



ELSEVIER

Contents lists available at ScienceDirect

Journal of Financial Economics

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

Evaluating asset pricing models using the second Hansen-Jagannathan distance[☆]

Haitao Li^a, Yuewu Xu^{b,*}, Xiaoyan Zhang^c

^a Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109, USA

^b Graduate School of Business, Fordham University, New York, NY 10023, USA

^c Johnson Graduate School of Management, Cornell University, Ithaca, NY 14850, USA

ARTICLE INFO

Article history:

Received 20 December 2007

Received in revised form

16 July 2008

Accepted 31 October 2008

Available online 10 March 2010

JEL classification:

C51

C52

G12

Keywords:

Stochastic discount factor

Specification test

Model selection

Hansen-Jagannathan distance

Arbitrage

ABSTRACT

We develop a specification test and a sequence of model selection procedures for non-nested, overlapping, and nested models based on the second Hansen-Jagannathan distance, which requires a good asset pricing model to not only have small pricing errors but also be arbitrage free. Our methods have reasonably good finite sample performances and are more powerful than existing ones in detecting misspecified models with small pricing errors but are not arbitrage-free and in differentiating models that have similar pricing errors of a given set of test assets. Using the Fama and French size and book-to-market portfolios, we reach dramatically different conclusions on model performances based on our approach and existing methods.

© 2010 Elsevier B.V. All rights reserved.

[☆] We thank Vikas Agarwal, Andrew Ang, Warren Bailey, Michael Brennan, Charles Cao, Sris Chatterjee, Jin-Chuan Duan, Jim Hodder, Kewei Hou, Jingzhi Huang, Jon Ingersoll, Ravi Jagannathan, Raymond Kan, Andrew Karolyi, Kenneth Kavajecz, Bob Kimmel, Tao Li, Peter C. B. Phillips, Mark Ready, Marcel Rindisbacher, Cesare Robotti, Tim Simin, Rene Stulz, Zhenyu Wang, Chu Zhang, Feng Zhao, Guofu Zhou, and seminar participants at the Chinese University of Hong Kong, Cornell University, Fordham University, Georgia Institute of Technology, Georgia State University, Ohio State University, Penn State University, Rutgers University, Singapore Management University, Syracuse University, the University of Hong Kong, the University of Toronto, the University of Virginia, the University of Wisconsin at Madison, Worcester Polytechnic Institute, the 2005 Western Finance Association Meeting, the 2006 China International Conference in Finance, and the Asset Pricing Symposium at Hong Kong University of Science and Technology for helpful comments. Haitao Li acknowledges financial support from the NTT research fellowship at the Mitsui Life Finance Center at the University of Michigan. We are responsible for any remaining errors.

* Corresponding author.

E-mail address: yuxu@fordham.edu (Y. Xu).

1. Introduction

The fundamental theorem of asset pricing, one of the cornerstones of neoclassical finance, establishes the equivalence between the absence of arbitrage and the existence of a positive stochastic discount factor (SDF) that correctly prices all assets.¹ The main purpose of this paper is to develop asset pricing tests that fully reflect the implications of the fundamental theorem of asset pricing. Specifically, we develop a systematic approach for estimating, testing, and comparing asset pricing models based on the second Hansen-Jagannathan distance (HJD).

¹ See Ross (2005) for a recent review of neoclassical finance. See Cochrane (2005) for a comprehensive treatment of both theoretical and empirical asset pricing based on the SDF approach and all the related references.

The first and second HJDs developed by Hansen and Jagannathan (1991, 1997) measure specification errors of SDF models by least squares distances between an SDF model and the set of admissible SDFs that can correctly price a set of test assets.² The first HJD considers the set of all admissible SDFs, which we denote as \mathcal{M} . The second HJD considers only the smaller set of strictly positive admissible SDFs, which we denote as \mathcal{M}_+ . The positivity constraint of the second HJD guarantees the admissible SDFs to be arbitrage-free and is important for pricing derivatives associated with the test assets. Hansen and Jagannathan (1997) show that, while the first HJD represents the maximum pricing error of a portfolio of the test assets with a unit norm, the second HJD represents the minimax bound of the pricing errors of a portfolio of both the test assets and their related derivatives with a unit norm. The second HJD represents a more stringent criterion for evaluating asset pricing models and is generally larger than the first HJD.

Using the second HJD in evaluating asset pricing models has an important advantage, although the existing empirical literature has mainly focused on the first HJD.³ Conceptually, the second HJD reflects more fully the implications of the fundamental theorem of asset pricing than the first HJD. While the first HJD tests only whether an SDF model can correctly price the test assets, the second HJD further tests whether the SDF model is strictly positive. As a result, the second HJD is more powerful than the first HJD in detecting misspecified SDF models that can price the test assets but are not strictly positive, a situation that is especially likely to happen to linear factor models.

Dybvig and Ingersoll (1982) show that linear asset pricing models are not arbitrage-free and are not appropriate for pricing derivatives because their SDFs take negative values in certain states of the world. Although linear factor models are seldom used directly to price derivatives, they have been widely used in performance evaluation of mutual funds and hedge funds. Mutual funds and hedge funds often employ dynamic trading strategies that generate option-like payoffs.⁴ Most hedge funds directly trade derivatives, and their returns exhibit option-like features.⁵ Grinblatt and Titman (1989), Glossten and Jagannathan (1994), Ferson and Khang (2002), and Ferson, Henry and Kisgen (2006), among others, emphasize the importance of the positivity constraint for mutual fund performance evaluation. The fast-growing

hedge fund industry and the need to evaluate hedge fund performances further highlight the significance of the second HJD for empirical asset pricing studies.

Even for applications that do not involve derivatives, using the second HJD for estimating and comparing asset pricing models can still be beneficial. For example, one is likely to obtain more robust and reliable parameter estimates using the second HJD than the first HJD, especially for linear factor models. The first HJD chooses the parameters of a linear model to minimize the pricing errors of the test assets. However, such estimated models could be far from \mathcal{M}_+ , because in many cases the estimated SDF models have to take negative values with high probabilities to price the test assets. Therefore, models estimated using the first HJD are likely to overfit the test assets and to perform poorly out of sample. The second HJD helps to overcome the overfitting problem because it chooses parameters to minimize the distance between an SDF model and \mathcal{M}_+ . As a result, the second HJD provides more realistic assessments of model performance and leads to estimated SDF models that are closer to \mathcal{M}_+ .

Moreover, the second HJD is more powerful than the first HJD in distinguishing the relative performances of models that have small pricing errors of the same set of test assets. According to Lewellen, Nagel, and Shanken (LNS, 2010), it is difficult to differentiate models that have been developed to explain the cross-sectional returns of the 25 size and book-to-market (BM) portfolios of Fama and French (1993) using traditional methods, because these models tend to have small pricing errors for the test assets by construction. While LNS (2010) suggest several interesting ways to improve the traditional methods, the second HJD represents a powerful measure of relative model performances. SDF models with similar pricing errors for the Fama and French portfolios could have very different probabilities in taking negative values and thus can be differentiated based on the second HJD. In addition, because the second HJD measures the distance between an SDF model and \mathcal{M}_+ , a model with a smaller second HJD is likely to be a better model. Given that most asset pricing models are approximations of reality and likely to be misspecified, it is important to have a powerful measure such as the second HJD to compare the relative performances of misspecified models.

Despite its many advantages, the second HJD has been rarely used in the existing literature. One main reason is that econometric analysis of the second HJD is much more difficult than that of the first HJD. Standard asymptotic analysis involves a Taylor series approximation of the second HJD near true parameter values. This procedure, however, breaks down for the second HJD, which involves a function that is not differentiable with respect to model parameters in the traditional sense. As a result, the standard econometric techniques used in Jagannathan and Wang (1996) for analyzing the first HJD cannot be applied to the second HJD.

To fully exploit the theoretical advantages of the second HJD for empirical asset pricing studies, we develop a systematic approach for evaluating asset pricing models

² Following Hansen and Jagannathan (1997), we refer to admissible models as models that can correctly price the set of test assets.

³ Studies that use the first HJD are Jagannathan and Wang (1996), Jagannathan, Kubota and Takehara (1998), Campbell and Cochrane (2000), Lettau and Ludvigson (2001), Hodrick and Zhang (2001), Dittmar (2002), and Kan and Zhou (2002), among others.

⁴ See Merton (1981), Dybvig and Ross (1985), and others for more detailed discussions on this issue.

⁵ TASS, a hedge fund research company, reports that more than 50% of the four thousand hedge funds it follows use derivatives. Fung and Hsieh (2001), Agarwal and Naik (2004), Ben Dor and Jagannathan (2002), and Mitchell and Pulvino (2001) show option-like features in hedge fund returns.

based on the second HJD and apply the new approach to empirical applications using the Fama and French 25 size/BM portfolios. By doing so, our paper makes the following methodological and empirical contributions to the existing literature.

First, we overcome the nondifferentiability issue of the second HJD by introducing the concept of differentiability in quadratic mean of Le Cam (1986), Pollard (1982), and Pakes and Pollard (1989). In particular, we develop a second-order stochastic representation of the second HJD based on the new concept of differentiation under general conditions (i.e., the SDF model can be either correctly specified or misspecified). The second-order stochastic representation forms the foundation of our econometric analysis of the specification test and model selection procedures based on the second HJD.

Second, we develop the asymptotic distribution of the second HJD under the null hypothesis of a correctly specified model, which has not been formally developed in the literature. The new asymptotic distribution makes it possible to conduct specification tests of SDF models based on the second HJD and to identify misspecified models that cannot be identified by the first HJD. We also develop the asymptotic distribution of model parameters, which provide diagnostic information on potential sources of model misspecifications.

Third, we develop a sequence of model selection procedures in the spirit of Vuong (1989) for non-nested, overlapping, and nested models based on the second HJD. Given that most asset pricing models are likely to be misspecified, it is also very important to compare the degrees of misspecifications of different models. We compare the relative performances of two models based on the asymptotic distribution of the difference between the second HJDs of the two models. One challenging aspect of the analysis is that the difference could have either normal or weighted χ^2 asymptotic distributions depending on the structures of the two models. Our model selection procedures represent probably the first attempt to formally compare the relative performances of asset pricing models based on the second HJD. Simulation evidence shows that both the specification test and model selection tests have reasonably good finite sample performances for sample sizes typically considered in the existing literature.

Finally, we demonstrate the advantages of the second HJD through empirical applications, in which we reach dramatically different conclusions on model performances based on our approach and existing methods. In particular, we evaluate several well-known asset pricing models that have been developed in the literature to explain the cross-sectional returns of the Fama and French 25 portfolios (the same set of models considered in LNS, 2010). Though certain models appear to have good performances in pricing the Fama and French portfolios according to the first HJD, their SDFs take negative values with high probabilities and are overwhelmingly rejected by the second HJD. The second HJD is also more powerful than the first HJD in distinguishing models that have similar pricing errors of the test assets but are not arbitrage-free.

Our paper extends Hansen, Heaton, and Luttmer (HHL, 1995) and Hansen and Jagannathan (1997) in important ways. Both papers show that the second HJD follows an asymptotic normal distribution under the null hypothesis that a given SDF model is misspecified. Their results, however, cannot be applied to our setting for at least two reasons. First, their asymptotic distribution becomes degenerate under the null hypothesis of a correctly specified model and therefore cannot be used for specification test. Second, their results cannot be used for formal model comparison, because they do not provide the distribution of the difference between the second HJDs of two models. One might be tempted to conclude that the difference between the two second HJDs should follow a normal distribution as well. Our model selection tests reveal the full complexities of the issue: the difference could follow either normal or weighted χ^2 distributions depending on model structures. Therefore, while Hansen and Jagannathan (1997) introduce the second HJD as an important theoretical measure of specification errors, our model selection procedures make it possible to formally compare the relative performances of misspecified models based on the second HJD in empirical studies.

Our paper is also related to Wang and Zhang (2005), one of the first papers after Hansen and Jagannathan (1997) that seriously studies the second HJD. Wang and Zhang (2005) develop a simulation-based Bayesian approach for inferences of the second HJD. Based on Markov Chain Monte Carlo simulation and additional assumptions on the data-generating process, they are able to obtain the posterior distribution of the second HJD and to demonstrate that the second HJD can make big differences in empirical analysis of asset pricing models. The Bayesian methodology of Wang and Zhang (2005), though a nice contribution to the literature, is very different from traditional methods in the literature, such as the generalized method of moments (GMM) test of Hansen (1982) or the Jagannathan and Wang (1996) test. In contrast, our paper provides a systematic approach for evaluating asset pricing models based on the second HJD within the established econometric framework in the existing literature.

Our model selection procedures differ from that of Vuong (1989) in several aspects. While Vuong (1989) compares relative model performance based on the Kullback and Leibler information criterion, a statistical measure, we compare model performance based on the second HJD, an economic criterion. Whereas Vuong (1989) considers only smooth likelihood functions, we have to deal with the nondifferentiability issue of the second HJD.

The rest of this paper is organized as follows. In Section 2, we discuss the advantages of the second HJD in evaluating asset pricing models. Section 3 develops a specification test and a sequence of model selection tests based on the second HJD. Section 4 provides simulation evidence on the finite sample performances of the new asset pricing tests. Section 5 contains the empirical results. Section 6 concludes, and the Appendix provides the mathematical proofs.

2. Advantages of the second HJD in evaluating asset pricing models

In this section, we first provide a brief introduction to the two HJDs. Then we discuss the advantages of using the second HJD as a criterion for evaluating asset pricing models.⁶

Let the uncertainty of the economy be described by a filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t \geq 0})$ for $t=0, 1, \dots, T$. Suppose we use n test assets with payoffs Y_t (an $n \times 1$ vector) at t in an empirical analysis of asset pricing models. Denote \underline{Y} , a subspace of L^2 , as the payoff space of all the test assets. In the absence of arbitrage, there must exist a strictly positive SDF that correctly prices all the test assets. That is, for all t , we have

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

where X_t , an $n \times 1$ vector, represents the prices of the n assets at t . The random variable m_{t+1} discounts payoffs at $t+1$ state by state to yield price at t and hence is called a stochastic discount factor. If the market is complete, then m_{t+1} is unique. Otherwise multiple m_{t+1} satisfy Eq. (1). Without loss of generality, for the rest of the paper, we focus our discussions on the unconditional version of the above pricing equation.⁷ That is,

$$\mathbb{E}[m_{t+1}Y_{t+1}] = \mathbb{E}[X_t], \quad m_{t+1} > 0, \quad \forall Y_{t+1} \in \underline{Y}. \quad (2)$$

In an important paper, Hansen and Jagannathan (1997) develop two measures of specification errors of SDF models. The first HJD, denoted as δ , measures the least squares distance or the L^2 -norm between a candidate SDF model H and \mathcal{M} . That is,

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

where $\mathcal{M} = \{m_{t+1} : \mathbb{E}[m_{t+1}Y_{t+1}] = \mathbb{E}[X_t] \text{ for } \forall Y_{t+1} \in \underline{Y}\}$ denotes the set of all admissible SDFs. The second HJD, denoted as δ^+ , is defined as

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

where $\mathcal{M}_+ = \{m_{t+1} : \mathbb{E}[m_{t+1}Y_{t+1}] = \mathbb{E}[X_t], m_{t+1} > 0, \forall Y_{t+1} \in \underline{Y}\}$ denotes the set of positive SDFs. In general, the second HJD is bigger than the first one, because \mathcal{M}_+ is a subset of \mathcal{M} . Often the SDF model H depends on some unknown parameters γ , and the two distances are defined as

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

and

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

⁶ The discussions in this section draw materials from Cochrane (2005), Dybvig and Ingersoll (1982), Hansen and Jagannathan (1997), and Wang and Zhang (2005).

⁷ 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.

In addition to interpretations as least squares distances, the two HJDs have interpretations as pricing errors. Hansen and Jagannathan (1997) show that the first HJD has an interpretation as the maximum pricing error of a portfolio of the test assets with a unit norm, i.e.,

$$\delta = \max_{\|Y\|=1} |\mathbb{E}(mY) - \mathbb{E}(HY)|, \quad \forall Y \in \underline{Y}. \quad (7)$$

Hansen and Jagannathan (1997) also show that the second HJD has an interpretation as the minimax bound on pricing errors of any payoff in L^2 with a unit norm, i.e.,

$$\delta^+ = \min_{m \in \mathcal{M}_+} \max_{\|Y\|=1} |\mathbb{E}(mY) - \mathbb{E}(HY)|, \quad \forall Y \in L^2. \quad (8)$$

While δ considers only pricing errors of the test assets, δ^+ considers pricing errors of both the test assets and payoffs that are not in \underline{Y} but in L^2 , which are derivatives that can be constructed from trading the test assets. Therefore, the positivity constraint on m in the definition of the second HJD is closely related to derivatives pricing: Only a strictly positive m is arbitrage-free and can price both the test assets and their related derivatives.

