Journal of Empirical Finance 19 (2012) 65–78

Contents lists available at SciVerse ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Empirical evaluation of asset pricing models: Arbitrage and pricing errors in contingent claims☆

Zhenyu Wang, Xiaoyan Zhang *

ARTICLE INFO

Article history:
Received 13 September 2007
Received in revised form 3 November 2011
Accepted 7 November 2011
Available online 12 November 2011

EL classification code:
G1
G12

Keywords:
Bayesian inference
Asset pricing
Pricing errors
Model comparison
Contingent claims

ABSTRACT

Hansen and Jagannathan (1997) have developed two measures of pricing errors for asset-pricing models: the maximum pricing error in all static portfolios of the test assets and the maximum pricing error in all contingent claims of the assets. In this paper, we develop simulation-based Bayesian inference for these measures. While the literature reports that the time-varying extensions substantially reduce pricing errors of classic models on the standard test assets, our analysis shows that the reduction is much smaller based on the second measure. Those time-varying models have large pricing errors on the contingent claims of the test assets because their stochastic discount factors are often negative and admit arbitrage opportunities.

1. Introduction

An asset-pricing model admits arbitrage opportunities for some contingent claims if the model's stochastic discount factor (SDF) is zero or negative with a positive probability (Hansen and Richard, 1987; Harrison and Kreps, 1979). For example, the SDF of the CAPM, as a linear function of the return on the market portfolio, can be negative and may thereby admits arbitrage opportunities for an index option on the market portfolio (Dybvig and Ingersoll, 1982). Linear asset-pricing models are not arbitrage free because their SDFs may take negative values. When a model admits arbitrage opportunities, derivative securities can be used to generate Jensen's alpha with respect to the model (Guasoni et al., 2011). Thus a linear model that prices all the test assets correctly can still have pricing errors on the derivatives of the assets.

Even for portfolios that do not directly contain derivative securities, models admitting arbitrage opportunities for contingent claims may still give incorrect valuations, according to Black and Scholes (1973), because dynamically managed portfolios can approximate contingent claims. Fung and Hsieh (1997) present empirical evidence showing the derivative-like behavior of hedge

☆ Wang is at the Federal Reserve Bank of New York. Zhang is at Purdue University. For helpful comments, we thank Tobias Adrian, Andrew Ang, Geert Bekaert, Tim Bollerslev, Mike Chernov, Gur Huberman, Paul Glasserman, Larry Glosten, Cam Harvey, Bob Hodrick, David Hsieh, Ravi Jagannathan, Michael Johannes, Jonathan Lewellen, Simon Potter, Tano Santos, Ernst Schaumburg, an anonymous referee, and the seminar participants in Rice University, Penn State University, Cornell University, University of Michigan at Ann Arbor, Case Western Reserve University, University of Texas at Austin, Columbia University, the annual meeting of Western Finance Association, Duke University, and the Federal Reserve Bank of New York. The authors acknowledge support from the Cornell Theory Center and research assistance from Alexis Iwaniszliw, Brian Kellar, Michal Lementowski, Kevin Pan, and Yongjun Tang. The views stated here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System.

* Corresponding author. Tel.: +765 496 7674.
E-mail address: zhang654@purdue.edu (X. Zhang).

0927-5398/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2011.11.001

funds. Fung and Hsieh (2001) further show that the trend-following strategies used by some hedge funds are akin to a look-back straddle. Mitchell and Pulvino (2001) demonstrate that the strategies of risk arbitrage funds are similar to an uncovered put. In addition, they construct stock portfolios based on merger announcements and show that such portfolios also behave like an uncovered put. Therefore, a model that works well on the test assets may have pricing errors on portfolios that are dynamically constructed from the assets (Glosten and Jagannathan, 1994). We therefore should consider pricing errors on all contingent claims or dynamic portfolios of the assets when evaluating asset-pricing models.

For model evaluation, Hansen and Jagannathan (1997) have developed two measures of pricing errors, referred to as HJ distances in the finance literature. The first HJ distance is the maximum pricing error on all static portfolios of a given set of assets, while second HJ distance is the maximum pricing error on all contingent claims on the assets. The difference between the two HJ distances indicates the additional pricing errors a model can have if contingent claims or dynamic portfolios are added to an empirical test of the model. Since it is difficult to test a model with all possible contingent claims or all possible dynamic portfolios, the second HJ distance provides a convenient and powerful tool for the evaluation of asset-pricing models.

There have been efforts to develop classic sampling distribution theories for the HJ distances. Hansen et al. (1995) made the first attempt on both HJ distances. Their sampling distribution theory assumes that the true distance is known and nonzero. This assumption is inconvenient because most applications do not provide a hypothesis about the magnitude of the distance. In the specification tests of HJ distances, the null hypothesis is that the distance is zero. Allowing for a nonzero distance, Jagannathan and Wang (1996) derived an asymptotic sampling distribution theory for the first HJ distance, and Li et al. (2010) did the same for the second HJ distance. Note that the second HJ distance is a complicated nonlinear function of asset returns and thus its sampling distribution is far more involved than the distribution of the first HJ distance. Although these sampling distribution theories allow for testing hypotheses with complicated combinations of the χ-squared distributions, a methodology that is more convenient for applications and allows for formal inferences on model comparisons will be useful.

We introduce a simulation-based Bayesian inference for the analysis of both HJ distances. Using the Bayesian inference we obtain the joint posterior distribution of the two HJ distances, which is convenient for formal inference in model comparisons. We also obtain the posterior distributions of many nonlinear measures of interest, such as the ratio of the second HJ distances of two models in a comparison. The methodology developed in this study allows for the comparison of models based on their pricing errors on either test assets or contingent claims. More important, we can use the methodology to compare performances of different models in all dynamic portfolios without actually constructing the portfolios. The simulation-based Bayesian inferenceoffers two advantages over the classic sampling theory. First, by conducting simulations, we overcome the small-sample bias of the asymptotic method. Second, based on the posterior distributions, we are able to conduct inference on many interesting measures for which asymptotic distributions are difficult.

Although the first HJ distance has gained popularity in empirical research, the second HJ distance has not been widely used in applications. Based on the first HJ distance, many researchers, such as Jagannathan and Wang (1996), Hodrick and Zhang (2001), and Lettau and Ludvigson (2002), report that time-varying linear models have substantially smaller pricing errors than the CAPM and consumption-based models. The analyses of these authors are based on only the first HJ distance. The empirical evaluation of these models in the literature, however, ignores the pricing errors on the contingent claims or dynamic portfolios of the test assets. This concern is especially serious for models whose SDFs often take negative values. According to Dybvig and Ross (1985) and Glosten and Jagannathan (1994), a model admitting arbitrage for contingent claims is likely to have large pricing errors on derivative securities on the models' factors. Therefore, the second HJ distance of a model measures the ability of the model to price dynamic portfolios of the test assets. Our paper, along with Li et al. (2010), seeks to fill that gap in the literature.

