Evaluating Asset Pricing Models in Absence of Arbitrage: Econometric Approach and Empirical Applications

Haitao Li[a], Yuewu Xu[b], and Xiaoyan Zhang[c]

January 2006

[a]Li is from the Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109. [b]Xu is from the School of Business, Fordham University, New York, NY 10017. [c]Zhang is from the Johnson Graduate School of Management, Cornell University, Ithaca, NY 14850. We thank Vikas Agarwal, Warren Bailey, Charles Cao, Jin-Chuan Duan, Jay Huang, Jon Ingersoll, Ravi Jagannathan, Raymond Kan, Peter Phillips, Marcel Rindisbacher, Tim Simin, 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.

Evaluating Asset Pricing Models in Absence of Arbitrage: Econometric Approach and Empirical Applications

ABSTRACT

We develop asset pricing tests that explicitly require that a true asset pricing model has to be arbitrage free, i.e., its stochastic discount factor has to be strictly positive. The no-arbitrage restriction is especially important in applications that involve dynamic trading strategies and derivatives. One prominent example is performance evaluation of actively managed portfolios, such as mutual funds and hedge funds. Empirical results show that the no-arbitrage restriction makes significant differences in asset pricing tests involving either the traditional Fama-French size and book-to-market portfolios or hedge fund style portfolios. Ignoring the no-arbitrage restriction, we reach the misleading conclusion that certain models can satisfactorily price both sets of assets. However, these models are overwhelmingly rejected by our tests because their stochastic discount factors take negative values with high probabilities. The no-arbitrage restriction also leads to better identified parameters, better behaved models, and more insights on potential sources of model misspecifications.

JEL Classification: C4, C5, G1

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

One central idea of modern asset pricing theory is the equivalence of absence of arbitrage and the existence of a positive stochastic discount factor (hereafter SDF) that correctly prices all assets.[1] This result implies that a good asset pricing model should have not only small pricing errors but also a stochastic discount factor that is strictly positive. Most existing empirical studies of asset pricing models, however, have mainly focused on the first aspect of model performance and ignored the second one. For example, the widely studied linear factor models have SDFs that are not strictly positive and hence are not arbitrage free. There are both statistical and economic reasons why it might be desirable to impose the no-arbitrage restriction in empirical analysis of asset pricing models.

First, ignoring the no-arbitrage restriction might lead to misleading conclusions in statistical inferences of asset pricing models. For example, certain models might have small pricing errors and thus are regarded as acceptable without the no-arbitrage restriction. However, the SDFs of such models might take negative values with high probabilities and would be regarded as unsatisfactory with the no-arbitrage restriction. Thus the no-arbitrage restriction helps to reveal true sources of model misspecifications which can be helpful for model improvements.

Second, an asset pricing model that allows arbitrage opportunities is not only theoretically unsound but also practically unreliable because it could lead to misleading results in certain applications. For example, Dybvig and Ingersoll (1982) show that although linear asset pricing models, such as the CAPM, might be able to price primary assets, they would have difficulties in pricing derivatives. Specifically, these models would assign negative values to options with positive payoffs in states where the SDFs take negative values and zero payoffs in other states. Though linear factor models are seldom used directly to price options, they have been widely used in important applications that implicitly involve derivatives.

One prominent example is performance evaluation of actively managed portfolios, such as mutual funds and hedge funds. The common practice in performance evaluation is to identify some "good" asset pricing models according to certain criteria and use these models to evaluate the performance of actively managed funds. Unfortunately, many mutual funds and most hedge funds employ dynamic trading strategies which, according to Merton (1981), Dybvig and Ross (1985), and others, could generate option-like payoffs. Many hedge funds directly trade derivatives[2] and it has been widely documented that hedge fund returns exhibit option-like features.[3] Grinblatt and Titman (1987) and Glosten and Jagannathan (1994) have emphasized that it is essential

[1] See Cochrane (2001), Duffie (2001), Dybvig and Ross (2003), Ingersoll (1987), and others for discussions of the fundamental theorem of asset pricing.

[2] TASS, a hedge fund research company, reports that more than 50 percent of the 4,000 hedge funds it follows use derivatives.

[3] See, for example, Fung and Hsieh (1997), Agarwal and Naik (2002), Ben Dor and Jagannathan (2002), and Mitchell and Pulvino (2001).

to impose the no-arbitrage restriction in asset pricing models when evaluating the performance of mutual funds with option-like returns. The fast-growing hedge fund industry and the need to evaluate hedge fund performance make this issue even more urgent in current asset pricing literature. Goetzmann, Ingersoll, Spiegel, and Welch (2002) show that the use of derivatives by hedge funds renders 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.

Even for applications that involve mainly primary assets, there are still important benefits of imposing the no-arbitrage restriction in testing linear factor models. This is because the no-arbitrage restriction significantly changes the objective function used for estimating model parameters. That is, parameters are chosen not solely to minimize pricing errors, but to make a linear factor model as close as possible to true asset pricing models with strictly positive SDFs. 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 restriction are likely to be better proxies of IMRS.

Incorporating the no-arbitrage restriction 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 results might be sensitive to which derivatives are used in estimation. And strictly 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. In addition, econometrically it is also rather difficult to impose the no-arbitrage restriction in traditional linear regressions that 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."

One of the main objectives of our paper is to develop asset pricing tests that explicitly incorporate the no-arbitrage restriction. Such tests would make it possible to fully reflect the important predictions of the asset pricing theory in empirical studies. As a result, the integrity of the theory will not be compromised due to the econometric difficulty in dealing with the no-arbitrage restriction. The approach developed in our paper is based on the pioneering work of Hansen and Jagannathan (1997). The Hansen-Jagannathan distances (hereafter HJ-distances) represent least-square distances between a model SDF and the set of admissible SDFs that correctly price all assets. The first HJ-distance considers all admissible SDFs without requiring correct SDFs to be strictly positive, and the second one considers only strictly positive admissible SDFs to avoid arbitrage opportunities.

There are two main advantages of adopting the HJ framework. First, as pointed out by Cochrane (2001), it is very convenient to impose the no-arbitrage restriction under the SDF framework.[4] Second, the second HJ-distance provides an economically meaningful measure of model misspecification and a natural objective function for estimating model parameters in absence of arbitrage. 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. Therefore, minimizing the second HJ-distance is equivalent to minimizing the pricing errors of all possible derivatives even though their prices are not used in estimation.

Unfortunately, the distribution of the second HJ-distance under the null hypothesis of a correctly specified model is still unknown. As a result, the second HJ-distance is seldom considered in the existing literature, in spite of the importance of the no-arbitrage restriction in many applications.[5] One main difficulty in econometric analysis of the second HJ-distance is that it 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. We overcome this obstacle by introducing the concept of "differentiation in quadratic mean" of Le Cam (1986), Pollard (1982), and Pakes and Pollard (1989).[6] Based on this technique, 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. This result allows us to conduct specification tests of asset pricing models in absence of arbitrage. We also provide the asymptotic distributions of the estimates of model parameters, Lagrangian multipliers, and pricing errors of individual assets for general nonlinear models. These statistics provide useful diagnostic information on model performance. For example, for linear factor models, the distributions of model parameters allow us to test whether the risk of a specific factor is priced. The estimates of Lagrangian multipliers help to reveal deviations of a given model from true asset pricing models. The pricing errors of individual assets reveal which assets are more difficult to price.

The new tests developed in our paper fill the gap in the literature and establish the theoretical foundation of econometric analysis of asset pricing models in absence of arbitrage. Jagannathan

[4] At the bottom of page 130 of Cochrane (2001), it is written, "It is trivally easy to graft it (absence of arbitrage) on to a discount factor model: just add $m > 0$."

[5] One exception is Wang and Zhang (2004) who examine the importance of the no-arbitrage restriction using Bayesian method. Bansal and Viswanathan (1993), Bansal, Hsieh, and Viswanathan (1993), and other studies also consider the no-arbitrage restriction by restricting the SDF of a model to be positive and evaluate model performance using GMM or the first HJ-distance.

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

and Wang (1996) develop the asymptotic distribution of the first HJ-distance for linear factor models when the model is correctly specified. Hansen, Heaton, and Luttmer (1995) develop the asymptotic distributions of the two HJ-distances when the asset pricing models are misspecified. Our tests are more powerful than that of Jagannathan and Wang (1996) and are able to reject certain misspecified models that the Jagannathan and Wang test cannot reject. These are models that can price test assets well but their SDFs are not strictly positive. In addition, we generally obtain very different estimates of model parameters and pricing errors using the two procedures because they focus on very different objective functions. We also extend the analysis of Hansen, Heaton, and Luttmer (1995) to models with unknown parameters that need to be estimated from the data. Thus, our paper completes the missing link in the econometric analysis of the HJ-distances and makes it convenient for empirical researchers to incorporate the no-arbitrage restriction in their studies of asset pricing models.

