

Econometric Evaluation of Asset Pricing Models with No-Arbitrage Constraint

HAITAO LI^a, YUEWU XU^b, AND XIAOYAN ZHANG^c

JUNE 2006

^aLi from the Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109. ^bXu is from School of Business, Fordham University, New York, NY 10017. ^cZhang is from the Johnson Graduate School of Management, Cornell University, Ithaca, NY 14850. We thank Vikas Agarwal, Warren Bailey, Charles Cao, Jin-Chuan Duan, Jingzhi Huang, Jon Ingersoll, Ravi Jagannathan, Raymond Kan, Peter Phillips, Marcel Rindisbacher, Tim Simin, Zhenyu Wang, Guofu Zhou, and seminar participants at Cornell University, Fordham University, Georgia Institute of Technology, Georgia State University, Penn State University, the University of Toronto, and the 2005 Western Finance Association Meeting for helpful comments. We are responsible for any remaining errors.

Econometric Evaluation of Asset Pricing Models with No-Arbitrage Constraint
ABSTRACT

We develop econometric methods for evaluating asset pricing models that explicitly require that a correct asset pricing model has to be arbitrage free. In particular, we develop the asymptotic distribution of the second Hansen-Jagannathan distance which measures the least-square distance between a given asset pricing model and a set of positive stochastic discount factors that correctly price all assets. Simulation evidence shows that our test has good finite sample performance for typical sample sizes considered in the literature. The no-arbitrage constraint makes significant differences in empirical studies of asset pricing models using the Fama-French size and book-to-market portfolios or hedge fund portfolios that exhibit option-like returns. Without the no-arbitrage constraint, we fail to reject certain models using existing methods. However, our test overwhelmingly rejects these models because their stochastic discount factors take negative values with high probabilities and therefore are not arbitrage free.

JEL Classification: C4, C5, G1

Keywords: Stochastic Discount Factor Models, Asset Pricing Tests, Hansen-Jagannathan Distances, Arbitrage.

In this paper, we develop econometric methods for evaluating asset pricing models that explicitly require that a correct asset pricing model has to be arbitrage free. The fundamental theorem of asset pricing asserts the equivalence of absence of arbitrage and the existence of a positive stochastic discount factor (hereafter SDF) that correctly prices all assets. Most existing empirical studies, however, have mainly focused on pricing errors and ignored the no-arbitrage constraint (i.e., the SDF has to be strictly positive) in evaluating asset pricing models. This practice has some undesirable consequences and could lead to misleading conclusions in empirical studies.

Dybvig and Ingersoll (1982) show that linear asset pricing models are not arbitrage free and are not appropriate for pricing derivatives because their SDFs take negative values in certain states of the world. Though linear factor models are seldom used directly to price derivatives, they have been widely used in important applications that implicitly involve derivatives. One prominent example is performance evaluation of actively managed mutual funds and hedge funds. Many mutual funds and most hedge funds employ dynamic trading strategies which generate option-like payoffs.¹ Many hedge funds directly trade derivatives and their returns exhibit option-like features.² Grinblatt and Titman (1987) and Glosten and Jagannathan (1994) emphasize the importance of imposing the no-arbitrage constraint when evaluating the performances of mutual funds. The fast growing hedge fund industry and the need to evaluate hedge fund performances make this issue even more urgent in current asset pricing literature.³

Even for applications that involve mainly primary assets, there are still important benefits of imposing the no-arbitrage constraint in evaluating linear factor models. With the no-arbitrage constraint, model parameters are chosen not solely to minimize pricing errors, but to make a linear factor model as close as possible to the true asset pricing model which by definition has a strictly positive SDF. As pointed out by Cochrane (2001), fundamentally all linear factor models are approximations of aggregate intertemporal marginal rate of substitution (IMRS hereafter), which is positive. Therefore, models estimated with the no-arbitrage constraint are likely to be better proxies of the IMRS.

Incorporating the no-arbitrage constraint in empirical analysis of asset pricing models, however, is not straightforward. One approach is to include returns on derivatives in model evaluation with the logic that the estimated model should be close to being arbitrage free if it can price both primary assets and derivatives well. However, this approach requires additional data and the

¹See Merton (1981), Dybvig and Ross (1985) and others for more detailed discussions on this issue.

²TASS, a hedge fund research company, reports that more than 50 percent of the 4,000 hedge funds it follows use derivatives. Fung and Hsieh (2001), Agarwal and Naik (2004), Ben Dor and Jagannathan (2002), and Mitchell and Pulvino (2001) have documented option-like features in hedge fund returns.

³Goetzmann, Ingersoll, Spiegel, and Welch (2002) show that the use of derivatives by hedge funds render the Sharpe ratio an inappropriate measure of hedge fund performance. While the Sharpe ratio adjusts for total risks, we are more interested in adjusting for systematic risks in evaluating hedge fund performance.

results might be sensitive to which derivatives are used in estimation.⁴ Econometrically it is also rather difficult to impose the no-arbitrage constraint in traditional linear regressions which have been widely used to analyze linear factor models. For example, Cochrane (2001) (p. 130) claims that “I do not know any way to cleanly graft absence of arbitrage on to expected return-beta models.”

Hansen and Jagannathan (1997) provide a theoretical framework within which the no-arbitrage constraint can be easily incorporated. The Hansen-Jagannathan distances (hereafter HJ-distances) measure least-square distances between a model SDF and the set of admissible SDFs that correctly price all assets. While the first HJ-distance considers all admissible SDFs, the second one considers only strictly positive admissible SDFs to avoid arbitrage opportunities. Hansen and Jagannathan (1997) show that the second HJ-distance represents the minimax bound of the pricing errors of all payoffs (including all derivatives) with a unit norm. This means that if a model has a zero second HJ-distance, then it can perfectly price both primary and derivatives assets, and is arbitrage free. Therefore, the second HJ-distance is a natural measure of model performance that explicitly imposes the no-arbitrage constraint.

Jagannathan and Wang (1996, hereafter JW) develop the asymptotic distribution of the first HJ-distance under the null hypothesis of a correctly specified model. This distribution has been widely used in the existing literature to conduct specification tests of asset pricing models. The exact distribution of the first HJ-distance developed by Kan and Zhou (2002) also simplifies empirical applications. So far the asymptotic distribution of the second HJ-distance under the null hypothesis of a correctly specified model has not been developed. This greatly hinders the applications of the second HJ-distance despite its many appealing features. Econometric analysis of the second HJ-distance is very challenging, because the second HJ-distance involves certain functions that are not pointwise differentiable with respect to model parameters. Standard asymptotic analysis involves Taylor series approximations of an appropriate objective function near true parameter values. This procedure breaks down if the objective function is not differentiable. As a result, the techniques of JW (1996) for analyzing the first HJ-distance cannot be applied to the second HJ-distance.

To overcome this difficulty, Wang and Zhang (2004) develop a simulation-based Bayesian approach based on the second HJ-distance. Using Markov Chain Monte Carlo simulation, they are able to obtain the posterior distributions of the two HJ-distances and many other related statistics. They then evaluate asset pricing models based on those distributions. They show that the two HJ-distances lead to dramatically different conclusions in evaluating time-varying and multi-factor models, especially using conditional portfolios. Though certain models have small

⁴Strictly speaking, to ensure that the estimated model is arbitrage free, one needs to incorporate all possible derivatives that can be formed from the primary assets, which can be empirically challenging to do.

first HJ-distances, they have much larger second HJ-distances because their SDFs take negative values with high probabilities. The Bayesian methodology of Wang and Zhang (2004), though a nice contribution to the literature, is very different from traditional methods in the literature, such as the GMM test of Hansen (1982) or the JW test.

Our paper complements Wang and Zhang (2004) by developing the asymptotic distribution of the second HJ-distance under the null hypothesis of a correctly specified model. We overcome the “non-differentiability” problem by introducing the concept of “differentiation in quadratic mean” of Le Cam (1986), Pollard (1982), and Pakes and Pollard (1989).⁵ Simulation studies show that our test based on the second HJ-distance has finite sample performance that is (i) comparable with that of the JW test; and (ii) reasonably good for sample sizes typically used in the existing literature. Our test based on the second HJ-distance has the advantage of being able to reject models with small first HJ-distances but whose SDFs take negative values with high probabilities. Our approach is a natural extension of the JW test and makes it very convenient to impose the no-arbitrage constraint within the established econometric framework in the existing literature.

In a related study, Hansen, Heaton, and Luttmer (1995) develop the asymptotic distributions of two HJ-distances under the null hypothesis that a given asset pricing model is misspecified. Their results, however, can not be applied to our setting because their asymptotic distributions become degenerate under the null hypothesis of a correctly specified model. In contrast, we can easily extend our analysis to obtain the results of Hansen, Heaton, and Luttmer (1995), even when parameters are not known and need to be estimated from the data.

To illustrate the importance of the no-arbitrage constraint, we apply our new methods to evaluate several well-known asset pricing models using the 25 Size/Book-to-Market (BM hereafter) portfolios of Fama and French (1993) and several popular hedge fund strategies that exhibit option-like returns. The no-arbitrage constraint makes significant differences in evaluating asset pricing models whether the testing assets explicitly involve derivatives or not. In both applications, we find substantial differences in estimated first and second HJ-distances for some models. Though certain models have relatively good performances in pricing both sets of assets measured by the first HJ-distance, their SDFs take negative values with high probabilities and are overwhelmingly rejected based on the second HJ-distance.

The rest of this paper is organized as follows. In section 1, we introduce the HJ-distances and discuss the importance of the no-arbitrage constraint. Section 2 develops the asymptotic distribution of the second HJ-distance and related statistics. Section 3 provides simulation evidence on the finite sample performance of the asymptotic theory. Sections 4 and 5 contain applications of the new econometric methods to the Fama-French 25 Size/BM portfolios and hedge fund returns,

⁵This means that although Taylor expansion does not work pointwise, to obtain the asymptotic result, we only need it to work in an average sense.

respectively. Section 6 concludes and the appendix provides the mathematical proofs.

1. No-Arbitrage Constraint and the Hansen-Jagannathan Distances

In this section, we discuss the importance of the no-arbitrage constraint in asset pricing applications under the SDF framework. We illustrate how the second HJ-distance, a measure of model performance, explicitly imposes the no-arbitrage constraint.⁶

Let the uncertainty of the economy be described by a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ for $t = 0, 1, \dots, T$. Suppose there are n assets with payoffs Y_t at t , where Y_t is an $n \times 1$ vector. Denote \underline{Y} as the payoff space of all the assets. In the absence of arbitrage, there must exist a strictly positive SDF that correctly prices all assets. That is, for all t , we have

$$\mathbb{E}[m_{t+1}Y_{t+1}|\mathcal{F}_t] = X_t, \quad m_{t+1} > 0, \quad \forall Y_{t+1} \in \underline{Y} \quad (1)$$

where X_t , an $n \times 1$ vector, represents the prices of the n assets at t . The random variable m_{t+1} discounts payoffs at $t+1$ state by state to yield price at t and hence is called a stochastic discount factor. If the market is complete, then m_{t+1} will be unique. Otherwise there will be multiple m_{t+1} 's that satisfy (1). For instance, in an economy with a representative investor, m_{t+1} measures aggregate IMRS:

$$m_{t+1} = \beta \frac{u_c(c_{t+1})}{u_c(c_t)},$$

where $u_c(\cdot)$ is the representative investor's marginal utility of consumption, c_t and c_{t+1} are the investor's consumptions at t and $t+1$ respectively, and β is the investor's subjective discount factor.

The pricing equation in (1) suggests that a good asset pricing model should have (i) small pricing errors and (ii) a SDF that is strictly positive. Existing empirical studies of asset pricing models, however, have mainly focused on the first aspect of model performance and ignored the second one. The no-arbitrage constraint is especially likely to be violated in linear factor models. According to Cochrane (2001), linear factor models identify economic factors whose linear combinations are good proxies for aggregate IMRS, i.e.,

$$\beta \frac{u_c(c_{t+1})}{u_c(c_t)} \approx a + b' f_{t+1}.$$

Due to its linear structure, however, the SDF proxy $a + b' f_{t+1}$ could take negative values under certain market conditions, even though the original SDF $\beta u_c(c_{t+1})/u_c(c_t)$ is always positive.

Ignoring the no-arbitrage constraint could have undesirable consequences in empirical analysis of asset pricing models. Ideally we want to choose parameters a and b such that $a + b' f_{t+1}$ is a close proxy of $\beta u_c(c_{t+1})/u_c(c_t)$. However, if we estimate a and b by solely minimizing pricing

⁶The discussions in this section draw materials from Cochrane (2001), Dybvig and Ingersoll (1982), Hansen and Jagannathan (1997), and Wang and Zhang (2004).

errors, then the estimated model $\hat{a} + \hat{b}'f_{t+1}$ could be far away from $\beta u_c(c_{t+1})/u_c(c_t)$ even though the estimated pricing errors are small. This is because in many cases the small pricing errors are obtained at the expense of violating the no-arbitrage constraint and the estimated model $\hat{a} + \hat{b}'f_{t+1}$ often takes negative values with high probabilities.

Violation of the no-arbitrage constraint could also lead to misleading conclusions in important applications. For example, the SDF of the CAPM equals

$$m_{t+1}^{CAPM} = a + br_{MKT,t+1}, \quad (2)$$

where $r_{MKT,t+1}$ is the excess return on the market portfolio at $t + 1$ and $b < 0$.⁷ Dybvig and Ingersoll (1982) show that m_{t+1}^{CAPM} takes negative values when $r_{MKT,t+1}$ is large enough. As a result, the CAPM would assign a negative price to an option that pays one dollar in the states where $m_{t+1}^{CAPM} < 0$ and zero otherwise, even though the option has non-negative payoff. Although the CAPM and other linear factor models are rarely used directly to price options, they have been widely used in important applications that implicitly involve options, such as performance evaluations of actively managed mutual funds and hedge funds. Models that admit arbitrage opportunities would give misleading results on the performances of these funds because most of them have option-like returns.⁸

Therefore, due to both statistical and economic concerns, it would be beneficial to impose the no-arbitrage constraint in empirical asset pricing studies regardless derivatives are explicitly involved or not. We develop econometric methods for evaluating asset pricing models based on the second HJ-distance. Below we give a brief introduction of the HJ-distances and explain the basic idea of our approach. In next section, we develop the econometric theory of our approach. Without loss of generality, we focus our discussions on the unconditional implication of (1)

$$\mathbb{E}[m_{t+1}Y_{t+1}] = \mathbb{E}[X_t].^9$$

The first HJ-distance, denoted as δ , measures the least-square distance or the L^2 norm between

⁷A negative b is consistent with a positive market risk premium.

⁸Suppose we use the CAPM to evaluate the performance of a mutual fund that invests in the option that pays one dollar when $m_{t+1}^{CAPM} > 0$ and zero otherwise. In general if the fund longs the above option, the CAPM would predict that the fund has a negative “alpha” even though all securities are fairly priced in the market. Because the CAPM assigns a negative value to the option, it underestimates the initial value of the portfolio and overestimates the expected return this portfolio should yield. Thus, the model-predicted expected return on the portfolio is higher than the actual expected return, which means that the fund has a negative “alpha.” Similar argument shows that the CAPM would predict a positive “alpha” for the fund that shorts the option, even though no abnormal performance exists.

⁹If we include enough scaled payoffs in our analysis, the unconditional pricing equation becomes the conditional pricing equation. For notational convenience, we omit time subscripts t whenever the meaning is obvious.

a candidate SDF model H and the set of SDFs that correctly price all assets:

$$\delta = \min_{m \in \mathcal{M}} \|H - m\| = \min_{m \in \mathcal{M}} \sqrt{\mathbb{E}(H - m)^2}, \quad (3)$$

where $\mathcal{M} = \{m_{t+1} : \mathbb{E}[m_{t+1}Y_{t+1}] = \mathbb{E}[X_t], \forall Y_{t+1} \in \underline{Y}\}$ is the set of SDFs that correctly price all assets.