Using the simulation-based Bayesian inference of the HJ distances, we investigate whether the pricing errors on contingent claims or dynamic portfolios substantially affect the evaluation of linear time-varying models. We find that the two HJ distances are about the same for static single-factor models, but that the two distances are drastically different for time-varying models. If we evaluate models by the first HJ distance, multifactor and time-varying models have substantially smaller pricing errors than static single-factor models. However, this result does not hold if we use the second HJ distance, mainly because time-varying models admit arbitrage opportunities for contingent claims.

The remainder of the paper is organized as follows. In Section 2, we lay out the econometric framework by reviewing the measures of pricing errors and describe the simulation-based Bayesian inference in Section 3. Then, we present the test assets and the models under examination in Section 4. Empirical results in Section 5 show how pricing errors on contingent claims affect model evaluation and comparison. We conclude in Section 6.

2. Measures of pricing errors

Suppose there are n assets. We use a $n\times 1$ vector r_t to denote the asset returns during period t. Suppose there are k observable factors and l state variables in the economy. At the end of period t, the vector of the factors is f_t,edand the vector of the state variables is x_t. Let $z_t=(r'_t,f'_t,x'_t)'$ and assume that z_t follows a stationary stochastic process with a finite second moment.

An asset-pricing model is represented by its stochastic discount factor, which is denoted by m_t. We assume $m_t\in L^2$, where L^2 is the space of random variables with finite second moments. If the asset-pricing model holds exactly on the assets, the SDF of the model satisfies

$$E_{t-1}[m_t r_t]=1_n, \tag{1}$$

where $E_{t-1}[\cdot]$ is the expectation under the conditional information in period *t-1*, and 1_n is an $n \times 1$ vector of 1 s. Under unconditional expectations, the moment restriction of the pricing model is

$$E[m_t r_t] = 1_n, \tag{2}$$

where $E[\cdot]$ is the unconditional expectation. Following Hansen and Jagannathan (1997), we define the set of all SDFs that satisfy the pricing restriction (2) as

$$\mathcal{M} = \left\{ m_t : m_t \in L^2, E[m_t r_t] = 1_n \right\}, \tag{3}$$

where L^2 is the space of random variables with finite second moments. If m_t is the SDF of an equilibrium model, it should not be negative because the SDF is the marginal rate of substitution of consumption between today and tomorrow. Following Hansen and Jagannathan again, we define

$$\mathcal{M}_+ = \left\{ m_t : m_t \in L^2, m_t \geq 0, E[m_t r_t] = 1_n \right\}. \tag{4}$$

This is the set of nonnegative SDFs that correctly values the scaled returns on average. We assume that $\mathcal{M}_+$ is nonempty. This assumption holds if the observed prices of the scaled portfolios do not allow arbitrage opportunities.

Let y_t be the SDF of an asset-pricing model that we want to evaluate empirically. In general, the prices assigned by y_t may not be consistent with Eq. (2). Hansen and Jagannathan (1997) introduce two measures of pricing errors. The first measure is the maximum pricing error on the static portfolios of the assets:

$$\delta = \max_{h_t \in P, \|h_t\| = 1} \left(\min_{m_t \in \mathcal{M}} |E[y_t h_t] - E[m_t h_t]| \right), \tag{5}$$

where P is the space of linear combinations of asset returns. The restriction $\|h_t\| = 1$ controls the second moment of h_t in the maximization. The second measure, denoted by δ_+, is the maximum pricing error in all contingent claims of the assets:

$$\delta_+ = \max_{h_t \in L^2, h_t = 1} \left(\min_{m_t \in \mathcal{M}_+} |E[y_t h_t] - E[m_t h_t]| \right). \tag{6}$$

The difference between the two measures is the payoff space: δ uses P whereas δ_+ uses L^2. Thus, the first HJ distance δ ignores the pricing errors on the contingent claims outside P. Hansen and Jagannathan show that these measures of pricing errors can be calculated using the following formulas:

$$\delta = \min_{m \in \mathcal{M}} \|y - m\| \text{ and } \quad \delta_+ = \min_{m \in \mathcal{M}_+} \|y - m\|. \tag{7}$$

It is clear that $\delta \leq \delta_+$.

HJ distances are related to Jensen's alpha, which is a common measure of pricing errors. Since alpha is the expected return of a portfolio that neutralizes the risk of the factors in a model, investors want to maximize the Sharpe ratio of the factor-neutral portfolio. Guasoni et al. (2011) Theorem 1 shows that the maximum Sharpe ratio of the factor-neutral portfolios is $S_{\max} = (1 + r_f)\delta$, where r_f is the risk-free interest rate. If the volatility of the factor-neutral portfolio is constrained to be σ, the maximum alpha of the portfolios is $\alpha_{\max} = (1 + r_f)\delta\sigma$. Since the gross return of the risk-free asset is approximately equal to 1, we can roughly interpret the first HJ distance as the maximum Sharpe ratio of the factor-neutral portfolios. We can also roughly interpret the first HJ distance as the maximum alpha of all the factor-neutral portfolios with unit volatility ($\sigma = 1$). If we expand the portfolios to include all the contingent claims on the test assets, the maximum Sharpe ratio and alpha of the factor-neutral portfolios are $S^+_{\max} = (1 + r_f)\delta_+$ and $\alpha^+_{\max} = (1 + r_f)\delta_+\sigma$, respectively. The interpretation of the second HJ distance in terms of the Sharpe ratio and alpha is similar to the first HJ distance.

In an empirical evaluation of an asset-pricing model, the model often has unknown parameters. A general form of the prespecified SDF is $y_t = g(\theta, f_t, z_{t-1})$. Note that we allow the SDF in a specified model to depend on lagged variables. The functional form $g(\cdot,\cdot,\cdot)$ is prespecified, and the vector of parameters, θ, is unknown but belongs to a set Θ. Therefore, the HJ distances of y_t are functions of θ and should be denoted by $\delta(\theta)$ and $\delta_+(\theta)$. Researchers usually choose parameters to minimize pricing errors. We define $\delta = min_\theta \delta(\theta)$ and $\delta_+ = min_\theta \delta_+(\theta)$. By $\hat{\theta}$ and $\hat{\theta}_+$, we denote the solutions to the two minimization problems, respectively. It follows that

$$0 \leq \delta_+\left(\hat{\theta}_+\right) - \delta\left(\hat{\theta}_+\right) \leq \delta_+ - \delta \leq \delta_+\left(\hat{\theta}\right) - \delta\left(\hat{\theta}\right). \tag{8}$$

Therefore, the difference $\delta_+ - \delta$ is always nonnegative and serves as a lower bound of $\delta_+\left(\hat{\theta}\right) - \delta\left(\hat{\theta}\right)$, which measures the model's additional pricing errors on the contingent claims beyond the static portfolios of the test assets.