To illustrate the importance of the no-arbitrage restriction, we apply our new tests to evaluate several well-known asset pricing models in pricing two sets of financial assets. The first one involves the traditional Fama-French 25 Size/Book-to-Market (BM hereafter) portfolios, and the second one involves several popular hedge fund strategies (trendfollowing, risk arbitrage, and equity derivative arbitrage) that have been shown in the literature to exhibit option-like returns. Our empirical results show that the no-arbitrage restriction could make significant differences in asset pricing tests whether the test assets explicitly involve derivatives or not. In both applications, we find substantial differences in the estimated first and second HJ-distances for certain models. We also reach dramatically different conclusions on model performance based on the two HJ-distances. Though certain models have relatively good performance 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 no-arbitrage restriction also leads to better identified parameters, better behaved models, and more insights on potential sources of model misspecifications.

The rest of this paper is organized as follows. In section I, we introduce the HJ-distances and the importance of the no-arbitrage restriction. Section II develops the econometric theories of asset pricing tests in absence of arbitrage based on the second HJ-distance. In section III, we apply the new tests to empirically examine several well-known asset pricing models using the 25 Size/BM portfolios of Fama and French (1993). In section IV, we apply the new tests to evaluate several popular hedge fund strategies that exhibit option-like returns. Section V concludes, and the appendix provides the mathematical proofs.

I. Arbitrage and the Hansen-Jagannathan Distances

In this section, we discuss the importance of the no-arbitrage restriction in asset pricing applications under the SDF framework. We also demonstrate that the second HJ-distance provides a natural approach of incorporating the no-arbitrage restriction in asset pricing tests.

Suppose the uncertainty of a discrete-time economy is described by a filtered probability space $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0}\right)$ for $t = 0, 1, ..., T$. Suppose there are n assets traded in the economy 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. The fundamental theorem of asset pricing asserts the equivalence of absence of arbitrage and the existence of a strictly positive SDF that correctly prices all assets. That is, for all t, we have

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

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 its name 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). We can derive (1) through the first order conditions of a representative investor's optimal consumption and portfolio choice. As a result, m_{t+1} measures aggregate intertemporal marginal rate of substitution:

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

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. As pointed out by Cochrane (2001), the general pricing relation in (1) encompasses most asset pricing models and is applicable to a wide range of assets such as stocks, bonds, options, and real assets.

The pricing equation in (1) suggests that a good asset pricing model should have (i) small pricing errors and (ii) an 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. For example, many empirical studies have used pricing errors of primary assets as moment conditions to estimate parameters and test model performance based on the Generalized Method of Moments (GMM) of Hansen (1982). But most of these studies have ignored *the no-arbitrage restriction* that m_{t+1}, the SDF of a true asset pricing model, has to be strictly positive. As a result, even though a given model might have small pricing errors, its SDF could take negative values with high probabilities. Thus "imposing the no-arbitrage restriction" in our paper means that we evaluate model performance by its deviations (in least-square sense) from true asset pricing models that can correctly price all assets and are arbitrage free.

The no-arbitrage restriction 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.,

$$\beta \frac{u_c(c_{t+1})}{u_c(c_t)} \approx a + b' f_{t+1}. \tag{2}$$

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 true SDF $\beta u_c(c_{t+1}) / u_c(c_t)$ is always positive.

Ignoring the no-arbitrage restriction 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 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 restriction and the estimated model $\hat{a} + \hat{b}' f_{t+1}$ often takes negative values with high probabilities. On the other hand, with the no-arbitrage restriction, parameters are chosen not to minimize pricing errors, but to minimize a model's deviations from true asset pricing models, whose SDFs are strictly positive. Hence, by incorporating the no-arbitrage restriction, we are able to reflect both important predictions of existing asset pricing theory in our empirical tests. We are also able to better identify model parameters and, as a result, the estimated models are likely to be better proxies of the aggregate marginal utility growth.

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

$$m_{t+1}^{CAPM} = a + b r_{MKT,t+1},$$

where $r_{MKT,t+1}$ is the excess return on the market portfolio at $t + 1$ and $b < 0$. 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 seldom used directly to price options, they have been widely used in important applications that implicitly involve options. Some mutual funds and most hedge funds use dynamic trading strategies and derivatives and thus exhibit option-like returns. Models that admit arbitrage opportunities would give misleading results on the performance of these funds.[7] Although the importance of the no-arbitrage restriction for mutual fund performance evaluation has been pointed out by Grinblatt and Titman (1987) and Glosten

[7] Suppose we use the CAPM to evalute 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

and Jagannathan (1994), the need for hedge fund performance evaluation makes this issue even more urgent in the current asset pricing literature. In fact, attempts have been made to deal with option-like returns of hedge funds. For example, Agarwal and Naik (2002) include option returns in linear factor models for hedge fund performance evaluation. Though a good improvement over the existing literature, these option-based models are still not arbitrage free and thus do not completely solve the problem.

Therefore, due to both statistical and economic concerns, it would be beneficial to impose the no-arbitrage restriction in empirical asset pricing studies regardless of whether derivatives are explicitly involved or not. In this paper, we develop asset pricing tests in absence of arbitrage based on the HJ-distances. Below we give a brief introduction of the two HJ-distances and explain the basic idea of our approach. In next section, we develop the econometric theory of our new asset pricing tests.

The HJ-distances are measures introduced in Hansen and Jagannathan (1997) for model misspecification. Specifically, the first HJ-distance measures the least-square distance or the L^2 norm between a candidate SDF model H and the set of SDFs that correctly price all assets but without requiring correct SDFs to be strictly positive, and the second HJ-distance considers only SDFs that correctly price all assets and are strictly positive. Without loss of generality, we focus our discussions on the unconditional implication of (1)

$$\mathbb{E}\left[m_{t+1}Y_{t+1}\right]=\mathbb{E}\left[X_t\right].$$[8]

The first HJ-distance, δ, is defined as

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

where $\mathcal{M}=\{m_{t+1}:\mathbb{E}\left[m_{t+1}Y_{t+1}\right]=\mathbb{E}\left[X_t\right],\ \forall Y_{t+1}\in\underline{Y}\}$ is the set of SDFs that correctly price all assets. The second HJ-distance, δ^+, is defined as

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

where $\mathcal{M}^+=\{m_{t+1}:\mathbb{E}\left[m_{t+1}Y_{t+1}\right]=\mathbb{E}\left[X_t\right],\ m_{t+1}>0,\ \forall Y_{t+1}\in\underline{Y}\}$ is the set of SDFs that correctly price all assets and are strictly positive. Often the candidate model H depends on some unknown parameters θ, and the two distances are defined as

$$\begin{aligned}\delta\left(\theta\right) &= \min_{\theta}\min_{m\in\mathcal{M}}\|H\left(\theta\right)-m\|=\min_{\theta}\min_{m\in\mathcal{M}}\sqrt{\mathbb{E}\left(H\left(\theta\right)-m\right)^2}, &(5)\\ \delta^+\left(\theta\right) &= \min_{\theta}\min_{m\in\mathcal{M}^+}\|H\left(\theta\right)-m\|=\min_{\theta}\min_{m\in\mathcal{M}^+}\sqrt{\mathbb{E}\left(H\left(\theta\right)-m\right)^2}. &(6)\end{aligned}$$

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.

[8]If 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.

Let $\hat{\theta} = \arg\min_\theta \delta(\theta)$ and $\hat{\theta}^+ = \arg\min_\theta \delta^+(\theta)$. The main difference between the two HJ-distances is that the second one explicitly requires that a true asset pricing model has to be arbitrage free. 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 that 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. The asymptotic distribution of estimated δ developed by Jagannathan and Wang (1996), which follows a weighted χ^2 distribution, makes it possible to conduct specification tests on asset pricing models. Hansen and Jagannathan (1997) also show that the first HJ-distance has a nice interpretation as the maximum pricing error of a portfolio 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 Jagannathan and Wang (1996) 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 small pricing errors, a good asset pricing model also should 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.,

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

Therefore, in pricing errors terms, the second HJ-distance differs from the first one by focusing on the pricing errors of not only primary assets but also all possible derivatives that can be formed from the primary assets. If the second HJ-distance of an SDF model is close to zero, then the model not only has small pricing errors but also is arbitrage free, a necessary condition to price derivatives.[9] Therefore, in general we will get very different parameter estimates using the two

[9] One advantage of the second HJ-distance is that even though we only use primary assets in our estimation, the objective function reflects pricing errors of all possible derivatives. This greatly simplifies our empirical analysis, because otherwise we would need to include all possible derivatives in our empirical estimation to fully reflect the no-arbitrage restriction. This can be difficult to do, because even though we might be able to obtain some derivative prices, we might not be able to obtain all.

