Gospodinov, Nikolay; Kan, Raymond; Robotti, Cesare

Working Paper

On the Hansen-Jagannathan distance with a no-arbitrage constraint

Working Paper, No. 2010-4

Provided in Cooperation with:
Federal Reserve Bank of Atlanta

Suggested Citation: Gospodinov, Nikolay; Kan, Raymond; Robotti, Cesare (2010) : On the Hansen-Jagannathan distance with a no-arbitrage constraint, Working Paper, No. 2010-4, Federal Reserve Bank of Atlanta, Atlanta, GA

This Version is available at:
https://hdl.handle.net/10419/70698

FEDERAL RESERVE BANK of ATLANTA

On the Hansen-Jagannathan Distance with a No-Arbitrage Constraint

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2010-4
March 2010

WORKING PAPER SERIES

FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

On the Hansen-Jagannathan Distance with a No-Arbitrage Constraint

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti

Working Paper 2010-4
March 2010

Abstract: We provide an in-depth analysis of the theoretical and statistical properties of the Hansen-Jagannathan (HJ) distance that incorporates a no-arbitrage constraint. We show that for stochastic discount factors (SDF) that are spanned by the returns on the test assets, testing the equality of HJ distances with no-arbitrage constraints is the same as testing the equality of HJ distances without no-arbitrage constraints. A discrepancy can exist only when at least one SDF is a function of factors that are poorly mimicked by the returns on the test assets. Under a joint normality assumption on the SDF and the returns, we derive explicit solutions for the HJ distance with a no-arbitrage constraint, the associated Lagrange multipliers, and the SDF parameters in the case of linear SDFs. This solution allows us to show that nontrivial differences between HJ distances with and without no-arbitrage constraints can arise only when the volatility of the unspanned component of an SDF is large and the Sharpe ratio of the tangency portfolio of the test assets is very high. Finally, we present the appropriate limiting theory for estimation, testing, and comparison of SDFs using the HJ distance with a no-arbitrage constraint.

JEL classification: G12, C12, C13

Key words: Hansen-Jagannathan distance, no-arbitrage constraint, stochastic discount factor, specification tests, model selection tests

---

The authors thank Esther Eiling, Wayne Ferson, Sergei Sarkissian, Jonathan Wright, Chu Zhang, Guofu Zhou, and seminar participants at Emory University and the University of Montreal for helpful discussions and comments. Kan gratefully acknowledges financial support from the National Bank Financial of Canada, the Social Sciences and Humanities Research Council of Canada, and the Center for Financial Innovation and Stability at the Federal Reserve Bank of Atlanta. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors' responsibility.

Please address questions regarding content to Nikolay Gospodinov, Associate Professor of Economics, Department of Economics, Concordia University, 1455 de Maisonneuve Boulevard, West Montreal, Quebec, Canada H3G 1M8, 514-848-2424, ext. 3935, nikolay.gospodinov@concordia.ca; Raymond M. Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6, 416-978-4291, kan@chass.utoronto.ca; or Cesare Robotti (corresponding author), Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, N.E., Atlanta, GA 30309-4470; 404-498-8543, cesare.robotti@atl.frb.org.

1. Introduction

Since all asset pricing models can be viewed as approximations of reality and are likely to be misspecified, researchers are often interested in evaluating and comparing their empirical performance. In order to perform these tasks, one has to take a stand on what measure of model misspecification to use. While there are many possible choices, Hansen and Jagannathan (1997, HJ hereafter) propose two interesting measures of model misspecification. The first one measures the distance between a proposed stochastic discount factor (SDF) and the set of admissible SDFs (i.e., the set of SDFs that price a given set of test assets correctly). The second one measures the distance between a proposed SDF and the set of nonnegative admissible SDFs. Since the first measure does not impose the nonnegativity constraint (no-arbitrage condition) on the set of admissible SDFs whereas the second one does, we refer to the first measure as the unconstrained HJ-distance and to the second one as the constrained HJ-distance.

While the unconstrained HJ-distance is analyzed and used in many studies (see, e.g., Bansal, Hsieh, and Viswanathan, 1993; Hansen, Heaton, and Luttmer, 1995; Jagannathan and Wang, 1996; Jagannathan, Kubota, and Takehara, 1998; Campbell and Cochrane, 2000; Lettau and Ludvigson, 2001; Hodrick and Zhang, 2001; Farnsworth, Ferson, Jackson, and Todd, 2002; Dittmar, 2002; Kan and Zhou, 2003; and Kan and Robotti, 2009, among others), the constrained HJ-distance is largely ignored in the literature. The short list of studies that have analyzed and used the constrained HJ-distance includes Hansen, Heaton, and Luttmer (1995), Bailey, Li, and Zhang (2004), Wang and Zhang (2005), Chen and Ludvigson (2009), Fletcher (2009), and Liu, Kuo, and Coakley (2009). We believe there are two reasons that contribute to the lack of popularity of the constrained HJ-distance. The first reason is that an explicit solution for the constrained HJ-distance is not readily available. The lack of an explicit solution hampers the theoretical analysis of the properties of the constrained HJ-distance and implies that researchers often have to solve a high-dimensional optimization problem in order to obtain an estimate of the constrained HJ-distance. The second reason is that the statistical inference theory for the sample constrained HJ-distance is not fully developed. Although Hansen, Heaton, and Luttmer (1995) derive some of the asymptotic theory for estimating and evaluating asset pricing models based on the constrained HJ-distance, the model selection tests are largely unavailable in the literature.

A recent paper by Li, Xu, and Zhang (2009, LXZ hereafter) aims at providing a more complete

econometric study of the constrained HJ-distance. Besides developing specification and model selection tests based on constrained HJ-distances, LXZ strongly advocate the use of the constrained HJ-distance in empirical work because they find that this metric is more powerful in detecting misspecified models, especially those that are not arbitrage free, and in differentiating between models that have similar pricing errors on a given set of test assets. Their recommendation, however, appears to be driven by empirical examples and is not based on an analysis of the theoretical and statistical properties of the unconstrained and constrained HJ-distances.[1]

Our paper has two main objectives. The first one is to better understand the merits and drawbacks of the constrained HJ-distance and the difference between this measure and its unconstrained counterpart. We point out that when the SDF is perfectly correlated with the returns on the test assets, the difference between the squared constrained and unconstrained HJ-distances is the same as the difference between the constrained and unconstrained Hansen-Jagannathan bounds (HJ-bounds, see Hansen and Jagannathan, 1991) constructed from the test assets. This suggests that the difference between the two HJ-distances is identical across all SDFs that are spanned by the returns. Therefore, for two spanned SDFs, testing the equality of unconstrained HJ-distances is the same as testing the equality of constrained HJ-distances. For the more general case in which the SDF is not spanned by the returns on the test assets, we derive an explicit solution of the constrained HJ-distance under the assumption that the SDF and the returns are jointly normally distributed. This allows us to show that nontrivial differences between the unconstrained and constrained HJ-distances can only arise when the volatility of the unspanned component of an SDF is large and the Sharpe ratio of the tangency portfolio of the test assets is very high. In addition, in the case of linear SDFs, we obtain analytical expressions of the SDF parameters that solve the constrained HJ-distance problem and use them to analyze the trade-off between the deterioration in the pricing ability of an SDF and its smaller probability of taking on negative values.

Our second objective is to provide an improved analysis of the sample constrained HJ-distance. We show that the sample constrained HJ-distance takes on the value of infinity with positive probability. As a result, the expectation of the sample constrained HJ-distance does not exist. We also show that the sample constrained HJ-distance takes on the value of infinity if and only if

[1]LXZ also provide some simulation evidence to support their conjectures, but these simulation experiments show that their model specification and selection tests based on the unconstrained and constrained HJ-distances have quite similar size and power properties (for a more detailed discussion on this, see Sections 5 and 6).

we can identify an in-sample arbitrage portfolio of the test assets. When an in-sample arbitrage portfolio is identified, it implies that all models will have a sample constrained HJ-distance of infinity, rendering it impossible to use the sample constrained HJ-distance for model comparison. Furthermore, we demonstrate that, contrary to LXZ's conjecture, the existing specification test based on the unconstrained HJ-distance is also applicable to the constrained HJ-distance. We also correct a common mistake in the literature (see, e.g., LXZ and Hodrick and Zhang, 2001) and show that when the asset pricing model is correctly specified, the vectors of sample Lagrange multipliers in the unconstrained and constrained HJ-distance problems are not always asymptotically normally distributed. Finally, we improve on the model selection tests proposed by LXZ along several dimensions. We show that for strictly non-nested models, LXZ's testing methodology is incomplete because it does not allow for the possibility that two strictly non-nested models are both correctly specified. In addition, unlike LXZ, our model comparison tests allow for autocorrelated errors and incorporate the proper null hypotheses. While the practice of not imposing the null hypotheses in constructing the test statistics can be justified based on asymptotic arguments, it produces the undesirable outcome of comparing test statistics that are positive by construction (as in the nested models case discussed in Section 5) to distributions that can take on negative values. In addition, incorporating the appropriate null hypotheses leads to simpler model comparison tests that require the estimation of far fewer parameters than the ones suggested by LXZ.

In light of our theoretical findings, we reexamine the empirical performance of the seven asset pricing models considered by LXZ. We find that LXZ's main conclusion that it is easier to differentiate between models when comparing them based on their sample constrained HJ-distances is heavily driven by the very high sample Sharpe ratio of the test assets that they use. When comparing their models on test assets that have a smaller sample Sharpe ratio, the model selection tests based on the unconstrained and constrained HJ-distances deliver very similar conclusions.

The rest of the paper is organized as follows. Section 2 presents a theoretical analysis of the unconstrained and constrained HJ-distances. Section 3 derives an analytical solution of the constrained HJ-distance under the normality assumption. Section 4 presents an econometrics analysis of the sample constrained HJ-distance. Section 5 develops our model selection tests based on sample constrained HJ-distances. Section 6 contains our empirical results. Some concluding remarks are provided in Section 7.

2. Unconstrained and constrained Hansen-Jagannathan distances

2.1. The setup

Following HJ, let $\mathcal{F}$ be the information that is observed at the date of the asset payoffs. Associated with $\mathcal{F}$ is the space L^2 of all random variables with finite second moments that are in the information set $\mathcal{F}$. This space is used as the collection of hypothetical claims that could be traded. However, for practical reasons, econometricians can only evaluate asset pricing models on a subspace of L^2. Let $\tilde{r} = [R_0,\ r']'$, where R_0 is the gross return on the risk-free asset, and r is a vector of excess returns (in excess of the risk-free rate) on N risky assets.[2] We assume that the payoff space used in an econometric analysis is given by the payoffs of portfolios of $\tilde{r}$:

$$\mathcal{P} \equiv \{w'\tilde{r} : w \in \Re^n\}, \tag{1}$$

where $n = N + 1$. In addition, we assume that $E[\tilde{r}\tilde{r}']$ is nonsingular so that none of the test assets is redundant.

An SDF is a random variable in L^2. We call m an admissible SDF if it prices the test assets correctly, i.e.,

$$E[\tilde{r}m] = q, \tag{2}$$

where $q = [1,\ 0_N']'$ and 0_N is an N-vector of zeros. Let $\mathcal{M}$ denote the set of all admissible SDFs. Although all SDFs in $\mathcal{M}$ can price the test assets correctly, some of them can take on negative values with positive probability and are not consistent with the absence of arbitrage opportunities on the space of hypothetical derivative claims. To eliminate these SDFs from consideration, HJ consider $\mathcal{M}^+$, which is the set of nonnegative admissible SDFs.

2.2. Pricing errors and Hansen-Jagannathan distances

Let y be a candidate stochastic discount factor. If y prices the n test assets correctly, then the vector of pricing errors, e, of the test assets is exactly zero:

$$e = E[\tilde{r}y] - q = 0_n. \tag{3}$$

[2]It can be readily shown that both the unconstrained and constrained HJ-distances and their SDF parameters are invariant to nonsingular transformations of the return data. Therefore, our results are the same regardless of whether we use excess returns or gross returns on the risky assets. For the case with no risk-free asset, the analysis is slightly more complicated and is available upon request.

However, the pricing errors are nonzero when the asset pricing model is misspecified. In this case, we are interested in measuring the degree of model misspecification. HJ suggest using

$$\delta = \min_{m \in \mathcal{M}} (E[(y-m)^2])^{\frac{1}{2}} \tag{4}$$

as a misspecification measure of y. In this paper, we refer to δ as the unconstrained HJ-distance.

It is possible for an SDF to price all the test assets correctly and yet to take on negative values with positive probability. Such an SDF does not necessarily rule out arbitrage opportunities and it could be problematic to use this SDF to price payoffs that are not in $\mathcal{P}$ (e.g., derivatives on the test assets). To deal with this issue, HJ provide a second model misspecification measure:

$$\delta_+ = \min_{m \in \mathcal{M}^+} (E[(y-m)^2])^{\frac{1}{2}}. \tag{5}$$

We refer to δ_+ as the constrained HJ-distance. Since $\mathcal{M}^+$ is a subset of $\mathcal{M}$, δ_+ cannot be smaller than δ.

Instead of solving the above primal problems to obtain δ and δ_+, HJ suggest that it is sometimes more convenient to solve the following dual problems:

$$\delta^2 = \max_{\lambda \in \Re^n} E[y^2 - (y - \lambda' \tilde{r})^2] - 2\lambda' q, \tag{6}$$

$$\delta_+^2 = \max_{\lambda \in \Re^n} E[y^2 - [(y - \lambda' \tilde{r})^+]^2] - 2\lambda' q, \tag{7}$$

where λ is a vector of Lagrange multipliers and $(a)^+ \equiv \max[a, 0]$.

When the candidate SDF y depends on some unknown parameters γ, it is customary to choose γ to minimize δ or δ_+, and the squared unconstrained and constrained HJ-distances are then defined as

$$\delta^2 = \min_{\gamma \in \Gamma} \min_{m \in \mathcal{M}} E[(y(\gamma) - m)^2] = \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[y(\gamma)^2 - (y(\gamma) - \lambda' \tilde{r})^2] - 2\lambda' q, \tag{8}$$

$$\delta_+^2 = \min_{\gamma \in \Gamma} \min_{m \in \mathcal{M}^+} E[(y(\gamma) - m)^2] = \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[y(\gamma)^2 - [(y(\gamma) - \lambda' \tilde{r})^+]^2] - 2\lambda' q, \tag{9}$$

where Γ is the parameter space of γ.

HJ provide a maximum pricing error interpretation of the two HJ-distances. Starting with the unconstrained HJ-distance, it is easy to show that for a given SDF y, the vector of Lagrange multipliers is given by

$$\lambda = U^{-1} e, \tag{10}$$

where $U = E[\tilde{r}\tilde{r}']$ is the second moment matrix of $\tilde{r}$. It follows that the squared unconstrained HJ-distance is given by

$$\delta^2 = e'U^{-1}e. \tag{11}$$

Consider a portfolio w with unit second moment, i.e., $w'Uw = 1$. By the Jensen's inequality, the squared pricing error of such a portfolio is

$$(w'e)^2 = (w'U^{\frac{1}{2}}U^{-\frac{1}{2}}e)^2 \leq (w'Uw)(e'U^{-1}e) = \delta^2. \tag{12}$$

Specifically, the portfolio $w = U^{-1}e/\delta$ has a pricing error δ. As a result,

$$\max_{w:\, w'Uw=1} |w'e| = \delta, \tag{13}$$

and we can interpret δ as the maximum pricing error that one can get from using y to price the test assets.

The constrained HJ-distance also has a pricing error interpretation. Consider $h \in L^2$ which can be a nonlinear function of $\tilde{r}$ (say payoff of an option) or the payoff of other primitive assets that are not used by the econometrician. Using Jensen's inequality, we obtain the following upper bound on the squared pricing error of h:

$$(E[yh] - E[mh])^2 = (E[(y-m)h])^2 \leq E[(y-m)^2]E[h^2], \tag{14}$$

where $m \in \mathcal{M}^+$. Since

$$\min_{m\in\mathcal{M}^+} E[(y-m)^2] = \delta_+^2, \tag{15}$$

we have

$$\min_{m\in\mathcal{M}^+,\, E[h^2]=1} |E[yh] - E[mh]| \leq \delta_+. \tag{16}$$

The upper bound in (16) is potentially attainable. Let $m_y^+ = (y - \lambda'\tilde{r})^+ \in \mathcal{M}^+$ be the nonnegative admissible SDF that is closest to y, where λ is the vector of Lagrange multipliers that solves the dual problem in (7). Suppose we choose

$$h = \frac{y - m_y^+}{\delta_+} = \frac{\min[y, \lambda'\tilde{r}]}{\delta_+}. \tag{17}$$

Note that this h is proportional to the payoff of an exchange option. Since $E[(y-m_y^+)^2] = \delta_+^2$, it is easy to see that $E[h^2] = 1$. If m_y^+ is in fact the SDF that the market uses to price h, then the pricing error of h is given by

$$E[yh] - E[m_y^+ h] = \frac{E[(y-m_y^+)^2]}{\delta_+} = \delta_+. \tag{18}$$

While δ is indeed equal to the actual maximum pricing error from using y to price the test assets, the pricing error interpretation of δ_+ is not as clear. The actual maximum pricing error from using y to price all the assets in the economy can be greater than or less than δ_+. There are two reasons why the actual maximum pricing error is not equal to δ_+. The first one is that the exchange option mentioned above may not be a tradable asset, especially if y is a function of non-traded factors. The second reason is that even when such an exchange option is traded, the SDF that the market uses to price this exchange option may not be m_y^+. The pricing error on this exchange option from using y can often be greater than δ_+ when the true SDF is used to price the exchange option. In fact, when the market is complete, δ_+ represents a lower bound on the maximum pricing error for payoffs in L^2. To see this, assume that $m^* \in \mathcal{M}^+$ is the true SDF. Then, using Jensen's inequality, the maximum squared pricing error for $h \in L^2$ with $E[h^2] = 1$ is given by

$$\max_{E[h^2]=1} (E[yh] - E[m^*h])^2 = \max_{E[h^2]=1} (E[(y-m^*)h])^2 = E[(y-m^*)^2], \tag{19}$$

when $h = (y-m^*)/E[(y-m^*)^2]^{\frac{1}{2}}$. Since

$$E[(y-m^*)^2] \geq \min_{m \in \mathcal{M}^+} E[(y-m)^2] = \delta_+^2, \tag{20}$$

the maximum pricing error is generally larger than δ_+. The only case in which we can interpret δ_+ as the maximum pricing error for payoffs in L^2 is when $m_y^+ = m^*$. However, it is hard to justify the maximum pricing error interpretation of δ_+ when multiple models are considered. The reason is that m_y^+ is model dependent and it is not possible that $m_y^+ = m^*$ for all models unless $\mathcal{M}^+$ contains only a single element.

From (19), the maximum pricing error of a model is equal to the distance between y and m^*. However, a model in $\mathcal{M}^+$ (i.e., $\delta_+ = 0$) can actually be further away from m^* than a model that is not in $\mathcal{M}^+$ (i.e., $\delta_+ > 0$). This makes it problematic to rank models by δ_+ because a model with a larger δ_+ can actually be closer to m^* and have a smaller maximum pricing error on payoffs in L^2. In particular, a model with a smaller δ_+ is not necessarily a better model for pricing derivatives.[3]

While it is desirable to consider SDFs that are strictly positive, most SDFs used in empirical work are typically misspecified and some of them can take on negative values. It is often believed

[3]The fact that different admissible SDFs can assign different prices to payoffs outside of the test assets is well known. Boyle, Feng, Tian, and Wang (2008) provide a robust approach for selecting admissible SDFs to price derivatives.

that a model with a smaller δ_+ has a smaller probability of taking on negative values because it is closer to $\mathcal{M}^+$. As it turns out, the probability for an SDF to take on negative values has very little to do with the magnitude of δ_+.

To illustrate this point, we adapt an example from LXZ. We consider an economy with two states (s_1 and s_2) that are equally likely to occur. The only test asset considered by the econometrician is risk-free with gross risk-free rate $R_0 = 1$, so that the payoff space of the test asset ($\mathcal{P}$) can be represented by the dashed line in Figure 1. For an SDF m to be admissible, it has to price the risk-free asset correctly, which implies

$$E[m] = 1 \Rightarrow \frac{1}{2} \times m_1 + \frac{1}{2} \times m_2 = 1 \Rightarrow m_2 = 2 - m_1, \tag{21}$$

where m_1 and m_2 are the values of m in states 1 and 2, respectively. As a result, the admissible set of SDFs ($\mathcal{M}$) is represented by the dotted line with a slope of -1. Since the probabilities of the two states are equal, the line $\mathcal{M}$ is perpendicular to the line $\mathcal{P}$. The part of $\mathcal{M}$ that represents the set of nonnegative SDFs ($\mathcal{M}^+$) is highlighted with a thick solid line.