Given an SDF y_t, the arbitrage opportunity is related to the probability that y_t takes negative values, i.e., Prob$\{y_t < 0\}$. When a set of SDFs is specih6.312r21.5(y)0(.)-247.2(I)0(99s)-2.ye *negativity rate* of $y_t = g(\theta, f_t, z_t)$. The negativity rate π indicates the probability that y_t will be negative after choosing the parameters θ to minimize y_t 's distance to .

It is necessary to point out that the focus of this paper is the HJ distances developed by Hansen and Jagannathan (1997), not the HJ bounds developed by Hansen and Jagannathan (1991). The latter can be characterized as a special case of the former only if all SDFs happen to have the same expected value. HJ distances are different from HJ bounds: the former focus on the pricing errors, whereas the latter focus on the volatility of the SDF. HJ bounds control for the expected value of the SDFs in a comparison of the volatilities. Since HJ distances are the subject of study in this paper, we should not restrict the expected values of the SDFs in a comparison of the maximum pricing errors. Instead, we need to obtain information about the expected values from data. Since the gross return on Treasury bills contains information about the expected value of the SDF, it is therefore important to include the gross return on Treasury bills in the analysis of HJ distances.

3. Simulation-based Bayesian inference

The basic idea of our simulation-based Bayesian inference is as follows. We assume that z_t follows a general stochastic process, depending on some unknown parameters Ψ, for which we specify a noninformative prior distribution. The likelihood of the data is the probability of Z conditioning on Ψ, denoted by $p(Z|\Psi)$. We want to obtain random draws from the posterior distribution of the parameters given Z, denoted by $p(\Psi|Z)$, and achieve this by using the Markov Chain Monte Carlo (MCMC) method. For a given set of SDFs in the form of $y_t = g(\theta, f_t, z_{t-1})$, the distribution of z_t conditioning on parameters Ψ should determine the negativity rates and HJ distances. That is, Ψ determinesπ, δ, and δ_+ for the given form of SDFs. Therefore, the random draws from the posterior distribution of Ψ allow us to calculate the random draws from the posterior distributions of π, δ, and δ_+. The posterior distributions, $p(\pi|Z)$, $p(\delta|Z)$, and $p(\delta_+|Z)$, can be estimated from the random draws and are all we need for the empirical analysis. The rest of this section details the idea just described.

We assume that z_t follows a vector autoregressive (VAR) process.[1] That is,

$$z_t = C + Az_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \ (0_m, \Omega), \tag{9}$$

where $m = n + k + l$ is the dimension of vector z_t, and Ω is an $m \times m$ positive definite matrix. The noise term ε_t is independent across time. Consequently, the unconditional distribution of z_t is a normal distribution with a mean equal to μ and a variance equal to Σ. The mean and variance are given by

$$\mu = (I_m - A)^{-1} C \tag{10}$$

$$\mathrm{vec}(\Sigma) = \left[I_{m^2} - A \otimes A \right]^{-1} \mathrm{vec}(\Omega), \tag{11}$$

where "vec" converts a matrix to a vector by stacking all the columns. The vector $v_t = (r_t, f_t, z_{t-1})'$, which is necessary for the calculation of HJ distances, is linearly related to z_{t-1} and ε_t in the following way:

$$v_t = \tilde{C} + \tilde{A} z_{t-1} + D\varepsilon_t, \tag{12}$$

for some vector $\tilde{C}$ and matrices $\tilde{A}$ and D. Therefore, the unconditional distribution of v_t is normal, and the mean and variance are, respectively, $\tilde{\mu} = \tilde{C} + \tilde{A}\mu$ and $\tilde{\Sigma} = \tilde{A}\Sigma\tilde{A}' + D\Omega D'$.

The unknown parameters in the data-generating process (9) are the initial value z_0, the coefficient $B = (C, A)'$ in the autoregressive regression, and the variance Ω of the noise term. Let $\Psi = (z'_0, \mathrm{vec}(B)', \mathrm{vech}(\Omega)')'$, which is the vector of parameters in the VAR process of z_t. (Here, "vech" converts the upper triangle of the symmetric matrix Ω to a vector.) We have T observations on z_t, and the set of observed data is $Z = (z_1, \cdots, z_T)'$. We treat z_0 as part of the unknown parameters because z_0 is not in our observed data Z.

For the purpose of computation, Hansen and Jagannathan (1997) show that the square of the first HJ distance can be written as the weighted average of squared pricing errors. Given SDF $y_t = g(\theta, f_t, z_{t-1})$, we have

$$\delta^2(\theta) = E\left[g(\theta, f_t, z_{t-1}) r_t - 1_n \right]' \left(E\left[r_t r_t' \right] \right)^{-1} E[g(\theta, f_t, z_{t-1}) r_t - 1_n]. \tag{13}$$

If g is a linear function, the above formula allows us to calculate $\delta(\theta)$ analytically for the given Ψ, because we can calculate the expectations in (13) analytically. If g is a nonlinear function, we must calculate the expectations numerically as described

[1] We can also assume a more complicated process like the multivariate ;riate ;riate8870lSc(hi1lTc(ID(v)Tj5.977406.578789249.39213748Tm(1)Tj6.g)-2050378j7.007.970

later. In order to calculate the second HJ distance, we can use the following formula, which is obtained by applying an equation for δ_+ derived by Hansen and Jagannathan:

$$\delta_+^2(\theta) = \max_{\lambda \in R^n} E\left\{ g^2(\theta, f_t, z_{t-1}) - \left(\left[g(\theta, f_t, z_{t-1}) - \lambda' r_t \right]^+ \right)^2 - 2\lambda' 1_n \right\}, \tag{14}$$

where R^n is the space of $n\times 1$ real vectors. The function $[\cdot]^+$ is defined as $[x]^+ = x$ if $x \geq 0$ and $[x]^+ = 0$ if $x<0$. In Eq. (14), we cannot analytically calculate the expectation for the given Ψ. In addition, the maximization in $\delta_+(\theta)$ must be computed numerically.