Fig. 1 provides a graphical illustration of the differences between the two HJDs in a one-period, two-state economy. The horizontal (vertical) axis represents payoffs when state one (two) occurs at the end of the period. The straight line going through the origin represents the payoff space of the test assets \underline{Y} . Let Y^* represent the SDF that is in the payoff space \underline{Y} and can correctly price all the test assets.⁸ Then \mathcal{M} is represented by the straight (dot-dash) line that intersects with \underline{Y} at Y^* and is orthogonal to \underline{Y} (the thin line), and \mathcal{M}_+ is represented by the segment of \mathcal{M} that is in the first quadrant (the thick line). Suppose $H(\gamma)$ is an SDF model that we want to evaluate. The dotted line represents $\delta(\gamma)$, the shortest distance between $H(\gamma)$ and \mathcal{M} , the dashed line represents $\delta^+(\gamma)$, the shortest distance between $H(\gamma)$ and \mathcal{M}_+ .⁹ One can reach very different conclusions on model performances and obtain very different parameter estimates using the two HJDs. While the first HJD chooses γ to minimize $\delta(\gamma)$, the second HJD chooses γ to minimize $\delta^+(\gamma)$. Therefore, certain models might have large second HJDs, even though they have small first HJDs.

The second HJD, as a criterion for evaluating asset pricing models, reflects more fully the implications of the fundamental theorem of asset pricing and therefore is more powerful in detecting misspecified models than the first HJD. While the first HJD requires only that m correctly prices all the test assets, the second HJD imposes the additional restriction of $m > 0$. As a result, the second HJD can detect misspecified SDF models that can price the test assets but are not strictly positive, a situation that is especially likely to happen to linear factor models.

The second HJD also leads to more reliable and economically more meaningful parameter estimates than the first HJD, especially for linear factor models. Consider

⁸ By the law of one price, Y^* always exists, and $Y^* + \varepsilon \in \mathcal{M}$, $\forall \varepsilon \perp \underline{Y}$.

⁹ Rigorously speaking, the solution to $\min_{m \in \mathcal{M}_+} \|H - m\|$ does not exist in our simple example, because \mathcal{M}_+ is not closed. To avoid this technical issue, we can re-define \mathcal{M}_+ as the set of non-negative admissible SDFs following Hansen and Jagannathan (1997).

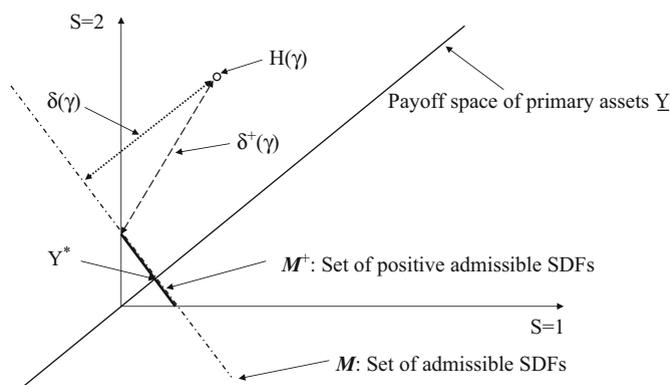


Fig. 1. The first and second Hansen-Jagannathan distances in a one-period, two-state economy. This graph illustrates the difference between the first and second HJDs in a one-period, two-state ($S=1$ or 2) economy. The stochastic discount factor is denoted by $H(\gamma)$. The dot-dash line represents the set \mathcal{M} of admissible stochastic discount factors (SDFs) that can correctly price the primary assets. The solid thick line segment represents the set \mathcal{M}^+ of admissible positive SDFs that can correctly price the primary assets. The dotted line segment represents the first HJD $\delta(\gamma)$, while the dash line segment represents the second HJD $\delta^+(\gamma)$. Y^* represents the SDF that is in the payoff space \underline{Y} and can correctly price all test assets.

a linear factor model $H(\gamma)$ in Fig. 1. If we choose γ to minimize the first HJD, the estimated model $H(\hat{\gamma})$ could be far away from \mathcal{M}_+ even though the estimated pricing errors for the test assets are small. This is because in many cases the small pricing errors are obtained at the expense of the estimated model $H(\hat{\gamma})$ taking negative values with high probabilities. As a result, the estimated model tends to overfit the test assets and is likely to perform poorly out of sample. In contrast, the second HJD mitigates the overfitting problem by choosing γ such that $H(\gamma)$ is as close as possible to \mathcal{M}_+ . Therefore, although it could appear that the second HJD unfairly punishes SDF models that can price the test assets well but are not strictly positive, the second HJD provides more realistic assessments of the performances of such models and leads to estimated SDF models that are closer to \mathcal{M}_+ .

The second HJD is a powerful measure of relative performances of misspecified SDF models and helps discriminate models that cannot be distinguished by the first HJD. For example, many models have been proposed in the literature to explain the cross-sectional returns of the Fama and French size/BM portfolios. LNS (2010) point out that because these models tend to have small pricing errors for the test assets by construction, it is very difficult to differentiate them using traditional methods, which mainly focus on pricing errors for model evaluation. Because the second HJD measures the distance between an SDF model and \mathcal{M}_+ , models with similar pricing errors for the Fama and French portfolios could have very different probabilities of taking negative values and thus can be differentiated based on the second HJD. Given that most linear factor models are misspecified, the ability to compare relative model performances using the second HJD is one of its important advantages for empirical asset pricing studies.

3. Asset pricing tests based on the second HJD

In this section, we first develop a second-order asymptotic representation of the second HJD, which forms

the foundation of our econometric analysis of the second HJD. Then we develop the asymptotic distribution of the second HJD under the null hypothesis of a correctly specified model, which can be used for specification tests of SDF models. Finally, we develop a sequence of model selection procedures in the spirit of Vuong (1989), which can compare the relative performances of SDF models based on the second HJD regardless whether the models are correctly specified or not.

3.1. Nondifferentiability and asymptotic representation of the second HJD

One technical difficulty we face in econometric analysis of the second HJD is that the second HJD involves a nonsmooth function that is not pointwise differentiable. We overcome this difficulty by introducing the concept of differentiability in quadratic mean which has been used in statistics and econometrics literature in dealing with nonsmooth objective functions. Based on the new concept of differentiation, we develop an asymptotic representation of the second HJD.

Following Hansen and Jagannathan (1997), our econometric analysis of the second HJD focuses on the conjugate representation of the minimization problem in Eq. (4):

$$[\delta^+]^2 = \max_{\lambda} \{ \mathbb{E}H^2 - \mathbb{E}[H - \lambda'Y]^+{}^2 - 2\lambda' \mathbb{E}X \}, \tag{9}$$

where λ is an $n \times 1$ vector of Lagrangian multipliers of the constraint that m has to be admissible in Eq. (4), and $[H - \lambda'Y]^+ = \max\{0, H - \lambda'Y\}$.¹⁰ The first-order condition of the above optimization problem is given as

$$\mathbb{E}X - \mathbb{E}[(H - \lambda'Y)^+ Y] = 0. \tag{10}$$

Suppose λ_0 solves Eq. (10), then $[H - \lambda_0'Y]^+ \in \mathcal{M}_+$. That is, $H - [H - \lambda_0'Y]^+$ is the necessary adjustment of H so that it

¹⁰ For more detailed discussions on the conjugate representation of the second HJD, see Hansen and Jagannathan (1997). Replacing $[H - \lambda'Y]^+$ by $[H - \lambda'Y]$ in Eq. (9), then Eq. (9) becomes the definition of $[\delta]^2$, and our entire analysis for the second HJD can be easily adapted to the first HJD.

becomes a member of \mathcal{M}_+ .¹¹ For SDF models that depend on unknown model parameters γ , the second HJD is defined as

$$[\delta^+]^2 = \min_{\gamma} \max_{\lambda} \mathbb{E}\phi(\theta), \tag{11}$$

where $\theta = (\gamma, \lambda)$ and $\phi(\theta) \equiv H(\gamma)^2 - [H(\gamma) - \lambda'Y]^+ - 2\lambda'X$.

Suppose we have the following time series observations of asset prices, payoffs, and an SDF model, $\{(X_{t-1}, Y_t, H_t(\gamma)) : t = 1, 2, \dots, T\}$, where γ is a $k \times 1$ vector of model parameters. Following Hansen and Jagannathan (1997), we use the empirical counterpart of $\mathbb{E}\phi(\theta) = \mathbb{E}\phi(\gamma, \lambda)$,

$$\mathbb{E}_T\phi(\theta) = \frac{1}{T} \sum_{t=1}^T \{H_t(\gamma)^2 - [H_t(\gamma) - \lambda'Y_t]^+ - 2\lambda'X_{t-1}\}, \tag{12}$$

in our econometric analysis of the second HJD, where $\mathbb{E}_T[X] = (1/T) \sum_{t=1}^T x_t$.¹² Therefore, the main objective of our asymptotic analysis is to characterize as $T \rightarrow \infty$ the behavior of

$$[\delta_T^+]^2 = \min_{\gamma} \max_{\lambda} \mathbb{E}_T\phi(\theta). \tag{13}$$

The standard approach for asymptotic analysis of $\mathbb{E}_T\phi(\theta)$ is to employ a pointwise quadratic Taylor expansion of the function $\phi(\theta)$ with respect to θ around true model parameters θ_0 , where $\theta_0 = (\gamma_0, \lambda_0) \equiv \arg \min_{\gamma} \max_{\lambda} \mathbb{E}\phi(\gamma, \lambda)$. That is,

$$\begin{aligned} \phi(\theta) &= \phi(\theta_0) + \frac{\partial\phi(\theta_0)}{\partial\theta'} \cdot (\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \cdot \frac{\partial^2\phi(\theta_0)}{\partial\theta\partial\theta'} \cdot (\theta - \theta_0) + o(\|\theta - \theta_0\|^2), \end{aligned} \tag{14}$$

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

$$\begin{aligned} \mathbb{E}_T\phi(\theta) &= \mathbb{E}_T\phi(\theta_0) + \mathbb{E}_T \frac{\partial\phi(\theta_0)}{\partial\theta'} \cdot (\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \cdot \mathbb{E}_T \frac{\partial^2\phi(\theta_0)}{\partial\theta\partial\theta'} \cdot (\theta - \theta_0) + o_p(\|\theta - \theta_0\|^2). \end{aligned} \tag{15}$$

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

$$\phi(\theta) \equiv H(\gamma)^2 - g(H(\gamma) - \lambda'Y) - 2\lambda'X, \tag{16}$$

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

$$g^{(1)}(x) = 2[x^+] = \begin{cases} 2x & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{17}$$

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

$$g^{(2)}(x) = \begin{cases} 2 & \text{if } x > 0, \\ \text{not exist} & \text{if } x = 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{18}$$

Therefore, for δ_T^+ , the function $[H_t(\gamma) - \lambda'Y_t]^+$ is not pointwise differentiable with respect to (γ, λ) for all $H_t(\gamma)$ and Y_t .¹³ That is, for a given (γ, λ) , there are combinations of $H_t(\gamma)$ and Y_t such that $H_t(\gamma) - \lambda'Y_t = 0$, which is the kink point of $[H_t(\gamma) - \lambda'Y_t]^+$. As a result, the derivatives of $\phi(\gamma, \lambda)$ with respect to (γ, λ) are not always well defined for those $H_t(\gamma)$ and Y_t .

The key to overcome this difficulty is that pointwise differentiability is not a necessary condition to obtain Eq. (15). All we need is a good approximation to $\mathbb{E}_T\phi(\theta)$ [but not $\phi(\theta)$ itself] around the true parameter value θ_0 . To this end, the notion of differentiability in quadratic mean in modern statistics (cf. Le Cam, 1986) plays an important role.¹⁴ In contrast to pointwise differentiability, which implies a good approximation to $\phi(\theta)$ for all H_t and Y_t , differentiability in quadratic mean implies that the error of approximating $\mathbb{E}_T\phi(\theta)$ is negligible in quadratic mean or $L^2(P)$ norm. In other words, all we need is an approximation of $\phi(\theta)$ that works well in an average sense. For further discussions of nondifferentiability issues, see Pollard (1982) and Pakes and Pollard (1989), among others.

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

$$\mathbb{E}_T\phi(\theta) = \mathbb{E}\phi(\theta) + (\mathbb{E}_T - \mathbb{E})\phi(\theta), \tag{19}$$

where $(\mathbb{E}_T - \mathbb{E})[\cdot] = \mathbb{E}_T[\cdot] - \mathbb{E}[\cdot]$. To obtain a quadratic representation such as Eq. (15), we consider a second-order approximation to the deterministic term $\mathbb{E}\phi(\theta)$ and a first-order approximation to the random component $(\mathbb{E}_T - \mathbb{E})\phi(\theta)$. We consider a lower-order approximation of the random component, because it is centered and thus in general one order smaller than the deterministic component. The above analysis leads to Lemma 1, which justifies a local asymptotic quadratic (LAQ) representation of $\mathbb{E}_T\phi(\theta)$.

Lemma 1. *Suppose Assumptions A.1 to A.7 in the Appendix hold. Then the following LAQ representation holds for $\mathbb{E}_T\phi(\theta)$ around θ_0 :*

$$\begin{aligned} \mathbb{E}_T\phi(\theta) &= \mathbb{E}\phi(\theta_0) + (\mathbb{E}_T - \mathbb{E})\phi(\theta_0) + A'(\theta - \theta_0) \\ &\quad + \frac{1}{2}(\theta - \theta_0)' \Gamma (\theta - \theta_0) + o(\|\theta - \theta_0\|^2) + o_p(\|\theta - \theta_0\|T^{-1/2}), \end{aligned} \tag{20}$$

¹¹ Replacing $[H - \lambda'Y]^+$ by $[H - \lambda'Y]$, the adjustment term $H - [H - \lambda'Y]^+$ simplifies to $\lambda'_0 Y$. That is, the random variable $\lambda'_0 Y$ represents the necessary adjustment of H so that it is a member of \mathcal{M} . Alternatively, $\lambda'_0 Y$ can be used to discount future payoffs state by state to yield current pricing errors of Y : $\mathbb{E}[(\lambda'_0 Y)Y] = \mathbb{E}[(H - m)Y]$, where $m \in \mathcal{M}$. Therefore, while δ measures average deviations of H from \mathcal{M} , $\lambda'_0 Y$ measures H 's deviations from \mathcal{M} in different states of the economy.

¹² For notational convenience, we suppress the dependence of $\phi(\theta)$ on t .

¹³ Let ω be a random variable. A function $f(\theta, \omega)$ is pointwise differentiable with respect to θ if the function has partial derivatives with respect to θ in the classical sense for all possible values of ω .

¹⁴ 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 $\mathbb{E}[(f(\theta, \omega) - f(\theta_0, \omega)) / (\theta - \theta_0) - \Delta(\omega)]^2 \rightarrow 0$ as $\theta \rightarrow \theta_0$. Similar ideas have been used by Pakes and Pollard (1989) and others to handle nondifferentiable criteria functions.

where $A \equiv (\mathbb{E}_T - \mathbb{E})\partial\phi(\theta_0)/\partial\theta$ and

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

Proof. See the Appendix.

We emphasize that the expectation of the second derivative is always well defined, even though the second derivative is well defined except on a set of zero probability. The first-order derivative for the deterministic term does not appear in Eq. (20), i.e., $\mathbb{E}\partial\phi(\theta_0)/\partial\theta = 0$, because $\theta_0 = (\gamma_0, \lambda_0)$ solves the population optimization problem $\min_{\gamma} \max_{\lambda} \mathbb{E}\phi(\theta)$. Compared with 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 $\mathbb{E}\phi(\theta)$.

Based on the LAQ representation of $\mathbb{E}_T\phi(\theta)$ in Lemma 1, we can solve for the minimax problem by first maximizing the objective function over λ for a given value of γ , then minimizing the objective function over γ . As a result, we obtain the following asymptotic representation of the second HJD.

Lemma 2. *Suppose Assumptions A.1 to A.9 in the Appendix hold. Then we have the following second-order asymptotic representation of the second HJD at estimated model parameter $\hat{\theta}$:*

$$[\hat{\delta}_T^+]^2 = \min_{\gamma} \max_{\lambda} \mathbb{E}_T \phi(\theta) = \mathbb{E}\phi(\theta_0) + (\mathbb{E}_T - \mathbb{E})\phi(\theta_0) - \frac{1}{2}A'\Gamma^{-1}A + o_p(T^{-1}). \tag{21}$$

Moreover, the optimizer $\hat{\theta} = (\hat{\gamma}, \hat{\lambda}) \equiv \arg \min_{\gamma} \max_{\lambda} \mathbb{E}_T \phi(\theta)$ equals

$$\hat{\theta} = \theta_0 - \Gamma^{-1}A + o_p(T^{-1/2}), \tag{22}$$

where A and Γ are defined in Lemma 1.

Proof. See the Appendix.

Therefore, the technique of differentiation in quadratic mean overcomes the nondifferentiability issue of the second HJD and makes asymptotic analysis of the second HJD feasible. In particular, our analyses on specification test and model selection procedures in later sections are all based on the second-order asymptotic representation of the second HJD in Lemma 2.