The second HJ-distance, denoted as δ^+ , is defined as:

$$\delta^+ = \min_{m \in \mathcal{M}^+} \|H - m\| = \min_{m \in \mathcal{M}^+} \sqrt{\mathbb{E}(H - m)^2}, \quad (4)$$

where $\mathcal{M}^+ = \{m_{t+1} : \mathbb{E}[m_{t+1}Y_{t+1}] = \mathbb{E}[X_t], m_{t+1} > 0, \forall Y_{t+1} \in \underline{Y}\}$. Thus it considers only SDFs that are in \mathcal{M} and are strictly positive.

Often the candidate model H depends on some unknown parameters θ , and the two distances are defined as

$$\delta(\theta) = \min_{\theta} \min_{m \in \mathcal{M}} \|H(\theta) - m\| = \min_{\theta} \min_{m \in \mathcal{M}} \sqrt{\mathbb{E}(H(\theta) - m)^2}, \quad (5)$$

$$\delta^+(\theta) = \min_{\theta} \min_{m \in \mathcal{M}^+} \|H(\theta) - m\| = \min_{\theta} \min_{m \in \mathcal{M}^+} \sqrt{\mathbb{E}(H(\theta) - m)^2}. \quad (6)$$

Let $\hat{\theta} = \arg \min_{\theta} \delta(\theta)$ and $\hat{\theta}^+ = \arg \min_{\theta} \delta^+(\theta)$. In general, the second HJ-distance is bigger than the first one, because \mathcal{M}_+ is a subset of \mathcal{M} .

The first HJ-distance has been widely used in the empirical asset pricing literature for model estimation and evaluation. Hansen and Jagannathan (1997) show that $\hat{\theta}$ is a GMM estimator with a weighting matrix equals the inverse of the second moment matrix of the payoffs:

$$\delta(\theta) = \mathbb{E}(H(\theta)Y - X)' \mathbb{E}(YY')^{-1} \mathbb{E}(H(\theta)Y - X).$$

The weighting matrix $\mathbb{E}(YY')^{-1}$ is model independent and thus simplifies model comparison. JW (1996) develop the asymptotic distribution of δ under the null hypothesis $\mathbb{H}_0: \delta = 0$. This distribution makes it possible to conduct specification tests based on the first HJ-distance. Hansen and Jagannathan (1997) also show that the first HJ-distance has a nice interpretation as the maximum pricing error of a portfolio of primary assets with a unit norm, i.e.,

$$\delta(\theta) = \min_{\theta} \max_{\|Y\|=1} |\mathbb{E}(mY) - \mathbb{E}(H(\theta)Y)|, \forall Y \in \underline{Y}.$$

Hence minimizing the first HJ-distance is equivalent to minimizing the pricing errors of primary assets. As a result, the estimated model $H(\hat{\theta})$ is not necessarily strictly positive and the JW test may not be able to reject a model if it has a small first HJ-distance but is not arbitrage free.

In contrast, the second HJ-distance explicitly requires that in addition to having small pricing errors, a good asset pricing model should also be arbitrage free. Hansen and Jagannathan (1997)

show that the second HJ-distance has a nice interpretation as the minimax bound on pricing errors of any payoff (including both primary and derivatives assets) in L^2 with a unit norm, i.e.,

$$\delta^+(\theta) = \min_{m \in \mathcal{M}_+} \max_{\|Y\|=1} |\mathbb{E}(mY) - \mathbb{E}(H(\theta)Y)|, \forall Y \in L^2.$$

In other words, the second HJ-distance differs from the first one by focusing on the pricing errors of not only primary assets but also of all possible derivatives that can be constructed from the primary assets. Hence if the second HJ-distance of a model equals zero, then the model can perfectly price all possible payoffs and also is arbitrage free, a necessary condition to price derivatives. As a result, in general we will get very different parameter estimates using the two HJ-distances. Though $\hat{\theta}$ minimizes the weighted pricing errors of primary assets, $\hat{\theta}^+$ minimizes the least-square distance between $H(\hat{\theta}^+)$ and \mathcal{M}_+ .

The second HJ-distance provides a natural approach to incorporate the no-arbitrage constraint in empirical analysis of asset pricing models. It provides an appropriate objective function to estimate model parameters and an economically meaningful measure of model misspecification. It also complements some existing approaches in dealing with the no-arbitrage constraint. For example, Bansal and Viswanathan (1993) develop a semi-nonparametric method to identify positive nonlinear SDFs that can correctly price all assets. They truncate a nonlinear SDF if it turns negative and evaluate model performance based on pricing errors via GMM. While the focus of Bansal and Viswanathan (1993) is to approximate true asset pricing models using flexible functional forms, the focus of our paper is to develop econometric methods to evaluate *all* asset pricing models, whether arbitrage free or not, by their second HJ-distances. Although truncation guarantees a model's SDF to be nonnegative, it still matters in practice how the truncated model is estimated and evaluated. Later empirical analysis shows that we obtain very different results when evaluating truncated models using the first and second HJ-distances.

2. Econometric Theory

In this section, we develop the asymptotic distributions of the second HJ-distance and related statistics under the null hypothesis of a correctly specified model. Following Hansen and Jagannathan (1997), our analysis focuses on the conjugate representations of the minimization problems in (3) and (4):

$$\delta^2 = \max_{\lambda} \left\{ \mathbb{E}H^2 - \mathbb{E}[H - \lambda'Y]^2 - 2\lambda'\mathbb{E}X \right\}, \quad (7)$$

$$[\delta^+]^2 = \max_{\lambda} \left\{ \mathbb{E}H^2 - \mathbb{E}[H - \lambda'Y]^{+2} - 2\lambda'\mathbb{E}X \right\}, \quad (8)$$

where λ is an $n \times 1$ vector of Lagrangian multipliers and $[H - \lambda'Y]^+ = \max[0, H - \lambda'Y]$. The first order conditions of the above two optimization problems are given as

$$\mathbb{E}X - \mathbb{E}[(H - \lambda'Y)Y] = 0, \quad \text{for } \delta^2, \quad (9)$$

$$\mathbb{E}X - \mathbb{E} \left[(H - \lambda'Y)^+ Y \right] = 0, \quad \text{for } [\delta^+]^2. \quad (10)$$

Suppose λ_0 and λ_0^+ solve (9) and (10), respectively, then $[H - \lambda_0'Y] \in \mathcal{M}$ and $[H - \lambda_0^{+'}Y]^+ \in \mathcal{M}_+$. That is, the random variable $\lambda_0'Y$ represents the necessary adjustments of H so that it can correctly price all assets. Or alternatively, $\lambda_0'Y$ can be used to discount future payoffs state by state to yield current pricing errors of Y : $\mathbb{E}[(\lambda_0'Y)Y] = \mathbb{E}[(H - m)Y]$, where $m \in \mathcal{M}$. Therefore, while δ measures average deviations of H from \mathcal{M} , $\lambda_0'Y$ measures H 's deviations from \mathcal{M} in different states of the economy. The interpretation of $\lambda_0^{+'}Y$, although similar to that of $\lambda_0'Y$, is more complicated due to the no-arbitrage constraint: $H - [H - \lambda_0^{+'}Y]^+$ is the necessary adjustments that make H a member of \mathcal{M}_+ .

For SDF models that depend on unknown model parameters θ , the two HJ-distances are defined as

$$\begin{aligned} [\delta]^2 &= \min_{\theta} \max_{\lambda} \mathbb{E}\phi(\theta, \lambda), \\ [\delta^+]^2 &= \min_{\theta} \max_{\lambda} \mathbb{E}\phi^+(\theta, \lambda), \end{aligned}$$

where

$$\begin{aligned} \phi(\theta, \lambda) &\equiv H(\theta)^2 - [H(\theta) - \lambda'Y]^2 - 2\lambda'X, \\ \phi^+(\theta, \lambda) &\equiv H(\theta)^2 - [H(\theta) - \lambda'Y]^+{}^2 - 2\lambda'X. \end{aligned}$$

In empirical applications, the population probability distribution is unobservable and we need to approximate expectations using time series averages. Suppose we have the following time series observations of asset prices, payoffs, and model SDFs, $\{(X_{t-1}, Y_t, H_t(\theta)) : t = 1, 2, \dots, T\}$, where θ is a k -dimensional parameter vector. Following Hansen and Jagannathan (1997), we use the empirical counterpart of $\mathbb{E}\phi^+(\theta, \lambda)$,

$$\mathbb{E}_T\phi^+(\theta, \lambda) = \frac{1}{T} \sum_{t=1}^T \{H_t(\theta)^2 - [H_t(\theta) - \lambda'Y_t]^+{}^2 - 2\lambda'X_{t-1}\}$$

in our econometric analysis of the second HJ-distance.¹⁰ Therefore, the main objective of our asymptotic analysis is to characterize the behavior of $[\delta_T^+]^2 = \min_{\theta} \max_{\lambda} \mathbb{E}_T\phi^+(\theta, \lambda)$, as $T \rightarrow \infty$.

The standard approach for an asymptotic analysis of $\mathbb{E}_T\phi^+(\theta, \lambda)$ would be to employ a point-wise quadratic Taylor expansion of the function $\phi^+(\theta, \lambda)$ with respect to (θ, λ) around true model

¹⁰From now on, we will focus our analysis on the second HJ-distance. For analysis of the first HJ-distance, see Jagannathan and Wang (1996).

parameters (θ_0, λ_0) :¹¹

$$\begin{aligned} \phi^+(\theta, \lambda) &= \phi^+(\theta_0, \lambda_0) + \left(\begin{array}{c} \frac{\partial}{\partial \theta} \phi^+ \\ \frac{\partial}{\partial \lambda} \phi^+ \end{array} \right)'_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} \\ &+ \frac{1}{2} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix}' \left(\begin{array}{cc} \frac{\partial^2}{\partial \theta \partial \theta'} \phi^+ & \frac{\partial^2}{\partial \theta \partial \lambda'} \phi^+ \\ \frac{\partial^2}{\partial \lambda \partial \theta'} \phi^+ & \frac{\partial^2}{\partial \lambda \partial \lambda'} \phi^+ \end{array} \right)_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} + o(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|^2), \end{aligned}$$

and then optimize the resulting quadratic representation with respect to θ and λ :

$$\begin{aligned} \mathbb{E}_T \phi^+(\theta, \lambda) &= \mathbb{E}_T \phi^+(\theta_0, \lambda_0) + \left(\begin{array}{c} \mathbb{E}_T \frac{\partial}{\partial \theta} \phi^+ \\ \mathbb{E}_T \frac{\partial}{\partial \lambda} \phi^+ \end{array} \right)'_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} \\ &+ \frac{1}{2} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix}' \left(\begin{array}{cc} \mathbb{E}_T \frac{\partial^2}{\partial \theta \partial \theta'} \phi^+ & \mathbb{E}_T \frac{\partial^2}{\partial \theta \partial \lambda'} \phi^+ \\ \mathbb{E}_T \frac{\partial^2}{\partial \lambda \partial \theta'} \phi^+ & \mathbb{E}_T \frac{\partial^2}{\partial \lambda \partial \lambda'} \phi^+ \end{array} \right)_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} + o_p(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|^2). \end{aligned} \quad (11)$$

However, standard Taylor expansion breaks down in our case because the function $\phi^+(\theta, \lambda)$ is not pointwise differentiable. To better illustrate this point, observe that $\phi^+(\theta, \lambda)$ can be written as

$$\phi^+(\theta, \lambda) \equiv H(\theta)^2 - g(H(\theta) - \lambda'Y) - 2\lambda'X,$$

where $g(x) = [\max(x, 0)]^2 \equiv [x^+]^2$. Observe that $g(x)$ is first order differentiable everywhere with first order derivative

$$g^{(1)}(x) = 2[x]^+ = \begin{cases} 2x & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

However, $g(x)$ does not have a second order derivative at $x = 0$, i.e., $g^{(1)}$ is no longer differentiable everywhere. The second order derivative of $g(x)$ equals

$$g^{(2)}(x) = \begin{cases} 2 & \text{if } x > 0, \\ \text{not exist} & \text{if } x = 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Therefore, for δ_T^+ , the function $[H_t(\theta) - \lambda'Y_t]^+$ is not *pointwise differentiable* with respect to (θ, λ) for all $H_t(\theta)$ and Y_t .¹² That is, for a given (θ, λ) , there are combinations of $H_t(\theta)$ and Y_t such that $H_t(\theta) - \lambda'Y_t = 0$, which is the kink point of $[H_t(\theta) - \lambda'Y_t]^+$. As a result, the derivatives of $\phi^+(\theta, \lambda)$ with respect to (θ, λ) are not always well defined for those $H_t(\theta)$ and Y_t .

The key to overcome this difficulty is that *pointwise differentiability* is not a necessary condition to obtain (11), because all we need is a good approximation to $\mathbb{E}_T \phi^+(\theta, \lambda)$ (but not $\phi^+(\theta, \lambda)$ itself)

¹¹The true parameters (θ_0, λ_0) solve the population optimization problem: $(\theta_0, \lambda_0) \equiv \arg \min_{\theta} \max_{\lambda} \mathbb{E} \phi^+(\theta, \lambda)$.

¹²Let ω be a random variable. A function $f(\theta, \omega)$ is pointwise differentiable with respect to θ means that the function has partial derivatives with respect to θ in the classical sense for all possible values of ω .

around true parameter values (θ_0, λ_0) . To this end, the notion of “*differentiability in quadratic mean*” in modern statistics (c.f. Le Cam (1986)) will play an important role.¹³ In contrast to “*pointwise differentiability*,” which implies a good approximation to $\phi^+(\theta, \lambda)$ for all H_t and Y_t , “*differentiability in quadratic mean*” implies that the error of approximating $\mathbb{E}_T \phi^+(\theta, \lambda)$ is negligible in quadratic mean or $L^2(P)$ norm. In other words, all we need is an approximation of $\phi^+(\theta, \lambda)$ that works well in an average sense. For further discussions of non-differentiability issues, see Pollard (1982), Pakes and Pollard (1989), and Hansen, Heaton, and Luttmer (1995), among others.

Our approach can be briefly described as follows and is along the lines of Pollard (1982). First we decompose $\mathbb{E}_T \phi^+(\theta, \lambda)$ into a deterministic component and a (centered) random component

$$\mathbb{E}_T \phi^+(\theta, \lambda) = \mathbb{E} \phi^+(\theta, \lambda) + (\mathbb{E}_T - \mathbb{E}) \phi^+(\theta, \lambda).$$

To obtain a quadratic representation like (11), we consider a second order approximation to the deterministic term to extract the curvature of $\mathbb{E} \phi^+(\theta, \lambda)$ and a first order approximation to the random component. Since the random component is centered, it is in general one order smaller than the deterministic component. This explains the difference in orders of approximation of the two components in the above equation.

The following Lemma justifies a local asymptotic quadratic (LAQ) expansion of the objective function $\mathbb{E}_T \phi^+(\theta, \lambda)$ along the lines of Pollard (1982).