Because we can calculate the expectations approximately, we can obtain approximations of HJ distances. Since the unconditional distribution of v_t is normal and has mean $\tilde{\mu}$ and variance $\tilde{\Sigma}$, we can generate independent draws of v_t from the unconditional distribution. Then it is easy to compute HJ distances. Let the independent draws be $v^{(j)} = (r^{(j)}, f^{(j)}, z^{(j)})'$ for $j = 1, \cdots,$. For a set of given parameters θ and a given function g in the given model $y_t = g(\theta, f_t, z_{t-1})$, we generate independent draws of $y^{(j)}$ by letting $y^{(j)} = g(\theta, f^{(j)}, z^{(j)})$.

We approximate $\delta(\theta)$ using the formula

$$\delta^2(\theta) = \hat{E}\left(y^{(j)} r^{(j)} - 1_n \right)' \left(\hat{E}\left[r^{(j)} r^{(j)\prime} \right] \right)^{-1} \hat{E}\left[y^{(j)} r^{(j)} - 1_n \right], \tag{15}$$

where $\hat{E}[\cdot]$ is defined as $^{-1}\sum_{j=1}[\cdot]$ and is a large integer. Similarly, we can approximate $\delta_+(\theta)$ using the formula

$$\delta_+^2(\theta) = \max_{\lambda \in R\tilde{n}} \hat{E}\left\{ \left(y^{(j)} \right)^2 - \left(\left[y^{(j)} - \lambda' r^{(j)} \right]^+ \right)^2 - 2\lambda' 1_n \right\}. \tag{16}$$

The convergence of the approximation can be established by the law of large numbers, and the precision of the approximation can be assessed by applying the central limit theorem. Note that the approximation can be arbitrarily precise by making large. The two HJ distances δ and δ_+ can then be obtained by minimizing $\delta(\theta)$ and $\delta_+(\theta)$ over all the choices of parameters θ. Using a simulation approach, we can also approximate the negativity rate using the formula

$$\pi = \lim_{\to +\infty} \hat{E}\left[I^-\left[g\left(\hat{\theta}, f_t^{(j)}, z_{t-1}^{(j)} \right) \right] \right], \tag{17}$$

where $I^-[x]$ equals 1 if $x<0$ and 0 otherwise, and $\hat{\theta}$ minimizes $\delta(\theta)$.

We assume the following standard noninformative prior distribution for Ψ in the data-generating process (9). The prior distributions of the three parts of Ψ are independent; i.e.,

$$p(\Psi) = p(z_0) p(B) p(\Omega), \tag{18}$$

where $p(z_0)$ and $p(B)$ are proportional to constants, and $p(\Omega)$ is proportional to $|\Omega|^{-(m+1)/2}$. The conditional structure of the posterior distribution is

$$z_0 | B, \Omega, Z \sim \left(A^{-1}(z_1 - C), A^{-1}\Omega A'^{-1} \right) \tag{19}$$

$$\Omega | z_0, Z \sim \mathrm{IW}\left(T\hat{\Omega}(z_0), T-1, m \right) \tag{20}$$

$$\mathrm{vec}(B)|\Omega, z_0, Z \sim \mathrm{Truncated}\left(\mathrm{vec}\left(\hat{B}(z_0) \right), \Omega \otimes \left(X(z_0)' X(z_0) \right)^{-1} \right), \tag{21}$$

where IW is the inverted Wishert distribution and the functions $\hat{B}(z_0)$, $\hat{\Omega}(z_0)$, and $X(z_0)$ are defined as

$$X(z_0) = \left(\left(1, z_0'\right)', \left(1, z_1'\right)', \cdots, \left(1, z_{T-1}'\right)' \right)', \tag{22}$$

$$\hat{B}(z_0) = \left[X(z_0)' X(z_0) \right]^{-1} X(z_0)' Z \tag{23}$$

$$\hat{\Omega}(z_0) = \frac{1}{T}\left[Z - X(z_0)\hat{B}(z_0) \right]' \left[Z - X(z_0)\hat{B}(z_0) \right]. \tag{24}$$

The normal distribution of $\mathrm{vec}(B)$ is truncated because the norm of the eigenvalues of A must be less than 1 for the VAR to be stationary.

It is analytically difficult to derive the posterior distribution of Ψ, and it is unknown how to derive the posterior distribution of the HJ distances δ and δ_+. The Markov Chain Monte Carlo (MCMC) simulation method provides a way to estimate the posterior distributions numerically. To estimate the posterior distributions of negativity rates and HJ distances, the MCMC procedure is as follows.

1. Start from an arbitrary $z_0^{(0)}$.
2. For $i = 1, \cdots, {}_0 +$, do the following:
   (a) Obtain the i^{th} sample of VAR parameters:
   - Draw $\Omega^{(i)}$ from IW $\left(T\hat{\Omega}\left(z_0^{(i-1)}\right), T-1, m\right)$,
   - Draw $\mathrm{vec}(B^{(i)})$ from

$$\text{truncated} \left(\mathrm{vec}\left(\hat{B}\left(z_0^{(i-1)}\right)\right), \Omega^{(i)} \otimes \left(X\left(z_0^{(i-1)}\right)' X\left(z_0^{(i-1)}\right)\right)^{-1}\right).$$

   - Draw $z_0^{(i)}$ from $([A^{(i)}]^{-1}(z_1 - C^{(i)}), [A^{(i)}]^{-1}\Omega^{(i)}[A^{(i)\prime}]^{-1})$.

   (b) Obtain the i^{th} sample of the unconditional mean and variance of z_t:

$$\mu^{(i)} = \left(I_m - A^{(i)}\right)^{-1} C^{(i)}$$
$$\mathrm{vec}\left(\Sigma^{(i)}\right) = \left(I_{m^2} - A^{(i)} \otimes A^{(i)}\right)^{-1} \mathrm{vec}\left(\Omega^{(i)}\right).$$

   (c) Obtain the i^{th} sample of the unconditional mean and variance of v_t:

$$\tilde{\mu}^{(i)} = \tilde{C}^{(i)} + \tilde{A}^{(i)}\mu^{(i)}$$
$$\tilde{\Sigma}^{(i)} = \tilde{A}^{(i)}\Sigma^{(i)}\tilde{A}^{(i)\prime} + D\Omega^{(i)}D',$$

   where $\tilde{C}^{(i)}$, $\tilde{A}^{(i)}$, and D are constructed from $C^{(i)}$ and $A^{(i)}$ in the same way as $\tilde{C}$, $\tilde{A}$ and D from C and A in Eq. (12).

   (d) Calculate the i^{th} samples, $\delta^{(i)}$, $\delta_+^{(i)}$, and $\pi^{(i)}$, with the help of Eqs. (15)–(17).
3. Discard the first ${}_0$ samples.
4. Approximate the posterior distributions of HJ distances, the negativity rates, and the model parameters by the distribution of the samples $\{\delta^{(i)}\}_{i=1}$, $\{\delta_+^{(i)}\}_{i=1}$, and $\{\pi^{(i)}\}_{i=1}$. These random draws can be used to obtain the posterior probability distribution of δ, δ_+ and π. The mean, standard deviation, median, and other statistics of the posterior distributions can be estimated by their sample analog.