Figure 1 about here

In Figure 1, we consider two competing SDFs, $y^{\mathcal{F}}$ and $y^{\mathcal{G}}$. The constrained HJ-distance of an SDF y that takes on the values of y_1 and y_2 in the two states is given by

$$\begin{aligned} \delta_+ &= \min_{m \in \mathcal{M}^+} E[(y-m)^2]^{\frac{1}{2}} \\ &= \min_{m \in \mathcal{M}^+} \left[\frac{1}{2} \times (m_1 - y_1)^2 + \frac{1}{2} \times (m_2 - y_2)^2 \right]^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{2}} \min_{m \in \mathcal{M}^+} [(m_1 - y_1)^2 + (m_2 - y_2)^2]^{\frac{1}{2}}. \end{aligned} \tag{22}$$

It follows that the shortest distance between $\mathcal{M}^+$ and y is equal to $\sqrt{2}\delta_+$.[4] In Figure 1, we observe that $y^{\mathcal{F}}$ is further away from $\mathcal{M}^+$ than $y^{\mathcal{G}}$. Consequently, we have $\delta_{\mathcal{F},+} > \delta_{\mathcal{G},+}$. Despite having a shorter constrained HJ-distance, $y^{\mathcal{G}}$ takes on negative values in both states. In contrast, $y^{\mathcal{F}}$ always takes on positive values. The important message here is that while $m \in \mathcal{M}^+$ is positive, the distance from $\mathcal{M}^+$ alone tells us little about the probability for an SDF to take on negative values.

2.3. Hansen-Jagannathan bounds and distances

[4] Although LXZ label the distance as δ_+ instead of $\sqrt{2}\delta_+$, the interpretation of their results is not affected.

To better understand the constrained and unconstrained HJ-distances, it proves advantageous to introduce the concept of HJ-bounds. Hansen and Jagannathan (1991) propose two volatility bounds on admissible stochastic discount factors. The unconstrained HJ-bound (σ_0^2) is the minimum variance that an SDF must have in order for it to be potentially admissible, and it is defined as

$$\sigma_0^2 = \min_{m\in\mathcal{M}} \text{Var}[m] = \min_{m\in\mathcal{M}} E[m^2] - E[m]^2 = \min_{m\in\mathcal{M}} E[m^2] - \frac{1}{R_0^2}. \tag{23}$$

The last equality follows because all $m \in \mathcal{M}$ price the risk-free asset correctly and hence $E[m] = 1/R_0$. In addition, Hansen and Jagannathan (1991) define the constrained HJ-bound (σ_c^2) as the minimum variance for the set of nonnegative admissible SDFs:

$$\sigma_c^2 = \min_{m\in\mathcal{M}^+} \text{Var}[m] = \min_{m\in\mathcal{M}^+} E[m^2] - E[m]^2 = \min_{m\in\mathcal{M}^+} E[m^2] - \frac{1}{R_0^2}. \tag{24}$$

Note that both HJ-bounds (and their difference) only depend on the choice of the test assets and are model independent.

For illustration purposes, we start with the case of a spanned SDF. We say that y is a spanned SDF when it can be perfectly mimicked by the returns on the test assets. For such an SDF, the difference between its squared constrained and unconstrained HJ-distances turns out to be equal to $\sigma_c^2 - \sigma_0^2$ — the difference between the two HJ-bounds. To see this, we write

$$\delta^2 = \min_{m\in\mathcal{M}} E[(y-m)^2] = E[y^2] - 2E[ym] + \min_{m\in\mathcal{M}} E[m^2]. \tag{25}$$

The last equality follows because y is the payoff of a portfolio of the test assets, so every $m \in \mathcal{M}$ assigns the same price to y. As a result, $E[ym]$ is constant across $m \in \mathcal{M}$. Similarly, we have

$$\delta_+^2 = \min_{m\in\mathcal{M}^+} E[(y-m)^2] = E[y^2] - 2E[ym] + \min_{m\in\mathcal{M}^+} E[m^2]. \tag{26}$$

It follows that

$$\delta_+^2 - \delta^2 = \min_{m\in\mathcal{M}^+} E[m^2] - \min_{m\in\mathcal{M}} E[m^2] = \sigma_c^2 - \sigma_0^2. \tag{27}$$

In establishing the above identity, we assume that the spanned SDF y is fixed. However, this identity continues to hold even when y depends on some unknown parameters γ. For example, if we assume $y(\gamma) = \gamma_0 + \gamma_1' f$, where f is a vector of returns on K mimicking portfolios, then we have

$$\delta^2 = \min_{\gamma} \min_{m\in\mathcal{M}} E[(y(\gamma)-m)^2] = \min_{\gamma}(E[y(\gamma)^2] - 2E[y(\gamma)m]) + \min_{m\in\mathcal{M}} E[m^2]. \tag{28}$$

The last equality follows because the last term is independent of the model and the middle term is the same for every $m \in \mathcal{M}$. Similarly, we have

$$\delta_+^2 = \min_{\gamma} \min_{m \in \mathcal{M}^+} E[(y(\gamma) - m)^2] = \min_{\gamma}(E[y(\gamma)^2] - 2E[y(\gamma)m]) + \min_{m \in \mathcal{M}^+} E[m^2]. \tag{29}$$

As a result, we have $\delta_+^2 - \delta^2 = \sigma_c^2 - \sigma_0^2$ and this difference is model independent.

The results above have two implications. The first one is that for a spanned SDF, the SDF parameters that minimize δ and δ_+ are identical because both of them are given by

$$\text{argmin}_{\gamma} E[y(\gamma)^2] - 2E[y(\gamma)m], \tag{30}$$

and they do not depend on whether $m \in \mathcal{M}$ or $\mathcal{M}^+$. This suggests that for δ and δ_+, one should not expect the corresponding SDFs to be any different, or take on negative values with different probabilities.

The second implication is that for two spanned SDFs, say $y^{\mathcal{F}}$ and $y^{\mathcal{G}}$, the difference between their unconstrained HJ-distances is the same as the difference between their constrained HJ-distances. This is because

$$\delta_{\mathcal{F},+}^2 - \delta_{\mathcal{G},+}^2 = \delta_{\mathcal{F}}^2 + (\sigma_c^2 - \sigma_0^2) - \delta_{\mathcal{G}}^2 - (\sigma_c^2 - \sigma_0^2) = \delta_{\mathcal{F}}^2 - \delta_{\mathcal{G}}^2. \tag{31}$$

This illustrates that for spanned SDFs, one should not expect the constrained HJ-distance to be better than the unconstrained HJ-distance in differentiating between competing models. The above two implications are based on analyses of the population HJ-distances of spanned SDFs. However, it can be easily shown that for spanned SDFs, these two implications also hold in sample provided that the sample estimate of σ_c^2 is finite.

Knowing that the choice of unconstrained or constrained HJ-distances does not affect a spanned SDF, we now turn our attention to SDFs that are not spanned by the returns on test assets. We can always decompose a candidate SDF y into two components:

$$y = y^* + z, \tag{32}$$

where y^* is the part of y that is spanned by the returns on the test assets and is given by

$$y^* = \mu_y + V_{ry}' V_{rr}^{-1}(r - \mu_r), \tag{33}$$

with $\mu_y = E[y]$, $\mu_r = E[r]$, $V_{rr} = \text{Var}[r]$, and $V_{ry} = \text{Cov}[r, y]$. It is easy to see that z has mean zero and is uncorrelated with r. Since y^* is spanned by the returns, $E[y^* m]$ is constant across $m \in \mathcal{M}$. It follows that the squared unconstrained HJ-distance of y is given by

$$\begin{aligned}
\delta^2 &= \min_{m \in \mathcal{M}} E[(y-m)^2] \\
&= \min_{m \in \mathcal{M}} E[(y^* + z - m)^2] \\
&= E[y^{*2}] + 2E[y^*(z-m)] + \min_{m \in \mathcal{M}} E[(z-m)^2] \\
&= E[y^{*2}] - 2E[y^* m] + \sigma_0^2 + \frac{1}{R_0^2}. \qquad (34)
\end{aligned}$$

The last equality follows because if $m \in \mathcal{M}$, then $\tilde{m} = m - z$ also prices all the test assets correctly and we have $\tilde{m} \in \mathcal{M}$. As a result, we have the following identity

$$\min_{m \in \mathcal{M}} E[(z-m)^2] = \min_{\tilde{m} \in \mathcal{M}} E[\tilde{m}^2] = \sigma_0^2 + \frac{1}{R_0^2}. \qquad (35)$$

However, the above equality does not hold if we replace $\mathcal{M}$ with $\mathcal{M}^+$. This is because when $m \in \mathcal{M}^+$, $\tilde{m} = m - z$ can take on negative values and is not always in $\mathcal{M}^+$. As a result, the derivation of the constrained HJ-distance is more complicated when the SDF is not spanned by the returns on the test assets. In general, the squared constrained HJ-distance of y is given by

$$\begin{aligned}
\delta_+^2 &= \min_{m \in \mathcal{M}^+} E[(y^* + z - m)^2] \\
&= E[y^{*2}] + 2E[y^*(z-m)] + \min_{m \in \mathcal{M}^+} E[(z-m)^2] \\
&= E[y^{*2}] - 2E[y^* m] + \min_{m \in \mathcal{M}^+} E[(z-m)^2]. \qquad (36)
\end{aligned}$$

It follows that the difference between the squared constrained and unconstrained HJ-distances of y is given by

$$\delta_+^2 - \delta^2 = \min_{m \in \mathcal{M}^+} E[(z-m)^2] - \sigma_0^2 - \frac{1}{R_0^2}. \qquad (37)$$

This result implies that only the unspanned component, z, of an SDF is responsible for determining the difference between δ_+^2 and δ^2. Therefore, if an SDF is a function of non-traded factors, it is possible that the SDF parameters differ across the constrained and unconstrained HJ-distances. Our analysis also suggests that $\delta_{\mathcal{F},+}^2 - \delta_{\mathcal{G},+}^2 \neq \delta_{\mathcal{F}}^2 - \delta_{\mathcal{G}}^2$ only when at least one of the two competing SDFs is not spanned. However, to deepen our understanding of the relation between the constrained and unconstrained HJ-distances, we first need to derive their analytical expressions. Obtaining an

analytical solution for δ_+ is a significant challenge, and we will take up this task (*albeit* with a distributional assumption) in the next section.

3. Analytical solution of the constrained Hansen-Jagannathan distance

While an explicit solution of the unconstrained HJ-distance is easy to obtain, the constrained HJ-distance problem is much harder and, to the best of our knowledge, an analytical solution is not available in the literature. To a large extent, the lack of an analytical expression has severely hampered our ability to understand the constrained HJ-distance. To overcome this problem, we make a joint distributional assumption on the SDF and the returns on the test assets. Throughout this section, we assume that the SDF and the returns are jointly normally distributed. Since a normally distributed SDF takes on negative values by construction, it cannot belong to $\mathcal{M}^+$ and our theoretical analysis of the constrained HJ-distance is clearly conducted under the hypothesis that the asset pricing model is misspecified. Note that we assume normality solely for the purpose of deriving an analytical solution for δ_+^2. As we demonstrate below, solving for δ_+^2 is nontrivial even under the normality assumption. However, we do not argue that this distributional assumption is always a good approximation of reality and whether normality provides a reasonable approximation or not depends on the problem at hand.

3.1. Stochastic discount factors without parameters

We start off with the case in which the SDF y does not depend on unknown parameters. As before, we decompose y into two components y^* (spanned) and z (unspanned) as in (32). For the unconstrained HJ-distance, the vector of pricing errors of $\tilde{r}$ is given by

$$e = E[\tilde{r}y] - q = \begin{bmatrix} R_0\mu_y - 1 \\ V_{ry} + \mu_r\mu_y \end{bmatrix}. \tag{38}$$

Using the partitioned matrix inverse formula, it is easy to rewrite the inverse of $U = E[\tilde{r}\tilde{r}']$ as

$$U^{-1} = \begin{bmatrix} R_0^2 & R_0\mu_r' \\ R_0\mu_r & V_{rr} + \mu_r\mu_r' \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1+a}{R_0^2} & -\frac{\mu_r' V_{rr}^{-1}}{R_0} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} & V_{rr}^{-1} \end{bmatrix}, \tag{39}$$

where $a = \mu_r' V_{rr}^{-1}\mu_r$ is the squared Sharpe ratio of the tangency portfolio of the N risky assets. It follows that the vector of Lagrange multipliers for the unconstrained HJ-distance is given by

$$\lambda = U^{-1}e = \begin{bmatrix} \frac{\mu_y - V_{ry}' V_{rr}^{-1}\mu_r}{R_0} - \frac{1+a}{R_0^2} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0}\right) \end{bmatrix} \tag{40}$$

and the admissible SDF that is closest to y is

$$m_y = y - \lambda'\tilde{r} = z + \frac{1}{R_0} - \frac{\mu_r' V_{rr}^{-1}(r - \mu_r)}{R_0}. \tag{41}$$

After simplification, the squared unconstrained HJ-distance of y is

$$\delta^2 = E[(y - m_y)^2] = \left(\mu_y - \frac{1}{R_0}\right)^2 + \left(V_{ry} + \frac{\mu_r}{R_0}\right)' V_{rr}^{-1} \left(V_{ry} + \frac{\mu_r}{R_0}\right). \tag{42}$$

Turning to the constrained HJ-distance case, the vector of Lagrange multipliers in (7) is given by

$$\tilde{\lambda} = \text{argmin}_\lambda E[(y - \lambda'\tilde{r})^{+2}] + 2\lambda' q, \tag{43}$$

and $\tilde{\lambda}$ can be obtained by solving the following first order condition:

$$E[\tilde{r}(y - \tilde{\lambda}'\tilde{r})^+] = q. \tag{44}$$

In principle, we can solve the n nonlinear equations $E[\tilde{r}(y - \tilde{\lambda}'\tilde{r})^+] = q$ to obtain the vector of Lagrange multipliers $\tilde{\lambda}$, but this can be very complicated. Instead, we simplify the problem so that we only need to solve one nonlinear equation to obtain $\tilde{\lambda}$. The following proposition shows that, under normality, it is possible to obtain explicit expressions for the Lagrange multipliers and the squared constrained HJ-distance.

Proposition 1. *Let η be the unique solution to*

$$g(u) = \left[a + \sigma_z^2 R_0^2 \Phi(u)^2\right]^{-\frac{1}{2}}, \tag{45}$$

where

$$g(u) = u + \frac{\phi(u)}{\Phi(u)}, \tag{46}$$

$\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution functions of a standard normal random variable, respectively, and $\sigma_z^2 = \text{Var}[z]$. The vector of Lagrange multipliers in the constrained HJ-distance case is given by

$$\tilde{\lambda} = \begin{bmatrix} \frac{\mu_y - V_{ry}' V_{rr}^{-1}\mu_r}{R_0} - \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \Phi(\eta)} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0 \Phi(\eta)}\right) \end{bmatrix}. \tag{47}$$

The squared constrained HJ-distance of an SDF y is given by

$$\delta_+^2 = \delta^2 + \sigma_z^2 \Phi(-\eta) + \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \Phi(\eta)} - \frac{1 + a}{R_0^2}, \tag{48}$$

where the expression of δ^2 is provided in (42).

Proof. See Appendix A.

The SDF in $\mathcal{M}^+$ that is closest to y is m_y^+, where

$$m_y = y - \tilde{\lambda}'\tilde{r} = z + \frac{1}{\Phi(\eta)}\left[\frac{\eta}{R_0 g(\eta)} - \frac{\mu_r' V_{rr}^{-1}(r-\mu_r)}{R_0}\right]. \tag{49}$$

Just like m_y in (41) for the unconstrained HJ-distance case, the m_y for the constrained HJ-distance also has two components. The first component, z, is uncorrelated with the returns, and the second component is a linear function of the excess return on the tangency portfolio of the test assets. Let $\Delta = \delta_+^2 - \delta^2$ be the difference between the squared constrained and unconstrained HJ-distances of y. We are interested in the determinants of Δ. Note that η, the solution to (45), depends on σ_z^2, a, and R_0. As a result, Δ also depends on these three parameters. Out of the three parameters, the only one that is related to the SDF is σ_z^2 — the variance of the unspanned component. When $\sigma_z^2 = 0$ (i.e., spanned SDF), we have $g(\eta) = 1/\sqrt{a}$ and hence

$$\Delta = \frac{a + \sqrt{a}\eta}{R_0^2\Phi(\eta)} - \frac{1+a}{R_0^2} = \left[\frac{\sqrt{a}(\sqrt{a}+\eta)}{R_0^2\Phi(\eta)} - \frac{1}{R_0^2}\right] - \frac{a}{R_0^2} = \sigma_c^2 - \sigma_0^2, \tag{50}$$

where the last equality is based on Proposition 1 of Kan and Robotti (2008). This confirms the result in Section 2.3 which suggests that $\delta_+^2 - \delta^2$ for a spanned SDF is equal to the difference between the constrained and unconstrained HJ-bounds.

The following lemma provides the comparative statics of Δ with respect to its three determinants.

Lemma 1. *The partial derivatives of* Δ *with respect to* $(\sigma_z^2,\ a,\ R_0)$ *are given by*

$$\frac{\partial \Delta}{\partial \sigma_z^2} = \Phi(-\eta) > 0, \tag{51}$$

$$\frac{\partial \Delta}{\partial a} = \frac{\Phi(-\eta)}{R_0^2\Phi(\eta)} > 0, \tag{52}$$

$$\frac{\partial \Delta}{\partial R_0} = \frac{2}{R_0^3}\left[1 - \frac{a\Phi(-\eta)}{\Phi(\eta)} - \frac{\eta}{\Phi(\eta)g(\eta)}\right]. \tag{53}$$

Proof. See Appendix A.

Lemma 1 shows that Δ is an increasing function of σ_z^2, which suggests that $\sigma_c^2 - \sigma_0^2$ is a lower bound for Δ. Intuitively, adding an unspanned component z to an SDF does not affect its ability to price the test assets, so the unconstrained HJ-distance (which is a measure of aggregate pricing

errors) of a model is unaffected by z. This explains why the expression of δ^2 in (42) is independent of σ_z^2. However, adding z to an SDF can affect its probability of taking on negative values and hence drives the SDF further away from $\mathcal{M}^+$. This explains why δ_+^2 and hence Δ is an increasing function of σ_z^2. Note that Δ does not depend on how good or bad a model is. It is only a function of the variance of its unspanned component. This suggests that for two different models, say $\mathcal{F}$ and $\mathcal{G}$, we can expect $\delta_{\mathcal{F},+}^2 - \delta_{\mathcal{G},+}^2$ to differ substantially from $\delta_{\mathcal{F}}^2 - \delta_{\mathcal{G}}^2$ only when the variances of the unspanned components across the two models are very different.

In addition, Lemma 1 shows that Δ is an increasing function of the Sharpe ratio of the tangency portfolio of the test assets. This result requires some explanation. Consider the case in which $\sigma_z^2 \to 0$. When this happens, $(\delta_+^2 - \delta^2) \to (\sigma_c^2 - \sigma_0^2)$ — the difference between the constrained and unconstrained HJ-bounds. Lemma 6 of Kan and Robotti (2008) shows that, under normality, $(\sigma_c^2 - \sigma_0^2) \to 0$ when $a \to 0$, $(\sigma_c^2 - \sigma_0^2) \to \infty$ when $a \to \infty$, and $\sigma_c^2 - \sigma_0^2$ is a strictly increasing function of a. Therefore, when a is small, we should not expect large differences between the constrained and unconstrained HJ-bounds and between the constrained and unconstrained HJ-distances. Intuitively, when a is close to zero, the weight of the risk-free asset in the minimum second moment portfolio is close to one, and the gross return on this portfolio has a very small probability of taking on a negative value. Since the minimum variance admissible SDF is proportional to the gross return on this portfolio, imposing the non-negativity constraint of Hansen and Jagannathan (1991) on it has almost no effect.