Lemma 1. Suppose Assumptions A.1 to A.6 in the appendix hold. Then we have the following local asymptotic quadratic (LAQ) representation for $\mathbb{E}_T \phi^+(\theta, \lambda)$ around (θ_0, λ_0) :

$$\begin{aligned} & \mathbb{E}_T \phi^+(\theta, \lambda) \\ &= \mathbb{E} \phi^+(\theta_0, \lambda_0) + (\mathbb{E}_T - \mathbb{E}) \phi^+(\theta_0, \lambda_0) + \begin{pmatrix} A \\ B \end{pmatrix}' \begin{pmatrix} U \\ V \end{pmatrix} + \frac{1}{2} \begin{pmatrix} U \\ V \end{pmatrix}' \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \\ &+ o(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|^2) + o_p(\|(\theta, \lambda) - (\theta_0, \lambda_0)\| T^{-1/2}), \end{aligned}$$

where $U \equiv (\theta - \theta_0)$, $V \equiv (\lambda - \lambda_0)$, $A \equiv (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \theta} \phi^+|_{(\theta_0, \lambda_0)}$, $B \equiv (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \lambda} \phi^+|_{(\theta_0, \lambda_0)}$, and

$$\Gamma \equiv \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \equiv \begin{pmatrix} \mathbb{E} \frac{\partial^2}{\partial \theta \partial \theta'} \phi^+ & \mathbb{E} \frac{\partial^2}{\partial \theta \partial \lambda'} \phi^+ \\ \mathbb{E} \frac{\partial^2}{\partial \lambda \partial \theta'} \phi^+ & \mathbb{E} \frac{\partial^2}{\partial \lambda \partial \lambda'} \phi^+ \end{pmatrix} \Big|_{(\theta_0, \lambda_0)}.$$

Proof. See the appendix.¹⁴

¹³A function $f(\theta, \omega)$ is differentiable in quadratic mean with respect to θ at θ_0 , if there exists a $\Delta(\omega)$ in L^2 such that $\mathbb{E} [(f(\theta, \omega) - f(\theta_0, \omega)) / (\theta - \theta_0) - \Delta(\omega)]^2 \rightarrow 0$ as $\theta \rightarrow \theta_0$. Similar ideas have been used by Pakes and Pollard (1989) and others to handle non-differentiable criteria functions.

¹⁴Note that since the second derivatives are well defined except on a set of probability zero, the expectations are indeed well defined.

Based on the LAQ of $\mathbb{E}_T \phi^+(\theta, \lambda)$ in Lemma 1, we develop the asymptotic distribution of the second HJ-distance in the following theorem. One of the assumptions we need is that a central limit theorem holds for the empirical process:

$$\sqrt{T} (\mathbb{E}_T - \mathbb{E}) (H(\theta_0) Y - X) \rightarrow Z \equiv N(0, \Lambda),$$

where $\Lambda = \mathbb{E}[H(\theta_0) Y - X][H(\theta_0) Y - X]'$.

Theorem 1. Suppose Assumptions A.1 to A.9 in the appendix hold. Then under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, $T [\delta_T^+]^2$ is asymptotically distributed as a weighted χ^2 distribution with $(n - k)$ degrees of freedom:

$$T [\delta_T^+]^2 \rightarrow Z' \Xi Z,$$

where $\Xi = D^{-1} - D^{-1} G (G' D^{-1} G)^{-1} G' D^{-1}$, $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$, $D \equiv \mathbb{E}[Y Y']$, and $\nabla H(\theta)$ is the gradient of $H(\theta)$ with respect to θ .

Proof. See the appendix.

The above result is a natural extension of the JW test based on the first HJ-distance. To impose the no-arbitrage constraint, all one needs to do is to change the objective function from δ to δ^+ and to use the asymptotic distribution in Theorem 1 to conduct specification tests. This makes it very convenient for researchers to incorporate the no-arbitrage constraint into their empirical studies of asset pricing models.

While JW test only considers linear factor models, our result is applicable to nonlinear SDF models where $H(\theta)$ depends on θ nonlinearly. These include asset pricing models with complicated IMRS. In a linear factor model, $H(\theta) = F'\theta$, where F is a vector of risk factors, $\nabla H(\theta) = F$ and $\nabla^2 H(\theta) = 0$. It is straightforward to show that under $\mathbb{H}_0: \delta^+ = 0$, the asymptotic distribution of δ^+ is identical to that of δ in JW (1996) for linear models. The reason is that $\delta^+ = 0$ necessarily implies $\delta = 0$ and as a result, δ and δ^+ have the same asymptotic distributions.

There are several important differences between our test and the JW test. First, the implementations of the two tests are very different. For the JW test, one estimates $(\hat{\theta}, \hat{\lambda})$ by minimizing δ ; for our approach, we estimate $(\hat{\theta}^+, \hat{\lambda}^+)$ by minimizing δ^+ . Except for the rare case in which $\delta^+ = 0$, the two estimated HJ-distances and their associated parameters are generally different from each other. Second, the two approaches have different powers in rejecting misspecified models. The JW test might accept SDF models that belong to \mathcal{M} but not \mathcal{M}_+ . Our test would reject those models because they are not arbitrage free. Third, the econometric techniques used in our analysis are quite different from that in the existing literature and can be useful in other finance applications that involve non-differentiable objective functions.

Hansen, Heaton, and Luttmer (1995) develop the asymptotic distributions of δ and δ^+ under the null hypothesis $\mathbb{H}_0: \delta \neq 0$ and $\delta^+ \neq 0$, respectively. However, their approach can not be applied to our setting because their asymptotic distributions become degenerate when $\delta = 0$ and

$\delta^+ = 0$ (see the last paragraph of p. 249). In contrast, we can easily extend our analysis in Theorem 1 to obtain the results of Hansen, Heaton, and Luttmer (1995). In fact, we can show that the results of Hansen, Heaton, and Luttmer (1995), derived for known parameters, still hold when parameters need to be estimated from the data.

In addition to the above result, we also develop the asymptotic distributions of $\hat{\theta}^+$, $\hat{\lambda}^+$, and pricing errors $H(\hat{\theta}^+)Y - X$, for general nonlinear SDF model, where $(\hat{\theta}^+, \hat{\lambda}^+) = \arg \min_{\theta} \max_{\lambda} \mathbb{E}_T \phi^+(\theta, \lambda)$.

Proposition 1. [Model Parameter] Suppose Assumptions A.1 to A.9 in the appendix hold. Then under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, $\sqrt{T}(\hat{\theta}^+ - \theta_0)$ is asymptotically normally distributed with mean zero and covariance matrix

$$(G'D^{-1}G)^{-1} G'D^{-1}\Lambda D^{-1}G (G'D^{-1}G)^{-1},$$

where $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E}[YY']$.

Proof. See the appendix.

The asymptotic distribution of parameter estimates provide useful information on model specification. For example, in a factor model, it allows us to examine the importance of a specific factor by testing whether the coefficient of the factor is significantly different from zero.

Proposition 2. [Lagrangian Multiplier] Suppose Assumptions A.1 to A.9 in the appendix hold. Then under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, $\sqrt{T}(\hat{\lambda}^+(\hat{\theta}^+) - \lambda_0)$ is asymptotically normally distributed with mean zero and covariance matrix

$$[D^{-1} - D^{-1}G (G'D^{-1}G)^{-1} G'D^{-1}]\Lambda [D^{-1} - D^{-1}G (G'D^{-1}G)^{-1} G'D^{-1}],$$

where $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E}[YY']$.

Proof. See the appendix.

The distribution of the Lagrangian multiplier provides directions for improvements of the model. If the multiplier of one particular asset is very large, then it means that the model SDF has to be significantly modified to correctly price this particular asset.

Proposition 3. [Pricing Errors] Suppose Assumptions A.1 to A.9 in the appendix hold. Then under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, the standardized pricing errors of individual assets, $\sqrt{T}\mathbb{E}_T(H(\theta)Y - X)|_{\theta=\hat{\theta}^+}$, have an asymptotic normal distribution with zero mean and covariance matrix

$$[I - G (G'D^{-1}G)^{-1} G'D^{-1}]\Lambda [I - D^{-1}G (G'D^{-1}G)^{-1} G'],$$

where $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E}[YY']$.

Proof. See the Appendix.

The distribution of the pricing errors helps to identify whether a given model has difficulties in pricing a specific asset.

3. Simulation Evidence on Finite Sample Performances

3.1 Simulation Designs

A good econometric test should have reliable finite sample performances. So before applying the above asymptotic results in empirical analysis, we provide simulation studies on the finite sample performances of δ - and δ^+ -based tests.

Suppose a SDF model has the following representation

$$H_t = b'F_t,$$

where b is an $K \times 1$ vector of market prices of risk and F_t is an $K \times 1$ vector of risk factors. We obtain simulated random samples of H_t and its associated asset returns, i.e., $D_t = (F_{it}, Y_{jt})'$, for $t = 1, \dots, T$, $i = 1, \dots, K$ (the number of factors), and $j = 1, \dots, N$ (the number of assets), from a $(K + N)$ -dimensional multivariate normal distribution

$$D_t \sim N(\mu_D, \Sigma_D),$$

where μ_D is an $(K + N) \times 1$ vector of the mean values of $(F_t, Y_t)'$ and Σ_D is an $(K + N) \times (K + N)$ covariance matrix of (F_t, Y_t) .

To make our simulation evidence empirically relevant, we choose simulation designs that are consistent with empirical studies in later sections. Specifically, we allow the market prices of risk b , the mean values of the risk factors, i.e., the first K elements of μ_D , and the covariance matrix Σ_D to be estimated from empirical data. On the other hand, the expected returns of the N assets are determined by the asset pricing model we choose. That is, if H_t can correctly price all test assets, i.e., $\mathbb{E}(H_t Y_t) = 1$, then the expected returns of the N assets can be written as:

$$\mathbb{E}(Y_t) = \frac{1 - \text{cov}(Y_t, H_t)}{\mathbb{E}(H_t)}.$$

Note that the above simulation procedure only guarantees that the expected returns of the test assets are determined by the pricing kernel we choose. The pricing kernel itself, however, can take negative values with positive probability.

Based on the above procedure, we generate 500 random samples of D_t with different combinations of N (number of assets) and T (number of time-series observations). We choose $N = 25$ (10) to mimic the Fama-French 25 size/BM portfolios (the ten hedge fund portfolios) considered in empirical analysis in Section 4 (5).¹⁵ We choose $T = 100, 300$, and 600, where 600 represents the typical number of monthly observations we have in standard empirical asset pricing studies.

¹⁵More detailed descriptions of the Fama-French 25 portfolios can be found in Section 4. The ten hedge fund portfolios are constructed in the following way. We first consider all hedge funds in the TASS database that follow a few strategies that exhibit option-like returns between January 1994 and September 2003. Then we group these funds into ten portfolios based on the ten broad investment styles they follow. More detailed discussions of the ten hedge fund portfolios can be found in Section 5.

For each simulated random sample, we estimate model parameters and conduct specification tests based on δ - and δ^+ -based tests. Then we report rejection rates based on the asymptotic critical values at the 1%, 5%, and 10% significance levels for the two tests. If the tests have good size performances, then the rejection rates for a correctly specified model at the above three critical values should be close to 1%, 5% and 10%, respectively. If the tests have good power performances, then the rejection rates for a misspecified model should be close to 1.

We examine the finite sample size performances of the two tests using the CAPM, whose SDF equals

$$H_t^{\text{CAPM}} = b_0 + b_1 r_{\text{MKT},t}. \quad (12)$$

One reason we choose the CAPM for our size simulations is that the SDFs of most other linear factor models we consider take negative values with positive probabilities at empirically estimated parameter values. In contrast, the probability that the SDF of the CAPM takes negative values is zero based on parameters values and historical data used in later two empirical applications.¹⁶ Therefore, we treat the CAPM as the “true” model in our size simulations because it is arbitrage free and by construction it correctly prices all the simulated assets. We choose the parameters of the CAPM to be $(b_0, b_1) = (1.01, -3.02)$ and $(1.04, -6.95)$ for $N = 25$ and $N = 10$, respectively. These are the parameter values estimated using the Fama-French 25 portfolios and the ten hedge fund portfolios in later two empirical sections, respectively.

To examine the finite sample power performances of the two tests, we consider the Fama-French three-factor model (FF3) with the following SDF

$$H_t^{\text{FF3}} = b_0 + b_1 r_{\text{MKT},t} + b_2 r_{\text{SMB},t} + b_3 r_{\text{HML},t}, \quad (13)$$

where $r_{\text{SMB},t}$ and $r_{\text{HML},t}$ are the return differences between small and big firms, and high and low BM firms, respectively. FF3 is a widely used model in the literature and its SDF at empirically estimated parameter values tends to take negative values with higher probabilities than that of the CAPM. Therefore, we use FF3 to examine the powers of the two tests in rejecting those models that have small pricing errors but are not arbitrage free. In our power simulations, we consider only $N = 10$ to mimic hedge fund returns. This is because the SDF of FF3 estimated using the Fama-French 25 portfolios takes negative values with zero probability based on historical returns of the three systematic risk factors. We choose the parameters of FF3 to be $(b_0, b_1, b_2, b_3) = (0.84, 1.04, 22.74, 29.70)$, parameter values estimated using monthly returns of the ten hedge fund portfolios between January 1994 and September 2003.

¹⁶Rigorously speaking, given that $r_{\text{MKT},t}$ follows a normal distribution in our simulation, there has to be a positive probability for H_t^{CAPM} to take negative values. However, given the parameter values we choose and the historical market returns in our two empirical applications, the probability that H_t^{CAPM} takes negative values is negligible for our simulation studies.

3.2 Finite Sample Size and Power Performances

Panel A of Table 1 reports the size performances of δ - and δ^+ -based tests based on simulated data that mimic the Fama-French 25 portfolios. Specifically, it reports rejection rates based on the asymptotic critical values at the 1%, 5%, and 10% significance levels for the two tests. If the two tests have good finite sample performances, then the rejection rates should be close to the asymptotic significance levels. It is clear that both tests have relatively poor finite sample performances for $T = 100$ months. The rejection rates for both tests are close to 20%, 35%, and 50% at the 1%, 5%, and 10% asymptotic critical values, respectively. The performances of both tests become much better for $T = 300$ months: The rejection rates for both tests are about 4%, 12%, and 20% at the 1%, 5%, and 10% asymptotic critical values, respectively. For $T = 600$ months, the sample size used in our empirical analysis in Section 4, the performances of both tests become reasonably good: The rejection rates for both tests are about 2%, 9%, and 15% at the 1%, 5%, and 10% asymptotic critical values, respectively. These rejection rates are fairly close to the asymptotic significance levels. Consistent with our theoretical predictions, we see that both tests have very similar size performances.

Panel B of Table 1 reports the size performances of δ - and δ^+ -based tests based on simulated data that mimic hedge fund returns. It reports the rejection rates for both tests based on the asymptotic critical values at the 1%, 5%, and 10% significance levels. With only ten assets involved, both tests have much better finite sample performances at each T than before. In fact, for $T = 100$ months, the sample size used in our empirical analysis in Section 4, the rejection rates for both tests are about 4%, 10%, and 20% at the 1%, 5%, and 10% asymptotic critical values, respectively. These rejection rates are fairly close to that in Panel A for $T = 300$ months. When $T = 300$ months, the performances of both tests become very good, with rejection rates equal to 1%, 5-6%, and 12% at the 1%, 5%, and 10% asymptotic critical values, respectively. When $T = 600$ months, the rejection rates for both tests are very close to the asymptotic significance levels.

Panel C of Table 1 reports the power performances of δ - and δ^+ -based tests based on simulated data that mimic hedge fund returns. It reports rejection rates based on the asymptotic critical values at the 1%, 5%, and 10% levels for the two tests. While the two tests have similar size performances, they have dramatically different powers in rejecting misspecified models. When T increases from 100 to 300 and 600 months, the rejection rates of the δ -based test become very close to the significance levels at their corresponding asymptotic critical values. Thus, the δ -based test fails to reject FF3 even though its SDF takes negative values with a high probability. In contrast, the rejection rates of the δ^+ -based test are much higher at each critical value and when $T = 300$ and 600 months, the rejection rates become close to 100%. The δ^+ -based test overwhelmingly rejects FF3 because it violates the no-arbitrage constraint.