3.2. Specification test based on the second HJD

In this subsection, based on the representation of $\hat{\delta}_T^+$ in Lemma 2, we develop the asymptotic distributions of the second HJD and parameter estimate $\hat{\theta}$ under the null hypothesis of a correctly specified model. We first discuss the intuition behind our asymptotic analysis and then present the formal results in Theorem 1 and Proposition 1. We also discuss the relations between our results and those in the existing literature.

Lemma 2 suggests that the behavior of $\hat{\delta}_T^+$ is determined by three terms: $\mathbb{E}\phi(\theta_0)$, $(\mathbb{E}_T - \mathbb{E})\phi(\theta_0)$, and the

quadratic form $\frac{1}{2}A'\Gamma^{-1}A$. Under the null hypothesis $\mathbb{H}_0: \delta^+ = 0$, $\mathbb{E}\phi(\theta_0) = 0$ by definition. Because $\phi(\theta_0)$ is non-negative, we must have $\phi(\theta_0) = 0$ almost everywhere to have $\mathbb{E}\phi(\theta_0) = 0$. As a result, we must have $(\mathbb{E}_T - \mathbb{E})\phi(\theta_0) = 0$ as well. Therefore, under $\mathbb{H}_0: \delta^+ = 0$, the asymptotic behavior of $\hat{\delta}_T^+$ is mainly determined by the quadratic form. From Assumption A.9, we have

$$\sqrt{T}A = \sqrt{T}(\mathbb{E}_T - \mathbb{E}) \frac{\partial\phi(\theta_0)}{\partial\theta} \rightsquigarrow N(0, A), \tag{23}$$

where \rightsquigarrow means convergence in distribution and $A = \mathbb{E}[\partial\phi(\theta_0)/\partial\theta \cdot \partial\phi(\theta_0)/\partial\theta']$. Therefore, $T[\hat{\delta}_T^+]^2$ should follow a weighted χ^2 distribution.

Theorem 1 (Specification test). *Suppose Assumptions A.1 to A.9 in the Appendix hold. Then under $\mathbb{H}_0: \delta^+ = 0$, $T[\hat{\delta}_T^+]^2$ has an asymptotic weighted χ^2 distribution, and the weights are the eigenvalues of the matrix*

$$-\frac{1}{2}[\Gamma_{22}^{-1} - \Gamma_{22}^{-1}\Gamma_{21}[\Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}]^{-1}\Gamma_{12}\Gamma_{22}^{-1}]A_{\lambda}, \tag{24}$$

where Γ_{12} , Γ_{21} , and Γ_{22} are defined in Lemma 1 and $A_{\lambda} = \mathbb{E}[\partial\phi(\theta_0)/\partial\lambda \cdot \partial\phi(\theta_0)/\partial\lambda']$.

Proof. See the Appendix.

We can use Theorem 1 to conduct specification tests of SDF models based on the second HJD. Next we develop the asymptotic distribution of $\hat{\theta} = (\hat{\gamma}, \hat{\lambda})$, which contains important diagnostic information on potential sources of model misspecifications.

Proposition 1 (Parameter estimation). *Suppose Assumptions A.1 to A.9 in the Appendix hold. Then the estimator of model parameters, $\hat{\gamma}$, has the following asymptotic distribution:*

$$\sqrt{T}(\hat{\gamma} - \gamma_0) \rightsquigarrow N(0, (J_{11} \ J_{12})\mathcal{A}(J_{11} \ J_{12})'), \tag{25}$$

and the estimator of Lagrangian multipliers, $\hat{\lambda}$, has the asymptotic distribution

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \rightsquigarrow N(0, (J_{21} \ J_{22})\mathcal{A}(J_{21} \ J_{22})'), \tag{26}$$

where

$$\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}^{-1}$$

and Γ_{ij} 's ($ij=1,2$) are defined in Lemma 1.

Proof. See the Appendix.¹⁵

For a linear SDF model, we can examine the importance of a specific factor by testing whether the coefficient of the factor is significantly different from zero using Proposition 1. The Lagrangian multipliers suggest directions for model improvements: If the multiplier of one particular asset is very large, then the SDF model has to be significantly modified to correctly price this particular asset.

The specification test in Theorem 1 differs from that of Jagannathan and Wang (1996), which is based on the first HJD, in several important ways. First, while Jagannathan

¹⁵ Proposition 1 can be greatly simplified under $\mathbb{H}_0: \delta^+ = 0$, even though it holds whether the model is correctly specified or not. The limiting normal distribution could be degenerate.

and Wang (1996) estimate model parameters by minimizing the first HJD, we estimate model parameters by minimizing the second HJD. Except for the rare case in which $\delta^+ = 0$, the two estimated HJDs and their associated parameters are generally different from each other. Second, the two methods have different powers in rejecting misspecified models with small first HJDs but large second HJDs. A typical example of such models is linear SDF models that have to take negative values with significant probabilities to fit the test assets. While such models could appear acceptable based on the first HJD, they could be rejected by the second HJD because they violate the positivity constraint. Third, while Jagannathan and Wang (1996) consider only linear factor models, our result is applicable to nonlinear SDF models in which $H(\gamma)$ depends on γ nonlinearly. These include asset pricing models in which investors have complicated utility functions. Finally, the econometric techniques used in our analysis are different from that of Jagannathan and Wang (1996) and can be useful in other finance applications that involve nondifferentiable objective functions.

Our results in this section are closely related to that of HHL (1995) and Hansen and Jagannathan (1997). Both papers show that the first and second HJDs follow asymptotic normal distributions for misspecified models.¹⁶ These distributions become degenerate when the models are correctly specified (i.e., the first or second HJD is identically zero).¹⁷ Hansen and Jagannathan (1997, p. 576) argue that the first HJD equals zero only when an SDF model correctly prices all the test assets and one can test this condition via GMM using the pricing errors of the test assets as moment conditions. This argument implies that, even if we were not aware of the asymptotic distribution of δ for correctly specified models, we can still test $\mathbb{H}_0: \delta = 0$ using GMM. It might be tempting to conclude that a similar result could hold for δ^+ as well. However, the second HJD equals zero only when the SDF model can correctly price all the test assets and their related derivatives. As a result, to test $\mathbb{H}_0: \delta^+ = 0$ using the GMM approach, we need data on all derivatives that can be constructed from trading the test assets. Such data could be impossible or at least very difficult to obtain. Therefore, the asymptotic distribution of the second HJD in Theorem 1 is essential for testing $\mathbb{H}_0: \delta^+ = 0$.

3.3. Model selection procedures based on the second HJD

While the specification test can tell whether a given asset pricing model is correctly specified, most SDF models in the existing literature are probably misspecified. In fact, one could argue that linear factor models are

misspecified by definition because their SDFs can take negative values with positive probabilities. Therefore, another important issue is how to compare the relative performances of potentially misspecified SDF models using the second HJD. In this subsection, building on the same technique in developing the specification test, we develop a sequence of model selection procedures in the spirit of Vuong (1989). Similar to that of Vuong (1989), our procedures apply to situations in which both, only one, or neither of the competing models could be correctly specified.

Consider two competing models \mathcal{F} and \mathcal{G} . We are interested in the following hypotheses:

- \mathbb{H}_0 : \mathcal{F} and \mathcal{G} are equally good, i.e., $\delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$;
- $\mathbb{H}_{\mathcal{F}}$: \mathcal{F} is better than \mathcal{G} , i.e., $\delta_{\mathcal{F}}^+ < \delta_{\mathcal{G}}^+$; and
- $\mathbb{H}_{\mathcal{G}}$: \mathcal{F} is worse than \mathcal{G} , i.e., $\delta_{\mathcal{F}}^+ > \delta_{\mathcal{G}}^+$.

In empirical studies, model comparison is conducted using empirical estimates of the second HJDs. Hence, we propose to test the above hypotheses using the following test statistic:

$$[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2, \tag{27}$$

where for convenience we denote $\hat{\delta}_{\mathcal{F}}^+ = \delta_T^+(\hat{\theta}_{\mathcal{F}})$, $\hat{\delta}_{\mathcal{G}}^+ = \delta_T^+(\hat{\theta}_{\mathcal{G}})$, $\hat{\theta}_{\mathcal{F}} = \arg \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} \mathbb{E}_T \phi(\theta_{\mathcal{F}})$, and $\hat{\theta}_{\mathcal{G}} = \arg \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} \mathbb{E}_T \phi(\theta_{\mathcal{G}})$. We also define $\theta_{0\mathcal{F}} = \arg \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} \mathbb{E} \phi(\theta_{\mathcal{F}})$ and $\theta_{0\mathcal{G}} = \arg \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} \mathbb{E} \phi(\theta_{\mathcal{G}})$ as the pseudo-true parameters for models \mathcal{F} and \mathcal{G} , respectively. Following the terminology of Vuong (1989), we say that the two models are observationally equivalent if $\phi(\theta_{0\mathcal{F}}) = \phi(\theta_{0\mathcal{G}})$ with probability one. For all practical purposes, observationally equivalent models are not distinguishable using any test statistics that are functions of ϕ .

The key to our analysis is to obtain the asymptotic representation of the difference between the second HJDs of the two models. From Lemma 2, we know that

$$[\hat{\delta}_{\mathcal{F}}^+]^2 = \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} \mathbb{E}_T \phi(\theta_{\mathcal{F}}) = \mathbb{E} \phi(\theta_{0\mathcal{F}}) + (\mathbb{E}_T - \mathbb{E}) \phi(\theta_{0\mathcal{F}}) - \frac{1}{2} A'_{\mathcal{F}} \Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} + o_p(T^{-1}) \tag{28}$$

and

$$[\hat{\delta}_{\mathcal{G}}^+]^2 = \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} \mathbb{E}_T \phi(\theta_{\mathcal{G}}) = \mathbb{E} \phi(\theta_{0\mathcal{G}}) + (\mathbb{E}_T - \mathbb{E}) \phi(\theta_{0\mathcal{G}}) - \frac{1}{2} A'_{\mathcal{G}} \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} + o_p(T^{-1}), \tag{29}$$

where $A_{\mathcal{F}} \equiv (\mathbb{E}_T - \mathbb{E}) \partial \phi(\theta_{0\mathcal{F}}) / \partial \theta_{\mathcal{F}}$, $A_{\mathcal{G}} \equiv (\mathbb{E}_T - \mathbb{E}) \partial \phi(\theta_{0\mathcal{G}}) / \partial \theta_{\mathcal{G}}$, $\Gamma_{\mathcal{F}} \equiv \mathbb{E} \partial^2 \phi(\theta_{0\mathcal{F}}) / \partial \theta_{\mathcal{F}} \partial \theta'_{\mathcal{F}}$, and $\Gamma_{\mathcal{G}} \equiv \mathbb{E} \partial^2 \phi(\theta_{0\mathcal{G}}) / \partial \theta_{\mathcal{G}} \partial \theta'_{\mathcal{G}}$. Consequently, we have the following asymptotic representation of the test statistic for model selection:

$$[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2 = \mathbb{E} \phi(\theta_{0\mathcal{F}}) - \mathbb{E} \phi(\theta_{0\mathcal{G}}) + (\mathbb{E}_T - \mathbb{E}) [\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})] - \frac{1}{2} A'_{\mathcal{F}} \Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} + \frac{1}{2} A'_{\mathcal{G}} \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} + o_p(T^{-1}). \tag{30}$$

One interesting as well as challenging aspect of our analysis is that $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ exhibits different asymptotic distributions depending on whether the two models are observationally equivalent. If $\phi(\theta_{0\mathcal{F}}) = \phi(\theta_{0\mathcal{G}})$ with probability one, then under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, we have $\mathbb{E} \phi(\theta_{0\mathcal{F}}) - \mathbb{E} \phi(\theta_{0\mathcal{G}}) = 0$ and $(\mathbb{E}_T - \mathbb{E}) [\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})] = 0$. As a

¹⁶ We can easily obtain this result from Lemma 2. Under $\mathbb{H}_0: \delta^+ \neq 0$, we know that $\mathbb{E} \phi(\theta_0) \neq 0$, $(\mathbb{E}_T - \mathbb{E}) \phi(\theta_0) \neq 0$, and the behavior of $\hat{\delta}_T^+$ is mainly determined by $(\mathbb{E}_T - \mathbb{E}) \phi(\theta_0)$ instead of the quadratic term. The second HJD has an asymptotic normal distribution centered at $\mathbb{E} \phi(\theta_0)$, because by central limit theorem, $(\mathbb{E}_T - \mathbb{E}) \phi(\theta_0)$ converges to a normal distribution as $T \rightarrow \infty$.

¹⁷ The approach of HHL (1995) and Hansen and Jagannathan (1997) is equivalent to a first-order Taylor approximation of the second HJD in our setting. However, a second-order approximation is needed to develop the asymptotic distribution of the second HJD under $\mathbb{H}_0: \delta^+ = 0$.

result, the asymptotic distribution of $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ is determined by the difference between the two quadratic forms and follows a weighted χ^2 distribution. However, if $\phi(\theta_{0\mathcal{F}}) \neq \phi(\theta_{0\mathcal{G}})$ with positive probability, then under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, though $\mathbb{E}\phi(\theta_{0\mathcal{F}}) - \mathbb{E}\phi(\theta_{0\mathcal{G}}) = 0$, $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ does not vanish. And the asymptotic distribution of $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ is determined by $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$, which follows a normal distribution.

To address this issue, we develop the asymptotic distributions of $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ for three different types of model structures of \mathcal{F} and \mathcal{G} . Specifically, following the terminology of [Vuong \(1989\)](#), we consider strictly non-nested, overlapping, and nested models. By strictly non-nested models, we mean that $\mathcal{F}_{\theta_{\mathcal{F}}} \cap \mathcal{G}_{\theta_{\mathcal{G}}}$ is an empty set, where $\mathcal{F}_{\theta_{\mathcal{F}}}$ and $\mathcal{G}_{\theta_{\mathcal{G}}}$ represent the entire families of models that we can obtain by considering all possible values of $\theta_{\mathcal{F}}$ and $\theta_{\mathcal{G}}$ in their parameter spaces, respectively. The definition implies that two non-nested models can never be observationally equivalent. Two models $\mathcal{F}_{\theta_{\mathcal{F}}}$ and $\mathcal{G}_{\theta_{\mathcal{G}}}$ are called overlapping if, and only if, $\mathcal{F}_{\theta_{\mathcal{F}}} \cap \mathcal{G}_{\theta_{\mathcal{G}}}$ is not empty and $\mathcal{F}_{\theta_{\mathcal{F}}} \not\subseteq \mathcal{G}_{\theta_{\mathcal{G}}}$ and $\mathcal{G}_{\theta_{\mathcal{G}}} \not\subseteq \mathcal{F}_{\theta_{\mathcal{F}}}$. Model $\mathcal{F}_{\theta_{\mathcal{F}}}$ is said to be nested by model $\mathcal{G}_{\theta_{\mathcal{G}}}$ if, and only if, $\mathcal{F}_{\theta_{\mathcal{F}}} \subseteq \mathcal{G}_{\theta_{\mathcal{G}}}$. Overlapping and nested models can be observationally equivalent for certain parameter values.

We first consider the cases of strictly non-nested and nested models, for which we know unambiguously whether the term $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ vanishes under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$. Then we consider the more difficult case of overlapping models, for which we do not know unambiguously whether the term $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ vanishes under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$.

Because strictly non-nested models can never be observationally equivalent, the term $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ never vanishes and is the dominating term in $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$. As a result, we obtain the following test for comparing strictly non-nested models based on the second HJD.

Theorem 2 (Model selection for non-nested models). Suppose models \mathcal{F} and \mathcal{G} are strictly non-nested and Assumptions A.1 to A.10 in the Appendix hold. Then

- under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, $\sqrt{T}([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) / \hat{\sigma}_T \rightarrow N(0,1)$;
- under $\mathbb{H}_{\mathcal{F}}: \delta_{\mathcal{F}}^+ < \delta_{\mathcal{G}}^+$, $\sqrt{T}([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) / \hat{\sigma}_T \rightarrow -\infty$; and
- under $\mathbb{H}_{\mathcal{G}}: \delta_{\mathcal{F}}^+ > \delta_{\mathcal{G}}^+$, $\sqrt{T}([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) / \hat{\sigma}_T \rightarrow +\infty$,

where $[\hat{\sigma}_T]^2$ is an estimator of $\text{Var}[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ and equals

$$[\hat{\sigma}_T]^2 = \mathbb{E}_T[\phi(\hat{\theta}_{\mathcal{F}}) - \phi(\hat{\theta}_{\mathcal{G}})]^2 - (\mathbb{E}_T[\phi(\hat{\theta}_{\mathcal{F}}) - \phi(\hat{\theta}_{\mathcal{G}})])^2. \quad (31)$$

Proof. See the Appendix.

Theorem 2 allows us to compare two non-nested SDF models based on their second HJDs. The implementation of Theorem 2 involves several steps. First, we solve the optimization problem in Eq. (13) for F and G to obtain $\hat{\theta}_{\mathcal{F}} = (\hat{\gamma}_{\mathcal{F}}, \hat{\lambda}_{\mathcal{F}})$ and $\hat{\theta}_{\mathcal{G}} = (\hat{\gamma}_{\mathcal{G}}, \hat{\lambda}_{\mathcal{G}})$. Then we compute $\hat{\sigma}_T$ to form the statistic $\sqrt{T}([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) / \hat{\sigma}_T$. Finally, we make inferences about the three hypotheses based on the

appropriate critical values of the standard normal distribution.