To gain some understanding of how σ_z and a affect Δ, Figure 2 plots Δ as a function of σ_z for three different values of the Sharpe ratio of the tangency portfolio ($\sqrt{a} = 0.25$, 0.5, and 0.75) with $R_0 = 1.005$ (the plot is not sensitive to other reasonable values of the gross risk-free rate).[5] As expected, Figure 2 reveals that Δ is an increasing function of σ_z. However, Δ is heavily influenced by the Sharpe ratio of the tangency portfolio. When $\sqrt{a} = 0.25$, the difference between δ_+^2 and δ^2 is indistinguishable from zero. For $\sqrt{a} = 0.5$, the difference between δ_+^2 and δ^2 is still quite small, even for relatively large σ_z. This suggests that for reasonable Sharpe ratio values, we should not expect to find a large difference between the constrained and unconstrained HJ-distances of a model, even if the model contains a large unspanned component.

[5] Although a Sharpe ratio of 0.75 may seem high, this is in line with the sample Sharpe ratio (0.71) of the tangency portfolio of the 25 Fama-French size and book-to-market portfolios used in the empirical application in Section 6.

Figure 2 about here

Our prediction that we should expect δ_+^2 to be close to δ^2 is based on the analysis of their population values, which are the quantities that researchers are typically interested in. However, even when the true Sharpe ratio of the tangency portfolio is 0.25, it is possible to find a large difference between the sample constrained and unconstrained HJ-distances in a given sample. Such difference could be partly due to sampling variation and partly due to an upward bias of the sample constrained HJ-distance.

3.2. Linear stochastic discount factors

In the previous subsection, we derived an explicit expression for δ_+^2 for the case in which the SDF does not depend on parameters. When the SDF depends on some parameters, we also need to solve the outer optimization problem in (9). For general nonlinear SDFs, it is hard to obtain explicit solutions for the SDF parameters, even for the unconstrained HJ-distance. Therefore, we focus on linear SDFs of the form

$$y(\gamma) = \gamma_0 + \gamma_1' f, \tag{54}$$

where f is a vector of K systematic factors, and $\gamma = [\gamma_0,\ \gamma_1']'$ is the vector of SDF parameters. In addition to facilitating the derivation of γ, linear SDFs deserve a thorough investigation because of their popularity in the literature.

Throughout this subsection, we assume that f and r are jointly normally distributed (which implies that $y(\gamma)$ and r are jointly normally distributed). Before presenting an analytical solution for γ, we take a short digression to discuss the issue of linear SDFs taking on negative values. Researchers are sometimes concerned that under the normality assumption, a linear SDF can take on negative values and hence is not arbitrage free. In the following lemma, we show that under normality, an SDF that takes on negative values can be converted into a positive SDF without affecting its ability to price the test assets.

Lemma 2. *Suppose that an SDF y and the returns are jointly normally distributed. Consider the following transformation of y:*

$$y_+ = \mu_y \exp\left(\frac{y}{\mu_y} - 1 - \frac{\sigma_y^2}{2\mu_y^2}\right), \tag{55}$$

where $\mu_y = E[y]$ *and* $\sigma_y^2 = \text{Var}[y]$. *Then, we have* $E[\tilde{r}y_+] = E[\tilde{r}y]$ *with* $y_+ > 0$ *when* $\mu_y > 0$.

Proof. See Appendix A.

Lemma 2 suggests that if one can find a linear SDF y that prices all the test assets correctly, then one should not be too concerned with the SDF's probability of taking on negative values because a simple transformation of y into y_+ would make it positive.[6] However, this is a mechanical transformation and there is no compelling reason to believe that simply because y_+ is positive, it can price derivatives better than y.

For a linear SDF, the unconstrained HJ-distance problem is easy to solve. Defining $\mu_f = E[f]$ and $V_{rf} = \text{Cov}[r, f']$, it can be readily shown that the $\gamma = [\gamma_0,\ \gamma_1']'$ that minimizes the unconstrained HJ-distance is

$$\gamma_1 = -\frac{1}{R_0}(V_{rf}' V_{rr}^{-1} V_{rf})^{-1}(V_{rf}' V_{rr}^{-1}\mu_r), \qquad \gamma_0 = \frac{1}{R_0} - \gamma_1'\mu_f. \tag{56}$$

As a result, the linear SDF that minimizes the unconstrained HJ-distance is

$$y = \frac{1}{R_0} + \gamma_1'(f - \mu_f). \tag{57}$$

In addition, defining $a_1 = \mu_r' V_{rr}^{-1} V_{rf}(V_{rf}' V_{rr}^{-1} V_{rf})^{-1} V_{rf}' V_{rr}^{-1}\mu_r$ as the squared Sharpe ratio of the tangency portfolio constructed from the K factor mimicking portfolios, the squared unconstrained HJ-distance for a linear SDF and the vector of Lagrange multipliers are given by

$$\delta^2 = \frac{a - a_1}{R_0^2} \tag{58}$$

and

$$\lambda = \begin{bmatrix} -\delta^2 \\ V_{rr}^{-1}\left(V_{rf}\gamma_1 + \frac{\mu_r}{R_0}\right) \end{bmatrix}, \tag{59}$$

respectively.

For the more difficult problem of the constrained HJ-distance, we first define the covariance matrix of the residuals from projecting the factors onto the returns as $V_{ff\cdot r} = V_{ff} - V_{rf}' V_{rr}^{-1} V_{rf}$, where $V_{ff} = \text{Var}[f]$. The following proposition presents the solution to the constrained HJ-distance problem.

[6]This lemma also suggests that while an SDF that is linear in the return on the market portfolio implies that the capital asset pricing model (CAPM) holds, the CAPM does not imply that the SDF can only be written as a linear function of the market portfolio.

Proposition 2. *Let η be the unique solution to*

$$g(u) = \left(a + \alpha' \left[\frac{1}{\Phi(u)} I_K - V_{ff}^{-\frac{1}{2}} V_{ff\cdot r} V_{ff}^{-\frac{1}{2}} \right]^{-2} \alpha \right)^{-\frac{1}{2}}, \tag{60}$$

where $\alpha = V_{ff}^{-1} V_{ff\cdot r}^{\frac{1}{2}} V_{rf}' V_{rr}^{-1} \mu_r$ and $g(u)$ is defined in (46). Then, the vector of SDF parameters that minimizes the constrained HJ-distance is given by $\tilde{\gamma} = [\tilde{\gamma}_0,\ \tilde{\gamma}_1']'$, where

$$\tilde{\gamma}_1 = -\frac{1}{R_0} [V_{ff} - \Phi(\eta) V_{ff\cdot r}]^{-1} V_{rf}' V_{rr}^{-1} \mu_r, \qquad \tilde{\gamma}_0 = \frac{1}{R_0} - \tilde{\gamma}_1' \mu_f, \tag{61}$$

and the SDF that minimizes the constrained HJ-distance is

$$\tilde{y} = \frac{1}{R_0} + \tilde{\gamma}_1' (f - \mu_f). \tag{62}$$

Furthermore, the squared constrained HJ-distance has the following expression:

$$\delta_+^2 \quad = \quad \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \Phi(\eta)} - \frac{(1 + \tilde{a}_1)}{R_0^2}, \tag{63}$$

where $\tilde{a}_1 = \mu_r' V_{rr}^{-1} V_{rf} [V_{ff} - \Phi(\eta) V_{ff\cdot r}]^{-1} V_{rf}' V_{rr}^{-1} \mu_r$, and the vector of Lagrange multipliers for the constrained HJ-distance is given by

$$\tilde{\lambda} = \begin{bmatrix} -\delta_+^2 \\ V_{rr}^{-1} \left(V_{rf} \tilde{\gamma}_1 + \frac{\mu_r}{R_0 \Phi(\eta)} \right) \end{bmatrix}. \tag{64}$$

Proof. See Appendix A.

Besides the simplicity of the expressions for $\tilde{\gamma}$, $\tilde{\lambda}$, and δ_+^2, a few interesting observations emerge from Proposition 2. First, the Lagrange multiplier on the risk-free asset is equal to $-\delta_+^2$ (expression (59) shows that a similar result holds for the unconstrained HJ-distance).[7] Second, in contrast to the SDF case without parameters, η does not depend on R_0 since $\tilde{y}$ prices the risk-free asset correctly. Third, when the factors are spanned by the returns (i.e., $V_{ff\cdot r} = 0_{K\times K}$), it can be readily shown that the difference between the squared constrained and unconstrained HJ-distances coincides with the difference between the constrained and unconstrained HJ-bounds. This result confirms our earlier findings for spanned SDFs in Section 2.3. Finally, when one or more factors are useless, i.e., they are uncorrelated with the returns, the SDF parameters that minimize the unconstrained

[7] This identity is not due to the normality assumption. We refer the readers to Lemma 4 in Section 4 for a more general version of this result.

HJ-distance are not identified since the matrix $V_{rf}'V_{rr}^{-1}V_{rf}$ is not of full rank and cannot be inverted. However, the SDF parameters that minimize the constrained HJ-distance are still well defined. For example, when all factors are useless, we have $\tilde{\gamma} = [1/R_0,\ 0_K']'$ and $\tilde{y} = 1/R_0$. In this case, δ_+^2 is equal to the constrained HJ-bound σ_c^2.

With the analytical solutions of the linear SDF parameters for the unconstrained and constrained HJ-distances, we can now answer two interesting questions. The first question is whether the linear SDF $\tilde{y}$ in (62) results in a lower probability of taking on negative values than the linear SDF y in (57). If this is the case, one can think of this as a potential benefit of using the constrained HJ-distance. The second question is whether there is a trade-off between getting the linear SDF closer to $\mathcal{M}^+$ and the ability of the SDF to price the test assets. For this purpose, we introduce an aggregate measure of pricing errors of $\tilde{y}$ as

$$\tilde{\delta}^2 = \tilde{e}'U^{-1}\tilde{e}, \tag{65}$$

where $\tilde{e} = E[\tilde{r}\tilde{y}] - q$ is the vector of pricing errors when we use $\tilde{y}$ to price the test assets. Just like the δ^2 measure, $\tilde{\delta}^2$ can be interpreted as the maximum squared pricing error of a portfolio of test assets when one uses $\tilde{y}$ as the SDF. Comparing $\tilde{\delta}^2$ with δ^2, we gain useful insights of the potential cost of using $\tilde{y}$ instead of y to price the test assets. The following lemma provides answers to these two questions.

Lemma 3. *Let y and $\tilde{y}$ be the linear SDFs that minimize the unconstrained and constrained HJ-distances, respectively. Then, we have*

$$P[y<0] - P[\tilde{y}<0] = \Phi\left(-\frac{1}{R_0\sqrt{\gamma_1'V_{ff}\gamma_1}}\right) - \Phi\left(-\frac{1}{R_0\sqrt{\tilde{\gamma}_1'V_{ff}\tilde{\gamma}_1}}\right) > 0, \tag{66}$$

and

$$\delta^2 \leq \tilde{\delta}^2 \leq \delta_+^2. \tag{67}$$

Proof. See Appendix A.

As shown in the proof of Lemma 3, (67) is a general result. It is not specific to the linear model and our proof does not rely on the normality assumption. However, we can only establish $P[y < 0] > P[\tilde{y} < 0]$ for the case of linear models and under the normality assumption. Whether this

inequality continues to hold for nonlinear models or without the normality assumption is an open question.[8]

Lemma 3 suggests that there are potential benefits and costs in choosing the SDF parameters to minimize the constrained HJ-distance as opposed to minimizing the unconstrained HJ-distance. On the one hand, $\tilde{y}$ is less likely than y to take on negative values. On the other hand, $\tilde{y}$ will price the test assets worse than y. Exactly how large is this cost-benefit trade-off depends on the parameters. For the one-factor case, we can show that

$$P[y<0]-P[\tilde{y}<0]=\Phi\left(-\frac{|\rho|}{\sqrt{a_1}}\right)-\Phi\left(-\frac{1-\Phi(\eta)(1-\rho^2)}{\sqrt{a_1}|\rho|}\right), \tag{68}$$

where η is the unique solution of

$$g(u)=\left[a+\frac{a_1\rho^2(1-\rho^2)\Phi(u)^2}{[1-\Phi(u)(1-\rho^2)]^2}\right]^{-\frac{1}{2}}, \tag{69}$$

and $\rho^2=V_{rf}'V_{rr}^{-1}V_{rf}/V_{ff}$ is the proportion of variability of the factor that is explained by the returns on the test assets. In addition, we have

$$\tilde{\delta}^2-\delta^2=\frac{a_1}{R_0^2}\left[\frac{\Phi(-\eta)(1-\rho^2)}{1-\Phi(\eta)(1-\rho^2)}\right]^2. \tag{70}$$

Note that both (68) and (70) depend on a, a_1 and ρ^2. In these expressions, a is the squared Sharpe ratio of the tangency portfolio of the test assets, which is a measure of the cross-sectional difference in expected excess returns across the test assets; a_1 measures how good the model is in explaining the expected returns on the test assets (recall that $\delta^2=(a-a_1)/R_0^2$); and, finally, ρ^2 measures how well the factor is spanned by the returns.

In Figure 3, we plot $P[y<0]-P[\tilde{y}<0]$ as a function of ρ^2 for three different values of the Sharpe ratio of the tangency portfolio ($\sqrt{a}=0.25$, 0.5, and 0.75). In each case, we assume $a_1=a/2$, so that the model explains about half of the cross-sectional variation in expected returns. From Figure 3, we can see that when $\rho^2\to 0$ (y is not defined when $\rho^2=0$), $P[y<0]-P[\tilde{y}<0]\to 0.5$. The reason is that when the unspanned component of the factor increases, y becomes more volatile (because γ_1 does not depend on the unspanned component of the factor) and behaves more like a useless factor. As a result, $P[y<0]\to 0.5$. However, as $\rho^2\to 0$, $\tilde{y}$ converges to $1/R_0$ and has almost zero probability of taking on negative values. In contrast, when $\rho^2\to 1$, the SDF behaves more like

[8]For example, many nonlinear SDFs are positive by construction. Therefore, the probability for these SDFs to take on positive values is always one regardless of whether we choose the parameters to minimize δ or δ_+.

a spanned SDF. For a spanned SDF, the SDF parameters and hence the probabilities of taking on negative values are the same for y and $\tilde{y}$. Finally, Figure 3 shows that the Sharpe ratio is important in determining $P[y < 0] - P[\tilde{y} < 0]$. For a given value of ρ^2, we can see that the difference between the two probabilities is an increasing function of a. The reason is that the spanned component of the SDF y is a linear function of the return on the factor mimicking portfolio. When a is small, a_1 is also small, so y puts relatively little weight on the factor mimicking portfolio and hence $P[y < 0]$ is small, leaving not much room for $\tilde{y}$ to improve.

Figure 3 about here

Using the same parameters as in Figure 3 and $R_0 = 1.005$, Figure 4 plots $\tilde{\delta}^2 - \delta^2$ as a function of ρ^2. Again, when $\rho^2 \approx 1$, the SDF is close to a spanned one. It follows that $y \approx \tilde{y}$, so they have roughly the same aggregate pricing errors and $\tilde{\delta}^2 - \delta^2 \to 0$. However, when $\rho^2 \to 0$, we have $\tilde{\delta}^2 = a/R_0^2$ (as $\tilde{y} \approx 1/R_0$ and $\tilde{y}$ does not explain any cross-sectional difference in expected excess returns). It follows that $\tilde{\delta}^2 - \delta^2 \to a_1/R_0^2$. Similar to Figure 3, we also find a to be quite important in determining $\tilde{\delta}^2 - \delta^2$. It is only when a is large (and hence a_1 is large) that we should expect a large difference between the aggregate measures of pricing errors of y and $\tilde{y}$.

Figure 4 about here

In summary, we should expect y and $\tilde{y}$ to behave differently if a is large and ρ^2 is small. In these situations, $P[\tilde{y} < 0]$ will be substantially smaller than $P[y < 0]$, but these are also situations in which $\tilde{y}$ will do substantially worse than y in pricing the test assets. Whether one should sacrifice the pricing of the test assets in exchange for a smaller SDF's probability of taking on negative values is not entirely clear. For example, when ρ^2 is small, $\tilde{y} \approx 1/R_0$ and $\tilde{y}$ is indeed almost always positive. However, this $\tilde{y}$ is unlikely to be a good SDF since it prices every asset by discounting the future asset payoffs using the risk-free rate.

4. Sample constrained Hansen-Jagannathan distance

Since the population constrained HJ-distance of a model is unobservable, researchers have to estimate it using data. In this section, we discuss issues related to the sample constrained HJ-distance as well as some of its finite sample and asymptotic properties. Our analysis can also be

easily adapted to the study of the sample unconstrained HJ-distance. Instead of considering test assets that consist of a gross risk-free rate and a vector of excess returns on N risky assets as in our theoretical section, here we allow for a slightly more general setup. We denote the vector of payoffs of n assets at the end of period t by x_t and the corresponding costs of these n assets at the end of period $t-1$ by q_{t-1}.[9] Besides being applicable to gross and excess returns, this setup can accommodate payoffs of trading strategies that are based on time-varying information.

4.1. Sample estimators

Let y_t and m_t be the realizations of a candidate SDF and an admissible SDF at time t, respectively. The sample squared constrained HJ-distance can be obtained as

$$\begin{aligned} \hat{\delta}_+^2 &= \min_{\{m_t,\, t=1,\ldots,T\}} \frac{1}{T}\sum_{t=1}^{T}(y_t - m_t)^2, \\ &\text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} m_t x_t = \bar{q}, \\ &\qquad m_t \geq 0, \;\; t = 1,\ldots,T, \end{aligned} \tag{71}$$

where $\bar{q} = \frac{1}{T}\sum_{t=1}^{T} q_{t-1}$. In the literature, $\hat{\delta}_+^2$ is often obtained by solving the dual problem:

$$\hat{\delta}_+^2 = \max_{\lambda \in \Re^n} \frac{1}{T}\sum_{t=1}^{T}\left[y_t^2 - [(y_t - \lambda' x_t)^+]^2\right] - 2\lambda'\bar{q}. \tag{72}$$

However, it is not entirely clear that solving the dual problem in (72) has a numerical advantage over solving the primal problem in (71). The reason is that the primal problem in (71) can be easily solved as a quadratic programming problem as we show in Appendix B.

When the candidate SDF depends on some parameters, say $y_t = y_t(\gamma)$, where $\gamma \in \Gamma$ is a k-vector of parameters, we need to solve an additional minimization problem. In this case, the primal problem is given by

$$\begin{aligned} \hat{\delta}_+^2 &= \min_{\gamma \in \Gamma} \min_{\{m_t,\, t=1,\ldots,T\}} \frac{1}{T}\sum_{t=1}^{T}(y_t(\gamma) - m_t)^2, \\ &\text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} m_t x_t = \bar{q}, \\ &\qquad m_t \geq 0, \;\; t = 1,\ldots,T. \end{aligned} \tag{73}$$

[9] When the SDF depends on some parameters, we assume that $E[q_{t-1}] \neq 0_n$. The reason is that when $E[q_{t-1}] = 0_n$, the mean of the SDF cannot be identified and researchers have to choose some normalization of the SDF (see, e.g., Kan and Robotti, 2008).

Similarly, the corresponding dual problem is given by

$$\hat{\delta}_+^2 = \min_{\gamma\in\Gamma}\max_{\lambda\in\Re^n}\frac{1}{T}\sum_{t=1}^{T}\left[y_t(\gamma)^2-[(y_t(\gamma)-\lambda' x_t)^+]^2\right]-2\lambda'\bar{q}. \tag{74}$$

In general, the outer minimization problem does not have a closed-form solution. However, for linear SDFs, the primal problem in (73) can also be solved using quadratic programming as we show in Appendix B. Appendix B also presents a fast numerical algorithm to solve (74) for linear SDFs. Our fast algorithms should prove useful to researchers who are interested in solving the constrained HJ-distance problem and in studying the sampling properties of various estimators associated with the constrained HJ-distance.

4.2. In-sample arbitrage opportunities

In this subsection, we document the surprising and important finding that, in finite samples, the sample constrained HJ-distance has a nonzero probability of taking on the value of infinity. This occurrence only depends on the return realizations and not on the choice of the model. There are two implications of this finding. The first one is that the moments of the sample constrained HJ-distance do not exist. The second one is that the sample constrained HJ-distance could fail to provide a ranking of models. Our results can be somewhat anticipated by similar findings related to the sample constrained HJ-bound. For example, Burnside (1994) finds that the sample constrained HJ-bound takes on the value of infinity quite frequently in his simulations. Kan and Robotti (2008) prove that this occurs with nonzero probability in any finite sample. What these earlier studies do not provide is the underlying reason for such an event to occur, a gap we try to fill in the following analysis.