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 restriction 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 prominent existing approaches in dealing with the no-arbitrage restriction. 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 an econometric framework within which we can evaluate *all* asset pricing models, whether arbitrage-free or not, by their deviations from true asset pricing models. Although truncation guarantees a model's SDF to be nonnegative, it still matters in practice how the truncated model is estimated. Later empirical analysis shows that we obtain very different results when estimating truncated models using the first and second HJ-distances. Of course it would be interesting to combine both approaches: We can use the method of Bansal and Viswanathan (1993) to identify some good proxies of asset pricing models and use our tests to evaluate such models.

II. Asset Pricing Tests in Absence of Arbitrage: Econometric Theory

In this section, we develop the econometric theory of asset pricing tests in absence of arbitrage based on the second HJ-distance. Specifically, we derive the asymptotic distributions of the second HJ-distance for general nonlinear asset pricing models. We also develop the asymptotic distributions of model parameters, Lagrangian multipliers, and pricing errors of individual assets based on the second HJ-distance. These statistics provide useful diagnostic information on potential sources of model misspecifications.

Hansen and Jagannathan (1997) show that it is much easier to consider the following conjugate representations of the minimization problems in (3) and (4):

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

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

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

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

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

Suppose λ_0 and λ_0^+ solve (9) and (10), respectively, then $[H - \lambda_0' Y] \in \mathcal{M}$ and $\left[H - \lambda_0^{+\prime} Y\right]^{+} \in \mathcal{M}^{+}$. 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 $Y : \mathbb{E}\left[\left(\lambda_0' Y\right) Y\right] = \mathbb{E}\left[\left(H - m\right) Y\right]$, 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 $\lambda_0^{+\prime} Y$, although similar to that of $\lambda_0' Y$, is more complicated due to the no-arbitrage restriction: $H - \left[H - \lambda_0^{+\prime} Y\right]^{+}$ is the necessary adjustments to make H a member of $\mathcal{M}^{+}$.

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

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

where

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

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, $\left\{\left(X_{t-1}, Y_{t,} H_t\left(\theta\right)\right) : t = 1, 2, ..., T\right\}$, where θ is a k-dimensional parameter vector. Following Hansen and Jagannathan (1997), we use the empirical counterpart of $\mathbb{E}\phi^{+}\left(\theta, \lambda\right)$,

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

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

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

[10] From 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]

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

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

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

However, standard Taylor expansion breaks down for 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

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

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

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

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)

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

[12] Let 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

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

Proof. See the appendix.[14]

[13] A 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 $\mathbb{E}\left[(f(\theta,\omega) - f(\theta_0,\omega))/(\theta-\theta_0) - \Delta(\omega)\right]^2 \to 0$ as $\theta\to\theta_0$. Similar ideas have been used by Pakes and Pollard (1989) and others to handle non-differentiable criteria functions.

[14] Note 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:

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

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

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

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

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

Proof. See the appendix.

The above result provides a basis of specification test of asset pricing models in the absence of arbitrage. Our result extends the Jagannathan and Wang test in two important ways. First, our test explicitly requires that a true asset pricing model has to be arbitrage free. Second, though Jagannathan and Wang (1996) only consider linear factor models, our test is applicable to SDF models $H\left(\theta\right)$ that have a general nonlinear dependence on θ. These include asset pricing models in which the representative investor has complicated utility functions.

In addition to the above test statistic, we also provide the asymptotic distributions of the estimators of model parameter $\hat{\theta}^+$, the Lagrangian multiplier $\hat{\lambda}^+$, and pricing errors of individual assets $H(\hat{\theta}^+)Y-X$, for general nonlinear SDF model $H\left(\theta\right)$, 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

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

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

Proof. See the appendix.

The asymptotic distribution of parameter estimates provide useful information on model specification. For example, in a linear 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

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

where $G \equiv \mathbb{E}\{Y[\triangledown H(\theta_0)]'\}$ 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

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

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

For the special case of linear factor models, the SDF has the form $H(\theta) = F'\theta$, where F is a vector of risk factors. Moreover, $\triangledown H(\theta) = F$ and $\triangledown^2 H(\theta) = 0$, which do not depend on θ at all. It is straightforward to show that under $\mathbb{H}_0$: $\delta^+ = 0$, our asymptotic distribution of δ^+ is identical to the asymptotic distribution of δ developed in Jagannathan and Wang (1996). The reason is that $\delta^+ = 0$ necessarily implies $\delta = 0$ and, as a result, both HJ-distances have the same asymptotic distributions. We emphasize that this fact, however, does not mean that our δ^+-based test is redundant for several important reasons. First, the implementations of the two tests are fundamentally different. While for the Jagannathan and Wang test we estimate $(\hat{\theta}, \hat{\lambda})$ by minimizing the first HJ-distance, we estimate $(\hat{\theta}^+, \hat{\lambda}^+)$ by minimizing the second HJ-distance for our test. Except for the rare case in which $\delta^+ = 0$, the estimated two HJ-distances and their associated parameters tend to be very different from each other. Second, because of the first reason, the two tests have very different powers in rejecting misspecified models. Even though $\delta^+ = 0$ implies $\delta = 0$, the reverse is not true. Consequently, although the Jagannathan and Wang 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, our econometric approach is quite different from that in the existing literature and can be useful in other finance applications that involve nondifferentiable objective functions. In fact, δ^+ has been rarely used in the literature mainly because its distribution is unknown, and studies that do consider δ^+, such as Wang and Zhang (2004), have to resort to simulation methods.[15]

[15] We believe that the similarity between our test and the existing ones is an advantage of our test: It makes it very convenient for empirical researchers, who have been familiar with the Jagannathan and Wang test, to incorporate the no-arbitrage restriction in their empirical evaluations of asset pricing models.

Hansen, Heaton, and Luttmer (1995) develop the asymptotic distributions of δ and δ^+ when the model is misspecified (i.e., $\delta \neq 0$ and $\delta^+ \neq 0$), respectively, with known model parameters. We can easily extend the above analysis to show that the results of Hansen, Heaton, and Luttmer (1995) still hold when parameters need to be estimated from the data.

III. Application I: Fama-French 25 Portfolios

Many linear factor models considered in the existing asset pricing literature are not arbitrage free because their SDFs can take negative values. To illustrate the importance of the no-arbitrage restriction in traditional asset pricing applications, we apply our new tests to evaluate several well-known asset pricing models using the Fama-French 25 portfolios. Although Hodrick and Zhang (2001) have conducted a comprehensive analysis of asset pricing models using the 25 portfolios based on the first HJ-distance, they do not explicitly address the importance of the no-arbitrage restriction.

A. Data and Asset Pricing Models

The data used in our analysis are monthly returns of the 25 portfolios in excess of one-month T-bill rates and one-month T-bill rates between January 1952 and December 2002. Therefore, the models are required to price both the equity premium and the cross-sectional differences in size and BM portfolios. We obtain the data from Kenneth French's website.

Table 1 provides summary statistics for the excess returns of the 25 portfolios. 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. Table 1 indicates that 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 those 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. Default premium (DEF hereafter), a commonly used conditioning variable, 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 prices. 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.[16]

The first model we consider is the well-known CAPM developed by Sharpe (1964) and Lintner (1965). This model is probably the most widely studied asset pricing model in finance literature. The SDF of this model is

$$H_t^{\text{CAPM}} = b_0 + b_1 r_{\text{MKT},t},$$

where $r_{\text{MKT},t}$ is the excess return on the market portfolio at t, and b_0 and b_1 are constant parameters. The CAPM is a one-period model with constant prices of risk. In a multiple-period setting, the market risk price could be time varying and depend on the state of the economy. 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 the Fama and French (1993) three-factor model, denoted as FF3. In addition to the market factor, FF3 also includes SMB and HML to capture the cross-sectional differences in average stock returns related to size and book-to-market ratio, respectively. FF3 has the following SDF,

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

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. We also consider two conditional versions of FF3, FF3+IP and FF3*IP,

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

with the following SDFs:

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

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

We extend FF3 further to include three additional factors and refer to the new model as FF6. The first additional factor is a momentum factor from Jegadeesh and Titman (1992). Momentum is one of the few anomalies that FF3 cannot capture. We use the momentum factor constructed by Kenneth French. The other two additional factors are TERM (yield spread between 30-year and 1-year government bonds) and DEF (yield spread between Baa and Aaa rated corporate bonds). Data on TERM are obtained from CRSP. Fama and French (1996) show that these bond market factors help explain cross-sectional stock returns.[17]

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 T-bill rate and its one-year backward moving average (from CRSP); and the term structure factor, TERM, the yield spread between 30-year and 1-year government bonds.[18] In the conditional version, we scale the original factors by the state variable IP.