Next we consider nested models. Without loss of generality we assume that model \mathcal{F} is nested by model \mathcal{G} . Because the bigger model is always at least as good as the smaller model (i.e., $\delta_{\mathcal{F}}^+ \geq \delta_{\mathcal{G}}^+$), under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, we must have $\phi(\theta_{0\mathcal{F}}) = \phi(\theta_{0\mathcal{G}})$ with probability one. That is, the two models must be observationally equivalent under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, which in turn implies that $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})] = 0$. As a result, $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ is asymptotically determined by the difference between the two quadratic forms in Eq. (30). From Assumption A.10, we have

$$\sqrt{T} \begin{pmatrix} A_{\mathcal{F}} \\ A_{\mathcal{G}} \end{pmatrix} = \sqrt{T} (\mathbb{E}_T - \mathbb{E}) \begin{pmatrix} \frac{\partial \phi(\theta_{0\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{\partial \phi(\theta_{0\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{pmatrix} \rightsquigarrow N(0, \Psi), \quad (32)$$

where

$$\Psi = \begin{pmatrix} A_{\mathcal{F}} & A_{\mathcal{F}\mathcal{G}} \\ A_{\mathcal{G}\mathcal{F}} & A_{\mathcal{G}} \end{pmatrix} = \mathbb{E} \begin{pmatrix} \frac{\partial \phi(\theta_{0\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{\partial \phi(\theta_{0\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{pmatrix} \begin{pmatrix} \frac{\partial \phi(\theta_{0\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{\partial \phi(\theta_{0\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{pmatrix}'$$

Then the asymptotic distribution of $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ should follow a weighted χ^2 distribution as given in Theorem 3.

Theorem 3 (Model selection for nested models). Suppose model \mathcal{F} is nested by model \mathcal{G} and Assumptions A.1 to A.10 in the Appendix hold. Then

under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2)$ has an asymptotic weighted χ^2 distribution, and the weights are the eigenvalues of the following matrix:

$$\frac{1}{2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} & -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{G}} \\ \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}\mathcal{F}} & \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} \end{pmatrix};$$

and

under $\mathbb{H}_{\mathcal{G}}: \delta_{\mathcal{F}}^+ > \delta_{\mathcal{G}}^+$, $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) \rightarrow +\infty$.

Proof. See the Appendix.

The case for overlapping models is much more complicated, because the term $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ might or might not vanish under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$. The two models could be observationally equivalent for certain parameter values under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$. In such cases, the term $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})] = 0$, and the test statistic $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ is determined by the two quadratic forms and follows a weighted χ^2 distribution. There is also the possibility that the two models are not observationally equivalent under $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$. In such cases, the term $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})] \neq 0$, and the test statistic $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ follows a normal distribution. Due to this additional complexity, we first need to test the possibility that the two models are observationally equivalent before we can decide which asymptotic distribution to use to test the null hypothesis $\mathbb{H}_0: \delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$. Following similar arguments in Theorem 3, we can show that if,

$\phi(\theta_{0\mathcal{F}}) = \phi(\theta_{0\mathcal{G}})$ with probability one, $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ should follow a weighted χ^2 distribution as given below.

Theorem 4 (Model selection for overlapping models). *Suppose models \mathcal{F} and \mathcal{G} are overlapping models and Assumptions A.1 and A.10 in the Appendix hold. Then*

under \mathbb{H}_0^* : $\phi(\theta_{\mathcal{F}0}) = \phi(\theta_{\mathcal{G}0})$ with probability one, $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2)$ has an asymptotic weighted χ^2 distribution and the weights are the eigenvalues of the matrix

$$\frac{1}{2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} & -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{G}} \\ \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}\mathcal{F}} & \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} \end{pmatrix};$$

and

under \mathbb{H}_A^* : $\phi(\theta_{\mathcal{F}0}) \neq \phi(\theta_{\mathcal{G}0})$ with positive probability, $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) \rightarrow \infty$ (either $+\infty$ or $-\infty$).

Proof. See the Appendix.

In summary, the comparison of overlapping models should be done according to the following procedures.

- Test whether the two models are observationally equivalent using Theorem 4,¹⁸
- If the test fails to reject \mathbb{H}_0^* , then the two models are indistinguishable given the test assets.
- If the test rejects \mathbb{H}_0^* , then test \mathbb{H}_0 : $\delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$ using the asymptotic distribution in Theorem 2.¹⁹

Although the asymptotic distributions in Theorems 3 and 4 appear to be similar, the focuses of the two theorems are totally different. Theorem 3 is a one-sided test of the hypothesis that two nested SDF models have the same second HJD. Theorem 4 is a two-sided test of the hypothesis that two overlapping SDF models are observationally equivalent. We obtain such similar results in Theorems 3 and 4, because we test both hypotheses using the same test statistic and two nested models are observationally equivalent if they have the same second HJDs.

Given that most models, especially linear factor models, are likely to be misspecified, the model selection procedures developed in this section make important methodological contributions to the asset pricing literature. Although HHL (1995) and Hansen and Jagannathan (1997) develop the asymptotic distributions of the two HJDs for misspecified models, their results can test only the null hypothesis that a given model has a fixed (nonzero) level of first or second HJD. Their results, however, cannot be used for formal model comparison because they do not provide the distribution of the

difference of the HJDs between two models.²⁰ Even though one can extend their analyses to study the difference of the HJDs between the two models, such an extension can cover only the case of strictly non-nested models. Only based on the second-order Taylor approximation of the HJDs can we develop the model selection tests for nested and overlapping models.²¹ Therefore, our model selection procedures fill an important gap in the literature by providing a systematic approach for comparing potentially misspecified SDF models based on the second HJD.

4. Finite sample performances of asset pricing tests

In this section, we provide simulation evidence on the finite sample performances of both the specification test and model selection tests. We first discuss the simulation designs and then report the simulation results. Overall, we find that our tests have reasonably good finite sample performances for sample sizes typically considered in the literature.

4.1. Simulation designs

We first discuss our simulation designs for the specification test. Suppose an SDF model has the representation

$$H_t = b'F_t, \tag{33}$$

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

$$D_t \sim N(\mu_D, \Sigma_D), \tag{34}$$

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

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

¹⁸ Vuong (1989) shows that under \mathbb{H}_0^* : $\phi(\theta_{\mathcal{F}0}) = \phi(\theta_{\mathcal{G}0})$ with probability one, $[\hat{\sigma}_{\mathcal{F}}]^2$ follows a weighted χ^2 distribution and chooses to test \mathbb{H}_0^* based on $[\hat{\sigma}_{\mathcal{F}}]^2$. However, our simulation results (not reported) show that our test in Theorem 4 has much better finite sample performances than Vuong's test.

¹⁹ For two overlapping but not observationally equivalent models \mathcal{F} and \mathcal{G} , $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2)$ has an asymptotic normal distribution as given in Theorem 2.

²⁰ For example, we can test whether the $\hat{\delta}^+$ of an SDF model is significantly different from a pre-specified level, say 0.5, using the results of HHL (1995) and Hansen and Jagannathan (1997). Suppose we find that $\hat{\delta}_{\mathcal{F}}^+ = 0.5$ and $\hat{\delta}_{\mathcal{G}}^+ = 1$ in our empirical analysis. We cannot test whether $\hat{\delta}_{\mathcal{G}}^+$ is significantly different from $\hat{\delta}_{\mathcal{F}}^+$ using the results of HHL (1995) and Hansen and Jagannathan (1997), because their tests do not simultaneously account for the estimation errors in both $\hat{\delta}_{\mathcal{F}}^+$ and $\hat{\delta}_{\mathcal{G}}^+$. In contrast, our model selection tests explicitly characterize the asymptotic behavior of $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$.

²¹ A second-order Taylor approximation of the second HJD is also needed to develop the model selection tests for overlapping and nested models.

expected returns of the N assets can be written as

$$\mathbb{E}(Y_t) = \frac{\mathbb{E}(X_{t-1}) - \text{cov}(Y_t, H_t)}{\mathbb{E}(H_t)}. \quad (35)$$

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

Based on the above procedure, we generate five hundred random samples of D_t with different number of time series observations T . In our main simulation, we choose $N=26$ to mimic the risk-free asset and the Fama and French 25 size/BM portfolios that have been widely used in the empirical literature.²² We choose $T=200,400$, and 600, where 200 (600) represents the typical number of quarterly (monthly) observations available in standard empirical asset pricing studies.

For each simulated random sample, we estimate model parameters and conduct specification tests based on the first and second HJDs. Then we report rejection rates based on the asymptotic critical values at the 10% and 5% significance levels of the two tests.²³ If the tests have good size performances, then the rejection rates for a correctly specified model at the above critical values should be close to 10% and 5%, respectively. If the tests have good power performances, then the rejection rates for a misspecified model should be close to one.

We examine the finite sample size performances of the specification tests using the Fama and French three-factor model (FF3) with the SDF

$$H_t^{\text{FF3}} = b_0 + b_1 r_{\text{MKT},t} + b_2 r_{\text{SMB},t} + b_3 r_{\text{HML},t}, \quad (36)$$

where $r_{\text{SMB},t}$ and $r_{\text{HML},t}$ are the return differences between small and big firms and high and low BM firms, respectively. FF3 is a widely used model in the literature. Although it is a linear factor model, its SDF at empirically estimated parameter values from historical Fama and French portfolios take negative values with negligible probability. Therefore, we treat FF3 as the true model in our size simulations. To examine the finite sample power performances of the specification tests, we consider a simple Capital Asset Pricing Model (CAPM). Because the data are generated from FF3, the CAPM should not be able to price the assets and should be rejected by the specification tests.

Next we discuss our simulation designs for the model selection tests. For all the tests we consider, we generate simulated data using the same FF3 model in Eq. (36). We face many choices in what types of models to use when testing the finite sample size and power performances of the model selection tests because of the different model structures. Given that we simulate data from FF3, we choose some simple deviations from FF3 in our simula-

tions. In particular, all deviations from FF3 are based on the following two redundant factors:

$$r_{1,t} \sim N(\mu_m, \sigma_m) \quad (37)$$

and

$$r_{2,t} \sim N(\mu_m, \sigma_m), \quad (38)$$

where $\mu_m(\sigma_m)$ is the mean (volatility) of excess market returns during our sample period and $r_{1,t}$ and $r_{2,t}$ are independent of each other. They are also independent of all the FF3 factors and the asset returns.

To examine the size of the test for non-nested models, we consider the two models

$$H_t^{r_1} = br_{1,t} \quad (39)$$

and

$$H_t^{r_2} = br_{2,t}. \quad (40)$$

We estimate the two models using simulated data and test the null hypothesis that the two models have the same first or second HJDs. Because the two models are equally wrong from FF3, the null hypothesis of equal first or second HJDs should hold. To examine the power of the test, we test whether $H_t^{r_1}$ and H_t^{FF3} have the same first or second HJDs, a hypothesis that should be rejected by the simulated data.

To examine the size of the test for nested models, we consider the model

$$H_t^{\text{FF3}+r_1} = b_0 + b_1 r_{\text{MKT},t} + b_2 r_{\text{SMB},t} + b_3 r_{\text{HML},t} + b_4 r_{1,t}. \quad (41)$$

We estimate $H_t^{\text{FF3}+r_1}$ and H_t^{FF3} using simulated data and test the null hypothesis that the two models have the same first or second HJDs. This hypothesis should hold, because $H_t^{\text{FF3}+r_1}$ nests H_t^{FF3} and $r_{1,t}$ is a redundant factor. To examine the power of the test, we test whether $H_t^{r_1}$ and $H_t^{\text{FF3}+r_1}$ have the same first or second HJDs. This hypothesis should be rejected by the simulated data because $H_t^{\text{FF3}+r_1}$ has smaller first and second HJDs than $H_t^{r_1}$.

To examine the size of the test for overlapping models, we compare $H_t^{\text{FF3}+r_1}$ with

$$H_t^{\text{FF3}+r_2} = b_0 + b_1 r_{\text{MKT},t} + b_2 r_{\text{SMB},t} + b_3 r_{\text{HML},t} + b_4 r_{2,t}. \quad (42)$$

We estimate $H_t^{\text{FF3}+r_1}$ and $H_t^{\text{FF3}+r_2}$ using the simulated data and test the null hypothesis that the two models are observationally equivalent. This hypothesis should hold because $H_t^{\text{FF3}+r_1}$ overlaps with $H_t^{\text{FF3}+r_2}$ and $r_{1,t}$ and $r_{2,t}$ are redundant factors. To examine the power of the test, we test whether $H_t^{\text{FF3}+r_1}$ is observationally equivalent to $H_t^{r_1+r_2}$, where

$$H_t^{r_1+r_2} = b_1 r_{1,t} + b_2 r_{2,t}. \quad (43)$$

The two models are not observationally equivalent, and $H_t^{\text{FF3}+r_1}$ should have smaller first and second HJDs than $H_t^{r_1+r_2}$.

4.2. Finite sample size and power performances

Panel A of Table 1 reports the size and power performances of the δ^- - and δ^+ -based specification tests using simulated data that mimic the risk-free asset and the Fama and French 25 portfolios. Specifically, it reports

²² We also consider simulation results not reported for the Fama and French nine portfolios. We find that most tests have better finite sample performances for a smaller number of assets.

²³ We obtain the asymptotic critical values at the 10% and 5% significance levels based on ten thousand simulated samples from the weighted χ^2 distributions in Theorem 1 for the first and second HJDs. We use similar procedures to obtain the asymptotic critical values for the model selection tests in Theorems 3 and 4 as well.

Table 1

Finite sample performances of specification and model comparison tests.

This table reports the finite sample size and power performances of the specification and model comparison tests based on the first and second Hansen–Jagannathan distances. In all simulations, we generate data according to the Fama and French three-factor model (FF3). That is, the expected returns of simulated assets are determined by FF3 and the covariance matrix of simulated asset returns mimics that of the Fama and French 25 portfolios between 1952 and 2000. We generate 500 simulated samples with sample sizes of 200, 400, and 600. For each simulated sample, we test the null hypothesis using either the specification or model comparison tests. We report rejection rates at the 10% and 5% asymptotic critical values over the 500 simulations. More details on simulation designs can be found in Section 4.1.

| Sample size | Using δ | | | | Using δ^+ | | | |
|---|----------------|-----|-------|------|------------------|-----|-------|------|
| | Size | | Power | | Size | | Power | |
| | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% |
| Panel A. Size and power performances of specification tests | | | | | | | | |
| 200 | 19% | 13% | 51% | 41% | 21% | 13% | 61% | 51% |
| 400 | 16% | 8% | 55% | 42% | 15% | 9% | 89% | 79% |
| 600 | 17% | 7% | 58% | 46% | 17% | 8% | 94% | 89% |
| Panel B. Size and power performances of model comparison tests for strictly non-nested models | | | | | | | | |
| 200 | 2% | 1% | 100% | 100% | 4% | 1% | 100% | 100% |
| 400 | 2% | 0% | 100% | 100% | 5% | 2% | 100% | 100% |
| 600 | 0% | 0% | 100% | 100% | 2% | 0% | 100% | 100% |
| Panel C. Size and power performances of model comparison tests for nested models | | | | | | | | |
| 200 | 5% | 3% | 99% | 99% | 5% | 3% | 100% | 100% |
| 400 | 4% | 1% | 100% | 100% | 4% | 1% | 100% | 100% |
| 600 | 7% | 4% | 99% | 99% | 6% | 4% | 100% | 100% |
| Panel D. Size and power performances of model comparison tests for overlapping models | | | | | | | | |
| 200 | 6% | 2% | 100% | 100% | 5% | 2% | 100% | 100% |
| 400 | 7% | 3% | 100% | 100% | 7% | 3% | 100% | 100% |
| 600 | 11% | 6% | 100% | 100% | 11% | 7% | 100% | 100% |

the rejection rates of FF3 (the null) and CAPM (the alternative) based on the asymptotic critical values at the 10% and 5% significance levels for the two tests.

Both tests clearly tend to over-reject the null hypothesis for $T=200$. The rejection rates for both tests are about 20% and 13% at the 10% and 5% asymptotic critical values, respectively. The performances of both tests become reasonably good for $T=400$ and 600. The rejection rates for both tests are about 13–15% and 7–8% at the 10% and 5% asymptotic critical values, respectively. Therefore, for typical sample sizes considered in the current literature, both δ - and δ^+ -based specification tests have similar and reasonably good size performances.

Both tests also have similar power performances in rejecting the misspecified CAPM. For $T = 200$, the rejection rates of both tests are about 80% and 75% at the 10% and 5% asymptotic critical values, respectively. As T increases to 400 and 600, the rejection rates of both tests are close to 100%.²⁴

Panel B of Table 1 reports both the size and power performances of the model selection tests for strictly non-nested models. Both δ - and δ^+ -based tests have very good size performances with rejection rates close to corresponding asymptotic critical values for all sample sizes. The tests also have excellent power in detecting models

with different HJDs: The rejection rates are close to 100% at all sample sizes.