To understand why $\hat{\delta}_+$ can take on the value of infinity with positive probability, we first define the set $\{m_t,\ t=1,\ldots,T\}$ that satisfies the constraints in (71) or (73) as $\hat{\mathcal{M}}^+$. $\hat{\mathcal{M}}^+$ is the sample counterpart of $\mathcal{M}^+$ and it is the set of T nonnegative random variables that can price the n assets correctly in sample. However, $\hat{\mathcal{M}}^+$ can be an empty set for some realizations of $\{x_t,\ t=1,\ldots,T\}$ even when the population $\mathcal{M}^+$ is not an empty set. To understand what are the random payoffs x_t that would lead to $\hat{\mathcal{M}}^+=\emptyset$, we invoke the Farkas' lemma, which suggests that the following two statements are equivalent:

1. There exists no $m_t \geq 0$, $t = 1, \ldots, T$, such that

$$\frac{1}{T}\sum_{t=1}^{T} x_t m_t = \bar{q}. \tag{75}$$

2. There exists an n-vector w such that $w'\bar{q} < 0$ and

$$w'x_t \geq 0, \quad t = 1, \ldots, T. \tag{76}$$

The equivalence of these two statements suggests that if one finds a portfolio w such that it has negative average cost (i.e., $w'\bar{q} < 0$) but nonnegative payoffs in every period of the sample, then $\hat{\mathcal{M}}^+$ is an empty set. When such an event (i.e., existence of an in-sample arbitrage portfolio) occurs, we show that $\hat{\delta}_+ = \infty$ for all models. Suppose $w'\bar{q} = -c$, where $c > 0$. Let $\{y_t,\ t = 1, \ldots, T\}$ be the realizations of a candidate SDF. Consider $\lambda = sw$, where s is a positive scalar. Since sw for $s \geq 0$ is only a subset of $\Re^n$, we have

$$\hat{\delta}_+^2 = \max_{\lambda \in \Re^n} \frac{1}{T}\sum_{t=1}^{T}\left[y_t^2 - [(y_t - \lambda' x_t)^+]^2\right] - 2\lambda'\bar{q} \geq \max_{s \geq 0} \frac{1}{T}\sum_{t=1}^{T}\left[y_t^2 - [(y_t - sw' x_t)^+]^2\right] + 2cs. \tag{77}$$

Note that since $sw'x_t \geq 0$, we have $(y_t - sw'x_t)^+ \leq |y_t|$ and $y_t^2 - [(y_t - sw'x_t)^+]^2 \geq 0$. Hence, it follows that

$$\hat{\delta}_+^2 \geq \max_{s \geq 0} 2cs = \infty. \tag{78}$$

Knowing that $\hat{\delta}_+$ can take on the value of infinity, it is of interest to understand how often such an event occurs. In general, the probability for an in-sample arbitrage portfolio to exist is nonzero because, for a finite T, there is always some probability that one asset will outperform another asset in every period of the sample. The exact probability for $\hat{\delta}_+ = \infty$ depends on the sample size as well as on the joint distribution of (x_t, q_{t-1}), $t = 1, \ldots, T$. For illustrative purposes, we consider a case in which x_t contains a constant gross risk-free rate and excess returns on N risky assets, i.e., $q = [1,\ 0_N']'$. When the excess returns are i.i.d. multivariate normally distributed, we are able to show that $P[\hat{\delta}_+ = \infty]$ is only a function of N, T, and a, where a is the squared Sharpe ratio of the tangency portfolio of the risky assets.[10]

Figure 5 about here

[10]The proof of this result is available upon request.

In Figure 5, we plot $P[\hat{\delta}_+ = \infty]$ as a function of T for $N = 5$, 10, 25, and 100 based on 10,000 simulations. For each N, we also consider three different values of the Sharpe ratio of the tangency portfolio of the risky assets: $\sqrt{a} = 0.25$, 0.5, and 0.75. As we can see from Figure 5, $P[\hat{\delta}_+ = \infty]$ is close to one when T is close to N. $P[\hat{\delta}_+ = \infty]$ declines as T increases, but the probability is still not negligible unless T is large relative to N. For a fixed $\sqrt{a}$, the probability is an increasing function of N. The reason is that it is easier to find an in-sample arbitrage opportunity when there are more assets to choose from. Finally, for a given N and T, the probability is an increasing function of a. This suggests that when the tangency portfolio of a set of assets has a high population Sharpe ratio, it is easier to observe an in-sample arbitrage opportunity.

It is worth noting that in-sample arbitrage portfolios can also be found in real world data, not only in our simulations. As an example, consider the return data used by LXZ in their empirical analysis (the quarterly returns on the three-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios). When splitting their sample of 195 observations into three subperiods, an in-sample arbitrage portfolio can be found in each subperiod. This shows that there are practical situations in which $\hat{\delta}_+ = \infty$ and the sample constrained HJ-distance fails to provide a ranking of models.

The result $P[\hat{\delta}_+ = \infty] > 0$ suggests that the finite sample moments of $\hat{\delta}_+$ do not exist. Therefore, the asymptotic theory can be inappropriate for finite sample inference, especially when N is large relative to T. To deal with this problem, we have two suggestions. The first one is that we should limit the number of test assets to reduce $P[\hat{\delta}_+ = \infty]$. The second one is that before using the dual problem (74) to obtain $\hat{\delta}_+$, we should first find out if the primal problem in (73) is feasible. An effective way to detect if $\hat{\mathcal{M}}^+ = \emptyset$ is to set up a linear programming problem that has the same constraints as in (74). We can then use standard linear programming routines (e.g., `linprog` in Matlab) to determine whether such a problem is feasible.[11]

4.3. Asymptotic distributions

In this subsection, we derive the asymptotic distributions of the sample constrained HJ-distance and the underlying parameter estimates under correctly specified and potentially misspecified mod-

[11] We implement the numerical procedure for estimating δ_+ and other associated parameters in a set of Matlab programs. The Matlab programs for this paper are available upon request.

els. We present results for the case in which the candidate SDF depends on some unknown parameters, but it is straightforward to adapt our analysis to the case in which the SDF does not depend on parameters. While some asymptotic results regarding the sample constrained HJ-distance are already available (Hansen, Heaton, and Luttmer, 1995; and LXZ), we contribute to the existing literature along several dimensions. First, we demonstrate that under the correctly specified model, the sample Lagrange multipliers do not always have an asymptotic normal distribution. This contradicts some of the claims in the current literature (see, e.g., footnote 15 of LXZ). For some special cases, we can explicitly show that some sample Lagrange multipliers have a limiting distribution that is in the form of a linear combination of independent chi-squared random variables with one degree of freedom. Second, we extend the results in LXZ to the case of possibly serially correlated errors, which can be important in empirical work, especially when models are misspecified.

We adopt the following definitions of correctly specified and misspecified models. An asset pricing model is correctly specified if there exists a $\gamma \in \Gamma$ such that $y_t(\gamma) \in \mathcal{M}^+$, which implies that $\lambda = 0_n$ and $\delta_+ = 0$. The model is misspecified if $y_t(\gamma) \notin \mathcal{M}^+$ for all $\gamma \in \Gamma$, which implies that $\delta_+ > 0$.

Let $\theta = [\gamma' \, , \lambda']'$ and denote by $\theta^* = [\gamma^{*\prime} \, , \lambda^{*\prime}]'$ the pseudo-true value that solves the population counterpart of the sample dual problem in (74):

$$\theta^* = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[\phi_t(\gamma, \lambda)], \tag{79}$$

where

$$\phi_t(\gamma, \lambda) = y_t(\gamma)^2 - [m_t(\theta)^+]^2 - 2\lambda' q_{t-1} \tag{80}$$

and

$$m_t(\theta) = y_t(\gamma) - \lambda' x_t. \tag{81}$$

When $y_t(\gamma^*) \in \mathcal{M}^+$, we have $\lambda^* = 0_n$ and we refer to γ^* as the true value.[12] The estimator of θ^* is often obtained as the solution to the sample dual problem:

$$\hat{\theta} = \left[\begin{array}{c} \hat{\gamma} \\ \hat{\lambda} \end{array} \right] = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} \frac{1}{T} \sum_{t=1}^{T} \phi_t(\gamma, \lambda). \tag{82}$$

[12]The optimization problem in (79) bears some resemblance to the structure of the generalized empirical likelihood (GEL) problem defined as $\min_\gamma \max_\lambda E[\mathbf{h}(\lambda, \gamma)]$, where the function $\mathbf{h}(\varsigma)$ indexes different members of the GEL class. For example, when $\mathbf{h}(\varsigma) = \log(1 - \varsigma)$, $\mathbf{h}(\varsigma) = -\frac{1}{2}\varsigma^2 - \varsigma$ and $\mathbf{h}(\varsigma) = 1 - \exp(\varsigma)$, the GEL problem reduces to the empirical likelihood, Euclidean likelihood (continuously-updated GMM), and exponential tilting estimators, respectively. See Almeida and Garcia (2009) for further discussion.

The major difficulty in developing the appropriate asymptotic theory for this problem is the nondifferentiability of $(y_t(\gamma) - \lambda' x_t)^+$ at the truncation point. To deal with the non-smoothness of the objective function, we appeal to the empirical process theory (Pollard, 1990; and Andrews, 1994) for deriving the consistency and the asymptotic distributions of the estimators. Heuristically, the main argument is based on the observation that while the function $\partial\phi_t(\theta)/\partial\theta$ may not be continuously differentiable in θ, its expectation typically is due to the smoothing properties of the expectation operator (Andrews, 1994). This allows us to apply the usual Taylor series expansion to the expected value of $\partial\phi_t(\theta)/\partial\theta$ and employ the limit theory for empirical processes.

The main regularity conditions for the consistency and the asymptotic distribution theory are listed as Assumptions A, B, and C in Appendix C.1. They include restrictions on the dependence of the data, identification conditions for the pseudo-true values, and some standard assumptions for deriving the limiting distributions.

Let $\overset{A}{\sim}$ stand for "asymptotically distributed as." The following proposition presents the asymptotic distributions of the sample squared constrained HJ-distance under correctly specified and misspecified models.

Proposition 3. *Under Assumptions A, B, and C,*

(a) if $\delta_+ = 0$,

$$T\hat{\delta}_+^2 \overset{A}{\sim} \sum_{i=1}^{n-k} \xi_i \nu_i, \tag{83}$$

where the ν_i*'s are independent chi-squared random variables with one degree of freedom and the* ξ_i*'s are the eigenvalues of*

$$A = P' U^{-\frac{1}{2}} S U^{-\frac{1}{2}} P, \tag{84}$$

with $S = \sum_{j=-\infty}^{\infty} E\left[(x_t y_t(\gamma^*) - q_{t-1})(x_{t+j} y_{t+j}(\gamma^*) - q_{t+j-1})'\right]$, $D = E\left[x_t \frac{\partial y_t(\gamma^*)}{\partial \gamma'}\right]$, $U = E\left[x_t x_t'\right]$, *and* P *being an* $n \times (n-k)$ *orthonormal matrix whose columns are orthogonal to* $U^{-\frac{1}{2}} D$.

(b) if $\delta_+ > 0$,

$$\sqrt{T}(\hat{\delta}_+^2 - \delta_+^2) \overset{A}{\sim} N(0, v), \tag{85}$$

where $v = \sum_{j=-\infty}^{\infty} E[(\phi_t(\theta^*) - \delta_+^2)(\phi_{t+j}(\theta^*) - \delta_+^2)]$.

Proof. See Appendix C.2.

While Theorem 1 of LXZ appears to present a rather different asymptotic distribution of $T\hat{\delta}_+^2$ under the correctly specified model, it can be shown that the nonzero eigenvalues of their matrix are the same as our ξ_i's once the null hypothesis $H_0 : \delta_+ = 0$ is imposed. It is also important to emphasize that the eigenvalues in part (a) of Proposition 3 coincide with the ones derived by Parker and Julliard (2005) for the unconstrained HJ-distance case. Hence, contrary to LXZ's claim, the existing specification test developed for the sample unconstrained HJ-distance is also applicable to the sample constrained HJ-distance.

To conduct inference, the variance matrices in Proposition 3 should be replaced by consistent estimators. In particular, in part (a), we can replace A with its sample analog

$$\hat{A} = \hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}, \tag{86}$$

where $\hat{U} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t'$ and $\hat{S}$ is obtained using a nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator (see, e.g., Newey and West, 1987; and Andrews, 1991). Similarly, in part (b) we can use a HAC estimator to estimate the variance v.

Proposition 4 below establishes the asymptotic normality of the estimates of the SDF parameters and of the Lagrange multipliers, $\hat{\theta}$, based on the constrained HJ-distance for the case of misspecified models.

Proposition 4. *Let*

$$G_+ = E\left[\left(y_t(\gamma^*) - m_t(\theta^*)^+\right)\frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'} + \frac{\partial y_t(\gamma^*)}{\partial\gamma}\frac{\partial y_t(\gamma^*)}{\partial\gamma'}I_{\{m_t(\theta^*)\leq 0\}}\right], \tag{87}$$

$$D_+ = E\left[x_t\frac{\partial y_t(\gamma^*)}{\partial\gamma'}I_{\{m_t(\theta^*)>0\}}\right], \tag{88}$$

$$U_+ = E\left[x_t x_t' I_{\{m_t(\theta^*)>0\}}\right], \tag{89}$$

where $I_{\{\cdot\}}$ denotes the indicator function. Under Assumptions A, B, and C, and if $\delta_+ > 0$,

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{k+n}, \Sigma_{\hat{\theta}}), \tag{90}$$

where $\Sigma_{\hat{\theta}} = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t\tilde{h}_{t+j}']$ with $\tilde{h}_t = [\tilde{h}_{1t}', \tilde{h}_{2t}']'$ and

$$\tilde{h}_{1t} = (G_+ + D_+'U_+^{-1}D_+)^{-1}\left[(y_t(\gamma^*) - m_t(\theta^*)^+)\frac{\partial y_t(\gamma^*)}{\partial\gamma} + D_+'U_+^{-1}[x_t m_t(\theta^*)^+ - q_{t-1}]\right] \tag{91}$$

$$\tilde{h}_{2t} = U_+^{-1}[D_+\tilde{h}_{1t} - x_t m_t(\theta^*)^+ + q_{t-1}]. \tag{92}$$

Proof. See Appendix C.2.

The variance matrix $\Sigma_{\hat{\theta}}$ in Proposition 4 can be consistently estimated using the sample analogs of (91) and (92). Tests of parameter restrictions based on the Wald or distance metric (likelihood ratio-type) statistics can be easily developed from the result in Proposition 4.

While the asymptotic distribution of $\hat{\gamma}$ in Proposition 4 is valid under the null as well as under the alternative hypotheses, the asymptotic distribution of some elements of $\hat{\lambda}$ is not always normal when $\delta_+ = 0$. To illustrate this, note that when $\delta_+ = 0$, we have $\lambda^* = 0_n$, $m_t(\theta^*)^+ = y_t(\gamma^*)$, $G_+ = 0_{k\times k}$, $D_+ = D$, and $U_+ = U$, where D and U are defined in Proposition 3. As a result, we can simplify $\tilde{h}_{1t}$ and $\tilde{h}_{2t}$ to

$$\tilde{h}_{1t} = (D'U^{-1}D)^{-1}D'U^{-1}(x_t y_t(\gamma^*) - q_{t-1}) \tag{93}$$

$$\tilde{h}_{2t} = [U^{-1}D(D'U^{-1}D)^{-1}D' - I_n]U^{-1}(x_t y_t(\gamma^*) - q_{t-1}). \tag{94}$$

Since $D'\tilde{h}_{2t} = 0_k$, the asymptotic covariance matrix of $\sqrt{T}\hat{\lambda}$ (which we denote by $\Sigma_{\hat{\lambda}}$) is singular when $\delta_+ = 0$. This implies that for a nonzero vector α in the span of the column space of D, $\sqrt{T}\alpha'\hat{\lambda}$ is not asymptotically normal because $\alpha'\tilde{h}_{2t} = 0$. In general, one cannot be sure that each diagonal element of $\Sigma_{\hat{\lambda}}$ is nonzero, so testing $H_0 : \lambda_i = 0$ using the asymptotic normal distribution can be problematic. In some cases, we can explicitly show that a particular diagonal element of $\Sigma_{\hat{\lambda}}$ is zero and one can no longer use the asymptotic normal distribution for testing $H_0 : \lambda_i = 0$.

To better illustrate this problem, we first state a lemma which shows the relation between the Lagrange multipliers and the constrained HJ-distance for linear SDF models.

Lemma 4. *For a linear SDF, we have*

$$E[q_{t-1}]'\lambda^* = -\delta_+^2, \qquad \bar{q}'\hat{\lambda} = -\hat{\delta}_+^2. \tag{95}$$

Proof. See Appendix C.2.

Lemma 4 shows that $\bar{q}'\hat{\lambda}$ is numerically identical to the negative of the sample squared constrained HJ-distance, so they should have the same asymptotic distribution. As an example, we consider a case with $q_{t-1} = q = [1,\ 0'_{n-1}]'$, so that the first payoff is a gross return and the rest of the payoffs are excess returns. In this case, $\hat{\lambda}_1 = q'\hat{\lambda} = -\hat{\delta}_+^2$, where $\hat{\lambda}_1$ is the first element of $\hat{\lambda}$. Therefore, when $\delta_+ = 0$ (or equivalently $\lambda_1 = 0$), $\sqrt{T}\hat{\lambda}_1$ cannot be asymptotically normally

distributed. Instead, $-T\hat{\lambda}_1 = T\hat{\delta}_+^2$ has a weighted chi-squared distribution as given in Proposition 3(a).[13] It can be easily shown that Lemma 4 also works for the unconstrained HJ-distance. This implies that some of the results in Table 5 of Hodrick and Zhang (2001) are incorrect because in their setup, the sample Lagrange multiplier associated with the risk-free asset is equal to the negative of the sample squared unconstrained HJ-distance, rendering the normal test, as suggested after their Equation (26), invalid.[14]

5. Model selection tests based on constrained Hansen-Jagannathan distances

In this section, we refine the asymptotic theory for model comparison tests for strictly non-nested, nested, and overlapping models. Our analysis is similar in spirit to Vuong (1989), Rivers and Vuong (2002), Golden (2003), Marcellino and Rossi (2008), and LXZ's model selection methodology. We provide several improvements upon the results derived in LXZ. First, in the case of nested models, the weights in the asymptotic weighted chi-squared distribution should be either all positive or all negative depending on which model is larger. This restriction is not satisfied by the general expression provided in LXZ which is expected to lead to size distortions in finite samples. In contrast, our limiting expressions and their sample counterparts are guaranteed to satisfy the restrictions imposed by the null hypothesis and the structure of the models. Second, we provide an asymptotic analysis that allows us to deal with the case of two strictly non-nested models that are both correctly specified. Such a case is not considered in LXZ. Third, for nested and overlapping models, we develop chi-squared versions of the model comparison tests that are easier to implement than the weighted chi-squared tests. Finally, our tests are robust to heteroskedasticity and serial correlation of unknown form.[15]

Define models

$$\mathcal{F} = \{y^{\mathcal{F}}(\gamma_{\mathcal{F}}) \, ; \gamma_{\mathcal{F}} \in \Gamma_{\mathcal{F}}\} \tag{96}$$

[13] Our result bears some resemblance to the results in Sims, Stock, and Watson (1990) on higher-order autoregressive (AR) models with a unit root (see also Inoue and Kilian, 2002). These authors find that while most linear combinations of the AR parameters are $\sqrt{T}$-consistent and asymptotically normally distributed, there exists a particular linear combination that converges to a non-normal limiting distribution at rate T. One important difference is that in our context we have k (instead of one) possible linear combinations of $\hat{\lambda}$ that are non-normally distributed and these linear combinations are, in general, not known a priori.

[14] It should be emphasized that when the SDF does not have parameters (as in the case of Proposition 4.1 of Hansen, Heaton, and Luttmer, 1995), then $\sqrt{T}\hat{\lambda}$ has an asymptotic normal distribution even when $\delta_+ = 0$.