In summary, our simulation evidence shows that for typical sample sizes considered in the current literature, both δ - and δ^+ -based tests have similar and reasonably good finite sample size performances. However, the two tests have very different powers in rejecting misspecified models. δ -based test fails to reject those models that have small pricing errors, but whose SDFs take negative values with high probabilities. In contrast, δ^+ -based test overwhelmingly rejects those models.

4. Empirical Application I: Fama-French 25 Portfolios

To illustrate the importance of the no-arbitrage constraint in traditional asset pricing applications, we evaluate several well-known asset pricing models using the Fama-French 25 portfolios using δ - and δ^+ -based tests. We do not consider consumption-based models because they perform quite poorly in capturing the cross-sectional differences in stock returns.

4.1 Data and Asset Pricing Models

Table 2 provides summary statistics for the monthly returns of the 25 portfolios in excess of one-month T-bill rates between January 1952 and December 2002. It is similar to Table 2 of Fama and French (1993), which covers a shorter period between January 1963 and December 1991. During our longer sample period, most average returns are higher, except that of low BM firms. Since the standard errors are smaller, the t -statistics are larger except for low BM firms. There are considerable dispersions in the average returns across the 25 portfolios. The average annualized returns range from 2.5% for the smallest firms with lowest BM ratios to 13.1% for the smallest firms with highest BM ratios. Within size quintiles, there is a nearly monotonic increase in average returns as BM increases. Within BM quintiles, the average returns of the smallest firms are larger than that of the largest firms, except for the lowest BM quintile. However, there is no monotonic relation in average returns across size quintiles.

To test the conditional implications of asset pricing models, we also examine scaled returns by multiplying the returns of the 25 portfolios by default premium (hereafter, DEF), a commonly used conditioning variable. DEF is defined as the yield spread between Baa and Aaa rated corporate bonds and is obtained from Federal Reserve Bank. Panel A of Figure 1 provides a time series plot of DEF.

We consider several widely studied asset pricing models and their conditional versions to capture time-varying risk premiums. Specifically, we use industrial production (IP hereafter) from Citibase as a state variable because it is a well-documented business cycle indicator. We apply the Hodrick and Prescott (1997) filter recursively to better measure the cyclical component of the IP series. We initiate the filter by using the first 5 years (1947-1951) of data. Consequently, the first element of our cycle is December 1951. We then use the procedure recursively on all available data to obtain the subsequent elements for the cyclical series. This method guarantees that each element is in the information set at t . Panel B of Figure 1 presents a time-series plot of

IP. Following Cochrane (1996), we scale the original factors by IP and in total we consider nine models in our analysis.¹⁷

The first model we consider is the CAPM whose SDF is given in (12). We consider two variations of the conditional version of the CAPM, which we denote as CAPM+IP and CAPM*IP to reflect the different ways in which the conditional information is introduced. The SDF of CAPM+IP is

$$H_t^{\text{CAPM+IP}} = b_0 + b_1 r_{\text{MKT},t} + c_0 z_{t-1},$$

where z_{t-1} is the realization of the state variable at $t - 1$, i.e., the cyclical component of IP. The SDF of CAPM*IP is

$$H_t^{\text{CAPM*IP}} = b_0 + b_1 r_{\text{MKT},t} + c_0 z_{t-1} + c_1 z_{t-1} r_{\text{MKT},t}.$$

This is equivalent to allowing b_0 and b_1 in H_t^{CAPM} to be linear functions of the state variable as suggested by Cochrane (1996). The above three versions of the CAPM are not arbitrage free because their SDFs can take negative values with positive probabilities.

The next model we consider is FF3 whose SDF is given in (13). We also consider two conditional versions of FF3, FF3+IP and FF3*IP, with the following SDFs:

$$\begin{aligned} H_t^{\text{FF3+IP}} &= b_0 + b_1 r_{\text{MKT},t} + b_2 r_{\text{SMB},t} + b_3 r_{\text{HML},t} + c_0 z_{t-1}, \\ H_t^{\text{FF3*IP}} &= b_0 + b_1 r_{\text{MKT},t} + b_2 r_{\text{SMB},t} + b_3 r_{\text{HML},t} \\ &\quad + c_0 z_{t-1} + c_1 z_{t-1} r_{\text{MKT},t} + c_2 z_{t-1} r_{\text{SMB},t} + c_3 z_{t-1} r_{\text{HML},t}. \end{aligned}$$

This type of extension of the Fama-French model has been explored by Kirby (1997).

Fama and French (1996) identify three bond market factors that help explain cross-sectional stock returns. These factors are STERM (yield spread between 1-year and 1-month government bonds), LTERM (yield spread between 30-year and 1-year government bonds) and DEF. We incorporate the three additional factors into FF3 to obtain a Fama-French six-factor model, denoted as FF6.¹⁸

Finally, we consider a linearized version of Campbell's (1996) log-linear asset pricing model, denoted as CAM, and its time-varying extension, CAM*IP. The intertemporal asset pricing model of Campbell (1996) allows for time-varying investment opportunities and variables that can predict market returns can be considered as risk factors. Other than the market factor, the original model includes four additional factors: the labor income factor, LBR, constructed as the monthly growth rate in real labor income (from Citibase); dividend yield on the market portfolio, DIV (from CRSP); the relative T-bill rate, RTB, calculated as the difference between the one-month

¹⁷Other than DEF and IP, we have used other popular conditioning variables and obtain similar results.

¹⁸The SDFs of FF6 and the remaining models have similar forms as that of CAPM and FF3 and are omitted to avoid repetition.

T-bill rate and its one-year backward moving average (from CRSP); and the term structure factor, LTERM.¹⁹ In the conditional version, we scale the original factors by the state variable IP.

An alternative approach of imposing the no-arbitrage constraint that has been considered in the literature is to truncate the SDF of a linear factor model when it turns negative. For example, the SDF of the truncated CAPM would be

$$H_t^{\text{Trun, CAPM}} = \max [0, b_0 + b_1 r_{\text{MKT},t}].$$

In our empirical analysis, we also evaluate the truncated versions of all nine models based on the two HJ-distances. Through this exercise, we examine whether we reach different conclusions for truncated models using the first and second HJ-distances.

4.2 Empirical Results

Table 3 reports results of specification tests of the nine models (both original and truncated versions) using the original 25 portfolios and the 25 portfolios scaled by DEF.²⁰ We first report $\delta_T(\hat{\theta})$ and $\delta_T^+(\hat{\theta}^+)$ and their differences. We then report the probability that $H_t(\hat{\theta})$ takes negative values, $p(H < 0)$. As a truncated model cannot take negative values by definition, we omit such information for truncated models. Finally, we report p -values of δ - and δ^+ -based tests.

Panel A of Table 3 reports results based on the original 25 portfolios. For CAPM and its two conditional variations, $\delta_T(\hat{\theta})$ and $\delta_T^+(\hat{\theta}^+)$ are very similar to each other: $\delta_T^+(\hat{\theta}^+)$ is about 2% larger than $\delta_T(\hat{\theta})$. For all three models, $p(H < 0)$ is close to zero. Thus, the no-arbitrage constraint does not make a big difference in inferences of the three CAPM models. Given the widely recognized failure of the CAPM in capturing the size and value premiums, it is not surprising that all three models are easily rejected by both tests.

The results of FF3 and its two time-varying extensions are quite similar to that of the three CAPM models, although the HJ-distances of the FF3 models are much smaller than that of the CAPM models. For these three models, $\delta_T^+(\hat{\theta}^+)$ is about 2 to 3% larger than $\delta_T(\hat{\theta})$ and $p(H < 0)$ is close to zero. All three models are also rejected by both tests. Again the no-arbitrage constraint does not significantly affect the inferences of the three FF3 models.

The no-arbitrage constraint makes a substantial difference in the inferences of FF6. For example, $\delta_T^+(\hat{\theta}^+)$ is about 11% bigger than $\delta_T(\hat{\theta})$, and the probability $H_t^{\text{FF6}}(\hat{\theta})$ takes negative values is about 12%. Most importantly, we reach very different conclusions on model performance using the two tests. Although the JW test fail to reject FF6 at conventional significance levels (p -value=11%), our test easily rejects the model (p -value=0%). Without the no-arbitrage constraint, we could have mistakenly concluded that the model does a good job in pricing the 25

¹⁹The original model of Campbell (1996) has a pricing proxy in the form of $y_t = \exp(-f_t'b)$. We consider its linearized version to illustrate the importance of the no-arbitrage constraint in studying linear factor models.

²⁰For brevity, we do not report the estimates of model parameters, Lagrangian multipliers, and pricing errors of individual assets. These results are available from the authors upon request.

portfolios based on $\delta_T(\hat{\theta})$. However, such performance is achieved at the expense of a relatively high probability of $H_t^{\text{FF6}}(\hat{\theta})$ taking negative values.

The no-arbitrage restriction is also important for CAM, and its conditional version CAM*IP. For example, $\delta_T^+(\hat{\theta}^+)$ is 25% and 30% bigger than $\delta_T(\hat{\theta})$ for the two models, respectively. The probabilities that $H_t^{\text{CAM}}(\hat{\theta})$ and $H_t^{\text{CAM*IP}}(\hat{\theta})$ take negative values are also quite substantial and equal to 17% and 20%, respectively. Most interestingly, neither models can be rejected by the JW test: the p -values are above 20% for both models. However, our test overwhelmingly rejects both models (the p -values are close to zero for both models). The pricing errors of the two models measured by the first HJ-distance are the smallest among the nine models. However, the “good” performances of the two models measured by $\hat{\delta}_T(\hat{\theta})$ are mainly driven by the fact that $H_t^{\text{CAM}}(\hat{\theta})$ and $H_t^{\text{CAM*IP}}(\hat{\theta})$ take negative values with high probabilities.

In Panel B of Table 3, we obtain qualitatively similar results using the 25 portfolios scaled by DEF. The no-arbitrage constraint significantly affects the inferences of FF6 and the two Campbell models. The second HJ-distances are much larger than the first HJ-distances for the three models (21%, 26%, and 37% larger, respectively). The SDFs of all models also have quite high probabilities of taking negative values. Although the JW test cannot reject the three models (p -values equal 7%, 42%, and 48%, respectively), the three models are strongly rejected by our test (p -values are all close to zero).

The corresponding results for the truncated versions of the nine models in Panels C and D of Table 3 are similar to that of the original models. For the truncated versions of FF6, CAM, and CAM*IP, the second HJ-distances are about 10 to 20% bigger than the first HJ-distances. Although certain models cannot be rejected by the first HJ-distance, all models are rejected based on the second HJ-distance. One thing worth mentioning is that though truncation guarantees a model’s SDF to be nonnegative, it does not guarantee that the model also has small pricing errors as shown by $\delta_T(\hat{\theta})$ s of FF6, CAM, and CAM*IP. This is because a truncated model is not as flexible as the original one and tends to have bigger pricing errors. Our results on truncated models illustrate the important point that truncation is different from our approach of imposing the no-arbitrage constraint: One can still reach misleading conclusions using the first HJ-distance for truncated models.

The two HJ-distances also lead to very different estimated model SDFs. We present time series plots of $H_t(\hat{\theta})$ and $H_t(\hat{\theta}^+)$ for FF6 and CAM*IP in Panels A and B of Figure 2, respectively. It is clear that $H_t(\hat{\theta})$ takes negative values much more frequently than $H_t(\hat{\theta}^+)$ for both models. In particular, $H_t(\hat{\theta})$ is negative around 1976, 1982, 1992 and after 2001, which coincide with NBER business cycle troughs. This suggests that both models have difficulties in pricing the Fama-French 25 portfolios during economic downturns because their SDFs have to take negative values to reduce the pricing errors. On the other hand, $H_t(\hat{\theta}^+)$ takes negative values only in

rare occasions, suggesting that for a given sample, linear factor models can be made closer to be arbitrage free if they are estimated using the right objective function.

Panels C and D of Figure 2 present time series plots of deviations from true asset pricing models for FF6 and CAM*IP, respectively. For each model, based on equation (10), we use $[H_t(\hat{\theta}^+) - \hat{\lambda}'Y_t]^+$ as a proxy for the true asset pricing model and denote it as m_t^* . Thus deviations from \mathcal{M}_+ for each model under δ and δ^+ become $m_t^* - H_t(\hat{\theta})$ and $m_t^* - H_t(\hat{\theta}^+)$, respectively. To become a member of \mathcal{M}_+ , both models need much more dramatic adjustments for $H_t(\hat{\theta})$ than $H_t(\hat{\theta}^+)$. Interestingly, for both models, even though $\delta(\hat{\theta})$ is much smaller than $\delta(\hat{\theta}^+)$, $H_t(\hat{\theta})$ is actually further away from \mathcal{M}_+ than $H_t(\hat{\theta}^+)$ across most states of the economy.

To summarize, the different objective functions, i.e., the first and second HJ-distances, used for model estimation and evaluation, could lead to dramatically different conclusions on model performance. Models estimated using the first HJ-distance tend to take negative values more frequently than those estimated using the second HJ-distance. More dramatic adjustments are also needed to make these models admissible than those estimated using the second HJ-distance. Most importantly, without the no-arbitrage constraint, we have the illusion that certain models can price the Fama-French 25 portfolios well based on the first HJ-distance.²¹ The analysis in this section shows that even for applications that do not explicitly involve derivatives, the no-arbitrage constraint could still make significant differences in empirical asset pricing studies.

5. Empirical Application II: Hedge Fund Returns

In this section, we examine the importance of the no-arbitrage constraint for asset pricing applications that involve derivatives. We focus on hedge fund returns because of frequent use of dynamic trading strategies and derivatives by hedge funds.

5.1 Hedge Fund Returns and Asset Pricing Models

We consider a few hedge fund strategies that exhibit option-like returns, specifically trend-following, risk arbitrage, and equity derivative arbitrage. Fung and Hsieh (2001) show that the returns of trendfollowing funds can be captured reasonably well by a lookback straddle. Mitchell and Pulvino (2001) show that risk arbitrage generates moderate returns in flat or rising market environments, but large negative returns in declining markets. In fact, risk arbitrage is akin to writing uncovered index put options. We also consider funds that are involved in equity derivative arbitrage because they obviously invest in derivatives.

The hedge fund data used in our analysis are obtained from TASS. Among all the datasets that have been used in existing hedge fund studies, the TASS database is probably the most comprehensive one. It covers over 4,000 funds from November 1977 to September 2003, which

²¹At a deeper level, all models are approximations of reality and therefore are wrong. However, the significance of the no-arbitrage constraint is not only that it allows us to reject certain models that previous tests fail to reject, but more importantly it provides more realistic assessments of the inadequacies of existing models. Such insights are helpful in guiding future research efforts in indentifying better models.

are classified into “live” and “graveyard” funds. The “graveyard” database did not exist before 1994. To mitigate the problem of survivorship bias, we include both “live” and “graveyard” funds and we restrict our sample to the period between January 1994 and September 2003. The database provides monthly net-of-fee returns and net asset values. In our sample, 557 funds use trendfollowing strategy, 559 funds use risk arbitrage, and 164 funds use equity derivative arbitrage.

It is numerically difficult to fit asset pricing models to returns of more than one thousand hedge funds. Instead we focus on the returns of hedge funds portfolios. The portfolio approach helps to remove noises in individual hedge funds returns, and increases the precision of parameter estimates and the power of statistical tests. Even though these funds follow one of the above three specific strategies, each of them also belongs to one of the eleven broad investment styles described by TASS. As documented by Brown and Goetzmann (2003), the most natural way to group hedge funds is by their styles because the styles greatly capture the cross-sectional return differences.