The approximation of the posterior distributions is more precise if the number of simulations, , is larger. We choose = 10,000 for this analysis. We discard the first ${}_0$ simulations as the usual MCMC practice to help the distribution of the draws converge to the posterior distribution. We choose ${}_0$ to be 1000.

In this simulation-based Bayesian approach, it is straightforward to conduct formal statistical inference on the comparison of the two HJ distances and the comparison of two different models. To compare the two HJ distances of a given model, we can examine the posterior distribution of the absolute difference, $\delta_+ - \delta$, or the relative difference, $\delta_+/\delta - 1$, of the two HJ distances. To compare two models based on the second HJ distance, for example, let δ_+^{A} and δ_+^{B} be the second HJ distance for the SDFs y_t^{A} and y_t^{B}, respectively. Suppose the question is whether y_t^{B} is an improvement over y_t^{A} because of its smaller pricing errors. We can examine the posterior distribution of the absolute improvement, $\delta_+^{\text{A}} - \delta_+^{\text{B}}$, or the relative improvement, $1 - \delta_+^{\text{B}}/\delta_+^{\text{A}}$.

4. Asset pricing models and data

Because the SDFs of linear asset-pricing models can be negative, the pricing errors on the contingent claims are the focus of this paper.[2] The classic linear-asset pricing model in finance is the CAPM developed by Sharpe (1964). The SDF of this model is

$$y_t^{\text{CAPM}} = b_0 + b_1 r_{\text{MKT},t}, \tag{25}$$

where $r_{\text{MKT},t}$ is the excess return on the market portfolio, and b_0 and b_1 are constant parameters in the model. The CAPM is often referred to as the unconditional or static CAPM because it is derived in a single-period setting.

Researchers extend the static CAPM to a multiperiod setting by adding state variables and their interactions with the model factors. For example, according to Jagannathan and Wang (1996), the conditional version of the CAPM implies that an unconditional expected return depends on the covariance of the market factors and the state variables. Cochrane (1996) adds the

[2] In contrast, the nonlinear models are usually derived as equilibrium restrictions of utility functions. Examples of such models include the power-utility model, the Abel (1990) model, and the Epstein and Zin (1989) model. Here, we focus on the linear models instead of the nonlinear models because SDFs of the nonlinear models are always positive by specification. Investigations of HJ distances of nonlinear models can be found in Wang and Zhang (2005).

interaction of instrument variables with the factors to make the CAPM varying over time. In general, the SDF of the time-varying CAPM is

$$y_t^{\mathrm{CAPM*IV}} = b_0 + b_1 r_{\mathrm{MKT},} \text{[redacted]} c'_0 x_{t-1} + c'_1 x_{t-1} r_{\mathrm{MKT},} \text{[redacted]} \quad \text{[redacted]} \quad \text{[redacted]} \tag{26}$$

where x_{t-1}, referred to as the instrument variable (IV), is the past realization of the vector of the state variables. For convenience, we denote this model by CAPM*IV. The CAPM and its time-varying extension are not arbitrage free because their SDFs are not restricted to be nonnegative.

In the finance literature, the equilibrium model with a power utility function is often approximated by a linear factor model with the growth rate of consumption as the factor. This model is studied in Breeden et al. (1989) and Chen et al. (1986). The SDF of the model is

$$y_t^{\mathrm{LCC}} = b_0 + b_1 ln(C_t/C_{t-1}). \tag{27}$$

where C_t/C_{t-1} is the growth rate of consumption. We refer to this model as LCC. The literature has also extended the LCC into a time-varying model by adding state variables and their interaction with the growth rate of consumption (see Hodrick and Zhang, 2001; Lettau and Ludvigson, 2002). The SDFs of these types of models are in the form of

Table 1
Summary statistics of data. This table reports the mean and standard deviations (in parenthesis) for the pricing factors, state variables, and base assets. Our sample period is 1964 to 2008, with 540 monthly observations. The test assets are 25 portfolios sorted by firm size and book-to-market ratios, and the data are obtained from Kenneth French's website. The pricing factors MKT (excess market return), SMB, and HML are size and book-to-market factors, and the data are obtained from French's website. The pricing factor consumption growth is consumption growth rate, and the data are obtained from the Bureau of Economic Analysis. The state variables are TRM (term spread between 10-year and one-year Treasury bonds); DEF (yield spread between Moody's Baa and Aaa corporate bonds); DIV (dividend yield of the S&P 500 index); HB3 (yield spread between three-month and one-month Treasury bills); and TBL (yield on one-month Treasury bill). The data for DIV is from the Center of Research on Stock Prices, and the data for all other state variables are from the website of the Federal Reserve Bank of New York.

A. Pricing factors

MKT	SMB	HML	LCC
0.37	0.26	0.42	0.17
(0.19)	(0.14)	(0.12)	(0.01)

B. Annualized state variables

TRM	DEF	DIV	HB3	TBL
0.85	1.02	3.12	0.32	5.52
(0.05)	(0.02)	(0.05)	(0.38)	(0.03)

C. The Fama–French portfolios

		B/M ratio				
		Low	ii	iii	iv	High
Size	Small	0.12	0.72	0.75	0.96	1.04
		(0.35)	(0.30)	(0.26)	(0.24)	(0.26)
	ii	0.33	0.59	0.84	0.88	0.95
		(0.32)	(0.26)	(0.23)	(0.23)	(0.25)
	iii	0.34	0.65	0.69	0.77	0.98
		(0.29)	(0.24)	(0.21)	(0.21)	(0.23)
	iv	0.44	0.44	0.61	0.73	0.74
		(0.26)	(0.22)	(0.22)	(0.21)	(0.23)
	Big	0.31	0.41	0.37	0.45	0.52
		(0.21)	(0.19)	(0.19)	(0.18)	(0.21)

5. Empirical results

By examining the posterior probability distribution functions (PDF) of the HJ distances, we evaluate the models' performance on the asset returns and state variables discussed in the previous section. To compare the two HJ distances of each model, we examine the posterior probability distributions of the absolute and relative differences between the two HJ distances (Section 1). To compare the performance of the two models, we examine the posterior distributions of the absolute and relative improvements of one model over another (Section 2).

5.1. Comparison of the two distances

The posterior probability distributions of the two HJ distances are very similar for each of the static models. These posterior distributions are plotted in Fig. 1. For each model, the two HJ distances, δ and δ_+, have almost identical posterior distributions in the range plotted. The slight difference is that δ_+ has a longer tail on the positive side for all three static models.