[15] Rivers and Vuong (2002), Golden (2003), and Marcellino and Rossi (2008) also propose inference procedures robust to heteroskedasticity and serial correlation in a general nonlinear setup.

and

$$\mathcal{G} = \{y^{\mathcal{G}}(\gamma_{\mathcal{G}}) \, ; \gamma_{\mathcal{G}} \in \Gamma_{\mathcal{G}}\}, \tag{97}$$

where $\gamma_{\mathcal{F}}$ and $\gamma_{\mathcal{G}}$ are k_1 and k_2 parameter vectors, respectively, and $\Gamma_{\mathcal{F}}$ and $\Gamma_{\mathcal{G}}$ denote their corresponding parameter spaces. The constrained HJ-distances for models $\mathcal{F}$ and $\mathcal{G}$ are given by

$$\delta^2_{\mathcal{F},+} = \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} E[\phi^{\mathcal{F}}_t(\theta_{\mathcal{F}})] \tag{98}$$

$$\delta^2_{\mathcal{G},+} = \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} E[\phi^{\mathcal{G}}_t(\theta_{\mathcal{G}})], \tag{99}$$

where $\theta_{\mathcal{F}} = [\gamma'_{\mathcal{F}}, \ \lambda'_{\mathcal{F}}]'$ and $\theta_{\mathcal{G}} = [\gamma'_{\mathcal{G}}, \ \lambda'_{\mathcal{G}}]'$. Denote by $\theta^*_{\mathcal{F}} = [\gamma^{*\prime}_{\mathcal{F}}, \ \lambda^{*\prime}_{\mathcal{F}}]'$ and $\theta^*_{\mathcal{G}} = [\gamma^{*\prime}_{\mathcal{G}}, \ \lambda^{*\prime}_{\mathcal{G}}]'$ the pseudo-true parameters of models $\mathcal{F}$ and $\mathcal{G}$, respectively. Note that model $\mathcal{F}$ can be correctly specified or misspecified depending on whether $y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) \in \mathcal{M}^+$ or $y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) \notin \mathcal{M}^+$. Similarly, $\mathcal{G}$ can be correctly specified or misspecified. If $\mathcal{F} \cap \mathcal{G} = \emptyset$, we have the case of strictly non-nested models. For nested models, we have $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$. Finally, if $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, and $\mathcal{G} \not\subset \mathcal{F}$, we refer to $\mathcal{F}$ and $\mathcal{G}$ as overlapping models.

5.1. The normal test

A simple way of testing $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$ is suggested by Hansen, Heaton, and Luttmer (1995, pp. 255–256) who establish that the difference between the sample squared constrained HJ-distances of two models under $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$ is asymptotically normally distributed:

$$\sqrt{T}(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} N(0, \sigma^2_d), \tag{100}$$

where

$$\sigma^2_d = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}], \tag{101}$$

and $d_t = (\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) - E[\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}})]) - (\phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}}) - E[\phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})])$.

It is worth emphasizing that the above result holds only if $\sigma^2_d \neq 0$. This variance is zero when $\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) = \phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})$. To determine whether the use of the normal test above is appropriate, one could do a pretest of $H_0 : \sigma^2_d = 0$ (see, e.g., Rivers and Vuong, 2002; Golden, 2003; and Marcellino and Rossi, 2008). In our analysis, we prefer to have separate tests of $\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) = \phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})$ since there are two possible reasons for $\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) = \phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})$: (i) the two SDFs are equal, i.e., $y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$, or (ii) the two SDFs are different and are both correctly specified, so that $\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) = 0$ and $\phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}}) = 0$,

which implies $\delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$. By separating the test of $\sigma^2_d = 0$ into these two cases, we provide more information on the underlying reason for the equality and simplify the implementation of the tests.

When the two models are strictly non-nested, we cannot have $y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})$. As a result, we only have to test $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$ before using the normal test. LXZ appear to have missed this possibility in the case of strictly non-nested models and proceed directly to the normal test. In fact, it is possible for two strictly non-nested models to be both correctly specified, and we offer such an example in Appendix C.2.[16] As a result, one cannot ignore this scenario.

When the two models are overlapping, it is possible that both models have the same SDF. Another possibility is that the two models have distinct SDFs but yet both models are correctly specified. As a result, we need to conduct two pre-tests before we can use the normal test.

Finally, for nested models, we only need to test whether the two SDFs are equal. The reason is that, in general, the larger model has a smaller constrained HJ-distance and the only way for the two models to have the same HJ-distance is when their SDFs are equal. We discuss each of these three cases in the following subsections.

5.2. Strictly non-nested models

To test $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$ for strictly non-nested models, we define $e^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) = x_t y^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}}) - q_{t-1}$, $e^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}}) = x_t y^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}}) - q_{t-1}$, and

$$\mathcal{S} \equiv \begin{bmatrix} S_{\mathcal{F}} & S_{\mathcal{F}\mathcal{G}} \\ S_{\mathcal{G}\mathcal{F}} & S_{\mathcal{G}} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E\left[\tilde{e}_t \tilde{e}'_{t+j}\right], \tag{102}$$

where $\tilde{e}_t = [e^{\mathcal{F}}_t(\gamma^*_{\mathcal{F}})',\ e^{\mathcal{G}}_t(\gamma^*_{\mathcal{G}})']'$. Also, let $P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ denote orthonormal matrices with dimensions $n \times (n - k_1)$ and $n \times (n - k_2)$ whose columns are orthogonal to $U^{-\frac{1}{2}} D_{\mathcal{F}}$ and $U^{-\frac{1}{2}} D_{\mathcal{G}}$, respectively, where $D_{\mathcal{F}}$ ($D_{\mathcal{G}}$) is the D matrix for model $\mathcal{F}$ ($\mathcal{G}$) defined in Proposition 3. The following proposition provides the appropriate asymptotic distribution of the difference in the sample squared constrained HJ-distances when both models are correctly specified.

Proposition 5. *Suppose that Assumptions A, B, and C hold for each model and* $y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) \neq y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$.

[16] In a likelihood framework (see Vuong, 1989), two strictly non-nested models cannot be both correctly specified. However, our correctly specified model is defined in terms of moment conditions, so it is possible for two strictly non-nested models to be both correctly specified. See Hall and Pelletier (2008) and Kan and Robotti (2009) for further discussion of this point.

Then, under $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$,

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} \sum_{i=1}^{2n-k_1-k_2} \xi_i \nu_i, \tag{103}$$

where the ν_i*'s are independent chi-squared random variables with one degree of freedom, and the* ξ_i*'s are the eigenvalues of the* $(2n - k_1 - k_2) \times (2n - k_1 - k_2)$ *matrix*

$$\begin{bmatrix} P'_{\mathcal{F}} U^{-\frac{1}{2}} S_{\mathcal{F}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P'_{\mathcal{F}} U^{-\frac{1}{2}} S_{\mathcal{FG}} U^{-\frac{1}{2}} P_{\mathcal{G}} \\ P'_{\mathcal{G}} U^{-\frac{1}{2}} S_{\mathcal{GF}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P'_{\mathcal{G}} U^{-\frac{1}{2}} S_{\mathcal{G}} U^{-\frac{1}{2}} P_{\mathcal{G}} \end{bmatrix}. \tag{104}$$

Proof. See Appendix C.2.

Since the eigenvalues ξ_i's can take on both positive and negative values, the test of the hypothesis $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$ should be two-sided.

In summary, our proposed test of equality of the squared constrained HJ-distances of two strictly non-nested models involves first testing whether the two models are both correctly specified using Proposition 5. If we reject, then we can perform the normal test in Section 5.1. Suppose that α_1 and α_2 are the asymptotic significance levels used in these two tests. Then, our sequential test has a significance level that is asymptotically bounded above by $\max[\alpha_1, \alpha_2]$. Thus, if $\alpha_1 = \alpha_2 = 0.05$, the significance level of this procedure, as a test of $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$, is asymptotically no larger than 5%.

5.3. Nested models

For nested models, σ^2_d is zero by construction under the null of equal constrained HJ-distances. Therefore, the normal test described in Section 5.1 cannot be used. In addition, for nested models, $\delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$ if and only if $y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$, so we can simply test $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$.

Let $H_{\mathcal{F}} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 E[\phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}})]}{\partial\theta_{\mathcal{F}} \partial\theta'_{\mathcal{F}}}$ ($H_{\mathcal{G}}$ is similarly defined) and $\Sigma_{\hat{\theta}_{\mathcal{FG}}} = \sum_{j=-\infty}^{\infty} E[\tilde{h}^{\mathcal{FG}}_t \tilde{h}^{\mathcal{FG}\prime}_{t+j}]$ with $\tilde{h}^{\mathcal{FG}}_t = [\tilde{h}^{\mathcal{F}\prime}_t, \tilde{h}^{\mathcal{G}\prime}_t]'$, where $\tilde{h}_t$ for each model is defined in Proposition 4. For nested and overlapping models, LXZ show that under $H_0 : \phi^{\mathcal{F}}_t(\theta^*_{\mathcal{F}}) = \phi^{\mathcal{G}}_t(\theta^*_{\mathcal{G}})$,

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} \sum_{i=1}^{2n+k_1+k_2} \xi_i \nu_i, \tag{105}$$

where the ξ_i's are the eigenvalues of[17]

$$\frac{1}{2}\begin{bmatrix} -H_{\mathcal{F}} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}} \end{bmatrix}\Sigma_{\hat{\theta}_{\mathcal{FG}}}. \tag{107}$$

In the ensuing analysis, we will show that under $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$, some of the ξ_i's are equal to zero and the nonzero ξ_i's have the same sign. Being unaware of such simplifications, LXZ use the sample counterpart of the matrix in (107) to estimate the ξ_i's and end up with more nonzero estimated ξ_i's than the theory suggests. In addition, their estimated ξ_i's do not have the same sign. This is problematic because for nested models, the larger model has a smaller constrained sample HJ-distance by construction. By not imposing the constraints that the ξ_i's should have the same sign, LXZ end up comparing $T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+})$ with a distribution that can take on both positive and negative values. This could result in serious finite sample distortions of the test.

Without loss of generality, we assume $\mathcal{F} \subset \mathcal{G}$. Suppose that the null hypothesis $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$ can be written as a parametric restriction of the form $H_0 : \psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_1}$ for model $\mathcal{G}$ against $H_1 : \psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) \neq 0_{k_2-k_1}$, where $\psi(\cdot)$ is a twice continuously differentiable function in its argument. Define

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma'_{\mathcal{G}}} \tag{108}$$

as a $(k_2-k_1)\times k_2$ derivative matrix of the parametric restrictions $\psi_{\mathcal{G}}$. For many models of interest, $y^{\mathcal{F}}(\gamma_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma_{\mathcal{G}})$ when a subset of the parameters of model $\mathcal{G}$ is equal to zero (or a constant vector c). In this case, we can rearrange the parameters such that $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1},\ I_{k_2-k_1}]\gamma_{\mathcal{G}} - c$. Then, $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1},\ I_{k_2-k_1}]$, which is a selector matrix that selects only the part of the parameter vector $\gamma_{\mathcal{G}}$ that is not contained in model $\mathcal{F}$. Also, let $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$ be the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{G}}$ given in Proposition 4, $M_{\mathcal{G}} \equiv \lim_{T\to\infty} \text{Var}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \phi_t^{\mathcal{G}}(\theta^*_{\mathcal{G}})}{\partial \theta_{\mathcal{G}}}\right]$, and $\tilde{H}_{\mathcal{G}} = (G^{\mathcal{G}}_{+} + D^{\mathcal{G}\prime}_{+}U^{-1}_{+}D^{\mathcal{G}}_{+})^{-1}$, where the matrices G_+, D_+, and U_+ are defined in Proposition 4.

Proposition 6 below presents the asymptotic distribution of $T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+})$ under the null hypothesis $H_0 : \psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_1}$.

[17] It is straightforward to verify that the eigenvalues of the matrix in (107) are identical to the ones of the matrix presented in LXZ. Alternatively, one can directly test $H_0 : \sigma^2_d = 0$. It can be shown that under $H_0 : \sigma^2_d = 0$, we have

$$T\hat{\sigma}^2_d \overset{A}{\sim} \sum_{i=1}^{2n+k_1+k_2} 4\xi_i^2\nu_i, \tag{106}$$

where $\hat{\sigma}^2_d$ is the sample version of σ^2_d. The proof of this result is available upon request.

Proposition 6. *Suppose that Assumptions A, B, and C hold and* $\mathcal{F} \subset \mathcal{G}$. *Then, under* H_0 : $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_1}$,

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} \sum_{i=1}^{k_2-k_1} \xi_i \nu_i, \tag{109}$$

where the ν_i*'s are independent chi-squared random variables with one degree of freedom, the* ξ_i*'s are the eigenvalues of* $(\Psi^{\mathcal{G}}_* \tilde{H}_{\mathcal{G}} \Psi^{\mathcal{G}\prime}_*)^{-1} \Psi^{\mathcal{G}}_* \Sigma_{\hat{\gamma}_{\mathcal{G}}} \Psi^{\mathcal{G}\prime}_*$, $\Psi^{\mathcal{G}}_* \equiv \Psi^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$, *and the* ξ_i*'s are all positive.*

Proof. See Appendix C.2.

Proposition 6 shows that, under $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$, only $k_2 - k_1$ of the eigenvalues of (107) are nonzero and they all have the same sign. In practice, we need to estimate the ξ_i's to construct the test. Using the sample version of the matrix in Proposition 6 instead of the sample version of the matrix in (107) to estimate the ξ_i's results in a substantial reduction of the number of estimated eigenvalues. In addition, the resulting estimated eigenvalues are guaranteed to be positive.

An alternative way of testing the equality of two nested SDFs is to test directly $H_0 : \psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_1}$ using the Wald test

$$T\hat{\psi}'_{\mathcal{G}}(\hat{\Psi}^{\mathcal{G}} \hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}} \hat{\Psi}^{\mathcal{G}\prime})^{-1} \hat{\psi}_{\mathcal{G}} \overset{A}{\sim} \chi^2_{k_2-k_1}, \tag{110}$$

where $\hat{\psi}_{\mathcal{G}} = \psi_{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})$, $\hat{\Psi}^{\mathcal{G}} = \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})$, and $\hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}}$ is a consistent estimator of the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{G}}$. The test in (110) is asymptotically pivotal and is easier to implement than the test in Proposition 6.

5.4. Overlapping models

For overlapping models, the variance σ^2_d can be zero when (i) $y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$ or (ii) both models are correctly specified.[18] Since Proposition 5 is applicable to this second scenario, here we only need to derive the test of $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$.

It is well known that for linear models, the equality of the SDFs implies zero restrictions on the parameter vectors (see, e.g., Lien and Vuong, 1987; and Kan and Robotti, 2009). Similar restrictions can also be obtained for nonlinear models. Let $y^{\mathcal{H}}(\gamma_{\mathcal{H}})$ be the SDF of model $\mathcal{H}$, where $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ and $\gamma_{\mathcal{H}}$ is a k_3-vector. Therefore, $y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$ implies $y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{H}}(\gamma^*_{\mathcal{H}})$ and $y^{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = y^{\mathcal{H}}(\gamma^*_{\mathcal{H}})$. Suppose that $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{H}}(\gamma^*_{\mathcal{H}})$ and $y^{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = y^{\mathcal{H}}(\gamma^*_{\mathcal{H}})$ can be written as

[18] Similar to the case of strictly non-nested models, it is possible for two overlapping positive SDFs to be both correctly specified. Examples are available upon request.

a parametric restriction of the form $H_0 : \psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$, where $\psi_{\mathcal{F}}(\cdot)$ and $\psi_{\mathcal{G}}(\cdot)$ are some twice continuously differentiable functions of their arguments. Let

$$\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{\partial \psi_{\mathcal{F}}(\gamma_{\mathcal{F}})}{\partial \gamma'_{\mathcal{F}}} \tag{111}$$

and

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma'_{\mathcal{G}}} \tag{112}$$

be $(k_1 - k_3) \times k_1$ and $(k_2 - k_3) \times k_2$ derivative matrices of the parametric restrictions $\psi_{\mathcal{F}}$ and $\psi_{\mathcal{G}}$, respectively. In many cases, $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{H}}(\gamma^*_{\mathcal{H}})$ implies that a subset of the parameters of model $\mathcal{F}$ is equal to zero, and $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{G}}) = y^{\mathcal{H}}(\gamma^*_{\mathcal{H}})$ implies that a subset of the parameters of model $\mathcal{G}$ is equal to zero. For such cases, we can arrange the parameters so that $\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = [0_{(k_1-k_3)\times k_3},\ I_{k_1-k_3}]$ and $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_3)\times k_3},\ I_{k_2-k_3}]$. Let $\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}$ the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{F}\mathcal{G}} = [\hat{\gamma}_{\mathcal{F}}',\ \hat{\gamma}_{\mathcal{G}}']'$, $\tilde{H}_{\mathcal{F}} = (G^{\mathcal{F}}_+ + D^{\mathcal{F}\prime}_+ U^{-1}_+ D^{\mathcal{F}}_+)^{-1}$, $\tilde{H}_{\mathcal{G}} = (G^{\mathcal{G}}_+ + D^{\mathcal{G}\prime}_+ U^{-1}_+ D^{\mathcal{G}}_+)^{-1}$, and

$$\Psi^{\mathcal{F}\mathcal{G}}_* \equiv \begin{bmatrix} \Psi^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi^{\mathcal{G}}(\gamma^*_{\mathcal{G}}) \end{bmatrix}. \tag{113}$$

The next proposition establishes the asymptotic distribution of $T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+})$ under $H_0 : \psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$.

Proposition 7. *Suppose that $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, $\mathcal{G} \not\subset \mathcal{F}$, and Assumptions A, B, and C hold. Then, under $H_0 : \psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$, we have*

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} \sum_{i=1}^{k_1+k_2-2k_3} \xi_i \nu_i, \tag{114}$$

where the ν_i's are independent chi-squared random variables with one degree of freedom and the ξ_i's are the eigenvalues of

$$\begin{bmatrix} -(\Psi^{\mathcal{F}}_* \tilde{H}_{\mathcal{F}} \Psi^{\mathcal{F}\prime}_*)^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi^{\mathcal{G}}_* \tilde{H}_{\mathcal{G}} \Psi^{\mathcal{G}\prime}_*)^{-1} \end{bmatrix} \Psi^{\mathcal{F}\mathcal{G}}_* \Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}} \Psi^{\mathcal{F}\mathcal{G}\prime}_*, \tag{115}$$

where $\Psi^{\mathcal{F}}_* \equiv \Psi^{\mathcal{F}}(\gamma^*_{\mathcal{F}})$ and $\Psi^{\mathcal{G}}_* = \Psi^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$

Proof. See Appendix C.2.

Unlike the case of nested models, the eigenvalues in Proposition 7 are not always positive because $\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}$ can take on both positive and negative values. As a result, we need to perform

a two-sided test of $H_0 : y^{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = y^{\mathcal{G}}(\gamma^*_{\mathcal{G}})$. Similarly to the nested case, an alternative way of testing the equality of two overlapping SDFs is to directly test $H_0 : \psi_{\mathcal{F}}(\gamma^*_{\mathcal{F}}) = 0_{k_1-k_3}$, $\psi_{\mathcal{G}}(\gamma^*_{\mathcal{G}}) = 0_{k_2-k_3}$ using the Wald test

$$T\hat{\psi}'_{\mathcal{FG}}(\hat{\Psi}^{\mathcal{FG}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}\hat{\Psi}^{\mathcal{FG}\prime})^{-1}\hat{\psi}_{\mathcal{FG}} \overset{A}{\sim} \chi^2_{k_1+k_2-2k_3}, \tag{116}$$

where $\hat{\psi}_{\mathcal{FG}} = [\psi_{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}})',\ \psi_{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})']'$,

$$\hat{\Psi}^{\mathcal{FG}} = \begin{bmatrix} \Psi^{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}}) & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}}) \end{bmatrix}, \tag{117}$$

and $\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}$ is a consistent estimator of $\Sigma_{\hat{\gamma}_{\mathcal{FG}}}$. This test is asymptotically pivotal and is easier to implement than the test in Proposition 7.

In summary, our proposed test of equality of the squared constrained HJ-distances of two overlapping models involves first testing whether the SDFs of the two models are equal using either Proposition 7 or (116). If we reject the null hypothesis, then we need to test whether the two models are both correctly specified using Proposition 5. Finally, if we still reject, we can perform the normal test in Section 5.1. The significance level of this procedure, as a test of $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$, is asymptotically bounded above by $\max[\alpha_1, \alpha_2, \alpha_3]$, where α_1, α_2, and α_3 are the asymptotic significance levels used in these three tests.