Panel A of Table 4 presents distributional information of the three strategies among ten broad investment styles.²² The trendfollowers are mainly concentrated in CTAs, although some belong to fund of funds, global macro, and long/short equity. The risk arbitrage funds are concentrated in event driven, and some belong to convertible arbitrage, fund of funds, long/short equity, emerging market and equity neutral. Equity derivative arbitrage is used by different styles. Panel B of Table 5 reports summary statistics of the style portfolio returns. The average monthly returns of the ten portfolios range from 0.97% of fixed-income arbitrage to 1.99% of global macro. Consistent with their investment approaches, the styles that take directional bets tend to have higher return volatility than those that take relative bets.

The asset pricing models we consider are similar to those in the previous section for the Fama-French 25 portfolios. However, since we only have ten portfolios to price, we choose the more parsimonious models to avoid overfitting the data. We only consider two CAPM models, CAPM and CAPM+IP, and two FF3 models, FF3 and FF3+IP. Following Agarwal and Naik (2004), to capture the option-like returns of hedge funds, we consider an option-based model, OPT, and its conditioning version, OPT+IP. In OPT, in addition to the market factor, we include two additional factors from the option market. The first factor is the return on at-the-money S&P 500 index straddles, with a time-to-maturity of 20 to 50 days. It captures the aggregate volatility risk (Ang, Hodrick, Xing, and Zhang 2006). The second factor is the return on out-of-the-money (moneyness 0.92) S&P 500 index puts that expire in 20 to 50 days. It captures the jump risk in market index. The data for option returns are obtained from Optionmetrics. The second model OPT+IP includes IP as a factor. Similar as before, we also consider the truncated versions of the above six models.

²²Dedicated short is excluded because none of the funds belongs to this category.

5.2 Empirical Results

Panel A (B) of Table 5 contains results of specification tests for the original (truncated) models.²³ We focus our discussions on Panel A, because we reach similar conclusions from both Panels. For CAPM and CAPM+IP, $\delta_T^+(\hat{\theta}^+)$ is much bigger than $\delta_T(\hat{\theta})$. The difference is especially large for CAPM+IP: $\delta_T^+(\hat{\theta}^+)$ is 67% bigger than $\delta_T(\hat{\theta})$. Although the JW test rejects CAPM (p -value=0%), it fails to reject CAPM+IP (p -value=57%). However, the “superior” performance of CAPM+IP comes at the expense that it allows arbitrage opportunities: The probability that $H_t^{\text{CAPM+IP}}(\hat{\theta})$ takes negative values is as high as 43%. As a result, CAPM+IP is easily rejected by our test. Therefore, without the no-arbitrage constraint, we could have reached the misleading conclusion that CAPM+IP is able to satisfactorily price hedge fund portfolio returns.

We obtain similar results for FF3 and FF3+IP: $\delta_T^+(\hat{\theta}^+)$ is 19% and 85% bigger than $\delta_T(\hat{\theta})$, respectively. The SDFs of both models take negative values with high probabilities: 17% and 43%, respectively. FF3+IP has $\delta_T(\hat{\theta})$ that is much smaller than that of all the other models and FF3+IP cannot be rejected by the JW test: the p -value is 56%. This result again gives the illusion that FF3+IP can satisfactorily explain the returns of the ten hedge fund portfolios. However, with the no-arbitrage constraint, we conclude that both models fail to explain hedge fund returns.

For the two option-based models, the probabilities that the SDFs take negative values are 17% and 40%, respectively, which are similar to that of the FF models. The estimated second HJ-distances are about 20% and 67% bigger than the first HJ-distances for the two models, respectively. Again, ignoring the no-arbitrage constraint, OPT+IP would be accepted by the JW test (p -value=31%). However, with the no-arbitrage constraint, both models are overwhelmingly rejected by our test. Recent studies, such as Agarwal and Naik (2004), include option returns in linear factor models to evaluate hedge fund performance. However, the results here show that this approach, although a good improvement over existing methods, does not completely resolve the problem, because there are still arbitrage opportunities in the option-based models.

We provide the time-series plots of $H_t(\hat{\theta})$ and $H_t(\hat{\theta}^+)$ for FF3+IP and OPT+IP in Panels A and B of Figure 3, respectively. It is clear that $H_t(\hat{\theta})$ takes negative values much more frequently than $H_t(\hat{\theta}^+)$ for both models. In particular, the $H_t(\hat{\theta})$ s of both models take negative values during the time periods of 1996-1998, 1999-2000, and after 2002. On the other hand, the $H_t(\hat{\theta}^+)$ s take negative values only in rare occasions, again suggesting that linear factor models can be made close to be arbitrage free if they are estimated using the right objective function.

Panels C and D of Figure 3 present time series plots of deviations from true asset pricing models for FF3+IP and OPT+IP, respectively. Again we use $m_t^* = [H_t(\hat{\theta}^+) - \hat{\lambda}'Y_t]^+$ as a proxy for the true asset pricing model and deviations from \mathcal{M}_+ for each model under δ and δ^+ become

²³For brevity, we do not report the estimates of model parameters, Lagrangian multipliers, and pricing errors of individual assets. These results are available from the authors upon request.

$m_t^* - H_t(\hat{\theta})$ and $m_t^* - H_t(\hat{\theta}^+)$, respectively. It is clear that both models need much more dramatic adjustments for $H_t(\hat{\theta})$ than $H_t(\hat{\theta}^+)$ to become a member of \mathcal{M}_+ . Interestingly, even though $\delta(\hat{\theta})$ is much smaller than $\delta(\hat{\theta}^+)$ for both models, $H_t(\hat{\theta})$ is actually further away from \mathcal{M}_+ than $H_t(\hat{\theta}^+)$ across most states of the economy.

For the Fama-French 25 portfolios, the no-arbitrage constraint makes a significant difference only for the most sophisticated models, such as FF6 and CAM+IP. However, for hedge fund portfolios, the no-arbitrage constraint makes a huge difference for most models. To price hedge fund returns, the $H_t(\hat{\theta})$ s of most models have to take negative values with high probabilities. Therefore, without the no-arbitrage constraint, one could reach very misleading conclusions on the performances of these asset pricing models in capturing hedge fund returns.

We emphasize that the main purpose of the exercise in this section is to illustrate the importance of the no-arbitrage constraint for applications that involve hedge fund returns. We do not draw any conclusion on whether hedge funds can deliver abnormal returns because such conclusions are obviously model dependent. Even though we use some of the most popular models in the literature, they are far from being perfect and some of them even fail to price the Fama-French 25 portfolios. It is very likely that we will get very different estimates of abnormal returns using other models. However, the bottom line here is that no matter what models we use, the no-arbitrage constraint is likely to be very important for evaluating hedge fund performances because of the option-like features of hedge fund returns. The new econometric methods developed here enable researchers to incorporate the no-arbitrage constraint into their future empirical studies.

6. Conclusion

In this paper, we have developed econometric methods for evaluating asset pricing models based on the second HJ-distance, which unlike the first HJ-distance, explicitly requires that a correct asset pricing model has to be arbitrage free. Our methods are natural extensions of existing ones based on the first HJ-distance and make it very convenient to incorporate the no-arbitrage constraint in empirical applications. Simulation studies show that the finite sample performance of our methods is (i) comparable with that of existing methods based on the first HJ-distance; and (ii) reasonably good for typical sample sizes considered in the literature. We illustrate the importance of the no-arbitrage constraint for evaluating asset pricing models using the Fama-French Size/BM portfolios and hedge fund portfolios that exhibit option-like returns. While we fail to reject certain models based on the first HJ-distance, we overwhelmingly reject these models based on the second HJ-distance because their SDFs take negative values with high probabilities and therefore violate the no-arbitrage constraint.

REFERENCES

- Agarwal, Vikas and Narayan Y. Naik, 2004, Risks and portfolio decisions involving hedge funds, *Review of Financial Studies* forthcoming.
- Ang, Andrew, Hodrick, Robert J., Xing, Yuhang, and Xiaoyan Zhang, 2006, The cross section of volatility and expected returns, *Journal of Finance* forthcoming.
- Bansal, Ravi and S. Viswanathan, 1993, No arbitrage and arbitrage pricing: A new approach, *Journal of Finance* 48, 1231-1262.
- Ben Dor, Arik and Ravi Jagannathan, 2002, Understanding mutual fund and hedge fund styles using return based style analysis, Northwestern University working paper.
- Brown, Stephen and William Goetzmann, 2003, Hedge funds with style, *Journal of Portfolio Management* 29, 101-112.
- Campbell, John, 1996, Understanding risk and return, *Journal of Political Economy* 104, 298-345.
- Cochrane, John, 2001, *Asset Pricing*, Princeton University Press.
- Dudley, R. M., 1981, Donsker classes of functions, *Statistics and Related Topics* 341-352, North-Holland, Amsterdam.
- Dybvig, Philip H. and Jonathan E. Ingersoll, 1982, Mean-variance theory in complete markets, *Journal of Business* 55, 233-251.
- Dybvig, Philip H. and Stephen A. Ross, 1985, Differential information and performance measurement using a security market line, *Journal of Finance* 40, 383-399.
- Dybvig, Philip H. and Stephen A. Ross, 2003, Arbitrage, state prices and portfolio theory, *Handbook of the Economics of Finance*, Elsevier, North Holland.
- Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Fung, William and David A. Hsieh, 2001, The risk in hedge fund strategies: Theory and evidence from trend followers, *Review of Financial Studies* 14, 313-341.
- Glosten, Lawrence R. and Ravi Jagannathan, 1994, A contingent claim approach to performance evaluation, *Journal of Empirical Finance* 1, 133-160.

- Goetzmann, William, Ingersoll, Jonathan, Spiegel, Matthew, and Ivo Welch, 2002, Sharpening sharpe ratios, NBER Working Paper 9116 (August).
- Grinblatt, Mark and Sheridan Titman, 1989, Portfolio performance evaluation: Old issues and new insights, *Review of Financial Studies* 2, 393–421.
- Hansen, Lars Peter, Large sample properties of generalized method of moments estimators, 1982, *Econometrica* 50, 1029-1054.
- Hansen, Lars Peter, John Heaton, and Erzo G. J. Luttmer, 1995, Econometric evaluation of asset pricing models, *Review of Financial Studies* 8, 237–274.
- 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.
- Hodrick, Robert J. and Edward C. Prescott, 1997, Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit and Banking* 29, 1–16.
- Jagannathan, Ravi and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.
- Jegadeesh, N. and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.
- Kan, Raymond and Guofu Zhou, 2002, Hansen-Jagannathan Distance: Geometry and Exact Distribution, Washington University in St. Louis working paper.
- Kirby, Chris, 1997, Measuring the predictable variation in stock and bond returns, *Review of Financial Studies* 10, 579–630.
- Le Cam L., 1986, *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York.
- Merton, Robert C., 1981, On market timing and investment performance I: An equilibrium theory of values for markets forecasts, *Journal of Business* 54, 363–406.
- Mitchell, Mark and Todd Pulvino, 2001, Characteristics of risk and return in risk arbitrage, *Journal of Finance* 56, 2135–2175.
- Pakes, Ariel and David Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica* 57, 1027-1057.

Pollard, David, 1982, A central limit theorem for k -means clustering, *Annals of Probability* 10, 919-926.

Pollard, David, 1989, Asymptotics via empirical processes, *Statistical Science* 4, 341-354.

Wang, Zhenyu and Xiaoyan Zhang, 2004, The role of arbitrage in the empirical evaluation of asset pricing models, Cornell University working paper.

Mathematical Appendix

In this appendix, we establish the econometric theory of asset pricing tests based on the second HJ-distance. Specifically, we develop the asymptotic distributions of the second HJ-distance and related statistics under the null hypothesis that a given asset pricing model is correctly specified.

Suppose the uncertainty of the economy is described by a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ for $t = 0, 1, \dots, T$. Suppose we also have the following time series observations of asset payoffs, prices, and model SDFs, $\{(Y_t, X_{t-1}, H_t(\theta))\}_{t=1}^T$, where Y_t denotes an n -dimensional vector of asset payoffs at time t , X_{t-1} denotes the corresponding n -dimensional asset price vector at time $t - 1$, and the stochastic discount factor $H_t(\theta)$ is assumed to be observable up to an unknown k -dimensional parameter vector θ .

The analysis in this section is based on the following assumptions.

Assumption A.1. The population optimization problem has a unique solution

$$(\theta_0, \lambda_0) \equiv \arg \min_{\theta} \max_{\lambda} \mathbb{E} \phi^+(\theta, \lambda).$$

Assumption A.2. $\mathbb{E} \|Y\|^2 < \infty$, $\mathbb{E} \|X\|^2 < \infty$, $\mathbb{E} \left[\max_{\|\theta - \theta_0\| < C} H(\theta)^2 \right] < \infty$ for some positive C .

Assumption A.3. The SDF $H(\theta)$ is twice differentiable in θ .

Assumption A.4. The set $\{H(\theta) - \lambda'Y = 0\}$ has probability zero under the true probability distribution.

Assumption A.5. The first order derivatives (which exist everywhere)

$$\begin{pmatrix} \frac{\partial}{\partial \theta} \phi^+(\theta, \lambda) \\ \frac{\partial}{\partial \lambda} \phi^+(\theta, \lambda) \end{pmatrix}$$

form a Donsker class for (θ, λ) in a neighborhood of (θ_0, λ_0) .

Assumption A6. The time series $(Y_t, X_{t-1}, H_t(\theta))$ is stationary and ergodic.

The above assumptions are somewhat standard in asymptotic analysis. Assumption A.1 is needed for identification purpose. Assumption A.2 requires all random variables to be square integrable. This is needed for the existence of the asymptotic covariance matrix of the second HJ-distance and exchanging differentiation and expectation operations. Assumption A.3 is a smoothness assumption needed for quadratic Taylor series expansion. Assumption A.4 guarantees that the set of non-differentiable points of the criterion function is not too big so that “*differentiation in quadratic mean*” will hold. It should hold for most models in the existing literature. Assumption A.5 ensures that central limit theorem holds for the first derivatives of ϕ^+ . A set \mathfrak{F} of functions is called a Donsker class for P if a functional central limit theorem holds for the sequence of empirical processes $\sqrt{T}(\mathbb{E}_T - \mathbb{E})f$ for $f \in \mathfrak{F}$ (see Dudley 1981). A key property of a Donsker class is that for every given $\varepsilon > 0$, $\eta > 0$, there exists a $\gamma > 0$ and an n_0 such that, for all $n > n_0$

$$P \left\{ \sup_{[\gamma]} |\sqrt{T}(\mathbb{E}_T - \mathbb{E})f_1 - \sqrt{T}(\mathbb{E}_T - \mathbb{E})f_2| > \eta \right\} < \varepsilon.$$

$[\gamma]$ means that the supremum runs over all pairs of functions f_1 and f_2 in \mathfrak{F} that are less than γ apart in $L^2(P)$ norm. This property is needed to justify the first order approximation to the random component $(\mathbb{E}_T - \mathbb{E})\phi^+(\theta, \lambda)$. Assumption A.6 enables inferences of population distribution using time series counterparts. In certain applications, we might have to transform the original price and/or payoff series to satisfy Assumption A.6. For example, although stock prices generally are not stationary and ergodic, stock returns generally are.

The following Lemma justifies local asymptotic quadratic (LAQ) expansion of the objective function $\mathbb{E}_T\phi^+(\theta, \lambda)$ along the lines of Pollard (1982).