An alternative approach of imposing the no-arbitrage restriction 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

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

In our empirical analysis, we also evaluate the truncated versions of all nine models based on the two HJ-distances. Though truncation guarantees a model's SDF to be nonnegative, it does not guarantee that the model also has small pricing errors (or smaller HJ-distances), because the truncated model could be even further away from true asset pricing models. Through this exercise,

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

[18]The original model of Campbell (1996) has a pricing proxy in the form of $y_t = \exp\left(-f_t{}'b\right)$. We consider its linearized version to illustrate the importance of the no-arbitrage restriction in studying linear factor models.

we examine whether we reach different conclusions using the two HJ-distances for truncated models as well.

B. Empirical Results

In this section, we report the empirical performance of the above models in pricing the Fama-French 25 portfolios. We estimate model parameters, Lagrangian multipliers, and pricing errors of individual assets of each model by minimizing the first and second HJ-distances, respectively.[19]

Table 2 reports specification test results of the nine models (both original and truncated versions) using the 25 portfolios and the 25 portfolios scaled by DEF. We first report the magnitude of the two HJ-distances, $\delta_T(\hat{\theta})$ and $\delta_T^+(\hat{\theta}^+)$, and their differences. If an asset pricing model does not allow arbitrage opportunities, then the two HJ-distances should be close to each other. Otherwise, the two distances could differ from each other dramatically. We then report the probabilities that $H_t(\hat{\theta})$ takes negative values, $p(H < 0)$. Since truncated models cannot take negative values by definition, we omit such information for truncated models. If a model takes negative values with high probability, it allows arbitrage opportunities and should not be considered as an economically sound model. We finally report specification test statistics based on two HJ-distances, $p(\delta = 0)$ from Jagannathan and Wang and $p(\delta^+ = 0)$ from our Theorem 1. If a model is not rejected by either specification test at the 95% confidence level, the corresponding p-value should be higher than 5%.

Panel A of Table 2 reports specification test results of the nine models using the 25 portfolios. For CAPM and its two conditional variations, CAPM+IP and CAPM*IP, the two HJ-distances are very similar to each other. The second HJ-distances $\delta_T^+(\hat{\theta}^+)$ of the three models are slightly larger (about 2%) than the first HJ-distances $\delta_T(\hat{\theta})$. The probabilities that the SDFs take negative values in the three models are close to zero. Therefore, it seems that the no-arbitrage restriction does not make a big difference in our inferences of the three CAPM models. Finally, 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 based on both HJ-distances.

The results of FF3 and its two time-varying extensions, FF3+IP and FF3*IP, 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 all three models, the estimated second HJ-distances $\delta_T^+(\hat{\theta}^+)$ are only slightly bigger (about 2 to 3%) than the first HJ-distances $\delta_T(\hat{\theta})$. The probabilities that the SDFs take negative values are also close to zero. All three models are rejected by the two HJ-distance tests. It is surprising to find that the Fama-French factors do not seem to be adequate to capture the cross-sectional return differences of the Fama-French portfolios. Again, it seems that the no-arbitrage restriction does not significantly affect the inference of the three FF3 models.

[19]In general $(\hat{\theta}, \hat{\lambda})$ and $(\hat{\theta}^+, \hat{\lambda}^+)$ differ from each other, except for the rare cases in which both the HJ-distances are identically zero.

In contrast to the previous six models, the no-arbitrage restriction has a much bigger impact on the inferences of the FF6 model. For example, the second HJ-distance $\delta_T^+(\hat{\theta}^+)$ is about 11% bigger than the first HJ-distance $\delta_T(\hat{\theta})$. There is also a nontrivial probability, about 12%, that $H_t^{\mathrm{FF6}}(\hat{\theta})$ takes negative values. Most important, the specification tests based on the two HJ-distances give very different conclusions on model performance. Although the Jagannathan and Wang test cannot reject FF6 at conventional significance levels (p-value=11%), our test easily rejects the model (p-value=0%). Therefore, the no-arbitrage restriction makes a substantial difference in our inferences of FF6. Without the no-arbitrage restriction, we could have mistakenly concluded that the model does a good job in pricing the 25 portfolios based on $\delta_T(\hat{\theta})$. However, such performance is achieved at the expense of a relatively high probability of $H_t^{\mathrm{FF6}}(\hat{\theta})$ 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 $H_t^{\mathrm{CAM}}(\hat{\theta})$ and $H_t^{\mathrm{CAM*IP}}(\hat{\theta})$ take negative values are also quite substantial and equal to 17% and 20%, respectively. Most interestingly, neither model can be rejected by the Jagannathan and Wang 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 $H_t^{\mathrm{CAM}}(\hat{\theta})$ and $H_t^{\mathrm{CAM*IP}}(\hat{\theta})$ take negative values with high probabilities.

In Panel B of Table 2, we obtain qualitatively similar results for the 25 portfolios scaled by default premium. The no-arbitrage restriction 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 Jagannathan and Wang 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 2 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. As expected, we find that the truncated models are not as flexible as the original ones and tend 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 restriction: One can still reach misleading conclusions using the first HJ-distance for truncated models.

The no-arbitrage restriction also could lead to very different estimated model parameters, Lagrangian multipliers, and pricing errors of individual assets. Table 3 reports model parameters in Panels A and B, Lagrangian multipliers in Panel C, and pricing errors of individual assets in Panel D, estimated using both the first and second HJ-distances. To save space, we only present results for FF6, CAM, and CAM*IP in Panels C and D. The patterns for the other three models are qualitatively similar.

Panels A and B of Table 3 do not show big differences between $\hat{\theta}$ and $\hat{\theta}^{+}$ for the first six models, but the differences become larger for FF6, CAM, and CAM*IP. For FF6, under δ, only the coefficient of the momentum factor (MOM) is significant; but under δ^{+}, the coefficients of both the market and the HML factors become significant, while the coefficient of MOM becomes insignificant. For CAM and CAM*IP, the coefficients of most factors become much smaller under δ^{+}.

Panel C of Table 3 reports the estimates of the Lagrangian multipliers using both HJ-distances. In general, we find more significant Lagrangian multipliers estimated under δ^{+} than under δ, suggesting that the first HJ-distance tends to underestimate the necessary adjustments a model needs to become admissible.

Panel D of Table 3 reports pricing errors of individual assets for the three models estimated using the two HJ-distances. We find that pricing errors of individual assets under δ to be smaller and less significant than that under δ^{+}. These results are consistent with our estimates of $\delta_T(\hat{\theta})$ and $\delta_T^{+}(\hat{\theta}^{+})$ and demonstrate that the first HJ-distance underestimates both overall and individual pricing errors.

The differences between $(\hat{\theta}, \hat{\lambda})$ and $(\hat{\theta}^{+}, \hat{\lambda}^{+})$ can be observed more directly from the estimated SDF models $H_t(\hat{\theta})$ and $H_t(\hat{\theta}^{+})$ and their deviations from true asset pricing models. 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 $H_t(\hat{\theta})$ is much more volatile and 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 a 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 admissible, i.e., 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 $\delta_T(\hat{\theta})$ is much smaller than

$\delta_T(\hat{\theta}^+)$, $H_t(\hat{\theta})$ 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 estimated parameters and conclusions on model performance. Models estimated using the first HJ-distance tend to be much more volatile and take negative values more frequently than those estimated using the second HJ-distance. More dramatic adjustments also are needed to make these models admissible than those estimated using the second HJ-distance. The pricing errors of individual assets are also bigger and more significant under the second HJ-distance. Most important, without the no-arbitrage restriction, we have the illusion that certain models can price the Fama-French 25 portfolios well based on the first HJ-distance.[20]

The analysis in this section illustrates an important point: Even for applications that do not explicitly involve derivatives, the no-arbitrage restriction could still make significant differences in empirical asset pricing studies.

IV. Application II: Hedge Fund Returns

The no-arbitrage restriction is crucial for applications that explicitly involve derivatives. One prominent example of such applications is hedge fund performance evaluation. Compared to regular mutual funds, hedge funds enjoy greater freedom in their investment strategies and frequently use dynamic trading strategies and derivatives. Therefore, models that allow arbitrage opportunities could easily lead to misleading conclusions on hedge fund performance.

A. Hedge Fund Returns and Asset Pricing Models

To examine the importance of the no-arbitrage restriction on hedge fund performance evaluation, we focus on a few specific hedge fund strategies that have been carefully studied in the literature and shown to exhibit option-like returns. Specifically, we study hedge funds which state that they follow any of the following three strategies: trendfollowing, risk arbitrage, and equity derivative arbitrage.