Panel C of Table 1 reports the performances of the model selection tests for nested models. The tests for nested models tend to slightly under reject the null hypothesis of equal HJDs between the two models. The rejection rates for both δ - and δ^+ -based tests are about 5–7% (2–3%) at the 10% (5%) asymptotic critical value for $T=600$. The δ^+ -based test for nested models also has 100% rejection rates for the two models with different HJDs.

Panel D of Table 1 reports the performances of the model selection tests for overlapping models. The tests for overlapping models tend to under reject the null hypothesis of two observationally equivalent models. The rejection rates of both δ - and δ^+ -based tests are 3–4% and 1% at the 10% and 5% asymptotic critical values for $T=600$, respectively. The δ^+ -based test has excellent power with 100% rejection rates for the two models that are not observationally equivalent.²⁵ In contrast, the power of the δ -based test is much worse with rejection rates in the range of 60–70%. One main reason for the different powers of the two tests is that the alternative model $H_t^{r_1+r_2}$ takes negative value 40% of the time in our simulation. Therefore, the δ^+ -based test is more powerful in differentiating $H_t^{FF3+r_1}$ from $H_t^{r_1+r_2}$, because the latter is not arbitrage-free. The above simulation results show that

²⁴ The δ^+ -based test should be more powerful than the δ -based test in rejecting misspecified models that have small pricing errors but are not arbitrage-free. This advantage, however, is not reflected here, because the SDF of CAPM rarely takes negative values in our simulation.

²⁵ The two models used in the power analysis of overlapping models have different HJDs. In results not reported here, these two models are also rejected at 100% level based on Theorem 2.

the model selection tests also have reasonably good finite sample performances for typical sample sizes considered in the current literature.

5. Empirical results

In this section, we provide empirical evidence on the advantages of the second HJD for empirical asset pricing studies. In particular, we examine several models that have been developed in the recent literature to explain the cross-sectional returns of the Fama-French size/BM portfolios. These models include that of Lettau and Ludvigson (2001), Lustig and Nieuwerburgh (2004), Santos and Veronesi (2006), Li, Vassalou, and Xing (2006), and Yogo (2006).²⁶ These models also have been considered by LNS (2010) due to both their importance in the literature and data availability. LNS (2010) show that these models pose serious challenges to existing asset pricing tests that have focused mainly on pricing errors because the models tend to have small pricing errors for the Fama and French portfolios by construction. We reach dramatically different conclusions on model performances using the second HJD than the first HJD.

5.1. Asset pricing models

The first model we consider is the conditional consumption CAPM of Lettau and Ludvigson (2001), in which the conditioning variable is the aggregate consumption-to-wealth ratio. The SDF of the model has the expression

$$H_t^{LL} = b_0 + b_1 cay_{t-1} + b_2 \Delta c_t + b_3 cay_{t-1} \Delta c_t, \quad (44)$$

where cay_{t-1} is the lagged consumption-to-wealth ratio and Δc_t is the log consumption growth rate.

The second model we consider is the conditional consumption CAPM of Lustig and Nieuwerburgh (2004), in which the conditioning variable is the housing collateral ratio. Following LNS (2010), we consider only their linear model with separate preferences. The SDF of the model has the expression

$$H_t^{LV} = b_0 + b_1 my_{t-1} + b_2 \Delta c_t + b_3 my_{t-1} \Delta c_t, \quad (45)$$

where my_{t-1} is the lagged housing collateral ratio based on mortgage data.

The third model we consider is the conditional CAPM of Santos and Veronesi (2006) with the SDF

$$H_t^{SV} = b_0 + b_1 r_{MKT,t} + b_2 s_{t-1}^{\omega} r_{MKT,t}, \quad (46)$$

where s_{t-1}^{ω} is the lagged labor income-to-consumption ratio.

The fourth model we consider is the sector investment model of Li, Vassalou, and Xing (LVX, 2006). The two versions of the model we consider, denoted as LVX1 and

LVX2, have the SDFs

$$H_t^{LVX1} = b_0 + b_1 \Delta I_{HH,t} + b_2 \Delta I_{Corp,t} + b_3 \Delta I_{Ncorp,t} \quad (47)$$

and

$$H_t^{LVX2} = b_0 + b_1 \Delta I_{HH,t} + b_2 \Delta I_{Corp,t} + b_3 \Delta I_{FCorp,t} + b_4 \Delta I_{Ncorp,t} + b_5 \Delta I_{FM,t}, \quad (48)$$

where $\Delta I_{HH,t}$, $\Delta I_{Corp,t}$, $\Delta I_{Ncorp,t}$, $\Delta I_{FCorp,t}$, and $\Delta I_{FM,t}$ represent log investment growth rates for households, nonfinancial corporations, the noncorporate sector, financial corporations, and the farm sector, respectively. While LVX2 is the original model considered in LVX (2006), LVX1 is the simpler version considered in LNS (2010).

The next model we consider is the durable-consumption CAPM of Yogo (2006), in which the factors are the growth of durable and nondurable consumption and the market return. The SDF of the model has the expression

$$H_t^{YOGO} = b_0 + b_1 \Delta c_{Ndur,t} + b_2 \Delta c_{Dur,t} + b_3 r_{MKT,t}, \quad (49)$$

where $\Delta c_{Ndur,t}$ and $\Delta c_{Dur,t}$ represent log consumption growth for nondurable and durable goods, respectively.

We obtain most of the factors from the corresponding authors' websites. Most of the models use consumption or investment as factors, which are typically available at only quarterly frequency. As a result, we estimate all the models using quarterly returns of the risk-free asset and the Fama and French 25 portfolios from 1952 to 2000. For comparison, we also consider the Fama and French three-factor model, H_t^{FF3} .²⁷

5.2. Empirical results for the Fama-French portfolios

Our empirical analysis is conducted in several steps. We first estimate all the models by minimizing their corresponding first and second HJDs.²⁸ We then conduct specification tests of all the models based on the first and second HJDs. Finally, we compare relative model performances using the model selection tests based on the first and second HJDs.

Panel A of Table 2 reports the results of specification tests of all the models. We first report the estimated first and second HJDs and their differences. We then report the probability that \hat{H}_t (estimated using the first HJD) takes negative values during the sample period, $\Pr(\hat{H}_t < 0)$. Finally, we report the p -values of the δ^- and δ^+ -based specification tests for all the models. We reach dramatically different conclusions on model performances based on the first and second HJDs. For example, LVX1 and LVX2 have the smallest first HJDs among all the models. In fact, the p -values of the δ^- -based specification test for the two models are 33% and 53%, respectively, while the p -values for all other models are zero. Therefore, the two models seem to capture the returns of the Fama and French portfolios reasonably well based on the first HJD. However, the probabilities that the estimated SDFs of the two models take negative values are 14% and 15%, respectively. Because

²⁶ Most of these models incorporate some kind of conditioning variables to improve the fit of the data. For issues related to conditional asset pricing models, see Ferson and Harvey (1999) and Farnsworth, Ferson and Jackson (2002), among others.

²⁷ The returns on all the test assets and the Fama and French factors were downloaded from Ken French's website on March 13, 2009.

²⁸ For brevity, we do not report the estimates of model parameters and Lagrangian multipliers. These results are available upon request.

Table 2

Asset pricing tests based on the first and second Hansen-Jagannathan distances for the risk-free asset and the Fama and French 25 portfolios.

This table provides empirical results on seven asset pricing models based on the two HJDs for the risk-free asset and the Fama and French 25 Size/BM portfolios from Q2.1952 to Q4.2000. The first data point is Q2.1952 because some of the factors are lagged by one quarter. The returns of all the test assets and the Fama and French factors are downloaded from Ken French's website on March 13, 2009. Panel A reports specification test results. The first (second) row of Panel A contains the estimated first (second) HJD. The third row of Panel A contains the percentage difference between the two HJDs. The fourth row reports the probabilities that SDF models estimated using the first HJD take negative values during the sample period. The last two rows of Panel A report the p -values of specification tests based on the first and second HJDs. Panel B contains the results on model comparison for overlapping models. The reported numbers are the p -values of the hypothesis that the model in the corresponding column is observationally equivalent to the model in the corresponding row. Panel C contains the results on model comparison for overlapping but not observationally equivalent models. The reported numbers are the t -statistics for the difference between the HJDs of the model in the corresponding column and the model in the corresponding row. If the column model is better than the row model at the 5% significance level, then the t -statistic should be smaller than -1.96 , and vice versa.

| Panel A: Results of specification tests using the risk-free asset and the Fama and French 25 portfolios | | | | | | | |
|---|-------|-------|-------|-------|-------|-------|-------|
| | LL | LV | SV | LVX1 | LVX2 | YOGO | FF3 |
| Δ | 0.643 | 0.643 | 0.642 | 0.580 | 0.546 | 0.651 | 0.582 |
| δ^+ | 0.685 | 0.700 | 0.667 | 0.691 | 0.684 | 0.673 | 0.607 |
| $(\delta^+ - \delta)/\delta$ | 6% | 9% | 4% | 19% | 25% | 3% | 4% |
| $p(H < 0)$ | 2% | 10% | 1% | 14% | 15% | 0% | 2% |
| $p(\delta = 0)$ | 0% | 0% | 0% | 33% | 53% | 0% | 0% |
| $p(\delta^+ = 0)$ | 0% | 0% | 0% | 0% | 0% | 0% | 0% |

| Panel B: Results of model comparison tests for overlapping models | | | | | | | | | | | | |
|---|----------------|-----|-----|------|------|------|------------------|-----|-----|------|------|------|
| Model | Using δ | | | | | | Using δ^+ | | | | | |
| | LL | LV | SV | LVX1 | LVX2 | YOGO | LL | LV | SV | LVX1 | LVX2 | YOGO |
| LV | 93% | | | | | | 42% | | | | | |
| SV | 90% | 85% | | | | | 87% | 30% | | | | |
| LVX1 | 7% | 10% | 5% | | | | 71% | 72% | 51% | | | |
| LVX2 | 5% | 6% | 3% | 16% | | | 80% | 63% | 63% | 34% | | |
| YOGO | 45% | 50% | 26% | 2% | 2% | | 67% | 73% | 26% | 96% | 84% | |
| FF3 | 23% | 36% | 0% | 18% | 11% | 0% | 7% | 1% | 0% | 1% | 2% | 0% |

| Panel C: Results of model comparison tests for overlapping but not observationally equivalent models | | | | | | | | | | | | |
|--|----------------|----|------|-------|-------|------|------------------|------|------|------|------|------|
| Model | Using δ | | | | | | Using δ^+ | | | | | |
| | LL | LV | SV | LVX1 | LVX2 | YOGO | LL | LV | SV | LVX1 | LVX2 | YOGO |
| LV | | | | | | | | | | | | |
| SV | | | | | | | | | | | | |
| LVX1 | | | 0.67 | | | | | | | | | |
| LVX2 | 0.82 | | 0.91 | | | | | | | | | |
| YOGO | | | | -0.79 | -1.00 | | | | | | | |
| FF3 | | | 1.83 | | | 2.03 | | 2.21 | 2.04 | 1.94 | 1.77 | 2.08 |

the second HJD explicitly requires a good model to be arbitrage-free, the two models are overwhelmingly rejected by the δ^+ -based specification test. All the other models are rejected by the second HJD as well.

In theory, the consumption-based models should have better performances than FF3 measured by the second HJD, because the former is designed to price both the primary and derivatives assets while FF3 focuses mainly on pricing the primary assets. However, FF3 has smaller second HJD than all the consumption-based models. One reason for this result is that the consumption-based models we consider are linearized versions of the original models and are not guaranteed to be arbitrage-free. Another and more fundamental reason is that the consumption-based models try to explain the cross-sectional returns of the Fama and French portfolios using fundamental economic variables, which is more

challenging than using factors extracted from stock returns, such as the Fama and French factors.

Next we consider the relative performances of these models using the model selection tests based on the first and second HJDs. Panel B of Table 2 reports the model comparison results based on the tests for overlapping models in Theorem 4, because all models we consider share at least one common constant factor and thus are overlapping models.²⁹ Each entry in Panel B represents the p -value of the hypothesis that the model in the corresponding column is observationally equivalent to the model in the corresponding row. Among all the pairs of models we consider, we conclude based on the first HJD

²⁹ Because LVX2 nests LVX1, we compare this pair of models using the tests in Theorem 3.

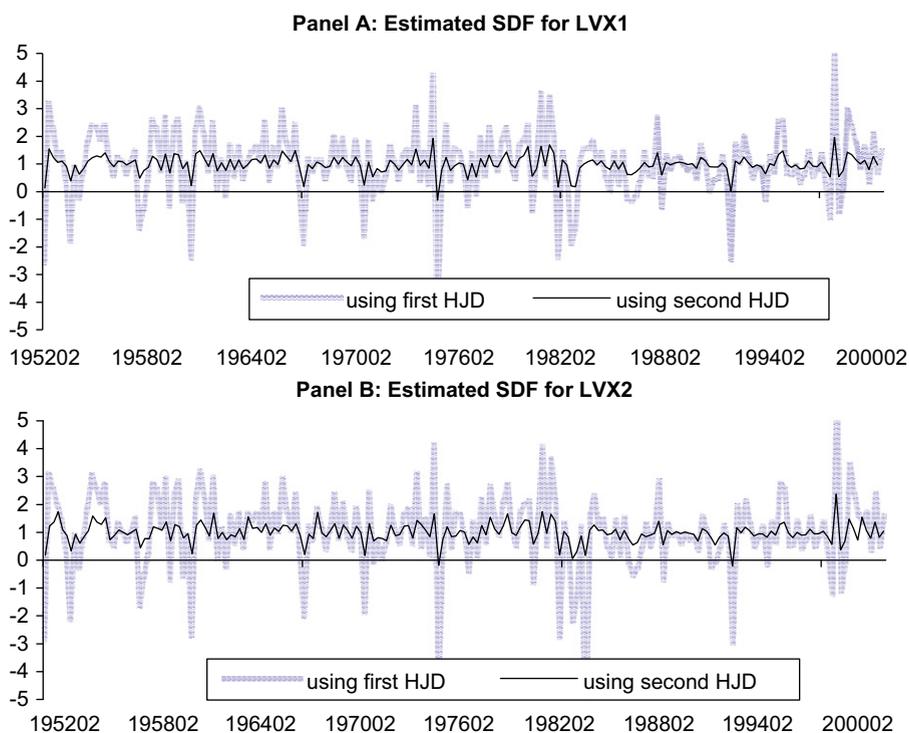


Fig. 2. Time series plots of two stochastic discount models estimated using the risk-free asset and the 25 Fama and French size/book-to-market portfolios. Panel A (B) contains time series plots of empirically estimated two stochastic discount factor models of Li, Vassalou, and Xing (2006) (denoted by LVX1 and LVX2) using the first and second HJDs. The details of the model can be found in Section 5.1.

that LVX2 is not observationally equivalent to LL, SV, and YOGO; LVX1 is not equivalent to SV to YOGO; and FF3 is not equivalent to SV and YOGO. In contrast, based on the second HJD, we conclude that FF3 is not observationally equivalent to almost all other models, which appear to be not distinguishable from each other. These results are consistent with that in Panel A of Table 2, where all models except FF3 have similar second HJDs.

For those models that are not observationally equivalent, in Panel C of Table 2, we test whether they have the same first or second HJDs based on the tests for non-nested models in Theorem 2.³⁰ Each entry in Panel C represents the test statistic of the hypothesis that the model in the corresponding column is better than the model in the corresponding row measured by the first or second HJD. We again reach dramatically different conclusions based on the first and second HJDs. Only YOGO has significantly bigger first HJD than FF3, while all other pairs of nonequivalent models do not exhibit significantly different first HJDs. In contrast, we find most other models have significantly bigger second HJDs than FF3, suggesting that none of them can capture the returns of the Fama and French portfolios as well as FF3.

The two HJDs also lead to very different estimated SDF models. We present time series plots of \hat{H}_t (estimated using the first HJD) and \hat{H}_t^+ (estimated using the second HJD) for LVX1 and LVX2 in Panels A and B of Fig. 2, respectively. It is clear that \hat{H}_t takes negative values much more frequently than \hat{H}_t^+ for both models, and \hat{H}_t^+ takes negative values only on rare occasions. Therefore, for a given sample, though linear factor models estimated using the first HJD can take negative values with high probabilities, they can be made closer to be arbitrage free when estimated using the second HJD.

To summarize, the first and second HJDs could lead to dramatically different conclusions on model performances. Even though based on the first HJD certain models appear to do a good job in explaining the Fama and French portfolios, all models are overwhelmingly rejected by the second HJD. Moreover, the second HJD is more powerful than the first HJD in distinguishing models that have similar pricing errors of the test assets but are not arbitrage-free. The analysis in this subsection shows that the second HJD can make significant differences in empirical asset pricing studies.

5.3. Further diagnostics

While we evaluate and compare asset pricing models using the second HJD in the previous subsection, we provide further diagnostics in this subsection to better

³⁰ The difference between the first or second HJDs of two overlapping but not observationally equivalent models has the same distribution as that between two strictly non-nested models.