Lemma 1. Suppose Assumptions A.1 to A.6 hold. Then we have the following local asymptotic quadratic representation for $\mathbb{E}_T\phi^+(\theta, \lambda)$ around (θ_0, λ_0) :

$$\begin{aligned} & \mathbb{E}_T\phi^+(\theta, \lambda) \\ &= \mathbb{E}\phi^+(\theta_0, \lambda_0) + (\mathbb{E}_T - \mathbb{E})\phi^+(\theta_0, \lambda_0) + \begin{pmatrix} A \\ B \end{pmatrix}' \begin{pmatrix} U \\ V \end{pmatrix} + \frac{1}{2} \begin{pmatrix} U \\ V \end{pmatrix}' \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \quad (\text{LAQ}) \\ &+ o(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|^2) + o_p(\|(\theta, \lambda) - (\theta_0, \lambda_0)\| T^{-1/2}), \end{aligned}$$

where $U \equiv (\theta - \theta_0)$, $V \equiv (\lambda - \lambda_0)$, $A \equiv (\mathbb{E}_T - \mathbb{E})\frac{\partial}{\partial\theta}\phi^+|_{(\theta_0, \lambda_0)}$, $B \equiv (\mathbb{E}_T - \mathbb{E})\frac{\partial}{\partial\lambda}\phi^+|_{(\theta_0, \lambda_0)}$, and

$$\Gamma \equiv \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \equiv \begin{pmatrix} \mathbb{E}\frac{\partial^2}{\partial\theta\partial\theta'}\phi^+ & \mathbb{E}\frac{\partial^2}{\partial\theta\partial\lambda'}\phi^+ \\ \mathbb{E}\frac{\partial^2}{\partial\lambda\partial\theta'}\phi^+ & \mathbb{E}\frac{\partial^2}{\partial\lambda\partial\lambda'}\phi^+ \end{pmatrix}_{|(\theta_0, \lambda_0)}.$$

Proof. Write

$$\mathbb{E}_T\phi^+(\theta, \lambda) = \mathbb{E}\phi^+(\theta, \lambda) + (\mathbb{E}_T - \mathbb{E})\phi^+(\theta, \lambda). \quad (\text{A.1})$$

First consider the first term. Since the SDF model $H(\theta)$ is twice differentiable in θ , it is clear that $\phi^+(\theta, \lambda)$ has continuous first derivatives

$$\begin{pmatrix} \frac{\partial}{\partial\theta}\phi^+(\theta, \lambda) \\ \frac{\partial}{\partial\lambda}\phi^+(\theta, \lambda) \end{pmatrix} = \begin{pmatrix} 2H(\theta)[\nabla H(\theta)] - 2[H(\theta) - \lambda'Y]^+[\nabla H(\theta)] \\ 2[H(\theta) - \lambda'Y]^+Y - 2X \end{pmatrix} \quad (\text{A.2})$$

everywhere, where $\nabla H(\theta)$ is the gradient of $H(\theta)$ with respect to θ . From Assumption A.2., we obtain immediately that

$$\begin{aligned} \frac{\partial}{\partial\theta}\mathbb{E}\phi^+(\theta, \lambda) &= \mathbb{E}\frac{\partial}{\partial\theta}\phi^+(\theta, \lambda), \\ \frac{\partial}{\partial\lambda}\mathbb{E}\phi^+(\theta, \lambda) &= \mathbb{E}\frac{\partial}{\partial\lambda}\phi^+(\theta, \lambda). \end{aligned}$$

Note that

$$\begin{pmatrix} \mathbb{E}\frac{\partial}{\partial\theta}\phi^+(\theta, \lambda) \\ \mathbb{E}\frac{\partial}{\partial\lambda}\phi^+(\theta, \lambda) \end{pmatrix}_{|(\theta_0, \lambda_0)} = 0,$$

because (θ_0, λ_0) solves the population optimization problem $\min_{\theta} \max_{\lambda} \mathbb{E}\phi^+(\theta, \lambda)$.

Next we demonstrate that although the first derivative of ϕ^+ in (A.2) are not differentiable everywhere, they are in fact differentiable in quadratic mean as in Pollard (1982, page 920). Recall that a random function

$f(\alpha)$ is differentiable in quadratic mean with respect to α at α_0 if there exists a random vector Δ such that

$$f(\alpha) = f(\alpha_0) + \Delta'(\alpha - \alpha_0) + \|\alpha - \alpha_0\| R$$

such that

$$\mathbb{E}(R)^2 \rightarrow 0, \text{ as } \alpha \rightarrow \alpha_0$$

Intuitively, this means that $f(\alpha_0) + \Delta'(\alpha - \alpha_0)$ is a good approximation to $f(\alpha)$ around α_0 “on average.” To demonstrate (A.2) is differentiable in quadratic mean, define the remainder term r by the following equation

$$\begin{aligned} & \begin{pmatrix} \frac{\partial}{\partial \theta} \phi^+(\theta, \lambda) \\ \frac{\partial}{\partial \lambda} \phi^+(\theta, \lambda) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial}{\partial \theta} \phi^+ \\ \frac{\partial}{\partial \lambda} \phi^+ \end{pmatrix}_{|(\theta_0, \lambda_0)} + \begin{pmatrix} \frac{\partial^2}{\partial \theta \partial \theta'} \phi^+ & \frac{\partial^2}{\partial \theta \partial \lambda'} \phi^+ \\ \frac{\partial^2}{\partial \lambda \partial \theta'} \phi^+ & \frac{\partial^2}{\partial \lambda \partial \lambda'} \phi^+ \end{pmatrix}_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} + \|\theta, \lambda) - (\theta_0, \lambda_0)\| r \end{aligned}$$

Note that although the second derivatives involved may not exist everywhere, the set of points for which they are not defined has probability zero due to Assumption A.4. Hence as a function in the Hilbert space $L^2(P)$, the remainder term r is well defined. The remainder term r can be shown to be dominated by a function in $L^2(P)$ and $r \rightarrow 0$ almost surely as $(\theta, \lambda) \rightarrow (\theta_0, \lambda_0)$. The argument for this assertion is similar to that of Lemma A in Pollard (1982). By the dominated convergence theorem, this implies differentiability in quadratic mean of the above first derivatives of ϕ^+ . Because $L^2(P)$ convergence implies $L^1(P)$ convergence, the quadratic mean differentiability for (A.2) implies that $\mathbb{E}\phi^+(\theta, \lambda)$ has traditional second derivatives given by

$$\begin{pmatrix} \mathbb{E} \frac{\partial^2}{\partial \theta \partial \theta'} \phi^+ & \mathbb{E} \frac{\partial^2}{\partial \theta \partial \lambda'} \phi^+ \\ \mathbb{E} \frac{\partial^2}{\partial \lambda \partial \theta'} \phi^+ & \mathbb{E} \frac{\partial^2}{\partial \lambda \partial \lambda'} \phi^+ \end{pmatrix}.$$

We emphasize again that the second derivatives inside the expectation operator is well defined except on a set of zero probability. It follows that the deterministic term $\mathbb{E}\phi^+(\theta, \lambda)$ has the following quadratic approximation,

$$\mathbb{E}\phi^+(\theta, \lambda) = \mathbb{E}\phi^+(\theta_0, \lambda_0) + \frac{1}{2} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix}' \begin{pmatrix} \mathbb{E} \frac{\partial^2}{\partial \theta \partial \theta'} \phi^+ & \mathbb{E} \frac{\partial^2}{\partial \theta \partial \lambda'} \phi^+ \\ \mathbb{E} \frac{\partial^2}{\partial \lambda \partial \theta'} \phi^+ & \mathbb{E} \frac{\partial^2}{\partial \lambda \partial \lambda'} \phi^+ \end{pmatrix}_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} + o(\|\theta, \lambda) - (\theta_0, \lambda_0)\|^2). \quad (\text{A.3})$$

Recall that the first order term vanishes because of vanishing first order derivatives at (θ_0, λ_0) . Therefore, compared to traditional Taylor expansions, the key point here is to justify that we can still use the second derivatives of ϕ^+ (which are not defined everywhere) to obtain an approximation to $\mathbb{E}\phi^+(\theta, \lambda)$.

The first order differentiability of $\phi^+(\theta, \lambda)$ with respect to (θ, λ) and Assumption A.5 on the first order derivatives guarantee the stochastic differentiability of the empirical process (see Pollard 1982, page 921, equation (4))

$$\begin{aligned} & (\mathbb{E}_T - \mathbb{E})\phi^+(\theta, \lambda) \\ &= (\mathbb{E}_T - \mathbb{E})\phi^+(\theta_0, \lambda_0) + \begin{pmatrix} (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \theta} \phi^+(\theta, \lambda) \\ (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \lambda} \phi^+(\theta, \lambda) \end{pmatrix}'_{|(\theta_0, \lambda_0)} \begin{pmatrix} \theta - \theta_0 \\ \lambda - \lambda_0 \end{pmatrix} + o_p(\|\theta, \lambda) - (\theta_0, \lambda_0)\| T^{-1/2}). \end{aligned} \quad (\text{A.4})$$

Combining (A.3) and (A.4), we obtain the LAQ for $\mathbb{E}_T\phi^+(\theta, \lambda)$. ■

Based on the LAQ of $\mathbb{E}_T\phi^+(\theta, \lambda)$ in Lemma 1, we develop the asymptotic distribution of the second HJ-distance in the following theorem with the following additional assumptions.

Assumption A.7. The estimator $(\hat{\theta}^+, \hat{\lambda}^+) \equiv \arg \min_{\theta} \max_{\lambda} \mathbb{E}_T\phi^+(\theta, \lambda)$ for (θ_0, λ_0) is consistent.

Assumption A.8. The matrix Γ in LAQ is nonsingular, with a positive definite Γ_{22} and a negative definite $[\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]$.

Assumption A.9. A central limit theorem holds for the empirical process:

$$\sqrt{T}(\mathbb{E}_T - \mathbb{E})(H(\theta_0)Y - X) \rightarrow Z \equiv N(0, \Lambda),$$

where $\Lambda = \mathbb{E}[H(\theta_0)Y - X][H(\theta_0)Y - X]'$.

We point out that the consistency condition in Assumption A.7 can be replaced by more primitive assumptions and see Hansen, Heaton and Luttmer (1995) for a detailed argument.

Theorem 1. Suppose Assumptions A.1 to A.9 hold. Then under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, $T[\delta_T^+]^2$ is asymptotically distributed as a weighted χ^2 distribution with $(n - k)$ degrees of freedom:

$$T[\delta_T^+]^2 \rightarrow Z' \Xi Z,$$

where $\Xi = D^{-1} - D^{-1}G(G'D^{-1}G)^{-1}G'D^{-1}$, $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$, $D \equiv \mathbb{E}[YY']$, and $\nabla H(\theta)$ is the gradient of $H(\theta)$ with respect to θ .

Proof. Based on the LAQ in Lemma 1, we have

$$\begin{aligned} & \mathbb{E}_T\phi^+(\theta, \lambda) \\ &= \mathbb{E}\phi^+(\theta_0, \lambda_0) + (\mathbb{E}_T - \mathbb{E})\phi^+(\theta_0, \lambda_0) + \underbrace{A'U + \frac{1}{2}V'\Gamma_{11}V + [B + \Gamma_{21}U]'V + \frac{1}{2}V'\Gamma_{22}V}_{\text{quadratic in } V} \\ &+ o(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|^2) + o_p(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|T^{-1/2}). \end{aligned}$$

For fixed θ , the quadratic in V in the above equation is maximized at

$$\hat{V} = \left(\hat{\lambda}^+ - \lambda_0\right) = -\Gamma_{22}^{-1}[B + \Gamma_{21}U] + o_p(T^{-1/2}), \quad (\text{A.5})$$

and its maximum at $V = \hat{V}$ equals

$$\begin{aligned} & A'U + \frac{1}{2}U'\Gamma_{11}U - \frac{1}{2}[B + \Gamma_{21}U]'\Gamma_{22}^{-1}[B + \Gamma_{21}U] \\ &= A'U + \frac{1}{2}U'\Gamma_{11}U - \frac{1}{2}B'\Gamma_{22}^{-1}B - B'\Gamma_{22}^{-1}\Gamma_{21}U - \frac{1}{2}U'[\Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]U + o_p(T^{-1}) \\ &= \underbrace{-\frac{1}{2}B'\Gamma_{22}^{-1}B + [A' - B'\Gamma_{22}^{-1}\Gamma_{21}]U + \frac{1}{2}U'[\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]U}_{\text{quadratic in } U} + o_p(T^{-1}). \end{aligned}$$

The above quadratic in U is minimized at

$$\hat{U} = \left(\hat{\theta}^+ - \theta_0 \right) = - \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \left[A - \Gamma_{12} \Gamma_{22}^{-1} B \right] + o_p(T^{-1/2}), \quad (\text{A.6})$$

and its minimum at $U = \hat{U}$ equals

$$-\frac{1}{2} B' \Gamma_{22}^{-1} B - \frac{1}{2} \left[A' - B' \Gamma_{22}^{-1} \Gamma_{21} \right] \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \left[A - \Gamma_{12} \Gamma_{22}^{-1} B \right] + o_p(T^{-1}).$$

Therefore, the two-step optimization leads to the following asymptotic representation of the objective function,

$$\begin{aligned} & \min_{\theta} \max_{\lambda} \mathbb{E}_T \phi^+ (\theta, \lambda) - \mathbb{E} \phi^+ (\theta_0, \lambda_0) \\ &= (\mathbb{E}_T - \mathbb{E}) \phi^+ (\theta_0, \lambda_0) \\ & - \frac{1}{2} B' \Gamma_{22}^{-1} B - \frac{1}{2} A' \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} A + A' \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \Gamma_{12} \Gamma_{22}^{-1} B \\ & - \frac{1}{2} B' \Gamma_{22}^{-1} \Gamma_{21} \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \Gamma_{12} \Gamma_{22}^{-1} B + o_p(T^{-1}) \\ &= (\mathbb{E}_T - \mathbb{E}) \phi^+ (\theta_0, \lambda_0) - \frac{1}{2} \begin{pmatrix} A \\ B \end{pmatrix}' \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} + o_p(T^{-1}), \end{aligned} \quad (\text{A.7})$$

where

$$\begin{aligned} J_{11} &= \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \\ J_{12} &= J_{21} = - \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \Gamma_{12} \Gamma_{22}^{-1} \\ J_{22} &= \Gamma_{22}^{-1} + \Gamma_{22}^{-1} \Gamma_{21} \left[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21} \right]^{-1} \Gamma_{12} \Gamma_{22}^{-1}. \end{aligned}$$

First we argue that $(\mathbb{E}_T - \mathbb{E}) \phi^+ (\theta_0, \lambda_0)$ vanishes under \mathbb{H}_0 : $\delta^+ = 0$. To see this, observe that (θ_0, λ_0) solves the population optimization problem $\min_{\theta} \max_{\lambda} \mathbb{E} \phi^+ (\theta, \lambda)$. Under the null hypothesis $\delta^+ = 0$, this means that $\mathbb{E} \phi^+ (\theta_0, \lambda_0) = 0$. Since $\phi^+ (\theta_0, \lambda_0)$ is nonnegative, we must have $\phi^+ (\theta_0, \lambda_0) = 0$ almost everywhere. Consequently, $(\mathbb{E}_T - \mathbb{E}) \phi^+ (\theta_0, \lambda_0) = 0$.