The summary statistics of these posterior distributions are presented in panel A of Table 2. Taking the Fama–French model as an example, the posterior means of δ and δ_+ are close: 0.637 and 0.671, respectively. Their medians are even closer: 0.583 for δ and 0.592 for δ_+. Their standard deviations are slightly different: 0.218 for δ and 0.309 for δ_+. The similarity of the posterior distributions of the FF's HJ distances is also reflected in the small means and medians of $\delta_+ - \delta$ and $(\delta_+ - \delta)/\delta$. We observe the same properties in the statistics for the CAPM and the linear consumption model (LCC). Consistent with the analysis of Guasoni et al. (2011), the similarity between the two HJ distances implies that one will not be able to generate substantial alpha by incorporating derivative securities or conditional information if the benchmark of performance is set by these models.

The choice of HJ distances does not affect the statistical inference on the significance of the pricing errors of the static models. For example, the fifth percentiles of the posterior distribution of δ and δ_+ are 0.415 and 0.416, respectively, for the Fama–French model. These high fifth percentiles indicate the statistical significance of the pricing errors. Thus the statistical significance of the model's pricing errors is the same regardless of the measure. Therefore, the pricing errors on contingent claims do not alter the evaluation of the model. This finding also holds for the CAPM and LCC.

In the case of time-varying models, the posterior probability distributions of the two HJ distances are significantly different. These posterior distributions are plotted in Fig. 2. For each model, the distribution of δ_+ spreads further to the right relative to the distribution of δ. The summary statistics of the posterior distributions are presented in panel B of Table 2. For the CAPM*IV, the posterior mean of δ_+ (0.610) is much larger than that of δ (0.442). The comparison of the posterior means for the LCC*IV is similar. For FF*IV, the posterior mean of δ_+ is 0.541, nearly twice that of the posterior mean of δ, which is only 0.249. The standard deviation of the posterior distribution of δ_+ is much larger than that of δ because the distribution of δ_+ spreads out to the positive side.

For time-varying models, the choice of an HJ measure affects the inference on the significance of the pricing errors. For each model, the fifth percentile of the posterior distribution of δ is drastically different from that of δ_+. Take the FF*IV as an example: the fifth percentile of the posterior distribution of δ is only 0.133, whereas the fifth percentile of δ_+ is nearly three times larger, indicating that the FF*IV has much higher errors on derivatives of the test assets. Therefore, although the statistical inferences about the pricing errors of the static models are not affected by the choice of HJ distances, this finding does not hold for the time-varying models.

The time-varying model has additional parameters, and thus may cause concerns regarding power in statistical inference (Ferson and Foerster, 1994). The Bayesian analysis has advantages here. Our analysis produces the entire posterior distribution that shows the uncertainty explicitly. If the addition of parameters increases uncertainty in statistical inference, the posterior distribution will be more dispersed. This uncertainty will be reflected in the fifth percentile, which corresponds to the hypothesis test of zero HJ distance at the 5 percent level. For the case of FF*IV, the difference (0.133 versus 0.319) between the fifth percentiles of the HJ distances confirms the significance of FF*IV's pricing errors on contingent claims, after taking the uncertainty of inference into account.

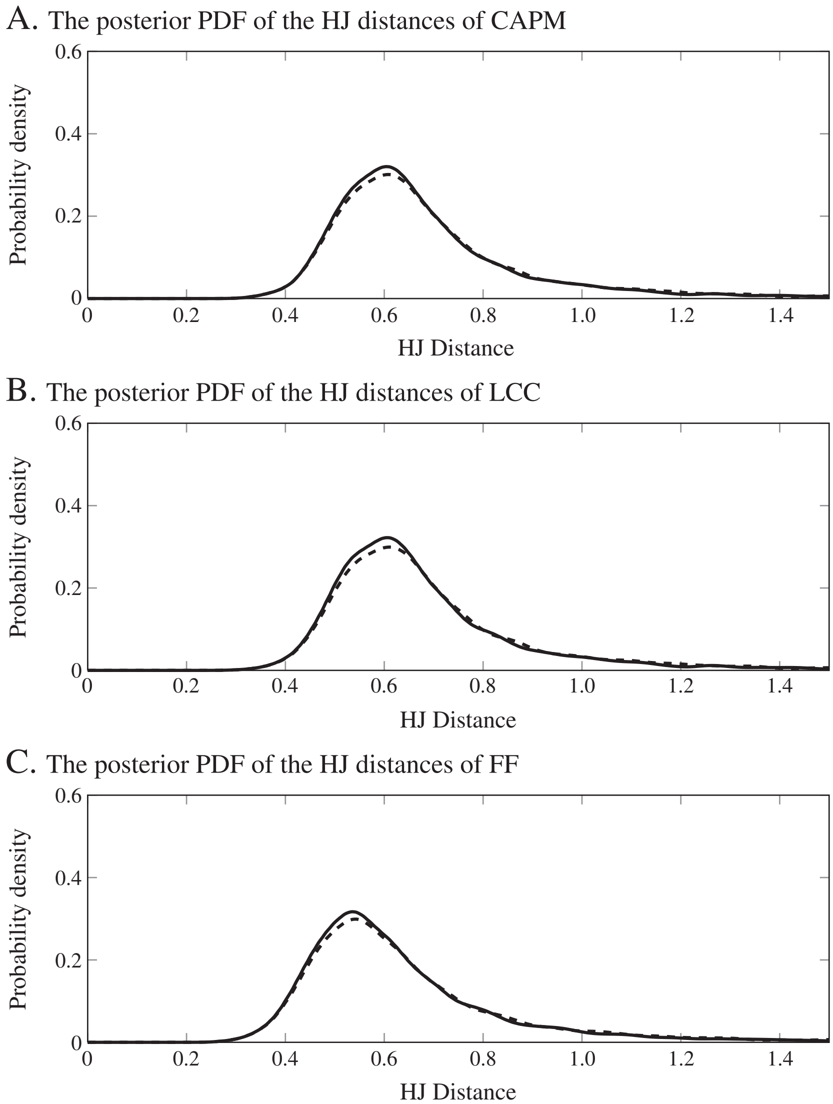

Fig. 1. Posterior distributions of HJ distances for static models. This figure presents the estimated posterior distributions of HJ distances for various models. The solid and dashed lines are the estimated posterior probability density function (PDF) of δ and δ_+, respectively.

Table 2
The posterior distributions of HJ distances. For each model, this table reports the summary statistics for the estimated posterior distributions of the first HJ distance (δ), second HJ distance (δ_+), their difference ($\delta_+ - \delta$), and their relative difference ($\delta_+/\delta - 1$). For each variable, we report its mean, standard deviation, median, and fifth percentile.

