1575–1617.

Fama, Eugene, and Kenneth French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–27.

Fama, Eugene, and Kenneth French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene, and Kenneth French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene, and Kenneth French, 1993. Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene, and Kenneth French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131–156.

Fama, Eugene, and Kenneth French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene, and William Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Fang, Hsing, and Tsong-Yue Lai, 1997, Co-kurtosis and capital asset pricing, *The Financial Review* 32, 293–307.

Farnsworth, Heber, Wayne Ferson, David Jackson, and Steven Todd, 1999, Performance evaluation with stochastic discount factors, Working paper, University of Washington.

Ferson, Wayne, 1989, Changes in expected security returns, risk, and the level of interest rates, *Journal of Finance* 44, 1191–1217.

Ferson, Wayne, and Stephen Foerster, 1994, Finite sample properties of the generalized method of moments estimator in tests of conditional asset pricing models, *Journal of Financial Economics* 36, 29–55.

Ferson, Wayne, and Campbell Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.

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

Hansen, Lars Peter, John Heaton, and Amir Yaron, 1996, Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics* 14, 262–280.

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

Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

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

Harvey, Campbell, 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–317.

Harvey, Campbell, and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance* 55, 1263–1295.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross section of expected returns, *Journal of Finance* 51, 3–54.

Kimball, Miles, 1993, Standard risk aversion, *Econometrica* 61, 589–611.

King, Benjamin, 1966, Market and industry factors in stock price behavior, *Journal of Business* 39, 139–190.

Lo, Andrew, and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–468.

Mayers, David, 1972, Nonmarketable assets and capital market equilibrium under uncertainty, in M. C. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).

Mehra, Rajnish, and Edward Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

Merton, Robert, 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Moskowitz, Tobias, and Mark Grinblatt, 1999, Do industries explain momentum? *Journal of Finance* 54, 1249–1290.

Newey, Whitney, and Kenneth West, 1987a, Hypothesis testing with efficient method of moments estimation, *International Economic Review* 28, 777–787.

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

Pratt, John, and Richard Zeckhauser, 1987, Proper risk aversion, *Econometrica* 55, 143–154.

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

Samuelson, Paul, 1963, Risk and uncertainty: A fallacy of large numbers, *Scientia* 98, 108–113.

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

Stambaugh, Robert, 1982, On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis, *Journal of Financial Economics* 10, 237–268.

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

Weil, Philippe, 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401–421.