Next we consider the second term in (A.7). Note that

$$\begin{aligned} A &= (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \theta} \phi^+ |_{(\theta_0, \lambda_0)} = (\mathbb{E}_T - \mathbb{E}) \left\{ 2H(\theta_0) [\nabla H(\theta_0)] - g^{(1)} [\nabla H(\theta_0)] \right\}, \\ B &= (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \lambda} \phi^+ |_{(\theta_0, \lambda_0)} = (\mathbb{E}_T - \mathbb{E}) [g^{(1)} Y - 2X], \end{aligned}$$

and

$$\begin{aligned} \Gamma_{11} &= \mathbb{E} \left\{ (2 - g^{(2)}) [\nabla H(\theta_0)] [\nabla H(\theta_0)]' + (2H(\theta_0) - g^{(1)}) [\nabla^2 H(\theta_0)] \right\}, \\ \Gamma_{12} &= \mathbb{E} \left\{ g^{(2)} Y [\nabla H(\theta_0)]' \right\}, \\ \Gamma_{22} &= \mathbb{E} [-g^{(2)} Y Y']. \end{aligned}$$

where $g^{(1)}$ and $g^{(2)}$ denote the first and second derivatives at points where they exist. These derivatives can be easily evaluated as

$$\begin{aligned} g &= g(H(\theta_0) - \lambda'_0 Y) = [(H(\theta_0) - \lambda'_0 Y)\mathbf{1}_{E(\theta_0, \lambda_0)}]^2, \\ g^{(1)} &= 2(H(\theta_0) - \lambda'_0 Y)\mathbf{1}_{E(\theta_0, \lambda_0)}, \\ g^{(2)} &= 2 \cdot \mathbf{1}_{E(\theta_0, \lambda_0)}, \end{aligned}$$

where $\mathbf{1}$ is an indicator function and $E(\theta_0, \lambda_0) = \{H(\theta_0) - \lambda'_0 Y > 0\}$. Under the null hypothesis $\mathbb{H}_0 : \delta^+ = 0$, we have $\lambda_0 = 0$ and hence we conclude that $E(\theta_0, \lambda_0)$ has full probability. Consequently

$$\begin{aligned} A &= (\mathbb{E}_T - \mathbb{E}) \frac{\partial}{\partial \theta} \phi^+|_{(\theta_0, \lambda_0)} \\ &= (\mathbb{E}_T - \mathbb{E}) \{2H(\theta_0) [\nabla H(\theta_0)] - [2(H(\theta_0) - \lambda'_0 Y)\mathbf{1}_{E(\theta_0, \lambda_0)}] [\nabla H(\theta_0)]\} \\ &= (\mathbb{E}_T - \mathbb{E}) \{2H(\theta_0) [\nabla H(\theta_0)] [1 - \mathbf{1}_{E(\theta_0, \lambda_0)}] [\nabla H(\theta_0)]\} \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} B &= (\mathbb{E}_T - \mathbb{E}) [g^{(1)} Y - 2X] \\ &= (\mathbb{E}_T - \mathbb{E}) 2[(H(\theta_0) - \lambda'_0 Y)\mathbf{1}_{E(\theta_0, \lambda_0)} Y - X] \\ &= (\mathbb{E}_T - \mathbb{E}) 2[(H(\theta_0) - \lambda'_0 Y)\mathbf{1}_{E(\theta_0, \lambda_0)} Y - X] \\ &= 2(\mathbb{E}_T - \mathbb{E}) [H(\theta_0) Y - X] \end{aligned}$$

$$\begin{aligned} \Gamma_{11} &= \mathbb{E} \left\{ (2 - g^{(2)}) [\nabla H(\theta_0)] [\nabla H(\theta_0)]' + (2H(\theta_0) - g^{(1)}) [\nabla^2 H(\theta_0)] \right\} \\ &= \mathbb{E} \left\{ (2 - 2 \cdot \mathbf{1}_{E(\theta_0, \lambda_0)}) [\nabla H(\theta_0)] [\nabla H(\theta_0)]' + [2H(\theta_0) - 2(H(\theta_0) - \lambda'_0 Y)\mathbf{1}_{E(\theta_0, \lambda_0)}] [\nabla^2 H(\theta_0)] \right\} \\ &= \mathbb{E} \left\{ 2 [1 - \mathbf{1}_{E(\theta_0, \lambda_0)}] [1 + H(\theta_0)] [\nabla H(\theta_0)] [\nabla H(\theta_0)]' \right\} \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Gamma_{21} &= \mathbb{E} \left\{ g^{(2)} Y [\nabla H(\theta_0)]' \right\} = \mathbb{E} \left\{ 2 \cdot \mathbf{1}_{E(\theta_0, \lambda_0)} Y [\nabla H(\theta_0)]' \right\} = \mathbb{E} \left\{ 2Y [\nabla H(\theta_0)]' \right\} \\ \Gamma_{22} &= \mathbb{E} [-g^{(2)} Y Y'] = \mathbb{E} [-2Y Y']. \end{aligned}$$

It follows that

$$\begin{pmatrix} A \\ B \end{pmatrix} = 2(\mathbb{E}_T - \mathbb{E}) \begin{pmatrix} 0 \\ H(\theta_0) Y - X \end{pmatrix}$$

and (from (A.7))

$$\begin{aligned} J_{11} &= 2^{-1} (G' D^{-1} G)^{-1} \\ J_{12} &= 2^{-1} (G' D^{-1} G)^{-1} G' D^{-1} \\ J_{22} &= -2^{-1} \left[D^{-1} - D^{-1} G (G' D^{-1} G)^{-1} G' D^{-1} \right], \end{aligned}$$

where $G \equiv \mathbb{E} \{Y [\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E} [Y Y']$.

Therefore, under $\mathbb{H}_0: \delta^+ = 0$, the second HJ-distance has the following representation

$$\begin{aligned}
& T [\delta_T^+]^2 \\
&= -2 \left[\sqrt{T} (\mathbb{E}_T - \mathbb{E}) \begin{pmatrix} 0 \\ H(\theta_0)Y - X \end{pmatrix} \right]' \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \left[\sqrt{T} (\mathbb{E}_T - \mathbb{E}) \begin{pmatrix} 0 \\ H(\theta_0)Y - X \end{pmatrix} \right] + o_p(1) \\
&= [\sqrt{T} (\mathbb{E}_T - \mathbb{E}) (H(\theta_0)Y - X)]' (-2) J_{22} [\sqrt{T} (\mathbb{E}_T - \mathbb{E}) (H(\theta_0)Y - X)] \\
&= [\sqrt{T} (\mathbb{E}_T - \mathbb{E}) (H(\theta_0)Y - X)]' [D^{-1} - D^{-1}G (G'D^{-1}G)^{-1} G'D^{-1}] [\sqrt{T} (\mathbb{E}_T - \mathbb{E}) ((H(\theta_0)Y - X))] + o_p(1).
\end{aligned}$$

By Assumption A.9, $\sqrt{T} (\mathbb{E}_T - \mathbb{E}) [H(\theta_0)Y - X]$ has an asymptotic limiting distribution $N(0, \Lambda)$. It follows immediately that under $\mathbb{H}_0: \delta^+ = 0$, the asymptotic distribution of $T [\delta_T^+]^2$ can be represented as $Z' \Xi Z$.

Next we show that $Z' \Xi Z$ follows a weighted χ^2 distribution with $(n - k)$ degrees of freedom. Observe that the matrix

$$I - D^{-1/2}G (G'D^{-1}G)^{-1} G'D^{-1/2} \quad (\text{A.7})$$

is symmetric and idempotent (a matrix A is idempotent iff $A^2 = A$). A simple computation shows that its trace is equal to $n - k$. Therefore this matrix has a rank of $(n - k)$. For a vector of standard normal random vector z , define

$$W = [I - D^{-1/2}G (G'D^{-1}G)^{-1} G'D^{-1/2}] D^{-1/2} \Lambda^{1/2} z.$$

Then $W'W$ has the same distribution as $Z' \Xi Z$, because $\Lambda^{1/2} z$ has the same distribution as Z . Since matrix (A.7) has rank $n - k$, the random variable W concentrates on a $n - k$ dimensional subspace. This finishes the proof. ■

The following propositions provide the asymptotic distributions for the estimates of the model parameters and the Lagrangian multipliers.

Proposition 1. [Model Parameter] Suppose Assumptions A.1 to A.9 hold. Then under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, $\sqrt{T}(\hat{\theta}^+ - \theta_0)$ is asymptotically normally distributed with mean zero and covariance matrix

$$(G'D^{-1}G)^{-1} G'D^{-1} \Lambda D^{-1} G (G'D^{-1}G)^{-1},$$

where $G \equiv \mathbb{E} \{Y[\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E} [YY']$.

Proof. Since $A = 0$ under the hypothesis $\mathbb{H}_0: \delta^+ = 0$, we obtain the following representation for $\hat{\theta}^+ - \theta_0$ from (A.6)

$$\begin{aligned}
\hat{\theta}^+ - \theta_0 &= - [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} [A - \Gamma_{12} \Gamma_{22}^{-1} B] + o_p(T^{-1/2}) \\
&= [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1} B + o_p(T^{-1/2}) \\
&= (G'D^{-1}G)^{-1} G'D^{-1} (\mathbb{E}_T - \mathbb{E}) [H(\theta_0)Y - X] + o_p(T^{-1/2}).
\end{aligned}$$

By Central Limit Theorem, $\sqrt{T} (\mathbb{E}_T - \mathbb{E}) [H(\theta_0)Y - X]$ has an asymptotic limiting distribution $N(0, \Lambda)$. It follows that $\sqrt{T}(\hat{\theta}^+ - \theta_0)$ is asymptotically normally distributed with mean zero and covariance matrix

$$(G'D^{-1}G)^{-1} G'D^{-1} \Lambda D^{-1} G (G'D^{-1}G)^{-1}. \blacksquare$$

Proposition 2. [Lagrangian Multiplier] Suppose Assumptions A.1 to A.9 hold. Then under the null hypothesis \mathbb{H}_0 : $\delta^+ = 0$, $\sqrt{T}(\hat{\lambda}^+(\hat{\theta}^+) - \lambda_0)$ is asymptotically normally distributed with mean zero and covariance matrix

$$[D^{-1} - D^{-1}G(G'D^{-1}G)^{-1}G'D^{-1}]\Lambda[D^{-1} - D^{-1}G(G'D^{-1}G)^{-1}G'D^{-1}],$$

where $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E}[YY']$.

Proof. We have the following asymptotic representation from (A.5) and (A.6)

$$\begin{aligned} \hat{\lambda}^+(\hat{\theta}^+) - \lambda_0 &= -\Gamma_{22}^{-1}[B + \Gamma_{21}(\hat{\theta}^+ - \theta_0)] + o_p(T^{-1/2}) \\ &= -\Gamma_{22}^{-1}(B - \Gamma_{21}[\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]^{-1}[-\Gamma_{12}\Gamma_{22}^{-1}B]) + o_p(T^{-1/2}) \\ &= -\Gamma_{22}^{-1}(I + \Gamma_{21}[\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]^{-1}\Gamma_{12}\Gamma_{22}^{-1})B + o_p(T^{-1/2}) \\ &= [D^{-1} - D^{-1}G(G'D^{-1}G)^{-1}G'D^{-1}](\mathbb{E}_T - \mathbb{E})[H(\theta_0)Y - X] + o_p(T^{-1/2}). \end{aligned}$$

The result follows immediately from the fact that $\sqrt{T}(\mathbb{E}_T - \mathbb{E})[H(\theta_0)Y - X]$ has an asymptotic limiting distribution $N(0, \Lambda)$. ■

Proposition 3. [Pricing Errors] Suppose assumptions A.1 to A.9 hold. Then under the null hypothesis \mathbb{H}_0 : $\delta^+ = 0$, the standardized pricing error $\sqrt{T}\mathbb{E}_T(H(\theta)Y - X)|_{\theta=\hat{\theta}^+}$ has an asymptotic normal distribution with zero mean and covariance matrix

$$[I - G(G'D^{-1}G)^{-1}G'D^{-1}]\Lambda[I - D^{-1}G(G'D^{-1}G)^{-1}G'],$$

where $G \equiv \mathbb{E}\{Y[\nabla H(\theta_0)]'\}$ and $D \equiv \mathbb{E}[YY']$.

Proof. The pricing error distribution $\mathbb{E}_T(H(\theta)Y - X)|_{\theta=\hat{\theta}^+}$ has the following representation

$$\begin{aligned} \mathbb{E}_T(H(\theta)Y - X)|_{\theta=\hat{\theta}^+} &= \mathbb{E}_T[H(\theta_0)Y - X] + [\mathbb{E}_T Y[\nabla H(\theta_0)]'](\hat{\theta}^+ - \theta_0) + o_p(T^{-1/2}) \\ &= \mathbb{E}_T[H(\theta_0)Y - X] + [\mathbb{E} Y[\nabla H(\theta_0)]'](\hat{\theta}^+ - \theta_0) + o_p(T^{-1/2}) \\ &= \mathbb{E}_T[H(\theta_0)Y - X] + G(\hat{\theta}^+ - \theta_0) + o_p(T^{-1/2}). \end{aligned}$$

Since

$$\begin{aligned} \hat{\theta}^+ - \theta_0 &= [\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]^{-1}\Gamma_{12}\Gamma_{22}^{-1}B + o_p(T^{-1/2}) \\ &= -(G'D^{-1}G)^{-1}G'D^{-1}\mathbb{E}_T[H(\theta_0)Y - X] + o_p(T^{-1/2}), \end{aligned}$$

we have

$$\mathbb{E}_T(H(\theta)Y - X)|_{\theta=\hat{\theta}^+} = [I - G(G'D^{-1}G)^{-1}G'D^{-1}]\mathbb{E}_T[H(\theta_0)Y - X] + o_p(T^{-1/2}).$$

The result follows immediately from the fact that $\sqrt{T}(\mathbb{E}_T - \mathbb{E})[H(\theta_0)Y - X]$ has an asymptotic limiting distribution $N(0, \Lambda)$. The proof of Theorem 1 indicates that this is a normal random variable that concentrates on $(n - k)$ dimensional subspace. ■

Table 1. Finite Sample Performances of Specification Tests Based on the Two HJ-Distances

This table reports the finite sample performances of specification tests based on the first and second HJ-distance. We examine the size (power) performances of the two tests using the CAPM (Fama-French three-factor model, or FF3). In Panel A, the parameters of the CAPM equal to those estimated using the monthly returns of the Fama-French 25 Size/BM portfolios between January 1952 and December 2002. In Panel B, the parameters of the CAPM equal to those estimated using the monthly returns of the ten hedge fund portfolios constructed in Section 4 between January 1994 and September 2003. In both cases, the SDF of the CAPM takes negative values with zero probability. In Panel C, the parameters of FF3 equal to those estimated using the monthly returns of the ten hedge fund portfolios between January 1994 and September 2003. Based on the chosen parameters and the historical returns of risk factors during the sample period, the probability that the SDF of FF3 takes negative values is 17%. In Panel A, for each sample size we generate 500 random samples of monthly returns of 25 assets whose expected returns are determined by the CAPM and whose covariance matrix equals to that of the Fama-French 25 Size/BM portfolios between January 1952 and December 2002. In Panel B, for each sample size we generate 500 random samples of monthly returns of ten assets whose expected returns are determined by the CAPM and whose covariance matrix equals to that of the ten hedge fund portfolios between January 1994 and September 2003. In Panel C, for each sample size we generate 500 random samples of monthly returns of ten assets whose expected returns are determined by FF3 and whose covariance matrix equals to that of the ten hedge fund portfolios between January 1994 and September 2003. The simulated systematic risk factors also have the same mean, variance, and covariances (with all other assets and factors) as the actual risk factors during their corresponding sample periods. In each panel, for each given sample size and each random sample, we estimate model parameters and conduct specification tests based on the first and second HJ-distance. For each given sample size, the rejection rates based on the asymptotic critical values at the 1%, 5%, and 10% significance levels for the two HJ-distances are reported.