Table 3

Correlation matrices for the estimated SDFs, the implied true SDFs, and the adjustment portfolios under the first and second HJDs.

This table provides correlation matrices for the estimated SDFs, the implied true SDFs, and the adjustment portfolios under the first and second HJDs using the risk-free asset and the Fama and French 25 Size/Book-to-Market portfolios from Q2.1952 to Q4.2000. The adjustment portfolios are the differences between the estimate SDFs and the implied true SDFs.

| Model | Using δ | | | | | | Using $\delta+$ | | | | | |
|--|----------------|------|-------|------|------|------|-----------------|------|------|------|------|------|
| | LL | LV | SV | LVX1 | LVX2 | YOGO | LL | LV | SV | LVX1 | LVX2 | YOGO |
| Panel A: Correlation matrix of the estimated SDFs under the first and second HJDs | | | | | | | | | | | | |
| LV | 0.56 | | | | | | 0.57 | | | | | |
| SV | 0.22 | 0.05 | | | | | 0.26 | 0.10 | | | | |
| LVX1 | 0.02 | 0.05 | 0.01 | | | | 0.09 | 0.12 | 0.06 | | | |
| LVX2 | 0.06 | 0.10 | -0.01 | 0.90 | | | 0.08 | 0.07 | 0.07 | 0.89 | | |
| YOGO | 0.26 | 0.23 | 0.65 | 0.05 | 0.01 | | 0.28 | 0.22 | 0.79 | 0.11 | 0.10 | |
| FF3 | 0.26 | 0.13 | 0.60 | 0.15 | 0.16 | 0.58 | 0.26 | 0.14 | 0.64 | 0.19 | 0.19 | 0.62 |
| Panel B: Correlation matrix of the implied true SDFs under the first and second HJDs | | | | | | | | | | | | |
| LV | 0.77 | | | | | | 0.94 | | | | | |
| SV | 0.76 | 0.63 | | | | | 0.89 | 0.93 | | | | |
| LVX1 | 0.40 | 0.36 | 0.47 | | | | 0.87 | 0.90 | 0.93 | | | |
| LVX2 | 0.37 | 0.34 | 0.40 | 0.92 | | | 0.85 | 0.88 | 0.92 | 0.99 | | |
| YOGO | 0.79 | 0.70 | 0.95 | 0.49 | 0.42 | | 0.91 | 0.95 | 0.98 | 0.94 | 0.93 | |
| FF3 | 0.80 | 0.67 | 0.96 | 0.52 | 0.45 | 0.99 | 0.91 | 0.94 | 0.99 | 0.94 | 0.93 | 1.00 |
| Panel C: Correlation matrix of the adjustment portfolios (the difference between estimated SDFs and the implied true SDFs) under the first and second HJDs | | | | | | | | | | | | |
| LV | 0.96 | | | | | | 0.97 | | | | | |
| SV | 0.94 | 0.89 | | | | | 0.95 | 0.93 | | | | |
| LVX1 | 0.78 | 0.81 | 0.77 | | | | 0.93 | 0.96 | 0.91 | | | |
| LVX2 | 0.74 | 0.76 | 0.71 | 0.94 | | | 0.93 | 0.95 | 0.90 | 0.99 | | |
| YOGO | 0.93 | 0.90 | 0.98 | 0.82 | 0.75 | | 0.95 | 0.94 | 0.99 | 0.92 | 0.91 | |
| FF3 | 0.87 | 0.82 | 0.92 | 0.75 | 0.73 | 0.89 | 0.87 | 0.85 | 0.91 | 0.84 | 0.84 | 0.90 |

understand the differences and similarities among all the models.

In Table 3, we report the correlation matrices of the estimated SDFs (\hat{H}_t and \hat{H}_t^+), the implied true SDFs ($\hat{H}_t - \hat{\lambda} Y_t$ and $[\hat{H}_t^+ - \hat{\lambda}^+ Y_t]^+$), and the adjustment portfolios ($\hat{\lambda} Y_t$ and $\hat{H}_t^+ - [\hat{H}_t^+ - \hat{\lambda}^+ Y_t]^+$) for all the models under the first and second HJDs. Panel A of Table 3 shows that the correlations of the estimated SDFs among most models are low under both the first and second HJDs. The only exception is the two investment-based models, whose correlation coefficient is 0.90 (0.89) under the first (second) HJD. There is not a uniform relation between the correlations under the first and second HJDs: For certain models the correlations are higher under the first HJD, but for other models, the opposite is true. These results are not surprising given that most models focus on different economic factors to explain asset prices. The implied true SDFs have to correctly price all the test assets by construction. As a result, their correlations are much higher than those of the estimated SDFs as shown in Panel B of Table 3. Moreover, because the true SDFs under the second HJD have to be arbitrage-free, their correlations are uniformly larger than that of the true SDFs under the first HJD. Panel C of Table 3 reports the correlations of the adjustment portfolios for all the models under the first and second HJDs. The low correlations of the estimated SDFs and the high correlations of the true SDFs and the adjustment portfolios of the consumption-based models suggest that these models fail to capture some common components that are important for pricing the Fama and French portfolios.

To have a better understanding of the missing components in these models, in Table 4 we regress the adjustment portfolios under the first and second HJDs on some well-known economic and stock market factors. The economic factors include gross domestic product growth (GDP), industrial production (IP) growth, credit spread (yield difference between BAA and AAA corporate bonds), term spread (yield difference between ten-year and one-year government bonds), quarterly risk-free rate, market risk premium, and market index volatility. The stock market factors include the six Fama and French size and BM benchmark portfolios, which help to better understand which dimension of the size and BM effects the models fail to capture. We report the ordinary least squares regression coefficients and the adjusted R^2 s in Table 4, where bold entries represent coefficients that are significant at 5% level.

Panels A and B of Table 4 report the regression results with the economic and stock market factors, respectively. Panel A shows that under both the first and second HJDs, term spread and risk-free rate are significant for all models, suggesting that all models need to better capture pricing information contained in the two factors to be admissible. Panel B shows that under both the first and second HJDs, the big-low (BL), small-low (SL), and small-high (SH) portfolios are significant for all models and that the big-high (BH) portfolio is significant for all models except LVX1 and LVX2. The R^2 s in Panel B are generally higher than that in Panel A, although the R^2 s are low in

Table 4

Regressions of the adjustment portfolios under the first and second Hansen-Jagannathan distances on economic and stock market factors.

This table provides regression results of the adjustment portfolios under the first and second HJDs on economic and stock market factors using the risk free-asset and the Fama and French 25 Size/Book-to-Market portfolios from Q2.1952 to Q4.2000. Panels A and B contain regression results based on economic and stock market factors, respectively. In Panel A, GDP is the quarterly growth rate of log gross domestic product growth, IP is the quarterly growth rate of industrial production, Credit is the quarterly yield spread between BAA corporate bond and AAA corporate bond, Term is the quarterly yield spread between ten-year T-bond and one-year T-bond, Vol is the quarterly volatility of the market return index, Rf is the quarterly risk-free rate, and Mkt is the quarterly excess return of the market index. In Panel B, BL, BM, BH, SL, SM, and SH represent the big/low (BL), big/median (BM), big/high (BH), small/low (SL), small/median (SM), and small/high (SH) Fama-French benchmark portfolios, respectively. Bold entries represent ordinary least squares regression coefficients that are significant at the 5% level.

| | Using δ | | | | | | | Using δ^+ | | | | | | |
|---|----------------|---------------|---------------|---------------|---------------|---------------|---------------|------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | LL | LV | SV | LVX1 | LVX2 | YOGO | FF3 | LL | LV | SV | LVX1 | LVX2 | YOGO | FF3 |
| Panel A: Regression results based on economic factors | | | | | | | | | | | | | | |
| GDP | 6.19 | 4.29 | 7.33 | 2.94 | 3.29 | 6.10 | 6.59 | 5.68 | 5.26 | 7.06 | 4.55 | 4.68 | 5.75 | 6.32 |
| IP | -3.80 | -3.12 | -3.86 | -2.09 | -2.14 | -3.52 | -3.98 | -3.15 | -3.73 | -3.13 | -2.60 | -2.53 | -3.16 | -3.56 |
| Credit | -4.84 | -0.16 | -21.71 | -13.16 | 1.45 | -25.66 | -23.00 | -10.07 | -21.23 | -24.58 | -24.77 | -25.32 | -35.27 | -32.11 |
| Term | 13.37 | 13.78 | 19.85 | 16.37 | 10.09 | 20.04 | 13.46 | 16.83 | 22.20 | 22.11 | 21.31 | 20.63 | 23.51 | 16.58 |
| Vol | 0.04 | -0.58 | -0.32 | -0.92 | -1.20 | -0.39 | -1.01 | -0.16 | -0.66 | -0.41 | -0.87 | -0.97 | -0.38 | -1.08 |
| Rf | -13.82 | -15.10 | -17.26 | -16.31 | -12.85 | -17.16 | -10.43 | -16.16 | -20.87 | -18.42 | -19.44 | -18.89 | -19.23 | -12.11 |
| Mkt | -0.05 | 0.05 | -0.24 | 0.29 | 0.26 | -0.15 | -0.21 | 0.13 | 0.33 | -0.16 | 0.60 | 0.52 | -0.09 | -0.15 |
| Adjusted R ² | 1.91% | 2.80% | 4.01% | 5.17% | 5.18% | 3.43% | 2.30% | 1.98% | 6.03% | 4.46% | 6.79% | 6.32% | 4.71% | 3.22% |
| Panel B: Regression results based on Fama-French six benchmark portfolios | | | | | | | | | | | | | | |
| BL | 3.41 | 4.12 | 3.16 | 4.64 | 4.20 | 2.82 | 5.34 | 3.77 | 4.44 | 3.18 | 4.50 | 4.66 | 2.97 | 5.46 |
| BM | 1.74 | 1.19 | 0.63 | 1.92 | 3.35 | 0.35 | -1.24 | 1.11 | 1.15 | 0.49 | 1.24 | 0.66 | -0.04 | -1.43 |
| BH | -4.96 | -4.23 | -5.54 | -1.23 | -2.38 | -3.49 | -6.25 | -4.40 | -3.62 | -5.05 | -2.72 | -2.96 | -3.34 | -6.24 |
| SL | -5.96 | -6.24 | -6.67 | -4.31 | -2.94 | -6.72 | -4.53 | -6.47 | -6.56 | -6.81 | -6.01 | -6.15 | -6.86 | -4.64 |
| SM | 1.37 | 1.23 | 1.91 | -4.22 | -5.28 | 1.19 | 2.02 | 1.81 | 1.70 | 1.91 | 0.70 | 0.88 | 1.67 | 2.30 |
| SH | 6.82 | 7.08 | 7.72 | 6.30 | 6.11 | 7.42 | 4.72 | 7.15 | 6.94 | 7.60 | 6.32 | 6.71 | 7.12 | 4.59 |
| Adjusted R ² | 19.44% | 24.47% | 22.92% | 23.14% | 20.68% | 24.46% | 11.84% | 22.42% | 27.17% | 21.72% | 23.83% | 24.31% | 22.59% | 11.05% |

both panels (the highest adjusted R^2 is less than 30%). This suggests that all the factors we consider still fail to adequately capture the missing components of the models, which we leave for future research.

6. Conclusion

In this paper, we develop a systematic approach for evaluating asset pricing models based on the second HJD, which unlike the first HJD, explicitly requires a good asset pricing model to be arbitrage free. We develop both a specification test and a sequence of model selection procedures in the spirit of Vuong (1989) for non-nested, overlapping, and nested models based on the second HJD. Compared with existing methods, our tests are more powerful in detecting misspecified models that have small pricing errors but are not arbitrage-free and in differentiating the relative performances of models that have similar pricing errors of a given set of test assets. Simulation studies show that our tests have reasonably good finite sample performance for typical sample sizes considered in the literature. Using the Fama and French size and book-to-market portfolios, we reach dramatically different conclusions on model performances using our approach and existing methods.

Appendix A

In this appendix, we provide the assumptions and detailed mathematical proofs of all the results in the paper.

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

$$\theta_0 = (\gamma_0, \lambda_0) \equiv \arg \min_{\gamma} \max_{\lambda} \mathbb{E} \phi(\gamma, \lambda),$$

which is an interior point of the parameter space of $\theta = (\gamma, \lambda)$.

Assumption A.2. $\mathbb{E} \|Y\|^2 < \infty$, $\mathbb{E} \|X\|^2 < \infty$, and $\mathbb{E} [\max_{\|\gamma - \gamma_0\| < c} H(\gamma)^2] < \infty$ for some $C > 0$.

Assumption A.3. The SDF model $H(\gamma)$ is twice continuously differentiable with respect to γ .

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

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

$$\frac{\partial \phi}{\partial \theta} = \begin{pmatrix} \frac{\partial \phi}{\partial \gamma} \\ \frac{\partial \phi}{\partial \lambda} \end{pmatrix}$$

form a Donsker class for (γ, λ) in a neighborhood of (γ_0, λ_0) .

Assumption A.6. The time series $(Y_t, X_{t-1}, H_t(\gamma))$ are stationary and ergodic.

Assumption A.7. The matrix Γ defined by

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

is nonsingular, with a positive definite Γ_{22} and a negative definite $[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]$. The second derivatives are well defined except on a set with zero probability.

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, which is needed for the existence of the asymptotic covariance matrix of the second HJD 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 second HJD is not too big so that differentiability in quadratic mean holds. 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 $\varsigma > 0$ and an T_0 such that, for all $T > T_0$

$$P \left\{ \sup_{\{f\}} |\sqrt{T}(\mathbb{E}_T - \mathbb{E})f_1 - \sqrt{T}(\mathbb{E}_T - \mathbb{E})f_2| > \eta \right\} < \varepsilon. \tag{50}$$

$[\varsigma]$ means that the supremum runs over all pairs of functions f_1 and f_2 in \mathfrak{F} that are less than ς 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 need to transform the original price or payoff series in order to satisfy Assumption A.6. For example, although stock prices generally are not stationary and ergodic, stock returns generally are. Assumption A.7 ensures that the optimization problem $\min_{\gamma} \max_{\lambda} \mathbb{E} \phi(\gamma, \lambda)$ is well defined.

Lemma 1. Suppose Assumptions A.1 to A.7 hold. Then the following local asymptotic quadratic representation holds for $\mathbb{E}_T \phi(\theta)$ around θ_0 :

$$\mathbb{E}_T \phi(\theta) = \mathbb{E} \phi(\theta_0) + (\mathbb{E}_T - \mathbb{E})\phi(\theta_0) + A'(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \Gamma (\theta - \theta_0) + o_p(\|\theta - \theta_0\|^2) + o_p(\|\theta - \theta_0\| T^{-1/2}),$$

where $A = (\mathbb{E}_T - \mathbb{E})\partial \phi(\theta_0) / \partial \theta$ and $\Gamma = \mathbb{E} \partial^2 \phi(\theta_0) / \partial \theta \partial \theta'$.

Proof. We decompose $\mathbb{E}_T \phi(\theta)$ into a deterministic term, $\mathbb{E} \phi(\theta_0)$, and a centered random term, $(\mathbb{E}_T - \mathbb{E})\phi(\theta)$,

$$\mathbb{E}_T \phi(\theta) = \mathbb{E} \phi(\theta) + (\mathbb{E}_T - \mathbb{E})\phi(\theta). \tag{51}$$

The LAQ representation of $\mathbb{E}_T \phi(\theta)$ is based on a second-order Taylor approximation of $\mathbb{E}_T \phi(\theta)$ and a first-order approximation of $(\mathbb{E}_T - \mathbb{E})\phi(\theta)$. The random component is

centered and in general one order smaller than the deterministic term.

We first consider the second-order approximation of $\mathbb{E}\phi(\theta)$. Because $H(\gamma)$ is twice differentiable in γ , $\phi(\theta)$ has the continuous first-order derivatives

$$\frac{\partial\phi(\theta)}{\partial\theta} = \begin{pmatrix} \frac{\partial\phi(\theta)}{\partial\gamma} \\ \frac{\partial\phi(\theta)}{\partial\lambda} \end{pmatrix} = \begin{pmatrix} 2H(\gamma)\frac{\partial H(\gamma)}{\partial\gamma} - 2[H(\gamma) - \lambda'Y] + \frac{\partial H(\gamma)}{\partial\gamma} \\ 2[H(\gamma) - \lambda'Y] + Y - 2X \end{pmatrix}. \tag{52}$$

From Assumption A.2, we have

$$\frac{\partial}{\partial\theta}\mathbb{E}\phi(\theta) = \mathbb{E}\frac{\partial\phi(\theta)}{\partial\theta}. \tag{53}$$

We now demonstrate that although the first-order derivatives of ϕ in Eq. (52) are not differentiable everywhere, they are in fact differentiable in quadratic mean as in Pollard (1982, p. 920). A random function $f(\alpha)$ is differentiable in quadratic mean with respect to α at α_0 if there exists a random vector A such that

$$f(\alpha) = f(\alpha_0) + A'(\alpha - \alpha_0) + \|\alpha - \alpha_0\|R, \tag{54}$$

where

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

Intuitively, this means that $f(\alpha_0) + A'(\alpha - \alpha_0)$ is a good approximation to $f(\alpha)$ around α_0 on average.