A. Static models					
		δ	δ_+	$\delta_+ - \delta$	$\frac{\delta_+ - \delta}{\delta}$
CAPM	Mean	0.687	0.719	0.035	0.033
	Stdev	0.219	0.307	0.125	0.076
	Median	0.634	0.643	0.009	0.014
	5th-pct	0.466	0.467	0.001	0.002
LCC	Mean	0.683	0.720	0.040	0.040
	Stdev	0.213	0.307	0.133	0.091
	Median	0.632	0.644	0.011	0.018
	5th-pct	0.466	0.468	0.001	0.003
FF	Mean	0.637	0.671	0.037	0.038
	Stdev	0.218	0.309	0.129	0.084
	Median	0.583	0.592	0.010	0.017
	5th-pct	0.415	0.416	0.001	0.002
B. Time-varying models					
		δ	δ_+	$\delta_+ - \delta$	$\frac{\delta_+ - \delta}{\delta}$
CAPM*IV	Mean	0.442	0.610	0.168	0.371
	Stdev	0.090	0.269	0.227	0.429
	Median	0.434	0.550	0.109	0.259
	5th-pct	0.310	0.391	0.028	0.064
LCC*IV	Mean	0.445	0.613	0.169	0.371
	Stdev	0.093	0.269	0.225	0.420
	Median	0.435	0.554	0.109	0.258
	5th-pct	0.310	0.394	0.034	0.077
FF*IV	Mean	0.249	0.541	0.292	1.289
	Stdev	0.079	0.269	0.249	1.127
	Median	0.242	0.479	0.233	1.010
	5th-pct	0.133	0.319	0.101	0.356

The reason why the two HJ distances differ for time-varying models is that their SDFs are likely to be negative. The likelihood of a negative SDF is measured by its negativity rate as discussed in Section 2. Table 3 presents the summary statistics of the posterior distributions of the negativity rates. The static models have low negativity rates; the posterior means are all below 0.09. In contrast, for the time-varying model, the posterior means of the negativity rates are all above 0.25. The medians and fifth percentiles of the negativity rates also show that the SDFs of the time-varying models are far more likely to be negative than the SDFs of the static models. To visualize the SDF of a model and the frequency with which it takes negative values, a particular path of the SDF is estimated using the posterior mean of the model parameters. The estimated paths for the SDFs of the FF and FF*IV models are presented in Fig. 3. Clearly, the SDF of the FF*IV model becomes negative more frequently than the SDF of the FF model.

5.2. Comparison of models

An advantage of HJ distances is their convenience for model comparison. To compare a pair of models (A and B) using the first HJ distance, we report the posterior means of the difference, $\delta^A - \delta^B$. A large difference indicates that model B is an improvement on model A. We also examine the posterior probability distribution of $1 - \delta^B/\delta^A$, which measures the improvement of model B as a percentage reduction in the pricing error of model A. More important, we can compare a pair of models using the second HJ distance. Thus, we look at the posterior distributions of $\delta_+^A - \delta_+^B$ and $1 - \delta_+^B/\delta_+^A$. The summary statistics of these posterior distributions are reported in Table 4.

The FF was originally suggested by Fama and French (1993) to explain the pricing errors of the CAPM. It is therefore natural to compare the FF with the CAPM. Given the well-known failure of the consumption model, we also compare FF with the LCC. We have observed that the two HJ distances of FF are about the same and that its SDF is rarely negative. Panel A of Table 4 shows that FF clearly has smaller pricing errors than the CAPM and LCC regardless of the measure of pricing errors. The posterior mean of $\delta^{CAPM} - \delta^{FF}$ is 0.051, and the posterior mean of $\delta_+^{CAPM} - \delta_+^{FF}$ is 0.048. Corresponding posterior means for the comparison of FF and LCC are 0.046 and 0.049. However, the fifth percentile of the improvement of FF over the CAPM and LCC is almost zero, using either HJ distance for the comparison. The posterior distribution of the relative improvement tells the same story: (1) the comparison is not affected by the choice of the HJ distance; and (2) the confidence of the improvement is not strong. Therefore, the improvement of the Fama–French model over the static single factor models is quite limited in terms of maximum pricing errors.

The comparison of static models with their time-varying extensions is significantly affected by the measure of pricing errors, as shown in panel B of Table 4. For example, pricing errors on contingent claims make a difference for the comparison between the FF and the FF*IV. We have 95% confidence that the time-varying extension reduces the error of the FF by 13.3% based on the first HJ distance. Based on the second HJ distance, however, the measure with 95-percent confidence reduces by only 5.8%. The inference about the magnitude of the improvement clearly depends on the measure of pricing errors. The posterior mean of $\delta^{\mathrm{FF}} - \delta^{\mathrm{FF*IV}}$ is 0.388, in contrast to the posterior mean of $\delta_+^{\mathrm{FF}} - \delta_+^{\mathrm{FF*IV}}$ estimated at 0.13. The posterior mean of $1 - \delta^{\mathrm{FF*IV}}/\delta^{\mathrm{FF}}$ is 0.59, whereas the mean of $1 - \delta_+^{\mathrm{FF*IV}}/\delta_+^{\mathrm{ectFF}}$ is only 0.196. Therefore, we are very confident that the switch from FF to FF*IV reduces the pricing error substantially if we measure pricing errors by the first HJ distance, but the error reduction is much smaller if we use the second HJ distance.

The measures of pricing errors affect the comparisons of the CAPM and LCC with their time-varying extensions (CAPM*IV and LCC*IV) similarly. The posterior mean of $\delta^{\mathrm{CAPM}} - \delta^{\mathrm{CAPM*IV}}$ is 0.246, but the mean of $\delta_+^{\mathrm{CAPM}} - \delta_+^{\mathrm{CAPM*IV}}$ is 0.109, less than half the former. The median and fifth percentile of $\delta^{\mathrm{CAPM}} - \delta^{\mathrm{CAPM*IV}}$ are also less than half the median and fifth percentile of $\delta_+^{\mathrm{CAPM}} - \delta_+^{\mathrm{CAPM*IV}}$. The

Table 4
Model comparison. For each pair of models, A and B, summary statistics are presented for the posterior distributions of reductions of pricing errors of model B over model A. The reduction of pricing errors is measured by $\delta^{B}-\delta^{A}$ and $1-\delta_{+}^{B}/\delta_{+}^{A}$.