Panel A. Size performances of specification tests based on the two HJ-distances using the CAPM as the null and simulated data that mimic the Fama-French 25 size/BM portfolios

N=25	First HJ-Distance			Second HJ-Distance		
	1%	5%	10%	1%	5%	10%
T=100	17%	32%	45%	21%	36%	48%
T=300	4%	12%	19%	4%	12%	20%
T=600	2%	9%	15%	2%	9%	15%

Panel B. Size performances of specification tests based on the two HJ-distances using the CAPM as the null and simulated data that mimic the ten hedge fund portfolios

N=10	First HJ-Distance			Second HJ-Distance		
	1%	5%	10%	1%	5%	10%
T=100	4%	10%	17%	4%	12%	19%
T=300	1%	5%	12%	1%	6%	12%
T=600	1%	5%	9%	1%	5%	8%

Panel C. Power performances of specification tests based on the two HJ-distances using FF3 as the null and simulated data that mimic the ten hedge fund portfolios

N=10	First HJ-Distance			Second HJ-Distance		
	1%	5%	10%	1%	5%	10%
T=100	6%	14%	22%	39%	57%	69%
T=300	2%	6%	12%	73%	89%	93%
T=600	2%	7%	14%	99%	100%	100%

Table 2. Summary Statistics for the Fama-French 25 Size/BM portfolios

This table provides summary statistics of monthly excess returns of the Fama-French 25 Size/BM portfolios from January 1952 to December 2002. The excess returns are constructed by subtracting one-month T-bill rates from total monthly returns. Portfolios are numbered ij with i indexing size increasing from one to five and j indexing book-to-market ratio increasing from one to five.

Portfolios	BM1	BM2	BM3	BM4	BM5
Panel A: Means					
SIZE 1	0.21	0.75	0.80	1.03	1.09
SIZE 2	0.36	0.67	0.87	0.92	1.02
SIZE 3	0.49	0.72	0.74	0.88	0.93
SIZE 4	0.54	0.57	0.79	0.81	0.87
SIZE 5	0.51	0.53	0.63	0.63	0.66
Panel B: Standard errors					
SIZE 1	7.82	6.69	5.68	5.32	5.62
SIZE 2	7.00	5.71	5.05	4.92	5.48
SIZE 3	6.43	5.13	4.72	4.62	5.16
SIZE 4	5.75	4.87	4.65	4.55	5.23
SIZE 5	4.71	4.40	4.17	4.28	4.83
Panel C: t-statistics					
SIZE 1	0.67	2.78	3.46	4.79	4.81
SIZE 2	1.26	2.90	4.26	4.64	4.61
SIZE 3	1.88	3.48	3.88	4.70	4.46
SIZE 4	2.34	2.89	4.22	4.40	4.11
SIZE 5	2.65	2.98	3.74	3.62	3.36

Table 4. Summary Statistics of Hedge Funds

The hedge fund data are obtained from TASS. The sample period is between January 1994 and September 2003, which yields 116 months of observations. We include both live funds and “graveyard” funds. Individual funds are included in our sample if they follow one of the following three strategies: trend follower, risk arbitrage and equity derivative arbitrage. Panel A reports the distribution of our sample hedge funds among the ten hedge fund investment styles defined by TASS. Panel B reports summary statistics of style portfolio monthly returns for our sample funds.

Panel A: Distribution of hedge funds among ten investment styles

Styles	Total Months	Mean	Std Dev	Minimum	Maximum
Convertible arbitrage	116	36	15	8	57
Emerging markets	116	13	5	4	19
Equity market neutral	116	10	7	2	26
Event driven	116	44	24	11	82
Fixed-income arbitrage	116	2	1	1	4
Fund of funds	116	150	47	59	241
Global macro	116	26	3	3	35
Long/short equity	116	40	16	16	65
Managed futures	116	101	10	66	114
Multi-strategy	116	17	6	5	29

Panel B: Summary statistics of style portfolio returns

Styles	Mean	Std Dev	t-Statistics
Convertible arbitrage	0.90	1.49	6.5
Emerging markets	0.80	4.28	2.0
Equity market neutral	0.80	0.63	13.6
Event driven	1.02	1.93	5.7
Fixed-income arbitrage	0.77	1.55	5.3
Fund of funds	0.80	1.93	4.5
Global macro	1.99	6.62	3.2
Long/short equity	1.09	2.60	4.5
Managed futures	0.99	3.97	2.7
Multi-strategy	1.06	4.30	2.7

Table 5. Asset Pricing Tests Based on the HJ-Distances for Ten Hedge Fund Portfolios

This table provides empirical results on specification tests of six asset pricing models and their truncated versions based on the two HJ-distances for the ten hedge fund portfolios from January 1994 to September 2003. Panels A and B contain results for original models and their truncated versions using the ten hedge fund portfolios. The first (second) row of each panel contains the estimated first (second) HJ-distances. The third row of each panel contains the percentage difference between the two HJ-distances. In Panel A, the fourth row reports the probabilities that model SDFs estimated using the first HJ-distance take negative values. The last two rows of each panel report the p-values of specification tests based on the first and second HJ-distances, respectively.

Panel A: Results for original models using ten hedge fund portfolios

	CAPM	CAPM+IP	FF3	FF3+IP	OPT	OPT+IP
δ	1.007	0.642	0.901	0.560	0.913	0.633
δ^+	1.131	1.070	1.074	1.037	1.098	1.057
$(\delta^+ - \delta) / \delta$	12%	67%	19%	85%	20%	67%
$p(H < 0)$	0%	43%	17%	43%	17%	40%
$p(\delta = 0)$	0%	57%	0%	56%	0%	31%
$p(\delta^+ = 0)$	0%	0%	0%	0%	0%	0%

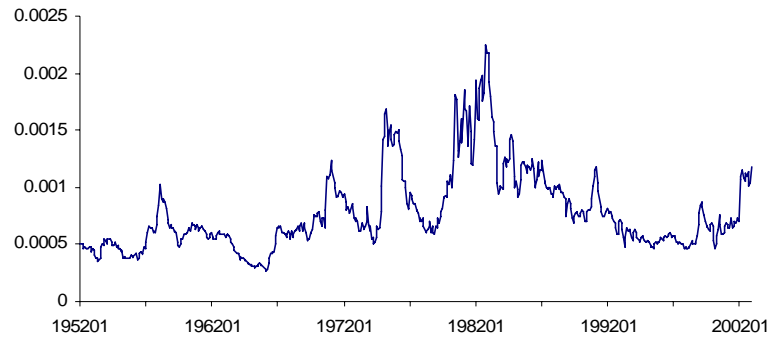
Panel B: Results for truncated models using ten hedge fund portfolios

	CAPM	CAPM+IP	FF3	FF3+IP	OPT	OPT+IP
δ	1.007	0.708	0.907	0.699	0.882	0.677
δ^+	1.131	1.069	1.066	1.025	1.090	1.048
$(\delta^+ - \delta) / \delta$	12%	51%	17%	47%	24%	55%
$p(\delta = 0)$	0%	10%	0%	3%	0%	8%
$p(\delta^+ = 0)$	0%	0%	0%	0%	0%	0%

Figure 1. Time series plots of conditioning variables.

Panel A contains yield spread between Baa and Aaa corporate bonds, and Panel B contains the cyclical components of natural logarithm of industrial production.

Panel A: Monthly default spread



Panel B: Monthly cycle (IP)

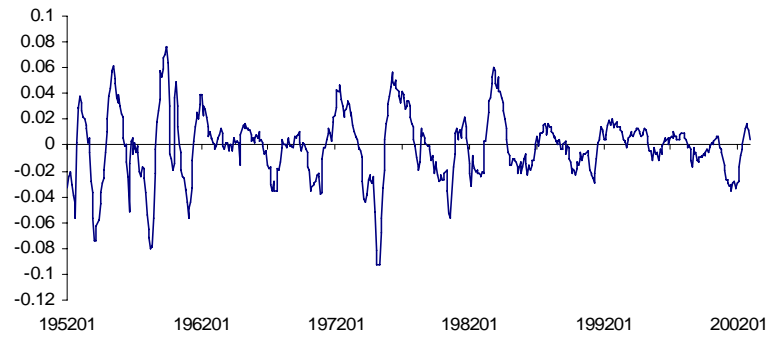


Figure 2. Time series plots of two asset pricing models SDF estimated using the 25 Fama-French size/BM portfolios.

Panel A and B contain time series plots of the estimated SDFs of the Fama-French six-factor model and the linearized Campbell model, using the first and second HJ-distances, respectively. Panel C and D contain time series plots of the deviations of the estimated SDFs from true SDF of the Fama-French six-factor model and the linearized Campbell model, respectively.

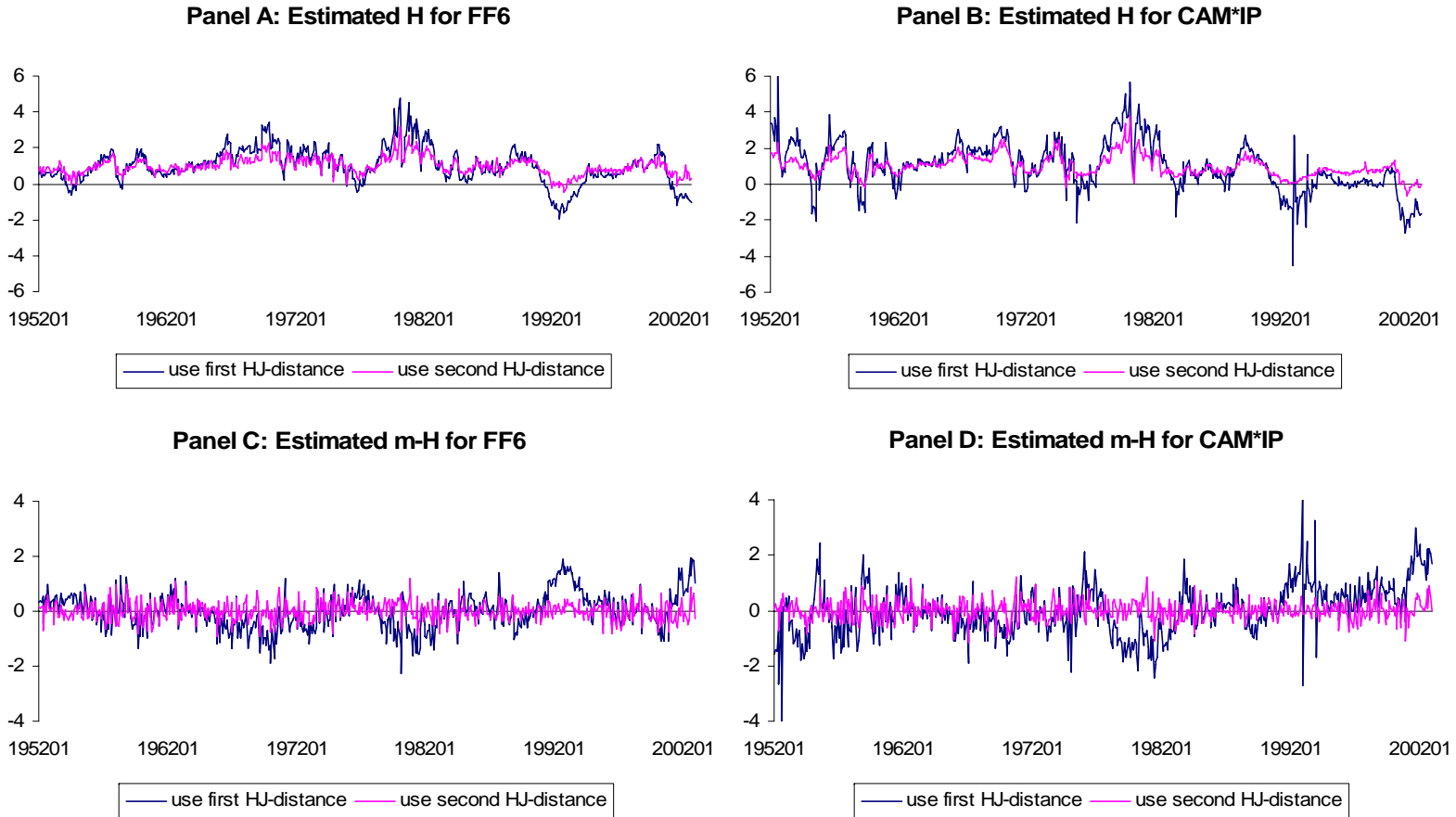
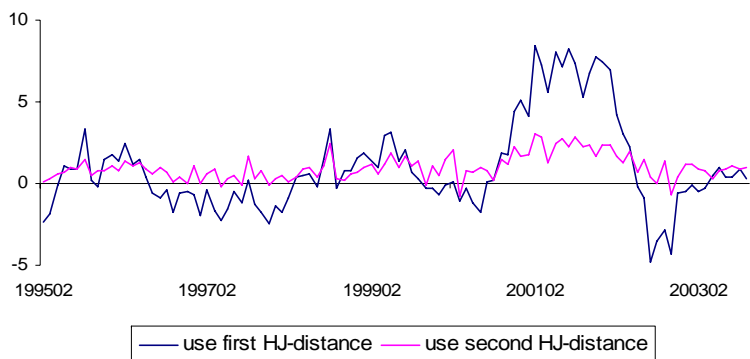


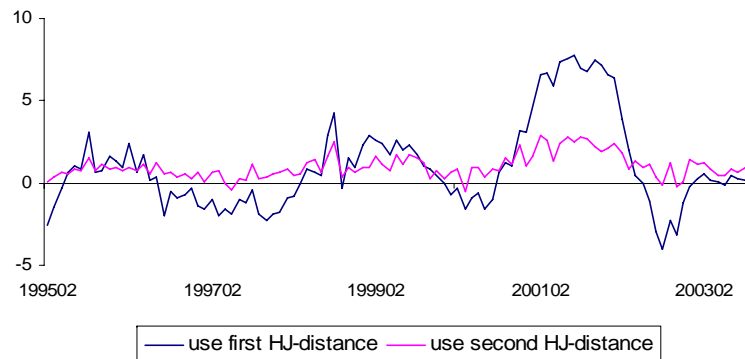
Figure 3. Time series plots of two asset pricing models' SDF estimated using the ten hedge fund portfolios.

Panel A and B contain time series plots of the estimated SDFs of the FF3+IP model and the OPT+IP model, using the first and second HJ-distances, respectively. Panel C and D contain time series plots of the deviations of the estimated SDFs from true SDF of the FF3+IP and the OPT+IP model, respectively.

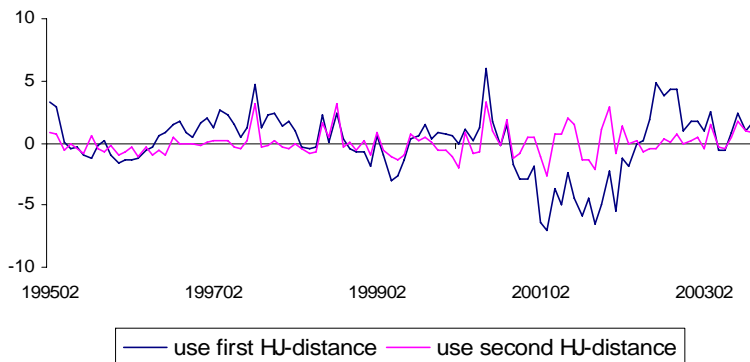
Panel A: Estimated H for FF3+IP



Panel B: Estimated H for OPT+IP



Panel C: Estimated m-H for FF3+IP



Panel D: Estimated m-H for OPT+IP