To demonstrate that Eq. (52) is differentiable in quadratic mean, define the remainder term r by the following equation:

$$\frac{\partial\phi(\theta)}{\partial\theta} = \frac{\partial\phi(\theta_0)}{\partial\theta} + \frac{\partial^2\phi(\theta_0)}{\partial\theta\partial\theta'}(\theta - \theta_0) + \|\theta - \theta_0\|r. \tag{55}$$

Although the second derivatives involved might not exist everywhere, the set of points for which they are not defined has probability zero due to Assumption A.4. Hence as a function in the Hilbert space $L^2(P)$, the remainder term r is well defined. The remainder term r can be shown to be dominated by a function in $L^2(P)$ and $r \rightarrow 0$ almost surely as $\theta \rightarrow \theta_0$. 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-order derivatives of ϕ . Because $L^2(P)$ convergence implies $L^1(P)$ convergence, the quadratic mean differentiability for Eq. (52) implies that $\mathbb{E}\phi(\theta)$ has traditional second derivatives given by $\mathbb{E}\partial^2\phi(\theta)/\partial\theta\partial\theta'$. We emphasize again that the second-order derivatives inside the expectation operator is well defined except on a set of zero probability.

It follows that the deterministic term $\mathbb{E}\phi(\theta)$ has the quadratic approximation

$$\mathbb{E}\phi(\theta) = \mathbb{E}\phi(\theta_0) + \frac{1}{2}(\theta - \theta_0)' \cdot \mathbb{E}\frac{\partial^2\phi(\theta_0)}{\partial\theta\partial\theta'} \cdot (\theta - \theta_0) + o(\|(\theta, \lambda) - (\theta_0, \lambda_0)\|^2). \tag{56}$$

Note that $\mathbb{E}\partial\phi(\theta_0)/\partial\theta = 0$, because $\theta_0 = (\gamma_0, \lambda_0)$ solves the population optimization problem $\min_{\gamma} \max_{\lambda} \mathbb{E}\phi(\theta)$. Therefore, compared with traditional Taylor expansions, the key point here is to justify that we can still use the second-order derivatives of ϕ (which are not defined everywhere) to obtain an approximation to $\mathbb{E}\phi(\theta)$.

Next we consider a first-order approximation of $(\mathbb{E}_T - \mathbb{E})\phi(\theta)$. The first-order differentiability of $\phi(\theta)$ with respect to θ and Assumption A.5 on the first-order derivatives guarantee the stochastic differentiability of the empirical process [see Pollard, 1982, p. 921, Eq. (4)]

$$\begin{aligned} (\mathbb{E}_T - \mathbb{E})\phi(\theta) &= (\mathbb{E}_T - \mathbb{E})\phi(\theta_0) + (\mathbb{E}_T - \mathbb{E})\frac{\partial\phi(\theta_0)}{\partial\theta'} \\ &\quad \cdot (\theta - \theta_0) + o_p(\|\theta - \theta_0\|T^{-1/2}). \end{aligned} \tag{57}$$

Combining Eq. (A.3) and (57), we obtain the LAQ representation for $\mathbb{E}_T\phi(\theta)$. □

Based on the LAQ representation of $\mathbb{E}_T\phi(\theta)$ in Lemma 1, we solve the minimax problem $\min_{\gamma} \max_{\lambda} \mathbb{E}_T\phi(\theta)$ to obtain the asymptotic representation of the second HJD at $\hat{\theta} = (\hat{\gamma}, \hat{\lambda}) \equiv \arg \min_{\gamma} \max_{\lambda} \mathbb{E}_T\phi(\theta)$. We first make the following additional assumptions.

Assumption A.8. The estimator $\hat{\theta} = (\hat{\gamma}, \hat{\lambda}) \equiv \arg \min_{\gamma} \max_{\lambda} \mathbb{E}_T\phi(\theta)$ for θ_0 is consistent.

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

$$\sqrt{T}(\mathbb{E}_T - \mathbb{E}) \begin{pmatrix} \frac{\partial\phi(\theta_0)}{\partial\gamma} \\ \frac{\partial\phi(\theta_0)}{\partial\lambda} \end{pmatrix} \rightsquigarrow N(0, A),$$

where

$$A = \mathbb{E} \left[\begin{pmatrix} \frac{\partial\phi(\theta_0)}{\partial\gamma} \\ \frac{\partial\phi(\theta_0)}{\partial\lambda} \end{pmatrix} \begin{pmatrix} \frac{\partial\phi(\theta_0)}{\partial\gamma} \\ \frac{\partial\phi(\theta_0)}{\partial\lambda} \end{pmatrix}' \right].$$

As suggested by HHL (1995), the consistency condition in Assumption A.8 can be replaced by more primitive assumptions.

Lemma 2. Suppose Assumptions A.1 to A.9 hold. Then we have the following asymptotic representation of the second HJD at estimated model parameter $\hat{\theta}$:

$$\begin{aligned} [\hat{\delta}_T^+]^2 &= \min_{\gamma} \max_{\lambda} \mathbb{E}_T\phi(\theta) = \mathbb{E}\phi(\theta_0) + (\mathbb{E}_T - \mathbb{E})\phi(\theta_0) \\ &\quad - \frac{1}{2}A'\Gamma^{-1}A + o_p(T^{-1}). \end{aligned}$$

Moreover, the optimizer $\hat{\theta} = (\hat{\gamma}, \hat{\lambda}) \equiv \arg \min_{\gamma} \max_{\lambda} \mathbb{E}_T\phi(\theta)$ equals

$$\hat{\theta} = \theta_0 - \Gamma^{-1}A + o_p(T^{-1/2}),$$

where A and Γ are defined in Lemma 1.

Proof. Denote $A_{\gamma} = (\mathbb{E}_T - \mathbb{E})\partial\phi(\theta_0)/\partial\gamma$, $A_{\lambda} = (\mathbb{E}_T - \mathbb{E})\partial\phi(\theta_0)/\partial\lambda$, $U = (\gamma - \gamma_0)$, and $V = (\lambda - \lambda_0)$. Based on the LAQ

representation in Lemma 1, we have

$$\begin{aligned} \mathbb{E}_T \phi(\theta) &= \mathbb{E} \phi(\theta_0) + (\mathbb{E}_T - \mathbb{E}) \phi(\theta_0) \\ &+ \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix}' \begin{pmatrix} U \\ V \end{pmatrix} + \frac{1}{2} \begin{pmatrix} U \\ V \end{pmatrix}' \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \begin{pmatrix} U \\ V \end{pmatrix} \\ &+ o(\|(\gamma, \lambda) - (\gamma_0, \lambda_0)\|^2) + o_p(\|(\gamma, \lambda) - (\gamma_0, \lambda_0)\| T^{-1/2}) \\ &= \mathbb{E} \phi(\theta_0) + (\mathbb{E}_T - \mathbb{E}) \phi(\theta_0) \\ &+ \underbrace{A_\gamma' U + \frac{1}{2} U' \Gamma_{11} U + [A_\lambda + \Gamma_{21} U]' V + \frac{1}{2} V' \Gamma_{22} V}_{\text{quadratic in } V} \\ &+ o(\|(\gamma, \lambda) - (\gamma_0, \lambda_0)\|^2) + o_p(\|(\gamma, \lambda) - (\gamma_0, \lambda_0)\| T^{-1/2}). \end{aligned}$$

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

$$\hat{V} = (\hat{\lambda} - \lambda_0) = -\Gamma_{22}^{-1} [A_\lambda + \Gamma_{21} U] + o_p(T^{-1/2}), \tag{58}$$

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

$$\begin{aligned} A_\gamma' U + \frac{1}{2} U' \Gamma_{11} U - \frac{1}{2} [A_\lambda + \Gamma_{21} U]' \Gamma_{22}^{-1} [A_\lambda + \Gamma_{21} U] \\ = A_\gamma' U + \frac{1}{2} U' \Gamma_{11} U - \frac{1}{2} A_\lambda' \Gamma_{22}^{-1} A_\lambda - A_\lambda' \Gamma_{22}^{-1} \Gamma_{21} U - \frac{1}{2} U' [\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}] U + o_p(T^{-1}) \\ = \underbrace{-\frac{1}{2} A_\lambda' \Gamma_{22}^{-1} A_\lambda + [A_\gamma - A_\lambda' \Gamma_{22}^{-1} \Gamma_{21}] U + \frac{1}{2} U' [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}] U}_{\text{quadratic in } U} + o_p(T^{-1}). \end{aligned} \tag{59}$$

The above quadratic in U is minimized at

$$\hat{U} = (\hat{\gamma} - \gamma_0) = -[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} [A_\gamma - \Gamma_{12} \Gamma_{22}^{-1} A_\lambda] + o_p(T^{-1/2}), \tag{60}$$

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

$$\begin{aligned} -\frac{1}{2} A_\lambda' \Gamma_{22}^{-1} A_\lambda - \frac{1}{2} [A_\gamma - A_\lambda' \Gamma_{22}^{-1} \Gamma_{21}] [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \\ [A_\gamma - \Gamma_{12} \Gamma_{22}^{-1} A_\lambda] + o_p(T^{-1}). \end{aligned} \tag{61}$$

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

$$\begin{aligned} \min_\gamma \max_\lambda \mathbb{E}_T \phi(\theta) - \mathbb{E} \phi(\theta_0) &= (\mathbb{E}_T - \mathbb{E}) \phi(\theta_0) \\ &- \frac{1}{2} A_\lambda' \Gamma_{22}^{-1} A_\lambda - \frac{1}{2} A_\gamma' [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} A_\gamma \\ &+ A_\gamma' [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1} A_\lambda \\ &- \frac{1}{2} A_\lambda' \Gamma_{22}^{-1} \Gamma_{21} [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1} A_\lambda + o_p(T^{-1}) \\ &= (\mathbb{E}_T - \mathbb{E}) \phi(\theta_0, \lambda_0) - \frac{1}{2} \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix}' \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix} + o_p(T^{-1}), \end{aligned} \tag{62}$$

where

$$J_{11} = [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1},$$

$$J_{12} = J_{21} = -[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1},$$

and

$$J_{22} = \Gamma_{22}^{-1} + \Gamma_{22}^{-1} \Gamma_{21} [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1}. \quad \square$$

Theorem 1 (Specification test). *Suppose Assumptions A.1 to A.9 hold. Then under $\mathbb{H}_0: \delta^+ = 0$, $T[\hat{\delta}_T^+]$ has an asymptotic weighted χ^2 distribution and the weights are the eigenvalues of the matrix*

$$-\frac{1}{2} [\Gamma_{22}^{-1} - \Gamma_{22}^{-1} \Gamma_{21} [\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1}] A_\lambda,$$

where Γ_{12} , Γ_{21} , and Γ_{22} are defined in Assumption A.7 and $A_\lambda = \mathbb{E}[\partial \phi(\theta_0) / \partial \lambda \partial \phi(\theta_0) / \partial \lambda']$.

Proof. From Lemma 2, we know that under $\mathbb{H}_0: \delta^+ = 0$

$$T[\hat{\delta}_T^+]^2 = (\mathbb{E}_T - \mathbb{E}) \phi(\gamma_0, \lambda_0) - \frac{1}{2} \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix}' \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix} + o_p(T^{-1}). \tag{63}$$

First, we argue that $(\mathbb{E}_T - \mathbb{E}) \phi(\gamma_0, \lambda_0) = 0$ under $\mathbb{H}_0: \delta^+ = 0$. The fact that $\delta^+ = 0$ means that $\mathbb{E} \phi(\gamma_0, \lambda_0) = 0$. Because $\phi(\gamma_0, \lambda_0)$ is nonnegative, we must have $\phi(\gamma_0, \lambda_0) = 0$ almost everywhere for $\mathbb{E} \phi(\gamma_0, \lambda_0)$ to be zero. Consequently, $(\mathbb{E}_T - \mathbb{E}) \phi(\gamma_0, \lambda_0) = 0$.

Next we consider the second term in Eq. (63). Under $\mathbb{H}_0: \delta^+ = 0$, we must have $\lambda_0 = 0$. Consequently, based on (A.2) and some simple calculations, we have

$$A_\gamma = \frac{\partial \phi(\theta_0)}{\partial \gamma} = 0 \tag{64}$$

and

$$\Gamma_{11} = \mathbb{E} \frac{\partial^2 \phi(\theta_0)}{\partial \gamma \partial \gamma'} = 0. \tag{65}$$

Substituting these representations into Eq. (63), we obtain that under $\mathbb{H}_0: \delta^+ = 0$

$$\begin{aligned} T[\hat{\delta}_T^+]^2 &= -\frac{1}{2} [\sqrt{T}(\mathbb{E}_T - \mathbb{E}) A_\lambda]' J_{22} [\sqrt{T}(\mathbb{E}_T - \mathbb{E}) A_\lambda] + o_p(1) \\ &= -\frac{1}{2} [\sqrt{T}(\mathbb{E}_T - \mathbb{E}) A_\lambda]' [\Gamma_{22}^{-1} - \Gamma_{22}^{-1} \Gamma_{21} [\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \\ &\quad \times \Gamma_{12} \Gamma_{22}^{-1}] [\sqrt{T}(\mathbb{E}_T - \mathbb{E}) A_\lambda] + o_p(1). \end{aligned} \tag{66}$$

Assumption A.9 implies that $\sqrt{T}(\mathbb{E}_T - \mathbb{E}) A_\lambda \rightarrow W$, where W is a multivariate normal random vector with mean zero and covariance matrix A_λ . It follows immediately that under $\mathbb{H}_0: \delta^+ = 0$, the asymptotic distribution of $T[\hat{\delta}_T^+]^2$ can be represented as $W' \Xi W = Z' A_\lambda^{1/2} \Xi A_\lambda^{1/2} Z$, where $\Xi = -\frac{1}{2} [\Gamma_{22}^{-1} - \Gamma_{22}^{-1} \Gamma_{21} [\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1}]$, Z is a vector of standard multivariate normal random vector, and $W = A_\lambda^{1/2} Z$. It can be easily shown that $A_\lambda^{1/2} \Xi A_\lambda^{1/2}$ has the same eigenvalues as ΞA_λ . Therefore, $T[\hat{\delta}_T^+]^2$ has an asymptotic distribution of weighted χ^2 , and the weights are the eigenvalues of the matrix $\Xi A_\lambda = -\frac{1}{2} [\Gamma_{22}^{-1} - \Gamma_{22}^{-1} \Gamma_{21} [\Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1}] A_\lambda$. \square

Proposition 1 (Parameter estimation). *Suppose Assumptions A.1 to A.9 hold. Then*

the estimator of model parameters, $\hat{\gamma}$, has the following asymptotic distribution:

$$\sqrt{T}(\hat{\gamma} - \gamma_0) \rightsquigarrow N(0, (J_{11} \ J_{12}) A (J_{11} \ J_{12})')$$

the estimator of Lagrangian multipliers, $\hat{\lambda}$, has the following asymptotic distribution:

$$\sqrt{T}(\hat{\lambda} - \lambda_0) \rightsquigarrow N(0, (J_{21} \ J_{22}) A (J_{21} \ J_{22})'),$$

where $\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} = (\Gamma_{21}^{-1} \ \Gamma_{22}^{-1})^{-1}$ and Γ_{ij} 's ($ij=1,2$) are defined in Assumption A.7.

Proof. From the proof of Lemma 2, we have

$$\begin{aligned} \hat{\gamma} - \gamma_0 &= -[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} [A_\gamma - \Gamma_{12} \Gamma_{22}^{-1} A_\lambda] + o_p(T^{-1/2}) \\ &= -[\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} (I \quad -\Gamma_{12} \Gamma_{22}^{-1}) \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix} \\ &+ o_p(T^{-1/2}) = -(J_{11} \quad J_{12}) \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix} + o_p(T^{-1/2}). \end{aligned}$$

From Assumption A.9, we know immediately that $\sqrt{T}(\hat{\gamma} - \gamma_0)$ is asymptotically normally distributed with mean zero and covariance matrix $(J_{11} \quad J_{12})A(J_{11} \quad J_{12})'$.