5.5. Discussion

The results in Propositions 6 and 7 (and the related chi-squared tests in (110) and (116)) offer substantial advantages over the inference procedure in (105)–(107). Imposing the parametric restrictions that directly arise from the structure of the model and the appropriate null hypotheses results in a drastic reduction of the number of weights that are used to compute the critical values of the tests. More specifically, the number of eigenvalues in the weighted chi-squared distribution is reduced from $2n + k_1 + k_2$ to $k_2 - k_1$ for nested and to $k_1 + k_2 - 2k_3$ for overlapping models. This proves to be particularly advantageous when the number of test assets n is large. The reduced dimensions of the matrices in Propositions 6 and 7 are expected to lead to improved finite sample (size and power) behavior of the model selection tests.

While our Sections 4 and 5 provide an exhaustive asymptotic treatment of specification and model selection tests based on sample constrained HJ-distances, it is beyond the scope of this paper to investigate the finite sample properties of such tests. However, it is worth mentioning that

some finite sample evidence is available in the current literature. In their simulation experiments, LXZ examine the size and power properties of model specification and selection tests. Overall, they find that the tests based on the unconstrained and constrained HJ-distances behave similarly, which provides little support to their claims that tests based on the constrained HJ-distance are more powerful than tests based on the unconstrained HJ-distance in detecting and distinguishing misspecified models. Nevertheless, it is important to mention that their simulation findings cannot be used to draw strong conclusions one way or the other since they suffer from the following problems. First, their analysis of the size properties of the specification test based on the constrained HJ-distance is incorrect because, by generating SDFs that are normally distributed, they do not impose the null of zero constrained HJ-distance. In fact, it is impossible to impose the null of zero constrained HJ-distance when the SDF has a normal distribution because the SDF takes on negative values by definition. Second, in all experiments, their power analyses under the alternative hypotheses are not size-adjusted, thus rendering the comparisons between the properties of tests based on the unconstrained and constrained HJ-distances problematic. Finally, in some of their simulation experiments, LXZ adopt a useless factor setting which is inappropriate for evaluating their methodologies.[19] In fact, for the unconstrained HJ-distance, the presence of a useless factor violates their Assumption A.1 which requires the parameters to be uniquely identified (see our discussion in Section 3.2).

6. Empirical analysis

In this section, we focus on linear asset pricing models because of their popularity in the literature and the fact that their SDFs can potentially take on negative values, making it interesting to study the difference between the unconstrained and constrained HJ-distances of these models.

6.1. Data and asset pricing models

For ease of comparison, we focus on the same asset pricing models considered by LXZ and use their data to perform our empirical analysis.[20] The return data consist of quarterly gross returns on the three-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios. The data are from 1952:2 to 2000:4 (195 quarterly observations). The seven models that are considered

[19] See, e.g., Kan and Zhang (1999a,b) and Burnside (2007) for analyses of asset pricing tests with useless factors.

[20] We thank LXZ for making their data available to us and refer to their paper for a more detailed description of the data.

are:

LL: the conditional consumption CAPM of Lettau and Ludvigson (2001)

$$y_t^{LL} = \gamma_0 + \gamma_1 cay_{t-1} + \gamma_2 c_t + \gamma_3 cay_{t-1} c_t,$$

where cay is the consumption-wealth ratio and c is the log consumption growth rate;

LV: a version of the conditional consumption CAPM of Lustig and Van Nieuwerburgh (2004)

$$y_t^{LV} = \gamma_0 + \gamma_1 my_{t-1} + \gamma_2 c_t + \gamma_3 my_{t-1} c_t,$$

where my is the housing collateral ratio;

SV: the conditional CAPM of Santos and Veronesi (2006)

$$y_t^{SV} = \gamma_0 + \gamma_1 r_{mkt,t} + \gamma_2 s_{t-1}^{\omega} r_{mkt,t},$$

where r_{mkt} is the excess return on the market portfolio and s^{ω} is the labor income-consumption ratio;

LVX1: the simple sector investment model of Li, Vassalou, and Xing (2006)

$$y_t^{LVX1} = \gamma_0 + \gamma_1 i_{hh,t} + \gamma_2 i_{corp,t} + \gamma_3 i_{ncorp,t},$$

where i_{hh}, i_{corp}, and i_{ncorp} are the log investment growth rates for households, non-financial corporations, and non-corporate sector, respectively;

LVX2: the extended sector investment model of Li, Vassalou, and Xing (2006)

$$y_t^{LVX2} = \gamma_0 + \gamma_1 i_{hh,t} + \gamma_2 i_{corp,t} + \gamma_3 i_{ncorp,t} + \gamma_4 i_{fcorp,t} + \gamma_5 i_{fm,t},$$

where i_{fcorp} and i_{fm} are the log investment growth rates for financial corporations and farm sector, respectively;[21]

[21] Li, Vassalou, and Xing (2006) eliminate the i_{fm} variable from their tests due to problems with missing observations from the data series. Although the i_{fm} series provided to us by LXZ has no missing observations, it contains large outliers. Therefore, in our empirical analysis we also consider a version of LVX2 without the i_{fm} variable. We find that all of our conclusions are unaffected by this alternative specification of LVX2. The results of this additional analysis are available upon request.

YOGO: the durable consumption CAPM of Yogo (2006)

$$y_t^{YOGO} = \gamma_0 + \gamma_1 c_{ndur,t} + \gamma_2 c_{dur,t} + \gamma_3 r_{mkt,t},$$

where c_{ndur} and c_{dur} denote the log consumption growth rates of non-durable and durable goods, respectively;

FF3: the three-factor model of Fama and French (1993)

$$y_t^{FF3} = \gamma_0 + \gamma_1 r_{mkt,t} + \gamma_2 r_{smb,t} + \gamma_3 r_{hml,t},$$

where r_{smb} is the return difference between portfolios of small and large stocks and r_{hml} is the return difference between portfolios of high and low book-to-market ratios.

6.2. Results

Table 1 presents the sample unconstrained ($\hat{\delta}$) and constrained ($\hat{\delta}_+$) HJ-distances (Panels A and B, respectively) of the seven linear asset pricing models considered.[22] The table also reports the standard errors of $\hat{\delta}$ and $\hat{\delta}_+$ (assuming $\delta > 0$ and $\delta_+ > 0$), the p-values for the tests of $H_0 : \delta = 0$ and $H_0 : \delta_+ = 0$, and the probability that the estimated SDF takes on negative values in the sample. For the specification test based on the sample constrained HJ-distance, we report two different p-values, p_n and p_a, where p_n stands for the p-value that imposes the null hypothesis $H_0 : \delta_+ = 0$, and p_a stands for the p-value that does not impose the null hypothesis (see Proposition 3(a) and related discussion). In order to allow for serial correlation in the data, the reported standard errors and p-values are constructed using the Newey and West (1994) HAC estimator with Bartlett kernel and automatic lag selection. In addition to the standard deviations of the estimated SDF and ϕ function, each panel also reports the centered R^2 from a linear regression of the estimated SDF on the returns on the test assets. The second last row of Panel B presents the percentage difference between the sample constrained and unconstrained HJ-distances of each model. Finally, in the last row of Panel B, we report $\hat{\tilde{\delta}}$ for the SDF that minimizes the constrained HJ-distance, where $\tilde{\delta}$, defined in (65), is a measure of the maximum pricing error on the test assets.

Table 1 about here

[22] For the unconstrained HJ-distance and related tests, we refer the readers to Kan and Robotti (2009).

While some of the results in Table 1 are already reported and discussed in LXZ, we would like to emphasize and reinterpret several important findings that naturally emerge from the predictions of our theoretical analyses in Sections 2 and 3. First, Table 1 clearly shows that the largest increases in the sample constrained HJ-distance over its unconstrained counterpart occur for models with high probabilities of taking on negative values (such as LV, LVX1, and LVX2). For these models, the probability for their SDF to take on negative values can be greatly reduced when the parameters are chosen to minimize the constrained HJ-distance. However, as discussed in Section 3.2, this reduction in probability generally comes at the cost of higher pricing errors on the test assets (a higher $\tilde{\delta}$). For models with large differences between $\hat{\delta}_+$ and $\hat{\delta}$ (such as LV, LVX1, and LVX2), we also see a significant difference between $\hat{\tilde{\delta}}$ and $\hat{\delta}$, indicating a substantial deterioration in the ability of the SDF to price the test assets when its parameters are chosen to minimize the constrained HJ-distance instead of the unconstrained HJ-distance. The deterioration in the pricing ability of LV, LVX1, and LVX2 is also reflected in the standard deviations of their SDFs, which significantly drop from 0.817, 1.229, and 1.478 in Panel A to 0.294, 0.318, and 0.350 in Panel B. This implies that it would be even harder for these models to satisfy the sample HJ-bounds if their parameters were chosen to minimize the constrained HJ-distance.

Second, the variation in the differences between the sample unconstrained and constrained HJ-distances across models deserves some remarks. As expected from our theoretical analysis, the differences are relatively small for SDFs that are close to being spanned by the returns on the test assets. For example, the percentage difference between the sample constrained and unconstrained HJ-distances of FF3 is only 4.2% since this model has a very high R^2 of 0.983. While we do not report the parameter estimates of the various models to preserve space, the pattern of the differences in the parameter estimates provides further support to our theoretical predictions: the largest differences in parameter estimates arise in models with non-traded factors and almost no differences arise in models with traded factors.[23] As a result, the different effects of imposing the no-arbitrage constraint across models appear to be driven by the underlying structure of the problem and characteristics of the factors (traded versus non-traded).

Finally, while choosing the SDF parameters to minimize $\hat{\delta}_+$ instead of $\hat{\delta}$ leads to a lower probability for the SDF to take on negative values and yields a less volatile SDF, the standard error of $\hat{\delta}_+$

[23]The full set of parameter estimates and their standard errors are available upon request.

is not always smaller than the standard error of $\hat{\delta}$. In fact, we find that for five out of seven models, the standard error of $\hat{\delta}_+$ is larger than the standard error of $\hat{\delta}$. This suggests that the sample unconstrained and constrained HJ-distances are similarly noisy measures of model misspecification and that it could be difficult to differentiate between models using either metric.

Table 2 about here

Table 2 presents our model comparison tests based on the unconstrained and constrained HJ-distances. Two of the considered models are nested (LVX1 and LVX2) and the remaining models are overlapping.[24] We use Propositions 5, 6, 7, and the normal test described in Section 5.1 to perform our sequential tests. It should be emphasized that Propositions 6 and 7 substantially reduce the dimension of the matrix used by LXZ to compute the eigenvalues for the weighted chi-squared tests. In particular, the number of eigenvalues is reduced from 59–62 (in the deficient rank matrix (107)) to 2–8. This is expected to improve considerably the finite sample properties of our tests and explains the (substantial, in some cases) differences between the p-values reported in Table 2 and those in LXZ.

Panel A of Table 2 shows that when comparing FF3 with other models, the test of $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$ based on the unconstrained HJ-distance rejects the equality of SDFs in all cases except two (LVX1 and LVX2). In the other model comparisons, the test based on the unconstrained HJ-distance detects no statistically significant differences across models. Consistent with our previous discussion, imposing the no-arbitrage constraint has the largest effect on models with a high probability of taking on negative values (LV, LVX1, and LVX2). Interestingly, the number of rejections of equality of two non-nested SDFs significantly increases when considering the test of $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$ based on the constrained HJ-distance. To gain some intuition of why this is the case, note that the test of $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$ is essentially a test of joint significance of the parameter vectors of the two models, as shown by the expression of the chi-squared test in (116). As discussed in Sections 2 and 3, imposing the no-arbitrage constraint on unspanned SDFs drastically reduces their volatility and makes them behave more like spanned SDFs by putting more weight on the risk-free asset and less weight on the unspanned factors (this is reflected in the higher R^2s of the estimated SDFs in Table 1.B). For such SDFs, it can be shown that the standard errors of the parameter estimates

[24] Since all models are linear and have a constant term in common, we do not have to deal with the case of strictly non-nested models.

shrink proportionately more than the estimates themselves, thus making the chi-squared test in (116) bigger and more likely to reject the null hypothesis $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$.

Since the test in Panel A only represents the first stage of our sequential test, in Panel B we test $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ ($H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$) for those pairs of non-nested models for which we reject $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$ at the 5% level. For the unconstrained HJ-distance, we find that the null is rejected at the 5% level in two out of four cases. For the constrained HJ-distance, the null is rejected at the 5% level for all models that are paired with FF3. For these pairs of models, we need to further investigate whether they have the same constrained HJ-distance using the normal test.

Panel C of Table 2 presents the results from the normal test in Section 5.1 for the pairs of models for which the null hypotheses in Panels A and B are rejected at the 5% level. This represents the last stage of our sequential test procedure. The normal test delivers only one rejection of the null hypothesis for both the unconstrained and constrained HJ-distances at the 5% level. In particular, FF3 only outperforms YOGO when using the unconstrained HJ-distance, while it only outperforms LV when using the constrained HJ-distance.[25] Although we find the same number of rejections for the unconstrained and constrained HJ-distances at the 5% level, there appears to be slightly stronger evidence that FF3 outperforms the other models when we do model comparison using the constrained HJ-distance. This is particularly the case if we increase the size of the test to 10%, which then leads to more rejections of equality in the constrained HJ-distance case (six) than in the unconstrained HJ-distance case (two). This result requires some explanation. First, as pointed out earlier, the differences between the sample squared HJ-distance of FF3 and the sample squared HJ-distances of the other models are larger in Table 1.B than in Table 1.A. Second, since the various SDFs are closer to being spanned in the constrained HJ-distance case, their ϕ's exhibit a stronger positive correlation with the ϕ of the almost spanned FF3. Given that the standard deviations of the ϕ's are similar across Panels A and B of Table 1, the standard errors of the differences between the sample squared HJ-distance of FF3 and the sample squared HJ-distances of the other six models tend to shrink when imposing the no-arbitrage constraint. This implies that in the constrained HJ-distance case, the t-ratios of these differences are higher and the corresponding p-values lower.

[25] Ignoring serial correlation in the data produces the same number of rejections as in LXZ. At the 5% level, we obtain three rejections of equality using the constrained HJ-distance and one rejection of equality using the unconstrained HJ-distance.

While our empirical results show that it is slightly easier to reject the hypothesis of equality of constrained HJ-distances than the hypothesis of equality of unconstrained HJ-distances for this dataset, it is important to understand how robust this finding is. Our theoretical results suggest that we can expect some meaningful differences between the unconstrained and constrained HJ-distances only when the Sharpe ratio of the tangency portfolio of the risky assets is very high. As it turns out, the tangency portfolio of the 25 Fama-French size and book-to-market portfolios used by LXZ has a relatively high sample Sharpe ratio (0.71). To understand whether this is an important reason for the difference in results between the two tests, we consider another set of test assets with a smaller sample Sharpe ratio. Specifically, we use quarterly gross returns on the three-month T-bill and ten size and 12 industry portfolios from Kenneth French's website.[26] The sample Sharpe ratio of the tangency portfolio of this new set of risky assets is 0.53. We then perform the same analyses as in Tables 1 and 2, leaving the sample period and the models unchanged (the results of this exercise are available upon request). Using this new set of test assets, we find that the specification tests based on the unconstrained and constrained HJ-distances reject all models but one (YOGO in the unconstrained case and LV in the constrained case). In addition, our model selection tests based on the unconstrained and constrained HJ-distances suggest that no model outperforms its competitors at any conventional significance level. Therefore, consistent with our theoretical results, a decrease in the Sharpe ratio of the tangency portfolio of the risky assets causes the model comparison tests based on unconstrained and constrained HJ-distances to behave similarly.

In summary, our empirical analysis suggests that one can find a nontrivial difference between the tests of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}}$ and $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$ only when at least one of the SDFs has a large unspanned component and the Sharpe ratio of the tangency portfolio of the test assets is high. Nevertheless, we would like to emphasize that since $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}}$ and $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+}$ are in general two different hypotheses, it is not meaningful to compare the power of the two tests. In particular, the choice of the model comparison test should depend on whether the research interest lies in the unconstrained or the constrained HJ-distances of a model, and should not be dictated by which test is more likely to reject its corresponding null hypothesis.

7. Conclusion

[26] Considering size and industry portfolios in addition to or instead of the 25 Fama-French size and book-to-market portfolios is consistent with one of the prescriptions of Lewellen, Nagel, and Shanken (2009).

This paper fills an important gap in the current literature by examining the population and sampling properties of the HJ-distance with a no-arbitrage constraint. We first clarify the maximum pricing error interpretation of the constrained HJ-distance. Unlike the unconstrained HJ-distance which is a measure of the maximum pricing error of an SDF on the test assets, the constrained HJ-distance does not represent the maximum pricing error of an SDF on all the tradable assets. In a complete market, the constrained HJ-distance is only a lower bound on the maximum pricing error.

Since a model with a smaller lower bound on the maximum pricing error does not necessarily have a smaller actual maximum pricing error, ranking models using the constrained HJ-distance can be problematic. However, when the SDF is spanned by the returns on the test assets, we show that ranking models using the constrained HJ-distance is the same as ranking models using the unconstrained HJ-distance. The reason is that in the spanned SDF case, the difference between the constrained and unconstrained HJ-distances becomes model independent and coincides with the difference between the constrained and unconstrained HJ-bounds. The rankings of models using the two HJ-distances can differ only when at least one of the SDFs is far from being spanned by the returns on the test assets.

When the SDF is not spanned by the returns on the test assets, we derive an analytical solution for the constrained HJ-distance, the associated Lagrange multipliers, and the SDF parameters in the case of linear SDFs under a joint normality assumption on the SDF and the returns. This allows us to show that nontrivial differences between the constrained and unconstrained HJ-distances can only arise when the volatility of the unspanned component of an SDF is large and the Sharpe ratio of the tangency portfolio of the test assets is very high. In addition, our analysis allows us to quantify the deterioration in the ability of a given linear SDF to price the test assets when imposing a no-arbitrage constraint.

In our econometric analysis, we document the surprising finding that in finite samples, the sample constrained HJ-distance takes on the value of infinity with positive probability. In particular, we show that the sample constrained HJ-distance takes on the value of infinity if and only if we can identify an in-sample arbitrage portfolio of the test assets. When such a portfolio is identified, the sample constrained HJ-distance cannot be used to rank models. Furthermore, we show that contrary to common belief, the asymptotic normality of the estimated Lagrange multipliers can

break down when the asset pricing model is correctly specified. Finally, we provide a complete limiting theory for estimation, testing, and comparison of SDFs using the constrained HJ-distance.

Our tests based on the unconstrained and constrained HJ-distances are used to analyze a variety of asset pricing models that have been proposed in the literature. With the exception of the three-factor model of Fama and French (1993), we do not find any evidence to conclude that one model outperforms the others. This is consistent with recent findings based on the sample unconstrained HJ-distance (Kan and Robotti, 2009) and the sample cross-sectional R^2 (Kan, Robotti, and Shanken, 2009). In addition, we reexamine the empirical results of Li, Xu, and Zhang (2009) in light of our new theoretical results. Their main conclusion that it is easier to differentiate between models when using the constrained HJ-distance instead of the unconstrained HJ-distance appears to be heavily driven by the very high Sharpe ratio of the test assets that they use.

Our analysis could be extended in a number of ways. For example, the issue of model selection when the SDF does not depend on parameters deserves further investigation. In addition, other metrics for comparing models could be considered. Finally, although we have made substantial progress in deriving asymptotic results, future research should also address the small sample properties of the various estimators and test statistics considered in this paper.

Appendix A

We first provide a lemma which will be repeatedly used in Appendix A.