Fung and Hsieh (2001) provide a thorough analysis of the trendfollowing strategy that is followed by the majority of commodity trading advisors (CTAs). Fung and Hsieh (1997) show that returns of trendfollowing funds are uncorrelated with standard equity, bond, currency, and commodity index returns, and they have option-like features—they tend to be large and positive during the best and worst performing months of the world equity markets. Fung and Hsieh (2001) provide the insightful observation that a lookback straddle provides a good characterization of the trendfollowing strategy.

[20] At a deeper level, all models are approximations of reality and therefore are wrong. However, the significance of the no-arbitrage restriction is that it allows us to reject certain models that previous tests fail to reject; more important, it also provides more realistic assessments of the inadequacies of existing models. Such insights are helpful in guiding future research efforts in indentifying better models.

Risk arbitrage, the second strategy we consider, involves simultaneously long and short in companies involved in merger and acquisition. The strategy longs the target company and shorts the acquiring company. Mitchell and Pulvino (2001) show that the returns to this strategy are correlated with market returns in a nonlinear way: Risk arbitrage generates moderate returns in flat or rising market environments, but large negative returns in declining markets. In fact, they demonstrate that risk arbitrage is akin to writing uncovered index put options.

In addition to the above two strategies, we also consider funds that are involved in equity derivative arbitrage, which obviously involves investment 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 more than 4,000 funds from November 1977 to September 2003, which 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 fund 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. As documented by Brown and Goetzmann (1999), the most natural way to group the funds is by their styles because the styles greatly capture the cross-sectional return differences.

Table 4 provides summary statistics of the hedge fund portfolios. Panel A presents distribution information of the three strategies among ten broad investment styles.[21] 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 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 used to evaluate hedge fund performance are similar to those in the previous section for the Fama-French 25 portfolios. However, since we only have ten portfolios to

[21] Dedicated short is excluded because none of the funds belong to this category.

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. 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 2004). 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 to before, we also consider the truncated versions of the above six models.

B. Empirical Results

Table 5 report specification test results of the six asset pricing models in pricing the ten hedge fund portfolios. Panel A contains results for the original models, while Panel B contains results for the truncated models.

In Panel A, 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 Jagannathan and Wang test rejects CAPM (p-value=0%), it fails to reject CAPM+IP (p-value=57%). However, the superior performance of CAPM+IP over CAPM according to the first HJ-distance is mainly because CAPM+IP 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 based on the second HJ-distance. Therefore, without the no-arbitrage restriction, we could have reached the misleading conclusion that CAPM+IP is able to satisfactorily price hedge fund portfolio returns. The negative SDF gives CAPM+IP greater freedom in fitting hedge fund returns and thus underestimates the magnitudes of the pricing errors.

The results of the two FF3 models are very similar. For both FF3 and FF3+IP, the second HJ-distances are much bigger than the first HJ-distances: The percentage differences are 19% and 85%, respectively. The SDFs of both models take negative values with high probabilities: 17% and 43%, respectively. The first HJ-distance of FF3+IP is much smaller than that of all the other models and FF3+IP cannot be rejected by the Jagannathan and Wang 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 and there is no strong evidence of abnormal performance. However, with the no-arbitrage restriction, we reach the dramatically different conclusion that neither model can satisfactorily explain hedge fund returns.

Finally, we examine the performance of the option-based models. 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 they still allow arbitrage opportunities. 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 restriction, OPT+IP would be accepted by the Jagannathan and Wang test (p-value=31%). However, with the no-arbitrage restriction, both models are overwhelmingly rejected by our test. We reach similar conclusions for the truncated models in Panel B, although the differences between the first and second HJ-distances are not as dramatic.

The no-arbitrage restriction also could lead to very different estimated model parameters, Lagrangian multipliers, and pricing errors of individual assets. Table 6 reports model parameters in Panels A and B, Lagrangian multipliers in Panel C, and pricing errors of individual assets in Panel D, estimated using both the first and second HJ-distances. To save space, we only present results for the conditional models in Panels C and D. The patterns for the other three models are qualitatively similar.

Panels A and B of Table 6 show that the coefficients of the SMB, HML, and IP factors under δ^+ are smaller and less significant than that under δ. Panel C of Table 6 reports the estimates of the Lagrangian multipliers using both HJ-distances. Though most Lagrangian multipliers under δ are not significant, most of them under δ^+ are significant, again suggesting that the first HJ-distance underestimates the necessary adjustments a model needs to become admissible. Panel D of Table 6 presents the pricing errors of individual assets for the three models. Under the first HJ-distance, among the three models, only one hedge fund style portfolio, event driven, has significant pricing errors. In contrast, under the second HJ-distance, many more hedge fund style portfolios have significant pricing errors. The magnitudes of pricing errors under δ^+ are also generally bigger than that under δ. These results suggest that by ignoring the no-arbitrage restriction, the first HJ-distance tends to underestimate the abnormal performance of hedge funds.

We illustrate the differences in estimated parameters using 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 $H_t(\hat{\theta})$ is much more volatile and takes negative values much more frequently than $H_t(\hat{\theta}^+)$ for both models. In particular, $H_t(\hat{\theta})$s of both models take negative values during the time periods of 1996-1998, 1999-2000, and after 2002. Interestingly, compared to $H_t(\hat{\theta})$s of the Fama-French 25 portfolios, $H_t(\hat{\theta})$s for hedge funds seem to be even more volatile, suggesting that it is much more difficult to price hedge fund returns. On the other hand, $H_t(\hat{\theta}^+)s$ 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 a true asset pricing model and 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. 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 $\delta_T(\hat{\theta})$ is much smaller than $\delta_T(\hat{\theta}^+)$ 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.

Although the no-arbitrage restriction is helpful in testing traditional asset pricing models using the Fama-French 25 portfolios, it is essential for evaluating the performance of hedge funds. For the Fama-French 25 portfolios, the no-arbitrage restriction makes a significant difference only for the more sophisticated models, such as FF6 and CAM+IP. However, for hedge funds, the no-arbitrage restriction makes a huge difference for most models. To price hedge fund returns, much more volatile SDFs are needed and these models take negative values with much higher probabilities. As a result, the models tend to underestimate model pricing errors and fund abnormal performances. Based on the first HJ-distance, several models suggest that the pricing errors of the ten hedge fund portfolios are not significantly different from zero and thus these funds do not have significant abnormal performances. However, based on the second HJ-distance, all models are overwhelmingly rejected and the magnitude of pricing errors are much bigger than before. Although these results are not definitive evidence that hedge funds can outperform the market, they nonetheless illustrate the crucial role played by the no-arbitrage restriction in evaluating hedge fund performance. Even though the models considered here are some of the most popular ones in the literature, there are still many other models we do not include in our analysis. It is possible that we might reach different conclusions using alternative models. However, the bottom line here is that no matter which model one uses, the new asset pricing tests developed in our paper enable empirical researchers to incorporate the no-arbitrage restriction in their studies.

V. Conclusion

According to modern asset pricing theory, a good asset pricing model should correctly price all assets and have an SDF that is strictly positive to be arbitrage free. Most existing asset pricing tests, however, have mainly focused on model pricing errors and ignored the no-arbitrage restriction. In this paper, we argue that it is important to reflect both theoretical predictions in empirical tests of asset pricing models. Although it may seem that an extra constraint would make it harder to fit the data, we argue that the no-arbitrage restriction leads to more appropriate measures of model performance and parameter estimates. It is also extremely important for applications that involve derivatives. We develop asset pricing tests that explicitly incorporate the no-arbitrage restriction under the SDF framework. Our new tests are natural extensions of existing tests and are just as easy to implement. Therefore, they make it straightforward to incorporate the no-arbitrage restriction in empirical asset pricing studies.

Empirically, we demonstrate the importance of the no-arbitrage restriction through two applications. In the first application, we evaluate several traditional asset pricing models using the Fama-French 25 Size/BM portfolios. In the second application, we evaluate several linear models using hedge fund strategies that explicitly exhibit option-like returns. Our empirical results show that the no-arbitrage restriction makes important differences in both applications, even though the first one does not explicitly involve derivatives. Without the no-arbitrage restriction, we could easily reach the misleading conclusion that certain models can satisfactorily explain the returns of the Fama-French portfolios or hedge fund returns. But, in fact, all these models have SDFs that take negative values with high probabilities and thus are not arbitrage free. The no-arbitrage restriction also leads to better identified model parameters, better behaved models, and provides more insights on potential sources of model misspecifications.

In the current asset pricing literature, it is often the case that different models are used to price primary and derivative assets. For example, we tend to use CAPM type models to price stocks and Black and Scholes (1973) type models to price options. We also use yield factor models to price bonds and Heath, Jarrow, and Morton (1992) type models to price interest rate derivatives. Due to both theoretical and practical considerations, it would be much more satisfying if we could develop models that can simultaneously price both primary and derivative assets. Obviously such models have to satisfy the minimum requirement of being arbitrage free. Therefore, it is our hope that the new tests developed in our paper would contribute to the developments of the asset pricing literature by making it possible to incorporate the no-arbitrage restriction in empirical asset pricing studies.