A. Comparing single- and multiple-factor models

Improvement	Mean	Stdev	Median	5th-pct
$\delta^{CAPM}-\delta^{FF}$	0.051	0.037	0.043	0.006
$\delta_{+}^{CAPM}-\delta_{+}^{FF}$	0.048	0.034	0.042	0.006
$\delta^{LCC}-\delta^{FF}$	0.046	0.042	0.042	−0.007
$\delta_{+}^{LCC}-\delta_{+}^{FF}$	0.049	0.037	0.043	0.003
$1-\delta^{FF}/\delta^{CAPM}$	0.078	0.055	0.066	0.009
$1-\delta_{+}^{FF}/\delta_{+}^{CAPM}$	0.074	0.054	0.063	0.008
$1-\delta^{FF}/\delta^{LCC}$	0.072	0.063	0.064	−0.010
$1-\delta_{+}^{FF}/\delta_{+}^{LCC}$	0.075	0.058	0.065	0.004

B. Comparing static and time-varying models

Improvement	Mean	Stdev	Median	5th-pct
$\delta^{CAPM}-\delta^{CAPM*IV}$	0.246	0.182	0.198	0.071
$\delta_{+}^{CAPM}-\delta_{+}^{CAPM*IV}$	0.109	0.069	0.093	0.035
$\delta^{LCC}-\delta^{LCC*IV}$	0.238	0.174	0.193	0.070
$\delta_{+}^{LCC}-\delta_{+}^{LCC*IV}$	0.106	0.069	0.091	0.030
$\delta^{FF}-\delta^{FF*IV}$	0.388	0.201	0.342	0.176
$\delta_{+}^{FF}-\delta_{+}^{FF*IV}$	0.130	0.070	0.113	0.054
$1-\delta^{CAPM*IV}/\delta^{CAPM}$	0.330	0.132	0.320	0.133
$1-\delta_{+}^{CAPM*IV}/\delta_{+}^{CAPM}$	0.150	0.066	0.142	0.058
$1-\delta^{LCC*IV}/\delta^{LCC}$	0.324	0.131	0.311	0.130
$1-\delta_{+}^{LCC*IV}/\delta_{+}^{LCC}$	0.146	0.067	0.137	0.051
$1-\delta^{FF*IV}/\delta^{FF}$	0.590	0.129	0.596	0.370
$1-\delta_{+}^{FF*IV}/\delta_{+}^{FF}$	0.196	0.070	0.188	0.096

C. Comparing single- and multiple-factor time-varying models

Improvement	Mean	StDev	Median	5th-pct
$\delta^{CAPM*IV}-\delta^{FF*IV}$	0.193	0.074	0.185	0.086
$\delta_{+}^{CAPM*IV}-\delta_{+}^{FF*IV}$	0.069	0.037	0.063	0.022
$\delta^{LCC*IV}-\delta^{FF*IV}$	0.196	0.084	0.191	0.069
$\delta_{+}^{LCC*IV}-\delta_{+}^{FF*IV}$	0.073	0.041	0.068	0.017
$1-\delta^{FF*IV}/\delta^{CAPM*IV}$	0.436	0.137	0.432	0.213
$1-\delta_{+}^{FF*IV}/\delta_{+}^{CAPM*IV}$	0.124	0.069	0.114	0.031
$1-\delta^{FF*IV}/\delta^{LCC*IV}$	0.435	0.154	0.439	0.179
$1-\delta_{+}^{FF*IV}/\delta_{+}^{LCC*IV}$	0.128	0.075	0.120	0.026

first HJ distance. In contrast, the reduction with the same confidence is only 2.6% if the pricing errors are measured by the second distance. These results demonstrate the importance of pricing errors on contingent claims.

6. Conclusion

In the literature, it has been unclear whether the second HJ distance is empirically important. In Hansen and Jagannathan's (1997) estimates, the second HJ distance is not very different from the first HJ distance because both focus on consumption-based nonlinear models that are arbitrage free by definition. However, Bansal and Viswanathan (1993) argue that the arbitrage-free requirement might be important when the focus is nonlinear APT models. The analysis presented in this paper conducts a formal statistical inference of model comparisons using both HJ distances and demonstrates the importance of the second HJ distance in the context of linear time-varying models.

A good asset-pricing model should have small pricing errors not only in test portfolios but also in the contingent claims of the portfolios. The requirement of zero pricing errors on contingent claims does not allow asset-pricing models to admit arbitrage opportunities and allows only positive stochastic discount factors. In this discussion, we emphasize the pricing errors on contingent claims, which have been ignored in a large body of the literature of empirical evaluation of asset-pricing models.

To show the importance of pricing errors on contingent claims, we focus on the comparison of static models to their time-varying extensions. According to our results, although the time-varying models are successful in explaining returns on the test assets, they are not arbitrage free and can thus have pricing errors on contingent claims of the test assets. In contrast, the static linear models are not successful in that regard, but their SDFs are mostly positive. Using the first HJ distance, which ignores pricing errors on contingent claims, a linear time-varying model can have substantially smaller pricing errors than a static single model. However, when using the second HJ distance, which does not ignore pricing errors on contingent claims, the linear time-varying model may not be a substantial improvement on the static single-factor model.

Therefore, although a linear time-varying model has small pricing errors measured by the first HJ distance, it is still possible to have large pricing errors on portfolios that are constructed with sophisticated rules. In searching for robust asset-pricing models, we should choose models that have small second HJ distances.

The issue investigated in this study is part of a larger issue of overfitting data with extended models that contains many variables and parameters. The special aspect of this problem focused on here is that the SDF, which has many unknown parameters and depends on conditional information, is likely to be negative because of the need to fit the data. This special overfitting problem has important economic implications because it leads to arbitrage and pricing errors on contingent claims of the test assets. The second HJ distance indicates the magnitude of the problem. Since the second HJ distance addresses the overfitting problem, it is a useful tool in search of robust asset-pricing models.

References

Abel, A.B., 1990. Asset prices under habit formation and catching up with the Joneses. Am. Econ. Rev. 80, 38–42.
Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: a new approach. J. Finance 48, 1231–1262.
Black, F., Scholes, M.S., 1973. The pricing of options and corporate liabilities. J. Polit. Econ. 81, 637–654.
Breeden, D.T., Gibbons, M.R., Litzenberger, R.H., 1989. Empirical tests of the consumption-oriented CAPM. J. Finance 44, 231–262.
Chen, N., Roll, R., Ross, S.A., 1986. Economic forces and the stock market. J. Bus. 59, 383–403.
Cochrane, J.H., 1996. A cross-sectional test of an investment-based asset pricing model. J. Polit. Econ. 104, 572–621.
Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross-sectional variation in stock returns. J. Finance 52, 1–33.
Dybvig, P., Ingersoll, J., 1982. Mean-variance theory in complete markets. J. Bus. 55, 233–251.
Dybvig, P., Ross, S.A., 1985. Differential information and performance measurement using a security market line. J. Finance 40, 383–399.