Lemma A.1 *Suppose* (u, v) *is bivariate normally distributed with* $\mu_u = E[u]$, $\mu_v = E[v]$, $\sigma_u^2 = \mathrm{Var}[u]$, $\sigma_v^2 = \mathrm{Var}[v]$, *and* $\sigma_{uv} = \mathrm{Cov}[u, v]$. *Then, we have*

$$E[uv^+] = E[uv]\Phi(\eta) + \mu_u\sigma_v\phi(\eta), \tag{A1}$$

where $\eta = \mu_v/\sigma_v$, *and* $\phi(\cdot)$ *and* $\Phi(\cdot)$ *are the density and cumulative distribution functions of a standard normal random variable, respectively.*

Proof of Lemma A.1: Lemma 4 of Kan and Robotti (2008) provides the following explicit expressions of $E[v^+]$ and $E[v^{+2}]$:

$$\begin{aligned} E[v^+] &= \mu_v\Phi(\eta) + \sigma_v\phi(\eta), & \text{(A2)} \\ E[v^{+2}] &= (\mu_v^2 + \sigma_v^2)\Phi(\eta) + \mu_v\sigma_v\phi(\eta). & \text{(A3)} \end{aligned}$$

Under the bivariate normality assumption, we have

$$E[u|v] = \mu_u + \frac{\sigma_{uv}}{\sigma_v^2}(v - \mu_v). \tag{A4}$$

It follows that

$$\begin{aligned} E[uv^+] &= E[E[u|v]v^+] \\ &= E\left[\left(\mu_u + \frac{\sigma_{uv}}{\sigma_v^2}(v - \mu_v)\right)v^+\right] \\ &= \left(\mu_u - \frac{\sigma_{uv}}{\sigma_v^2}\mu_v\right)E[v^+] + \frac{\sigma_{uv}}{\sigma_v^2}E[v^{+2}] \\ &= \left(\mu_u - \frac{\sigma_{uv}}{\sigma_v^2}\mu_v\right)[\mu_v\Phi(\eta) + \sigma_v\phi(\eta)] + \frac{\sigma_{uv}}{\sigma_v^2}\left[(\mu_v^2 + \sigma_v^2)\Phi(\eta) + \mu_v\sigma_v\phi(\eta)\right] \\ &= E[uv]\Phi(\eta) + \mu_u\sigma_v\phi(\eta). \end{aligned} \tag{A5}$$

Note that the above expression can also be rewritten as

$$E[uv^+] = \sigma_{uv}\Phi(\eta) + \mu_u E[v^+]. \tag{A6}$$

This completes the proof.

Proof of Proposition 1: Let $m_y = y - \tilde{\lambda}'\tilde{r} \sim N(\mu_m, \sigma_m^2)$. Invoking Lemma A.1, we obtain

$$E[\tilde{r}m_y^+] = E[\tilde{r}m_y]\Phi(\eta) + E[\tilde{r}]\sigma_m\phi(\eta) = (E[\tilde{r}y] - U\tilde{\lambda})\Phi(\eta) + E[\tilde{r}]\sigma_m\phi(\eta), \tag{A7}$$

where $\eta = \mu_m/\sigma_m$. Using the first order condition $E[\tilde{r}m_y^+] = q$ and the expression of U^{-1} in (39), we obtain $U^{-1}E[\tilde{r}] = q/R_0$ and

$$\begin{aligned} \tilde{\lambda} &= U^{-1}E[\tilde{r}y] + \frac{\phi(\eta)}{\Phi(\eta)}\sigma_m U^{-1}E[\tilde{r}] - \frac{1}{\Phi(\eta)}U^{-1}q \\ &= \begin{bmatrix} \frac{\mu_y - V_{ry}'V_{rr}^{-1}\mu_r}{R_0} \\ V_{rr}^{-1}V_{ry} \end{bmatrix} + \begin{bmatrix} \frac{\sigma_m\phi(\eta)}{R_0\Phi(\eta)} \\ 0_N \end{bmatrix} - \frac{1}{\Phi(\eta)}\begin{bmatrix} \frac{1+a}{R_0^2} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} \end{bmatrix}. \end{aligned} \tag{A8}$$

Instead of directly solving this system of n nonlinear equations to obtain $\tilde{\lambda}$, we seek a way to first solve for η. Using (A8) and after simplification, we have

$$m_y = y - \tilde{\lambda}'\tilde{r} = z + \frac{1}{\Phi(\eta)}\left[\frac{1}{R_0} - \frac{\mu_r'V_{rr}^{-1}(r - \mu_r)}{R_0}\right] - \frac{\sigma_m\phi(\eta)}{\Phi(\eta)}, \tag{A9}$$

and the variance of m_y is given by

$$\sigma_m^2 = \sigma_z^2 + \frac{a}{R_0^2\Phi(\eta)^2}. \tag{A10}$$

In addition, we can use (A2) to express $E[m_y^+]$ as

$$E[m_y^+] = \mu_m\Phi(\eta) + \sigma_m\phi(\eta) = \eta\sigma_m\Phi(\eta) + \sigma_m\phi(\eta) = \sigma_m\Phi(\eta)g(\eta). \tag{A11}$$

Since m_y^+ prices R_0 correctly, we have $E[m_y^+] = 1/R_0$, which implies

$$g(\eta) = \frac{1}{\sigma_m R_0\Phi(\eta)}. \tag{A12}$$

Substituting σ_m from (A10) into this expression, we obtain the following first order condition:

$$g(\eta) = [a + \sigma_z^2 R_0^2\Phi(\eta)^2]^{-\frac{1}{2}}. \tag{A13}$$

As shown in Lemma 5 of Kan and Robotti (2008), $g(\eta)$ is increasing in η and goes from zero to infinity as η increases. Since the right hand side of (A13) is positive and decreasing in η, the solution to this equation is unique.

Using (A12), we can express the vector of Lagrange multipliers as

$$\tilde{\lambda} = \begin{bmatrix} \frac{\mu_y - V_{ry}'V_{rr}^{-1}\mu_r}{R_0} + \frac{\phi(\eta)}{g(\eta)R_0^2\Phi(\eta)^2} - \frac{1+a}{R_0^2\Phi(\eta)} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0\Phi(\eta)}\right) \end{bmatrix} = \begin{bmatrix} \frac{\mu_y - V_{ry}'V_{rr}^{-1}\mu_r}{R_0} - \frac{a + \frac{\eta}{g(\eta)}}{R_0^2\Phi(\eta)} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0\Phi(\eta)}\right) \end{bmatrix}, \tag{A14}$$

where the last equality is obtained by using $\phi(\eta)/\Phi(\eta) = g(\eta) - \eta$.

The nonnegative admissible SDF that is closest to y is m_y^+, where m_y is defined in (A9). It follows that the squared constrained HJ-distance of y is given by

$$\delta_+^2 = E[(y - m_y^+)^2] = E[y^2] - 2E[ym_y^+] + E[m_y^{+2}]. \tag{A15}$$

It is straightforward to show that

$$E[y^2] = \sigma_z^2 + \mu_y^2 + V_{ry}' V_{rr}^{-1} V_{yr}. \tag{A16}$$

Using (A6) and the fact that $E[m_y^+] = 1/R_0$, we obtain

$$E[ym_y^+] = \left(\sigma_z^2 - \frac{V_{ry}' V_{rr}^{-1}\mu_r}{R_0\Phi(\eta)}\right)\Phi(\eta) + \frac{\mu_y}{R_0}, \tag{A17}$$

$$E[m_y^{+2}] = E[m_y m_y^+] = \sigma_m^2\Phi(\eta) + \frac{\mu_m}{R_0}. \tag{A18}$$

With these expressions, we obtain

$$\begin{aligned}
\delta_+^2 &= \sigma_z^2 + \mu_y^2 + V_{ry}' V_{rr}^{-1} V_{yr} - 2\left(\sigma_z^2\Phi(\eta) - \frac{V_{ry}' V_{rr}^{-1}\mu_r}{R_0}\right) - \frac{2\mu_y}{R_0} + \sigma_m^2\Phi(\eta) + \frac{\mu_m}{R_0} \\
&= \left(\mu_y - \frac{1}{R_0}\right)^2 + \left(V_{ry} + \frac{\mu_r}{R_0}\right)' V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0}\right) \\
&\quad + \sigma_z^2 - 2\sigma_z^2\Phi(\eta) + \sigma_m^2\Phi(\eta) + \frac{\mu_m}{R_0} - \frac{1+a}{R_0^2} \\
&= \delta^2 + \sigma_z^2\Phi(-\eta) + (\sigma_m^2 - \sigma_z^2)\Phi(\eta) + \frac{\eta\sigma_m}{R_0} - \frac{1+a}{R_0^2} \\
&= \delta^2 + \sigma_z^2\Phi(-\eta) + \frac{a + \frac{\eta}{g(\eta)}}{R_0^2\Phi(\eta)} - \frac{1+a}{R_0^2},
\end{aligned} \tag{A19}$$

where the last equality follows from (A10) and (A12). This completes the proof.

Proof of Lemma 1: We first show that

$$\begin{aligned}
\frac{\partial\Delta}{\partial\eta} &= -\sigma_z^2\phi(-\eta) - \frac{a\phi(\eta)}{R_0^2\Phi(\eta)^2} + \frac{\Phi(\eta)g(\eta) - \eta[\phi(\eta)g(\eta) + \Phi(\eta)g'(\eta)]}{R_0^2\Phi(\eta)^2 g(\eta)^2} \\
&= -\sigma_z^2\phi(\eta) - \frac{a\phi(\eta)}{R_0^2\Phi(\eta)} + \frac{g(\eta) - \eta}{R_0^2\Phi(\eta)g(\eta)^2} \\
&= -\frac{\phi(\eta)}{R_0^2\Phi(\eta)^2 g(\eta)^2} + \frac{\phi(\eta)}{R_0^2\Phi(\eta)^2 g(\eta)^2} \\
&= 0.
\end{aligned} \tag{A20}$$

The second equality follows because $g'(\eta) = 1 - g(\eta)\phi(\eta)/\Phi(\eta)$. The third equality follows because

$$a + \sigma_z^2 R_0^2 \Phi(\eta)^2 = \frac{1}{g(\eta)^2}. \tag{A21}$$

This suggests that as far as the partial derivatives of Δ with respect to $(\sigma_z^2,\ a,\ R_0)$ are concerned, we can treat η as a constant. Thus, it follows that

$$\begin{aligned}
\frac{\partial \Delta}{\partial \sigma_z^2} &= \Phi(-\eta) > 0, && \text{(A22)}\\
\frac{\partial \Delta}{\partial a} &= \frac{1}{R_0^2 \Phi(\eta)} - \frac{1}{R_0^2} = \frac{\Phi(-\eta)}{R_0^2 \Phi(\eta)} > 0, && \text{(A23)}\\
\frac{\partial \Delta}{\partial R_0} &= -\frac{2}{R_0^3}\left[\frac{a}{\Phi(\eta)} + \frac{\eta}{\Phi(\eta) g(\eta)} - 1 - a\right] = \frac{2}{R_0^3}\left[1 - \frac{a\Phi(-\eta)}{\Phi(\eta)} - \frac{\eta}{\Phi(\eta) g(\eta)}\right]. && \text{(A24)}
\end{aligned}$$

This completes the proof.

Proof of Lemma 2: Since $y/\mu_y \sim N(1, \sigma_y^2/\mu_y^2)$, using the mean of a lognormal distribution yields $E[\exp(y/\mu_y)] = \exp(1 + \sigma_y^2/(2\mu_y^2))$. Then, it follows that

$$E\left[\exp\left(\frac{y}{\mu_y} - 1 - \frac{\sigma_y^2}{2\mu_y^2}\right)\right] = 1. \tag{A25}$$

Therefore, $E[R_0 y_+] = R_0 E[y_+] = R_0 \mu_y = E[R_0 y]$ and both y_+ and y assign the same price to the risk-free asset. For risky assets, we use Stein's lemma to obtain[27]

$$\mathrm{Cov}[r, y_+] = E\left[\exp\left(\frac{y}{\mu_y} - 1 - \frac{\sigma_y^2}{2\mu_y^2}\right)\right] \mathrm{Cov}[r, y] = \mathrm{Cov}[r, y]. \tag{A26}$$

Together with the fact that $E[y_+] = E[y]$, this implies $E[r y_+] = E[ry]$ and y_+ and y assign the same price to the risky assets. This completes the proof.

Proof of Proposition 2: Let $\tilde{y} = \tilde{\gamma}' \tilde{f}$ and $m_{\tilde{y}} = \tilde{y} - \tilde{\lambda}' \tilde{r}$, where $\tilde{f} = [1,\ f']'$. In addition, let $C = E[\tilde{f}\tilde{f}']$ and $D = E[\tilde{r}\tilde{f}']$. Differentiating

$$\delta_+^2 = E[\tilde{y}^2] - E[m_{\tilde{y}}^{+2}] - 2\tilde{\lambda}' q \tag{A27}$$

with respect to $\tilde{\gamma}$ and $\tilde{\lambda}$, we obtain the following first order conditions:

$$\begin{aligned}
C\tilde{\gamma} - E[\tilde{f} m_{\tilde{y}}^+] &= 0_{K+1}, && \text{(A28)}\\
E[\tilde{r} m_{\tilde{y}}^+] &= q. && \text{(A29)}
\end{aligned}$$

[27] Stein's lemma suggests that if u and v are jointly normally distributed, we have $\mathrm{Cov}[f(u), v] = E[f'(u)]\mathrm{Cov}[u, v]$, where $f(u)$ is a differentiable function of u.

Let $\mu_m = E[m_{\tilde{y}}]$ and $\sigma_m^2 = \text{Var}[m_{\tilde{y}}]$. Invoking Lemma A.1, we have

$$E[\tilde{f}m_{\tilde{y}}^+] = \Phi(\eta)E[\tilde{f}m_{\tilde{y}}] + \sigma_m\phi(\eta)E[\tilde{f}] = \Phi(\eta)(C\tilde{\gamma} - D'\tilde{\lambda}) + \sigma_m\phi(\eta)E[\tilde{f}], \tag{A30}$$

$$E[\tilde{r}m_{\tilde{y}}^+] = \Phi(\eta)E[\tilde{r}m_{\tilde{y}}] + \sigma_m\phi(\eta)E[\tilde{r}] = \Phi(\eta)(D\tilde{\gamma} - U\tilde{\lambda}) + \sigma_m\phi(\eta)E[\tilde{r}], \tag{A31}$$

where $\eta = \mu_m/\sigma_m$. Putting the above expressions into the first order conditions, we obtain

$$\begin{bmatrix} \Phi(-\eta)C & \Phi(\eta)D' \\ \Phi(\eta)D & -\Phi(\eta)U \end{bmatrix} \begin{bmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \sigma_m\phi(\eta)E[\tilde{f}] \\ q - \sigma_m\phi(\eta)E[\tilde{r}] \end{bmatrix}. \tag{A32}$$

Let $H = [C + \Phi(\eta)(D'U^{-1}D - C)]^{-1}$. We can use the partitioned matrix inverse formula to write

$$\begin{bmatrix} \Phi(-\eta)C & \Phi(\eta)D' \\ \Phi(\eta)D & -\Phi(\eta)U \end{bmatrix}^{-1} = \begin{bmatrix} H & HD'U^{-1} \\ U^{-1}DH & -\frac{1}{\Phi(\eta)}U^{-1} + U^{-1}DHD'U^{-1} \end{bmatrix}. \tag{A33}$$

Using (39), we can easily verify that $U^{-1}E[\tilde{r}] = q/R_0$ and hence

$$D'U^{-1}E[\tilde{r}] = \frac{1}{R_0}D'q = \frac{1}{R_0}E[R_0\tilde{f}] = E[\tilde{f}]. \tag{A34}$$

Using this identity, we can then show that

$$\tilde{\gamma} = HD'U^{-1}q + \sigma_m\phi(\eta)H(E[\tilde{f}] - D'U^{-1}E[\tilde{r}]) = HD'U^{-1}q, \tag{A35}$$

$$\begin{aligned} \tilde{\lambda} &= \sigma_m\phi(\eta)U^{-1}DHE[\tilde{f}] - \left[\frac{1}{\Phi(\eta)}U^{-1} - U^{-1}DHD'U^{-1}\right](q - \sigma_m\phi(\eta)E[\tilde{r}]) \\ &= U^{-1}D\tilde{\gamma} - \frac{1}{\Phi(\eta)}U^{-1}q + \frac{\sigma_m\phi(\eta)}{R_0\Phi(\eta)}q. \end{aligned} \tag{A36}$$

From the partitioned matrix inverse formula and after some algebra, we can simplify the H matrix as

$$H = \begin{bmatrix} 1 + \mu_f'P\mu_f & -\mu_f'P \\ -P\mu_f & P \end{bmatrix}, \tag{A37}$$

where $P = [V_{ff} - \Phi(\eta)V_{ff\cdot r}]^{-1}$. Using this expression and (39), we can then rewrite (A35) as

$$\tilde{\gamma} = \begin{bmatrix} 1 + \mu_f'P\mu_f & -\mu_f'P \\ -P\mu_f & P \end{bmatrix} \begin{bmatrix} R_0 & \mu_r' \\ \mu_f R_0 & V_{rf}' + \mu_f\mu_r' \end{bmatrix} \begin{bmatrix} \frac{(1+a)}{R_0^2} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} \end{bmatrix}. \tag{A38}$$

After some algebra, we can express $\tilde{\gamma} = [\tilde{\gamma}_0,\ \tilde{\gamma}_1']'$ as

$$\tilde{\gamma}_1 = -\frac{1}{R_0}PV_{rf}'V_{rr}^{-1}\mu_r, \qquad \tilde{\gamma}_0 = \frac{1}{R_0} - \tilde{\gamma}_1'\mu_f. \tag{A39}$$

As a result, we can write $\tilde{y} = (1/R_0) + \tilde{\gamma}_1'(f - \mu_f)$ and

$$D\tilde{\gamma} = E[\tilde{r}\tilde{y}] = \begin{bmatrix} 1 \\ \frac{\mu_r}{R_0} + V_{rf}\tilde{\gamma}_1 \end{bmatrix}. \tag{A40}$$

Then, using (39), we can simplify $\tilde{\lambda}$ to

$$\tilde{\lambda} = U^{-1}\begin{bmatrix} 1 \\ \frac{\mu_r}{R_0} + V_{rf}\tilde{\gamma}_1 \end{bmatrix} - \begin{bmatrix} \frac{1+a}{R_0^2\Phi(\eta)} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0\Phi(\eta)} \end{bmatrix} + \begin{bmatrix} \frac{\sigma_m\phi(\eta)}{R_0^2\Phi(\eta)} \\ 0_K \end{bmatrix} = \begin{bmatrix} \frac{1+\tilde{a}_1}{R_0^2} - \frac{1+a}{R_0^2\Phi(\eta)} + \frac{\sigma_m\phi(\eta)}{R_0\Phi(\eta)} \\ V_{rr}^{-1}\left(V_{rf}\tilde{\gamma}_1 + \frac{\mu_r}{R_0}\right) \end{bmatrix}. \tag{A41}$$

To obtain explicit expressions for $\tilde{\gamma}$ and $\tilde{\lambda}$, we first need to solve for η. Defining $\epsilon = (f - \mu_f) - V_{rf}'V_{rr}^{-1}(r - \mu_r)$ as the unspanned components of the factors, we can write

$$m_{\tilde{y}} = \tilde{\gamma}'\tilde{f} - \tilde{\lambda}'\tilde{r} = \tilde{\gamma}_1'\epsilon + \frac{1}{\Phi(\eta)}\left[\frac{1}{R_0} - \frac{\mu_r'V_{rr}^{-1}(r - \mu_r)}{R_0}\right] - \frac{\sigma_m\phi(\eta)}{\Phi(\eta)}. \tag{A42}$$

Using $E[\epsilon] = 0_K$ and $\text{Var}[\epsilon] = V_{ff\cdot r}$, we have

$$\mu_m = \frac{1}{\Phi(\eta)}\left[\frac{1}{R_0} - \sigma_m\phi(\eta)\right], \tag{A43}$$

$$\sigma_m^2 = \tilde{\gamma}_1'V_{ff\cdot r}\tilde{\gamma}_1 + \frac{a}{R_0^2\Phi(\eta)^2}. \tag{A44}$$

Since $m_{\tilde{y}}^+$ prices the risk-free asset correctly, we have

$$E[m_{\tilde{y}}^+] = \sigma_m g(\eta)\Phi(\eta) = \frac{1}{R_0}. \tag{A45}$$

Then, plugging the expression of σ_m^2 from (A44) into (A45), we obtain

$$g(\eta) = \left[a + \tilde{\gamma}_1'V_{ff\cdot r}\tilde{\gamma}_1 R_0^2\Phi(\eta)^2\right]^{-\frac{1}{2}}. \tag{A46}$$

Using the expression for $\tilde{\gamma}_1$ in (A39) and rearranging terms, we can see that η satisfies the following equation:

$$g(u) = \left(a + \alpha'\left[\frac{1}{\Phi(u)}I_K - V_{ff}^{-\frac{1}{2}}V_{ff\cdot r}V_{ff}^{-\frac{1}{2}}\right]^{-2}\alpha\right)^{-\frac{1}{2}}, \tag{A47}$$

where $\alpha = V_{ff}^{-1}V_{ff\cdot r}^{\frac{1}{2}}V_{rf}'V_{rr}^{-1}\mu_r$. Since the left hand side is increasing in u from zero to infinity and the right hand side is decreasing in u and is positive (because all the eigenvalues of $V_{ff}^{-\frac{1}{2}}V_{ff\cdot r}V_{ff}^{-\frac{1}{2}}$ are less than or equal to one), (A47) has a unique solution. Using Lemma A.1, it is straightforward to obtain

$$E[m_{\tilde{y}}^{+2}] = E[m_{\tilde{y}}m_{\tilde{y}}^+] = \sigma_m^2\Phi(\eta) + \frac{\mu_m}{R_0}, \tag{A48}$$

$$E[\tilde{y}m_{\tilde{y}}^+] = \text{Cov}[\tilde{y}, m_{\tilde{y}}]\Phi(\eta) + \frac{1}{R_0^2} = \left[\tilde{\gamma}_1'V_{ff\cdot r}\tilde{\gamma}_1 + \frac{\tilde{a}_1}{R_0^2\Phi(\eta)}\right]\Phi(\eta) + \frac{1}{R_0^2}. \tag{A49}$$

The squared constrained HJ-distance is then given by

$$\begin{aligned}
\delta_+^2 &= E[(\tilde{y} - m_{\tilde{y}}^+)^2] \\
&= E[\tilde{y}^2] - 2E[\tilde{y}m_{\tilde{y}}^+] + E[m_{\tilde{y}}^{+2}] \\
&= \frac{1}{R_0^2} + \tilde{\gamma}_1' V_{ff}\tilde{\gamma}_1 - 2\left[\tilde{\gamma}_1' V_{ff\cdot r}\tilde{\gamma}_1 + \frac{\tilde{a}_1}{R_0^2\Phi(\eta)}\right]\Phi(\eta) - \frac{2}{R_0^2} + \sigma_m^2\Phi(\eta) + \frac{\mu_m}{R_0} \\
&= \tilde{\gamma}_1' V_{ff}\tilde{\gamma}_1 - 2\tilde{\gamma}_1' V_{ff\cdot r}\tilde{\gamma}_1\Phi(\eta) - \frac{1+2\tilde{a}_1}{R_0^2} + \left[\tilde{\gamma}_1' V_{ff\cdot r}\tilde{\gamma}_1 + \frac{a}{R_0^2\Phi(\eta)^2}\right]\Phi(\eta) + \frac{\eta\sigma_m}{R_0} \\
&= \tilde{\gamma}_1'(V_{ff} - \Phi(\eta)V_{ff\cdot r})\tilde{\gamma}_1 - \frac{1+2\tilde{a}_1}{R_0^2} + \frac{a}{R_0^2\Phi(\eta)} + \frac{\eta}{R_0^2\Phi(\eta)g(\eta)} \\
&= \frac{a + \frac{\eta}{g(\eta)}}{R_0^2\Phi(\eta)} - \frac{(1+\tilde{a}_1)}{R_0^2},
\end{aligned} \tag{A50}$$

where the second last equality is obtained using $\sigma_m = 1/[R_0\Phi(\eta)g(\eta)]$ from (A45). Finally, using this expression of σ_m and $\phi(\eta)/\Phi(\eta) = g(\eta) - \eta$, we can see that the first element of $\tilde{\lambda}$ is equal to $-\delta_+^2$. This completes the proof.