REFERENCES

Agarwal, Vikas and Narayan Y. Naik, 2004, Risks and portfolio decisions involving hedge funds, *Review of Financial Studies,* forthcoming.

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2004, The cross section of volatility and expected returns, *Journal of Finance,* forthcoming.

Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637-654.

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, Armsterdam.

Duffie, Darrell, 2001, *Dynamic Asset Pricing Theory*, Princeton University Press, New Jersey.

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, 1997, Empirical characteristics of dynamic trading strategies: The case of hedge funds, *Review of Financial Studies* 10, 275-302.

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, Jonathan Ingersoll, Matthew Spiegel, 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.

Heath, D., R. Jarrow, and A. Morton, 1992, Bond pricing and the term structure of interest rates: A new methodology, *Econometrica* 60, 77-105.

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.

Hodrick, Robert J. and Xiaoyan Zhang, 2001, Evaluating the specification errors of asset pricing models, *Journal of Financial Economics* 62, 327-376.

Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield, Maryland.

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.

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 market 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.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Wang, Zhenyu and Xiaoyan Zhang, 2004, The role of arbitrage in the empirical evaluation of asset pricing models, unpublished 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 $\left(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0}\right)$ for $t = 0, 1, ..., T$. 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) :

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

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

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

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

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

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

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

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 $r \to 0$ almost surely as $(\theta,\lambda) \to (\theta_0,\lambda_0)$. 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,

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

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))

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

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:

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

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

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 $\mathbb{H}_0$: $\delta^+=0$, $T\left[\delta_T^+\right]^2$ is asymptotically distributed as a weighted χ^2 distribution with $(n-k)$ degrees of freedom:

$$T\left[\delta_T^+\right]^2\to Z'\Xi Z,$$

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

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

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

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

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

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

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

The above quadratic in U is minimized at

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

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

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

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

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

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 $\left(\mathbb{E}_T - \mathbb{E}\right)\phi^{+}(\theta_0,\lambda_0)$ 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, $\left(\mathbb{E}_T - \mathbb{E}\right)\phi^{+}(\theta_0,\lambda_0) = 0$.

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

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

and

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

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

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

and

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

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

$$\Gamma_{21} = \mathbb{E}\left\{g^{(2)} Y[\triangledown H(\theta_0)]'\right\} = \mathbb{E}\left\{2 \cdot \mathbf{1}_{E(\theta_0,\lambda_0)} Y[\triangledown H(\theta_0)]'\right\} = \mathbb{E}\left\{2Y[\triangledown H(\theta_0)]'\right\}$$

$$\Gamma_{22} = \mathbb{E}[-g^{(2)} YY'] = \mathbb{E}[-2YY'].$$

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))

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

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

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

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

By Assumption A.9, $\sqrt{T}\left(\mathbb{E}_T-\mathbb{E}\right)\left[H\left(\theta_0\right)Y-X\right]$ has an asymptotic limiting distribution $N\left(0,\Lambda\right)$. It follows immediately that under $\mathbb{H}_0$: $\delta^+ = 0$, the asymptotic distribution of $T\left[\delta_T^+\right]^2$ 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

$$I - D^{-1/2}G\left(G'D^{-1}G\right)^{-1}G'D^{-1/2} \tag{A.7}$$

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

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

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

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

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

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)

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

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

$$\left(G'D^{-1}G\right)^{-1}G'D^{-1}\Lambda D^{-1}G\left(G'D^{-1}G\right)^{-1}. ■$$

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

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

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

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

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

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

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

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

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

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

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

Since

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

we have

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

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

Table 1. Summary Statistics for 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 2. Asset Pricing Tests Based on the HJ-Distances for the Fama-French 25 Portfolios

This table provides empirical results on specification tests of nine asset pricing models and their truncated versions based on the two HJ-distances for the Fama-French 25 Size/BM portfolios from January 1952 to December 2002. Panels A and B contain results for original models using the 25 portfolios and the 25 portfolios scaled by default premium (DEF), respectively. DEF is the yield difference between Baa and Aaa rated corporate bonds. Panels C and D contain corresponding results for the truncated versions of the nine models 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 Panels A and B, 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 Fama-French 25 portfolios

	CAPM	CAPM+IP	CAPM*IP	FF3	FF3+IP	FF3*IP	FF6	CAM	CAM*IP
δ	0.417	0.416	0.416	0.364	0.364	0.361	0.308	0.295	0.276
δ^+	0.425	0.423	0.422	0.373	0.372	0.371	0.343	0.369	0.357
$(\delta^+ - \delta)/\delta$	2%	2%	2%	2%	2%	3%	11%	25%	30%
p(H<0)	0%	0%	0%	0%	0%	0%	12%	17%	20%
$p(\delta=0)$	0%	0%	0%	0%	0%	0%	11%	26%	21%
$p(\delta^+=0)$	0%	0%	0%	0%	0%	0%	0%	0%	0%

Panel B: Results for original models using Fama-French 25 portfolios scaled by DEF

	CAPM	CAPM+IP	CAPM*IP	FF3	FF3+IP	FF3*IP	FF6	CAM	CAM*IP
δ	0.368	0.368	0.367	0.325	0.324	0.320	0.242	0.253	0.222
δ^+	0.387	0.387	0.387	0.340	0.340	0.334	0.293	0.320	0.305
$(\delta^+ - \delta)/\delta$	5%	5%	5%	4%	5%	4%	21%	26%	37%
p(H<0)	0%	0%	0%	0%	0%	1%	12%	14%	17%
$p(\delta=0)$	0%	0%	0%	0%	0%	0%	7%	42%	48%
$p(\delta^+=0)$	0%	0%	0%	0%	0%	0%	0%	0%	0%

Panel C: Results for truncated models using Fama-French 25 portfolios

	CAPM	CAPM+IP	CAPM*IP	FF3	FF3+IP	FF3*IP	FF6	CAM	CAM*IP
δ	0.417	0.416	0.416	0.366	0.366	0.365	0.316	0.315	0.315
δ^+	0.425	0.423	0.422	0.372	0.371	0.370	0.338	0.362	0.347
$(\delta^+ - \delta)/\delta$	2%	2%	2%	2%	2%	2%	7%	15%	10%
$p(\delta=0)$	0%	0%	0%	0%	0%	0%	2%	12%	0%
$p(\delta^+=0)$	0%	0%	0%	0%	0%	0%	0%	0%	0%

Panel D: Results for truncated models using Fama-French 25 portfolios scaled by DEF

	CAPM	CAPM+IP	CAPM*IP	FF3	FF3+IP	FF3*IP	FF6	CAM	CAM*IP
δ	0.368	0.368	0.367	0.328	0.326	0.319	0.251	0.267	0.245
δ^+	0.387	0.387	0.387	0.339	0.338	0.333	0.290	0.319	0.298
$(\delta^+ - \delta)/\delta$	5%	5%	5%	3%	4%	4%	16%	20%	21%
$p(\delta=0)$	0%	0%	0%	0%	0%	0%	2%	18%	6%
$p(\delta^+=0)$	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 3. Estimates of Model Parameter, Lagrangian Multipliers, and Pricing Errors of Individual Assets Based on the Two HJ-Distances for the Fama-French 25 Portfolios

This table provides estimates of model parameters, Lagrangian multipliers, and pricing errors of individual assets of the nine asset pricing models based on the two HJ-distances for the Fama-French 25 portfolios from January 1952 to December 2002. Panels A and B report the estimates of the parameters of the nine models based on the first and second HJ-distances, respectively. Panel C reports the estimates of the Lagrangian multipliers and Panel D reports the pricing errors for the 25 portfolios based on the two HJ-distances.