Similarly, we have the following asymptotic representation for $\hat{\lambda} - \lambda_0$,

$$\begin{aligned} \hat{\lambda} - \lambda_0 &= -\Gamma_{22}^{-1} [A_\lambda + \Gamma_{21} (\hat{\gamma} - \gamma_0)] + o_p(T^{-1/2}) \\ &= -\Gamma_{22}^{-1} (A_\lambda - \Gamma_{21} [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} [A_\gamma - \Gamma_{12} \Gamma_{22}^{-1} A_\lambda]) \\ &\quad + o_p(T^{-1/2}) = -\Gamma_{22}^{-1} (A_\lambda - \Gamma_{21} [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} A_\gamma \\ &\quad + \Gamma_{21} [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1} A_\lambda) + o_p(T^{-1/2}) \\ &= \Gamma_{22}^{-1} [\Gamma_{21} [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} A_\gamma - (I + \Gamma_{21} \\ &\quad \times [\Gamma_{11} - \Gamma_{12} \Gamma_{22}^{-1} \Gamma_{21}]^{-1} \Gamma_{12} \Gamma_{22}^{-1} A_\lambda)] + o_p(T^{-1/2}) \\ &= -(J_{21} \quad J_{22}) \begin{pmatrix} A_\gamma \\ A_\lambda \end{pmatrix} + o_p(T^{-1/2}) \end{aligned} \tag{67}$$

From Assumption A.9, we know immediately that $\sqrt{T}(\hat{\lambda} - \lambda_0)$ is asymptotically normally distributed with mean zero and covariance matrix $(J_{21} \quad J_{22})A(J_{21} \quad J_{22})'$. \square

Assumption A.10. A central limit theorem holds for the empirical process

$$\sqrt{T}(\mathbb{E}_T - \mathbb{E}) \begin{pmatrix} \frac{\partial \phi(\theta_{0\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{\partial \phi(\theta_{0\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{pmatrix} \rightsquigarrow N(0, \Psi),$$

where

$$\Psi = \begin{pmatrix} A_{\mathcal{F}} & A_{\mathcal{F}\mathcal{G}} \\ A_{\mathcal{G}\mathcal{F}} & A_{\mathcal{G}} \end{pmatrix} = \mathbb{E} \begin{pmatrix} \frac{\partial \phi(\theta_{0\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{\partial \phi(\theta_{0\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{pmatrix} \begin{pmatrix} \frac{\partial \phi(\theta_{0\mathcal{F}})}{\partial \theta_{\mathcal{F}}} \\ \frac{\partial \phi(\theta_{0\mathcal{G}})}{\partial \theta_{\mathcal{G}}} \end{pmatrix}'.$$

Theorem 2 (Model selection for non-nested models). Suppose models \mathcal{F} and \mathcal{G} are strictly non-nested and Assumptions A.1 to A.10 hold. Then

- under \mathbb{H}_0 : $\delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, $\sqrt{T}([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2) / \hat{\sigma}_T \rightsquigarrow N(0, 1)$;
- under $\mathbb{H}_{\mathcal{F}}$: $\delta_{\mathcal{F}}^+ < \delta_{\mathcal{G}}^+$, $\sqrt{T}([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2) / \hat{\sigma}_T \rightarrow -\infty$; and
- under $\mathbb{H}_{\mathcal{G}}$: $\delta_{\mathcal{F}}^+ > \delta_{\mathcal{G}}^+$, $\sqrt{T}([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2) / \hat{\sigma}_T \rightarrow +\infty$,

where $[\hat{\sigma}_T]^2$ is an estimator of $\text{Var}[\phi(\theta_{0\mathcal{F}}) - \phi(\theta_{0\mathcal{G}})]$ and equals

$$[\hat{\sigma}_T]^2 = \mathbb{E}_T[\phi(\hat{\theta}_{\mathcal{F}}) - \phi(\hat{\theta}_{\mathcal{G}})]^2 - (\mathbb{E}_T[\phi(\hat{\theta}_{\mathcal{F}}) - \phi(\hat{\theta}_{\mathcal{G}})])^2.$$

Proof. From Lemma 2, we have

$$[\hat{\delta}_{\mathcal{F}}^+]^2 = \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{G}}} \mathbb{E}_T \phi(\theta_{\mathcal{F}}) = \mathbb{E} \phi(\theta_{\mathcal{F}0}) + (\mathbb{E}_T - \mathbb{E}) \phi(\theta_{\mathcal{F}0}) - \frac{1}{2} A'_{\mathcal{F}} J_{\mathcal{F}} A_{\mathcal{F}} + o_p(T^{-1}), \tag{68}$$

$$[\hat{\delta}_{\mathcal{G}}^+]^2 = \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} \mathbb{E}_T \phi(\theta_{\mathcal{G}}) = \mathbb{E} \phi(\theta_{\mathcal{G}0}) + (\mathbb{E}_T - \mathbb{E}) \phi(\theta_{\mathcal{G}0}) - \frac{1}{2} A'_{\mathcal{G}} J_{\mathcal{G}} A_{\mathcal{G}} + o_p(T^{-1}), \tag{69}$$

and

$$[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2 = \mathbb{E} \phi(\theta_{\mathcal{F}0}) - \mathbb{E} \phi(\theta_{\mathcal{G}0}) + (\mathbb{E}_T - \mathbb{E}) [\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})] - \frac{1}{2} A'_{\mathcal{F}} J_{\mathcal{F}} A_{\mathcal{F}} + \frac{1}{2} A'_{\mathcal{G}} J_{\mathcal{G}} A_{\mathcal{G}} + o_p(T^{-1}). \tag{70}$$

Because two non-nested models can never be observationally equivalent, the empirical process $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})]$ does not degenerate. Under \mathbb{H}_0 : $\delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, $\mathbb{E} \phi(\theta_{\mathcal{F}0}) - \mathbb{E} \phi(\theta_{\mathcal{G}0}) = 0$ and the second-order terms in Eq. (A.11) are $O_p(T^{-1})$ and negligible. As a result, $\sqrt{T}([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2)$ is asymptotically equivalent to $\sqrt{T}(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})]$, which has an asymptotic normal distribution with mean zero and variance $\sigma^2 = \text{Var}[\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})]$. This proves the first assertion. The assertion for $\mathbb{H}_{\mathcal{F}}$ and $\mathbb{H}_{\mathcal{G}}$ follows immediately by observing that $\mathbb{E} \phi(\theta_{\mathcal{F}0}) - \mathbb{E} \phi(\theta_{\mathcal{G}0})$ is then the dominating term in Eq. (A.11). \square

Theorem 3 (Model selection for nested models). Suppose that model \mathcal{F} is nested by model \mathcal{G} and that Assumptions A.1 to A.10 hold. Then

under \mathbb{H}_0 : $\delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, $T([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2)$ has an asymptotic weighted χ^2 distribution and the weights are the eigenvalues of the matrix

$$\frac{1}{2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} & -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{G}} \\ \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}\mathcal{F}} & \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} \end{pmatrix};$$

and

under $\mathbb{H}_{\mathcal{G}}$: $\delta_{\mathcal{F}}^+ > \delta_{\mathcal{G}}^+$, $T([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2) \rightarrow +\infty$.

Proof. From the proof of Theorem 2, we have

$$[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2 = \mathbb{E} \phi(\theta_{\mathcal{F}0}) - \mathbb{E} \phi(\theta_{\mathcal{G}0}) + (\mathbb{E}_T - \mathbb{E}) [\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})] - \frac{1}{2} A'_{\mathcal{F}} J_{\mathcal{F}} A_{\mathcal{F}} + \frac{1}{2} A'_{\mathcal{G}} J_{\mathcal{G}} A_{\mathcal{G}} + o_p(T^{-1}).$$

Under \mathbb{H}_0 : $\delta_{\mathcal{F}}^+ = \delta_{\mathcal{G}}^+$, because \mathcal{F} is nested by \mathcal{G} , we must have $\phi(\theta_{\mathcal{F}0}) = \phi(\theta_{\mathcal{G}0})$ with probability one. Consequently, $\mathbb{E} \phi(\theta_{\mathcal{F}0}) - \mathbb{E} \phi(\theta_{\mathcal{G}0}) = 0$ and $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})] = 0$. Therefore, we obtain

$$\begin{aligned} [\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2 &= -\frac{1}{2} A'_{\mathcal{F}} J_{\mathcal{F}} A_{\mathcal{F}} + \frac{1}{2} A'_{\mathcal{G}} J_{\mathcal{G}} A_{\mathcal{G}} + o_p(T^{-1}) \\ &= \frac{1}{2} \begin{pmatrix} A_{\mathcal{F}} \\ A_{\mathcal{G}} \end{pmatrix}' \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \begin{pmatrix} A_{\mathcal{F}} \\ A_{\mathcal{G}} \end{pmatrix} + o_p(T^{-1}). \end{aligned}$$

From Assumption A.10, we have $\sqrt{T} \begin{pmatrix} A_{\mathcal{F}} \\ A_{\mathcal{G}} \end{pmatrix} \rightarrow W$, where W is a multivariate normal random vector with mean zero and covariance matrix Ψ .

Rewrite W as $W = \Psi^{1/2} Z$, where Z is a vector of standard multivariate normal random vector. Then the limiting distribution of $T([\delta_{\mathcal{F}}^+]^2 - [\delta_{\mathcal{G}}^+]^2)$ can be represented as

$$\frac{1}{2} Z' \left[\Psi^{1/2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \Psi^{1/2} \right] Z.$$

Because an orthogonal transformation of a standard normal random vector is still a standard normal random vector, the theorem is proved if we can show that the eigenvalues of

$$\left[\Psi^{1/2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \Psi^{1/2} \right]$$

are the same as that of

$$\begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} & -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{G}} \\ \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}\mathcal{F}} & \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} \end{pmatrix}.$$

It is easy to see that matrix

$$\Psi^{1/2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \Psi^{1/2}$$

has the same eigenvalues as matrix

$$\begin{aligned} \left[\begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \Psi^{1/2} \right] \Psi^{1/2} &= \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \Psi \\ &= \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} & -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{G}} \\ \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}\mathcal{F}} & \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} \end{pmatrix}. \end{aligned}$$

Therefore, it follows that, under \mathbb{H}_0 , $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2)$ is asymptotically distributed as a weighted χ^2 distribution.

Under $\mathbb{H}_{\mathcal{G}}$, the first two terms in the representation of $[\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2$ are the leading terms, and hence the conclusion of the theorem under $\mathbb{H}_{\mathcal{G}}$ follows. This completes the proof of Theorem 3. \square

Theorem 4 (Model selection for overlapping models). *Suppose models \mathcal{F} and \mathcal{G} are overlapping models and Assumptions A.1 and A.10 hold. Then*

under \mathbb{H}_0^ : $\phi(\theta_{\mathcal{F}0}) = \phi(\theta_{\mathcal{G}0})$ with probability one, $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2)$ has an asymptotic weighted χ^2 distribution, and the weights are the eigenvalues of the matrix*

$$\frac{1}{2} \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}} & -\Gamma_{\mathcal{F}}^{-1} A_{\mathcal{F}\mathcal{G}} \\ \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}\mathcal{F}} & \Gamma_{\mathcal{G}}^{-1} A_{\mathcal{G}} \end{pmatrix};$$

and

under \mathbb{H}_A^ : $\phi(\theta_{\mathcal{F}0}) \neq \phi(\theta_{\mathcal{G}0})$ with positive probability, $T([\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2) \rightarrow \infty$ (either $+\infty$ or $-\infty$).*

Proof. From the proof of Theorem 3, we have

$$\begin{aligned} [\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2 &= \mathbb{E}\phi(\theta_{\mathcal{F}0}) - \mathbb{E}\phi(\theta_{\mathcal{G}0}) + (\mathbb{E}_T - \mathbb{E})[\phi(\theta_{\mathcal{F}0}) \\ &\quad - \phi(\theta_{\mathcal{G}0})] - \frac{1}{2} A_{\mathcal{F}}' J_{\mathcal{F}} A_{\mathcal{F}} + \frac{1}{2} A_{\mathcal{G}}' J_{\mathcal{G}} A_{\mathcal{G}} + o_p(T^{-1}). \end{aligned}$$

Under \mathbb{H}_0^* : $\phi(\theta_{\mathcal{F}0}) \neq \phi(\theta_{\mathcal{G}0})$ with positive probability, $\mathbb{E}\phi(\theta_{\mathcal{F}0}) - \mathbb{E}\phi(\theta_{\mathcal{G}0}) = 0$ and $(\mathbb{E}_T - \mathbb{E})[\phi(\theta_{\mathcal{F}0}) - \phi(\theta_{\mathcal{G}0})] = 0$. Therefore, we obtain

$$\begin{aligned} [\hat{\delta}_{\mathcal{F}}^+]^2 - [\hat{\delta}_{\mathcal{G}}^+]^2 &= -\frac{1}{2} A_{\mathcal{F}}' J_{\mathcal{F}} A_{\mathcal{F}} + \frac{1}{2} A_{\mathcal{G}}' J_{\mathcal{G}} A_{\mathcal{G}} + o_p(T^{-1}) \\ &= \frac{1}{2} \begin{pmatrix} A_{\mathcal{F}} \\ A_{\mathcal{G}} \end{pmatrix}' \begin{pmatrix} -\Gamma_{\mathcal{F}}^{-1} & \\ & \Gamma_{\mathcal{G}}^{-1} \end{pmatrix} \begin{pmatrix} A_{\mathcal{F}} \\ A_{\mathcal{G}} \end{pmatrix} + o_p(T^{-1}). \end{aligned}$$

The rest of the proof is similar to that of Theorem 3 and thus is omitted. \square

References

Agarwal, V., Naik, N., 2004. Risks and portfolio decisions involving hedge funds. *Review of Financial Studies* 17, 63–98.

Ben Dor, A., Jagannathan, R., 2002. Understanding mutual fund and hedge fund styles using return based style analysis. NBER Working Paper No. 9111, National Bureau of Economic Research, Cambridge, MA.

Campbell, J., Cochrane, J., 2000. Explaining the poor performance of consumption-based asset pricing models. *Journal of Finance* 55, 2863–2878.

Cochrane, J., 2005. *Asset Pricing*. Princeton University Press, Princeton, NJ.

Dittmar, R., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance* 57, 369–403.

Dudley, R., 1981. Donsker classes of functions. In: *Statistics and Related Topics*. North-Holland, Amsterdam, pp. 341–352.

Dybvig, P., Ingersoll, J., 1982. Mean-variance theory in complete markets. *Journal of Business* 55, 233–251.

Dybvig, P., Ross, S., 1985. Differential information and performance measurement using a security market line. *Journal of Finance* 40, 383–399.

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

Farnsworth, H., Ferson, W., Jackson, D., Todd, S., 2002. Performance evaluation with stochastic discount factors. *Journal of Business* 75, 473–504.

Ferson, W., Harvey, C.R., 1999. Conditioning variables and cross-section of stock returns. *Journal of Finance* 54, 1325–1360.

Ferson, W., Khang, K., 2002. Conditional performance measurement using portfolio weights: evidence for pension funds. *Journal of Financial Economics* 65, 249–282.

Ferson, W., Henry, T., Kisgen, D., 2006. Evaluating government bond fund performance with stochastic discount factors. *Review of Financial Studies* 19, 423–456.

Fung, W., Hsieh, D., 2001. The risk in hedge fund strategies: theory and evidence from trend followers. *Review of Financial Studies* 14, 313–341.

Glosten, L., Jagannathan, R., 1994. A contingent claim approach to performance evaluation. *Journal of Empirical Finance* 1, 133–160.

Grinblatt, M., Titman, S., 1989. Portfolio performance evaluation: old issues and new insights. *Review of Financial Studies* 2, 393–421.

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

Hansen, L.P., Heaton, J., Luttmer, E., 1995. Econometric evaluation of asset pricing models. *Review of Financial Studies* 8, 237–274.

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

Hansen, L.P., Jagannathan, R., 1997. Assessing specification errors in stochastic discount factor models. *Journal of Finance* 52, 557–590.

Hodrick, R., Zhang, X., 2001. Evaluating the specification errors of asset pricing models. *Journal of Financial Economics* 62, 327–376.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51, 3–53.

Jagannathan, R., Kubota, K., Takehara, H., 1998. Relationship between labor income risk and average return: empirical evidence from the Japanese stock market. *Journal of Business* 71, 319–348.

Kan, R., Zhou, G., 2002. Hansen-Jagannathan Distance: Geometry and Exact Distribution. Unpublished working paper. Washington University in St. Louis., St. Louis, MO.

Le Cam, L., 1986. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. *Journal of Political Economy* 109, 1238–1287.

Li, Q., Vassalou, M., Xing, Y., 2006. Sector investment growth rates and the cross-section of equity returns. *Journal of Business* 79, 1637–1665.

Lustig, H., Nieuwerburgh, S., 2004. Housing collateral, consumption insurance, and risk premia. *Journal of Finance* 60, 1167–1221.

- Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics* 96, 175–194.
- Merton, R., 1981. On market timing and investment performance I: an equilibrium theory of values for markets forecasts. *Journal of Business* 54, 363–406.
- Mitchell, M., Pulvino, T., 2001. Characteristics of risk and return in risk arbitrage. *Journal of Finance* 56, 2135–2175.
- Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
- Pollard, D., 1982. A central limit theorem for k -means clustering. *Annals of Probability* 10, 919–926.
- Ross, S., 2005. *Neoclassical Finance*. Princeton University Press, New Jersey.
- Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. *Review of Financial Studies* 19, 1–44.
- Vuong, Q., 1989. Likelihood ratio tests for model selection and non nested hypotheses. *Econometrica* 57, 307–333.
- Wang, Z., Zhang, X., 2005. The role of arbitrage in the empirical evaluation of asset pricing models. Unpublished working paper, Cornell University, Ithaca, NY.
- Yogo, M., 2006. A consumption-based explanation of expected stock returns. *Journal of Finance* 61, 539–580.