Panel A: Model parameters estimated based on the first HJ-distance (bold entries indicate that the parameters are significant at the 5% level)

	CNST	MKT						
CAPM	**1.01**	**-3.02**						
	CNST	MKT	IP					
CAPM+IP	**1.01**	**-3.29**	-6.27					
	CNST	MKT	IP	MKT*IP				
CAPM*IP	**1.01**	**-3.19**	-6.56	18.00				
	CNST	MKT	SMB	HML				
FF3	**1.06**	**-4.59**	-2.10	**-8.11**				
	CNST	MKT	SMB	HML	IP			
FF3+IP	**1.05**	**-4.68**	-2.24	**-8.03**	-2.96			
	CNST	MKT	SMB	HML	IP	MKT*IP	SMB*IP	HML*IP
FF3*IP	**1.05**	**-5.49**	-3.02	**-9.12**	0.02	-0.90	-1.60	-1.00
	CNST	MKT	SMB	HML	MOM	TERM	DEF	
FF6	0.67	-1.74	-0.74	-3.36	**-9.14**	1.96	9.71	
	CNST	MKT	LBR	DY	TRB	TERM		
CAM	-0.47	-0.01	-2.84	**55.51**	24.42	**-8.45**		
	CNST	MKT	LBR	DY	TRB	TERM		
CAM*IP	-0.27	-0.27	-19.55	52.49	130.22	-7.60		
	IP	MKT*IP	LBR*IP	DY*IP	TRB*IP	TERM*IP		
CAM*IP	0.11	-1.90	-21.60	-5.10	-76.70	6.80		

Panel B: Model parameters estimated based on the second HJ-distance (bold entries indicate that the parameters are significant at the 5% level)

	CNST	MKT						
CAPM	**1.01**	**-3.04**						
	CNST	MKT	IP					
CAPM+IP	**1.01**	**-3.30**	-5.99					
	CNST	MKT	IP	MKT*IP				
CAPM*IP	**1.01**	**-3.13**	-7.39	36.90				
	CNST	MKT	SMB	HML				
FF3	**1.06**	**-4.61**	-2.08	**-8.08**				
	CNST	MKT	SMB	HML	IP			
FF3+IP	**1.05**	**-4.72**	-2.28	**-7.97**	-3.99			
	CNST	MKT	SMB	HML	IP	MKT*IP	SMB*IP	HML*IP
FF3*IP	**1.05**	**-4.82**	-2.53	**-8.09**	-0.03	0.30	-1.30	0.40
	CNST	MKT	SMB	HML	MOM	TERM	DEF	
FF6	0.79	**-3.61**	-1.73	**-6.55**	-3.81	-0.18	6.26	
	CNST	MKT	LBR	DY	TRB	TERM		
CAM	0.52	-1.69	-0.77	19.26	90.02	-3.35		
	CNST	MKT	LBR	DY	TRB	TERM		
CAM*IP	0.67	-1.37	1.71	15.92	65.99	-3.46		
	IP	MKT*IP	LBR*IP	DY*IP	TRB*IP	TERM*IP		
CAM*IP	-0.11	-0.10	-1.30	-0.30	-52.60	48.80		

Panel C: Estimates of Lagrangian multipliers (bold entries indicate that the estimates are significant at the 5% level)

	Use δ			Use δ+		
	FF6	CAM	CAM*IP	FF6	CAM	CAM*IP
11	**-4.95**	-3.62	-3.75	**-7.62**	**-7.49**	**-7.19**
12	-1.31	-1.19	0.18	-1.30	-1.01	-0.85
13	-3.99	-5.65	-5.17	-2.32	-3.45	-3.48
14	**14.73**	7.16	7.12	**15.59**	**14.04**	**14.56**
15	-4.00	-1.02	-2.32	0.78	3.09	1.74
21	0.06	-0.72	0.40	-3.32	-4.51	-3.73
22	3.06	2.40	-0.80	0.03	-1.81	-1.47
23	2.23	2.70	3.06	4.30	5.10	4.54
24	0.58	1.99	4.46	2.04	2.72	3.92
25	3.66	6.07	4.85	0.64	3.81	2.91
31	-5.19	-0.20	-0.99	-2.76	-1.33	-1.54
32	3.45	1.81	3.59	1.88	0.00	0.63
33	-8.18	-5.27	-3.41	-7.56	-6.82	**-8.44**
34	0.38	1.34	-2.55	-0.37	0.36	-0.38
35	0.55	0.08	0.36	-0.19	0.18	1.06
41	7.09	4.26	5.35	**9.54**	**8.63**	**8.40**
42	**-12.18**	**-14.33**	**-13.52**	**-10.16**	**-10.40**	**-10.75**
43	3.54	6.10	3.24	3.74	4.67	4.15
44	1.06	-2.02	-2.89	1.32	-0.56	-0.35
45	-1.17	-1.14	1.14	-1.16	-0.69	0.22
51	4.36	1.61	-0.04	3.51	0.71	-0.38
52	-1.52	-3.10	-2.53	-1.10	-2.37	-0.36
53	4.58	6.43	**7.35**	4.94	**6.55**	**7.04**
54	-2.32	-0.54	0.56	-3.76	-2.56	-2.93
55	-2.36	-1.40	-2.24	-3.17	-2.64	**-3.44**

Panel D: Estimates of pricing errors of individual assets (bold entries indicate that the estimates are significant at the 5% level)

	Use δ			Use δ+		
	FF6	CAM	CAM*IP	FF6	CAM	CAM*IP
11	**-0.34%**	-0.41%	-0.31%	**-0.48%**	-0.47%	-0.41%
12	-0.07%	-0.13%	-0.03%	-0.08%	0.01%	0.03%
13	-0.08%	-0.13%	-0.09%	-0.04%	0.07%	0.08%
14	0.11%	0.06%	0.08%	0.15%	**0.31%**	0.31%
15	0.00%	0.08%	0.06%	0.05%	**0.34%**	**0.31%**
21	-0.14%	-0.21%	-0.10%	-0.23%	**-0.33%**	**-0.28%**
22	-0.02%	0.00%	-0.01%	-0.07%	0.01%	0.02%
23	0.02%	**0.08%**	0.09%	0.05%	0.20%	0.19%
24	0.03%	0.13%	0.15%	0.04%	0.26%	0.26%
25	0.07%	0.25%	0.20%	0.02%	**0.35%**	0.32%
31	-0.10%	-0.11%	-0.04%	-0.08%	-0.20%	-0.18%
32	0.00%	0.03%	0.08%	0.00%	0.06%	0.07%
33	-0.12%	0.01%	0.02%	-0.08%	0.10%	0.07%
34	0.01%	0.16%	0.09%	0.01%	**0.25%**	0.22%
35	-0.03%	0.12%	0.09%	-0.06%	**0.26%**	0.23%
41	0.05%	-0.07%	0.00%	0.05%	-0.12%	-0.10%
42	-0.20%	-0.19%	-0.16%	-0.16%	-0.11%	-0.11%
43	-0.01%	0.11%	0.06%	0.01%	0.17%	0.14%
44	0.00%	0.07%	0.04%	-0.01%	0.18%	0.16%
45	-0.08%	0.09%	0.10%	-0.11%	0.19%	0.19%
51	0.11%	-0.03%	-0.04%	0.11%	-0.07%	-0.09%
52	-0.07%	-0.07%	-0.05%	-0.05%	-0.05%	-0.01%
53	0.01%	0.10%	0.13%	0.01%	0.11%	0.14%
54	-0.08%	0.09%	0.08%	-0.12%	0.11%	0.09%
55	-0.18%	0.01%	-0.06%	-0.24%	0.08%	0.04%

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. Evaluation of Hedge Fund Returns Based on the Two HJ-Distances

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%

Table 6. Estimates of Model Parameters, Lagrangian Multipliers, and Pricing Errors for Individual Assets Based on the Two HJ-Distances for the Ten Hedge Fund Portfolios

This table provides estimates of model parameters, Lagrangian multipliers, and pricing errors of individual assets of the six asset pricing models based on the two HJ-distances for the ten hedge fund portfolios from January 1994 to September 2003. Panels A and B report estimates of the parameters of the six models based on the first and second HJ-distances, respectively. Panel C reports the estimates of the Lagrangian multipliers and Panel D reports the pricing errors for the ten hedge fund portfolios based on the two HJ-distances.

Panel A: Model parameters estimated based on the first HJ-distance (bold entries indicate that the parameters are significant at the 5% level)

	CNST	MKT	SMB	HML	STR	PUT	IP
CAPM	**1.04**	**-6.95**					
CAPM+IP	**0.07**	-3.16					**-2.36**
FF3	**0.84**	1.04	**22.74**	**29.70**			
FF3+IP	-0.03	4.46	14.80	23.89			**-2.20**
OPT	**0.86**	-22.22			-1.00	-0.36	
OPT+IP	0.05	-9.83			-0.20	-0.18	**-2.21**

Panel B: Model parameters estimated based on the second HJ-distance (bold entries indicate that the parameters are significant at the 5% level)

	CNST	MKT	SMB	HML	STR	PUT	IP
CAPM	**1.05**	**-8.10**					
CAPM+IP	**0.83**	-6.96					-0.51
FF3	**1.00**	-8.50	11.06	5.46			
FF3+IP	**0.85**	-7.91	8.55	3.22			-0.41
OPT	**1.04**	-9.06			-0.55	0.05	
OPT+IP	**0.87**	-7.28			-0.37	0.04	-0.43

Panel C: Estimates of Lagrangian multipliers (bold entries indicate that the estimates are significant at the 5% level)

	Use δ			Use δ+		
λ	CAPM+IP	FF3+IP	OPT+IP	CAPM+IP	FF3+IP	OPT+IP
convertible arbitrage	27.8	25.5	26.4	**64.4**	**60.9**	**60.6**
emerging markets	10.1	9.4	9.6	3.7	5.2	3.9
equity market neutral	61.1	23.8	53.4	**210.1**	**196.2**	**208.2**
event driven	-10.6	11.3	-4.9	**42.1**	**50.4**	**49.7**
fixed income	14.3	14.5	11.4	**39.3**	**29.0**	**34.2**
fund of funds	**-86.8**	**-80.6**	**-86.2**	**-180.1**	**-174.3**	**-182.1**
global macro	4.3	4.3	5.0	**13.7**	**11.9**	**13.9**
long/short equity	16.0	4.7	13.3	**49.2**	**46.9**	**47.2**
managed futures	4.9	5.8	5.9	**19.1**	**22.0**	**20.6**
multi-strategy	17.3	14.4	18.1	**11.3**	**12.5**	**12.5**

Panel D: Estimates of pricing errors of individual assets (bold entries indicate that the estimates are significant at the 5% level)

	Use δ			Use δ+		
	CAPM+IP	FF3+IP	OPT+IP	CAPM+IP	FF3+IP	OPT+IP
convertible arbitrage	0.16%	0.24%	0.17%	**0.52%**	**0.59%**	**0.49%**
emerging markets	0.43%	0.74%	0.52%	-0.13%	0.17%	-0.04%
equity market neutral	0.15%	0.04%	0.13%	**0.35%**	**0.30%**	**0.33%**
event driven	0.00%	**0.27%**	0.10%	0.31%	**0.54%**	**0.40%**
fixed income	0.02%	0.09%	0.05%	0.25%	0.20%	0.25%
fund of funds	-0.10%	0.01%	0.01%	0.14%	0.27%	0.20%
global macro	0.63%	0.82%	1.12%	1.19%	1.17%	**1.45%**
long/short equity	0.09%	0.10%	0.15%	0.30%	**0.58%**	0.35%
managed futures	-0.17%	-0.20%	-0.04%	0.79%	0.74%	**0.82%**
multi-strategy	0.64%	0.75%	0.94%	0.28%	0.59%	0.51%

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.

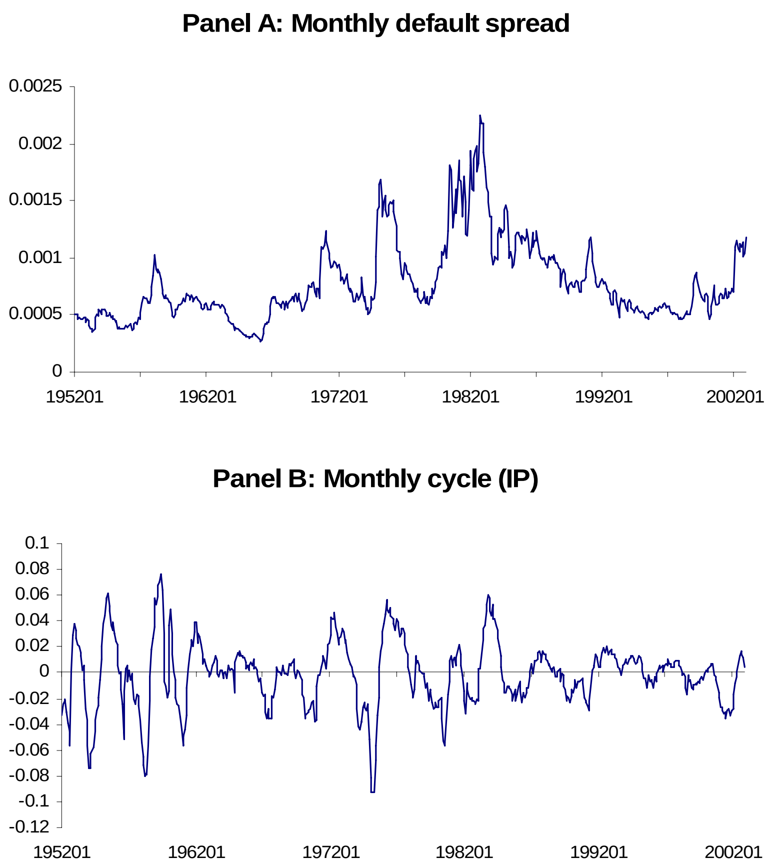

Figure 2. Time series plots of the SDFs of two asset pricing models estimated using the 25 Fama-French size/BM portfolios.

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 SDFs of the Fama-French six-factor model and the linearized Campbell model, respectively.

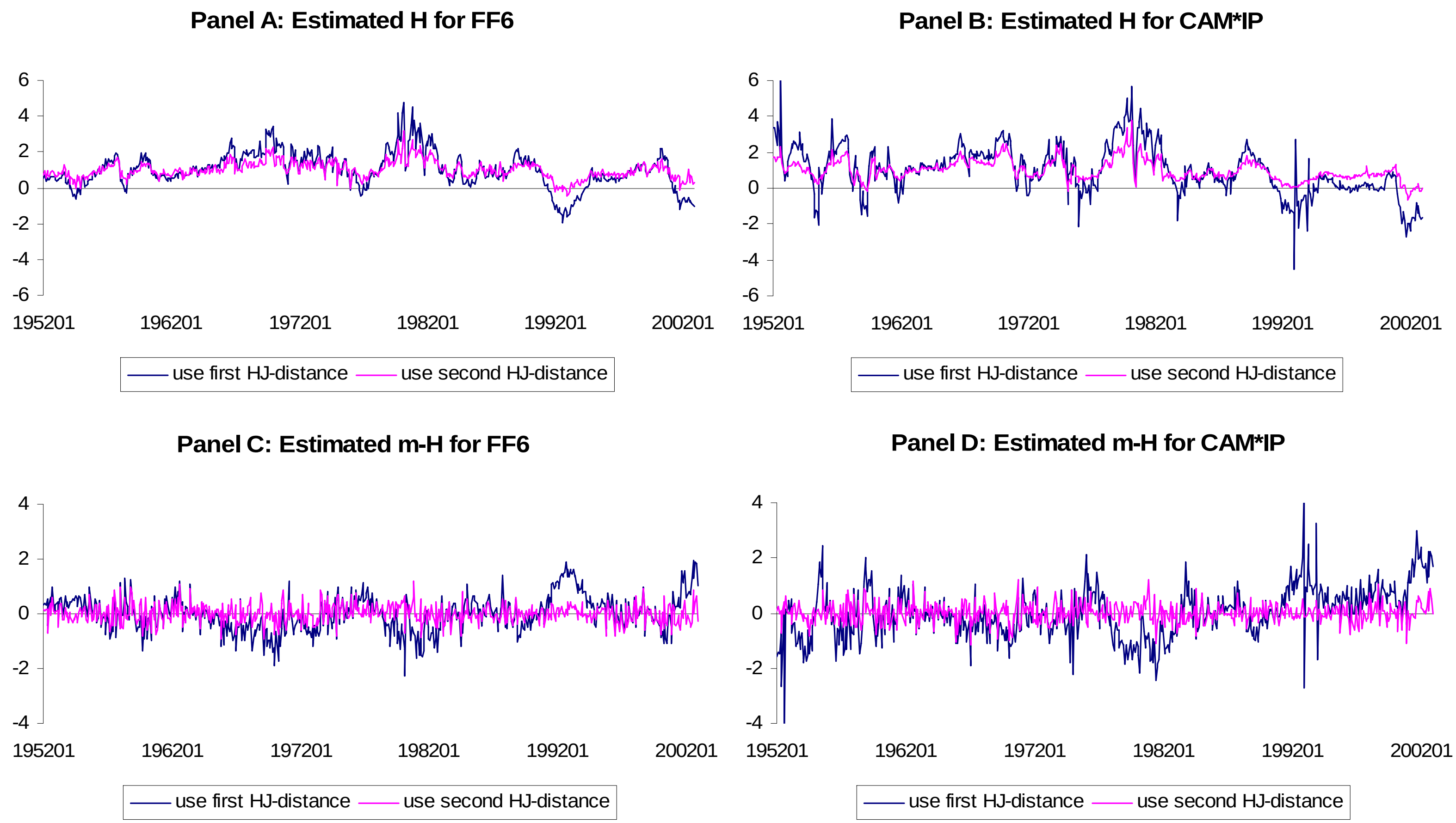

Figure 3. Time series plots of the SDFs of two asset pricing models estimated using the ten hedge fund portfolios.

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 SDFs of the FF3+IP model and the OPT+IP model, respectively.