Proof of Lemma 3: The probability for y to take on negative values is equal to

$$P[y < 0] = \Phi\left(-\frac{\mu_y}{\sigma_y}\right) = \Phi\left(-\frac{1}{R_0\sigma_y}\right) = \Phi\left(-\frac{1}{R_0\sqrt{\gamma_1' V_{ff}\gamma_1}}\right), \tag{A51}$$

where σ_y is the standard deviation of the SDF y. In contrast, the probability for $\tilde{y}$ to take on negative values is equal to

$$P[\tilde{y} < 0] = \Phi\left(-\frac{\mu_{\tilde{y}}}{\sigma_{\tilde{y}}}\right) = \Phi\left(-\frac{1}{R_0\sigma_{\tilde{y}}}\right) = \Phi\left(-\frac{1}{R_0\sqrt{\tilde{\gamma}_1' V_{ff}\tilde{\gamma}_1}}\right). \tag{A52}$$

The inequality holds because

$$\begin{aligned}
R_0^2\tilde{\gamma}_1' V_{ff}\tilde{\gamma}_1 &= \mu_r' V_{rr}^{-1}V_{rf}(V_{ff} - \Phi(\eta)V_{ff\cdot r})^{-1}V_{ff}(V_{ff} - \Phi(\eta)V_{ff\cdot r})^{-1}V_{rf}'V_{rr}^{-1}\mu_r \\
&= \mu_r' V_{rr}^{-1}V_{rf}V_{ff}^{\frac{1}{2}}\left(I_K - \Phi(\eta)V_{ff}^{-\frac{1}{2}}V_{ff\cdot r}V_{ff}^{-\frac{1}{2}}\right)^{-2}V_{ff}^{\frac{1}{2}}V_{rf}'V_{rr}^{-1}\mu_r \\
&\leq \mu_r' V_{rr}^{-1}V_{rf}V_{ff}^{\frac{1}{2}}(I_K - V_{ff}^{-\frac{1}{2}}V_{ff\cdot r}V_{ff}^{-\frac{1}{2}})^{-2}V_{ff}^{\frac{1}{2}}V_{rf}'V_{rr}^{-1}\mu_r \\
&= \mu_r' V_{rr}^{-1}V_{rf}(V_{rf}'V_{rr}^{-1}V_{rf})^{-1}V_{ff}(V_{rf}'V_{rr}^{-1}V_{rf})^{-1}V_{rf}'V_{rr}^{-1}\mu_r \\
&= R_0^2\gamma_1' V_{ff}\gamma_1.
\end{aligned} \tag{A53}$$

For (67), the first inequality, $\delta^2 \leq \tilde{\delta}^2$, is obvious since γ is chosen to minimize $\delta^2 = e'U^{-1}e$ but $\tilde{\gamma}$ is not. For the second inequality, $\tilde{\delta}^2 \leq \delta_+^2$, note that for every $h \in L^2$ with $E[h^2] = 1$, we have

$$\min_{m\in\mathcal{M}^+}(E[\tilde{y}h] - E[mh])^2 \leq \delta_+^2. \tag{A54}$$

Consider a portfolio w with unit second moment, i.e., $w'Uw = 1$. When $\tilde{y}$ is the SDF, the squared pricing error of the portfolio is $(w'\tilde{e})^2$, and it is maximized when w is chosen to be

$$w^* = \frac{U^{-1}\tilde{e}}{(\tilde{e}'U^{-1}\tilde{e})^{\frac{1}{2}}}. \tag{A55}$$

Let $h = w^{*\prime}\tilde{r}$. Since h is a linear combination of $\tilde{r}$, $E[mh] = w^{*\prime}E[m\tilde{r}] = w^{*\prime}q$ and the price of h is the same for every $m \in \mathcal{M}^+$. It follows that

$$\delta_+^2 \geq \inf_{m\in\mathcal{M}^+}(E[\tilde{y}h] - E[mh])^2 = (E[\tilde{y}h] - E[mh])^2 = (w^{*\prime}(E[\tilde{y}\tilde{r}] - q))^2 = (w^{*\prime}\tilde{e})^2 = \tilde{\delta}^2. \tag{A56}$$

This completes the proof.

Appendix B

In this Appendix, we present a discussion of the numerical methods for obtaining the sample constrained HJ-distance and its associated parameters. We first show that the primal problem for solving the sample constrained HJ-distance for a linear SDF can be written as a quadratic programming problem. The sample constrained HJ-distance for a fixed y_t can also be obtained using quadratic programming but, since the setup is similar to the linear SDF case, we do not provide a separate discussion of this case.

Let $y_t = \tilde{f}_t'\gamma$, where $\tilde{f}_t$ is a k-vector of systematic factors at time t with its first element typically being equal to one. From (73), the squared constrained HJ-distance of the model is given by

$$\begin{aligned} \hat{\delta}_+^2 &= \min_{\gamma} \min_{\{m_t, t=1,\ldots,T\}} \frac{1}{T}\sum_{t=1}^{T}(\tilde{f}_t'\gamma - m_t)^2, \\ &\text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} m_t x_t = \bar{q}, \\ &\qquad m_t \geq 0, \;\; t = 1,\ldots,T. \end{aligned} \tag{B1}$$

Let $z = [\gamma',\ m_1, m_2,\ \ldots, m_T]'$ and

$$A = \frac{1}{T}[\tilde{F},\ -I_T]'[\tilde{F},\ -I_T], \tag{B2}$$

where $\tilde{F}$ is a $T \times k$ matrix with typical row $\tilde{f}_t'$. We can then rewrite (B1) as the following quadratic programming problem:

$$\min_{z} z'Az \quad \text{s.t.} \quad \frac{1}{T}\sum_{t=1}^{T} m_t x_t = \bar{q}, \quad m_t \geq 0, \;\; t = 1,\ldots,T. \tag{B3}$$

Besides the fact that well developed procedures for quadratic programming are widely available (e.g., `quadprog` in Matlab), solving the primal problem has the additional advantage that we do not need to worry that the λ in the dual problem may not be unique.

Instead of using the quadratic programming method to solve the primal problem, in the case of linear SDFs we can also solve the dual problem using a fast iterative method. Specifically, the

first order conditions for the dual problem are given by

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t[\tilde{f}_t'\gamma-(\tilde{f}_t'\gamma-\lambda' x_t)^+] = 0_k, \tag{B4}$$

$$\frac{1}{T}\sum_{t=1}^{T}x_t(\tilde{f}_t'\gamma-\lambda' x_t)^+ = \bar{q}. \tag{B5}$$

More compactly, we can write these first order conditions as

$$\hat{C}_-\gamma+\hat{D}_+'\lambda = 0_k, \tag{B6}$$

$$\hat{D}_+\gamma-\hat{U}_+\lambda = \bar{q}, \tag{B7}$$

where

$$\hat{C}_- = \frac{1}{T}\sum_{t=1}^{T}\tilde{f}_t\tilde{f}_t' I_{\{m_t\leq 0\}}, \tag{B8}$$

$$\hat{D}_+ = \frac{1}{T}\sum_{t=1}^{T}x_t\tilde{f}_t' I_{\{m_t>0\}}, \tag{B9}$$

$$\hat{U}_+ = \frac{1}{T}\sum_{t=1}^{T}x_t x_t' I_{\{m_t>0\}}, \tag{B10}$$

and $m_t=\tilde{f}_t'\gamma-\lambda' x_t$. When $\hat{U}_+$ is nonsingular (i.e., when $\sum_{t=1}^{T}I_{\{m_t>0\}}\geq n$), we can solve (B6) and (B7) conditional on $\hat{C}_-$, $\hat{D}_+$, and $\hat{U}_+$. This gives us the following closed-form solution for $\hat{\gamma}$ and $\hat{\lambda}$:

$$\hat{\gamma} = (\hat{C}_-+\hat{D}_+'\hat{U}_+^{-1}\hat{D}_+)^{-1}\hat{D}_+'\hat{U}_+^{-1}\bar{q}, \tag{B11}$$

$$\hat{\lambda} = \hat{U}_+^{-1}[\hat{D}_+(\hat{C}_-+\hat{D}_+'\hat{U}_+^{-1}\hat{D}_+)^{-1}\hat{D}_+'\hat{U}_+^{-1}\bar{q}-\bar{q}]. \tag{B12}$$

Note that (B11) and (B12) only provide $\hat{\gamma}$ and $\hat{\lambda}$ conditional on a given set of indicator functions $I_{\{m_t>0\}}$. To obtain the optimal $\hat{\gamma}$ and $\hat{\lambda}$, we use the following iterative method. We start off with the estimates of γ and λ from the unconstrained HJ-distance problem to compute the initial set of $I_{\{m_t>0\}}$. Then, we obtain $\hat{\gamma}$ and $\hat{\lambda}$ using (B11) and (B12). If the m_t's computed based on these updated $\hat{\gamma}$ and $\hat{\lambda}$ do not alter $I_{\{m_t>0\}}$, we have found our optimal solution. If not, we compute $\hat{C}_-$, $\hat{D}_+$, $\hat{U}_+$ using the updated $I_{\{m_t>0\}}$ and obtain a new set of $\hat{\gamma}$ and $\hat{\lambda}$. We repeat this process until the first order conditions are satisfied. In our experience, this iterative approach is extremely efficient and typically delivers the final $\hat{\gamma}$ and $\hat{\lambda}$ with only a few iterations.

Appendix C

C.1. Assumptions

We first introduce regularity conditions to ensure the stochastic equicontinuity of the sample constrained HJ-distance and establish the consistency of $\hat{\theta}$. Let $|w| = (w'w)^{\frac{1}{2}}$ denote the Euclidean norm of a vector w and $||A|| = \sqrt{\text{tr}(A'A)}$ be the Euclidean norm of a matrix A, where $\text{tr}(\cdot)$ is the trace operator. In addition, denote the indicator function by $I_{\{\cdot\}}$.

Assumption A. *Assume that*

(i) $\phi_t(\theta)$ *is* m*-dependent,*

(ii) *the parameter space* Θ *is compact,*

(iii) $\phi_t(\theta)$ *is continuous in* $\theta \in \Theta$ *almost surely,*

(iv) $|\phi_t(\theta_1) - \phi_t(\theta_2)| \leq A_t |\theta_1 - \theta_2| \ \forall \ \theta_1, \theta_2 \in \Theta$*, where* A_t *is a bounded random variable that satisfies* $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[|A_t|^{2+\omega}] < \infty$ *for some* $\omega > 0$,

(v) $\sup_{\theta\in\Theta} E[|\phi_t(\theta)|^{2+\omega}] < \infty$ *for some* $\omega > 0$,

(vi) *the population dual problem (79) has a unique solution* θ^* *which is in the interior of* Θ.

Assumptions A(i)–A(v) ensure the stochastic equicontinuity of $\phi_t(\theta)$ (see Andrews, 1994; and Stock and Wright, 2000) and imply that

$$\sup_{\theta\in\Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \phi_t(\theta) - E[\phi_t(\theta)] \right| \xrightarrow{p} 0. \tag{C1}$$

The m-dependence can be relaxed although results for empirical processes with more general dependence structure are still limited (see, for instance, Andrews, 1993; and Andrews and Pollard, 1994). Assumption A(vi) is an identification condition that ensures the uniqueness of the pseudo-true value θ^*. The uniform convergence in (C1) and Assumption A(vi) are sufficient for establishing the consistency of $\hat{\theta}$:

$$\hat{\theta} \xrightarrow{p} \theta^*. \tag{C2}$$

Let

$$H \equiv \begin{bmatrix} H_{\gamma\gamma} & H'_{\lambda\gamma} \\ H_{\lambda\gamma} & H_{\lambda\lambda} \end{bmatrix} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 E[\phi_t(\theta^*)]}{\partial\theta\partial\theta'}, \tag{C3}$$

$$M \equiv \begin{bmatrix} M_{\gamma\gamma} & M'_{\lambda\gamma} \\ M_{\lambda\gamma} & M_{\lambda\lambda} \end{bmatrix} = \lim_{T\to\infty} \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \phi_t(\theta^*)}{\partial\theta} \right], \tag{C4}$$

and $h_t(\theta) = \partial\phi_t(\theta)/\partial\theta$. The next assumption provides conditions for the existence and uniform convergence of the limiting matrices in (C3) and (C4).

ASSUMPTION B. *Let $\mathcal{N}(\theta^*)$ be a neighborhood of θ^*. Assume that*

(i) $P[y_t(\gamma) = \lambda' x_t] = 0$ *almost surely for* $\theta \in \mathcal{N}(\theta^*)$,

(ii) $E[\phi_t(\theta)]$ *is twice continuously differentiable in* θ *for* $\theta \in \mathcal{N}(\theta^*)$,

(iii) $\sup_{\theta\in\mathcal{N}(\theta^*)} \left\| \frac{\partial^2 E[\phi_t(\theta)]}{\partial\theta\partial\theta'} \right\| < \infty$ *and* H *is of full rank,*

(iv) M *is a finite positive definite matrix when* $\delta_+ > 0$, *or* $M_{\lambda\lambda}$ *is a finite positive definite matrix when* $\delta_+ = 0$.

Assumption B(i) implies that $h_t(\theta)$ is well defined for θ close to θ^* and the consistent estimator $\hat{\theta}$ satisfies the first order conditions

$$\bar{h}_T(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^{T} h_t(\hat{\theta}) = 0_{n+k} \tag{C5}$$

with probability approaching one as $T \to \infty$. Following Andrews (1994), let

$$\bar{h}^*_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} E[h_t(\theta)] \tag{C6}$$

and define an empirical process by $\sqrt{T}\bar{v}_T(\theta)$, where

$$\bar{v}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} v_t(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} (h_t(\theta) - E[h_t(\theta)]). \tag{C7}$$

ASSUMPTION C. *Assume that $v_t(\theta)$ satisfies the conditions*

(i) $|v_t(\theta_1) - v_t(\theta_2)| \leq B_t |\theta_1 - \theta_2| \ \forall\ \theta_1, \theta_2 \in \Theta$, *where B_t is a bounded random variable that satisfies* $\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[|B_t|^{2+\omega}] < \infty$ *for some* $\omega > 0$,

(ii) $\sup_{\theta\in\Theta} E[|v_t(\theta)|^{2+\omega}] < \infty$ *for some* $\omega > 0$.

Assumption C imposes sufficient conditions for the stochastic equicontinuity of the empirical process $\sqrt{T}\bar{v}_T(\theta)$. Together with Assumption B, these conditions ensure that $\sqrt{T}\bar{v}_T(\theta)$ obeys the central limit theorem.

Below we provide explicit expressions for the partitioned matrices in (C3) and (C4). Using the fact that

$$\frac{\partial \phi_t(\theta^*)}{\partial \gamma} = 2[y_t(\gamma^*) - m_t(\theta^*)^+]\frac{\partial y_t(\gamma^*)}{\partial \gamma}, \tag{C8}$$

$$\frac{\partial \phi_t(\theta^*)}{\partial \lambda} = 2[x_t m_t(\theta^*)^+ - q_{t-1}], \tag{C9}$$

and under Assumptions A, B, and C, we can write

$$H_{\gamma\gamma} = 2E\left[\left(y_t(\gamma^*) - m_t(\theta^*)^+\right)\frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'} + \frac{\partial y_t(\gamma^*)}{\partial\gamma}\frac{\partial y_t(\gamma^*)}{\partial\gamma'}I_{\{m_t(\theta^*)\le 0\}}\right] \equiv 2G_+, \tag{C10}$$

$$H_{\lambda\gamma} = 2E\left[x_t\frac{\partial y_t(\gamma^*)}{\partial\gamma'}I_{\{m_t(\theta^*)>0\}}\right] \equiv 2D_+, \tag{C11}$$

$$H_{\lambda\lambda} = -2E\left[x_t x_t' I_{\{m_t(\theta^*)>0\}}\right] \equiv -2U_+, \tag{C12}$$

and

$$M_{\lambda\lambda} = 4\sum_{j=-\infty}^{\infty} E\left[(x_t m_t(\theta^*)^+ - q_{t-1})(x_{t+j}m_{t+j}(\theta^*)^+ - q_{t+j-1})'\right] \equiv 4S_+. \tag{C13}$$

If the model is correctly specified, we have $\lambda^* = 0_n$ and $y_t(\gamma^*) = m_t(\theta^*)^+$. Then, it follows that $G_+ = 0_{k\times k}$, $D_+ = D \equiv E\left[x_t\frac{\partial y_t(\gamma^*)}{\partial\gamma'}\right]$, $U_+ = U \equiv E[x_t x_t']$, and $S_+ = S \equiv \sum_{j=-\infty}^{\infty} E[(x_t y_t(\gamma^*) - q_{t-1})(x_{t+j}y_{t+j}(\gamma^*) - q_{t+j-1})']$. Furthermore, we have $\partial\phi_t(\theta^*)/\partial\gamma = 0_k$ which yields $M_{\gamma\gamma} = 0_{k\times k}$ and $M_{\lambda\gamma} = 0_{n\times k}$. This is the reason why Assumption B(iv) requires only $M_{\lambda\lambda}$, and not M, to be positive definite when $\delta_+ = 0$.

The next lemma develops an expansion of the sample constrained HJ-distance that will be used in the proofs of the subsequent propositions for model specification and model selection tests.

Lemma C.1. *Under Assumptions A, B, and C,*

$$\hat{\delta}_+^2 - \delta_+^2 = \frac{1}{T}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)]) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right). \tag{C14}$$

Proof. We start by expanding $E[\phi_t(\theta^*)] = \delta_+^2$ about $\hat{\theta}$. Since $\frac{1}{T}\sum_{t=1}^{T} \partial\phi_t(\hat{\theta})/\partial\theta = 0_{n+k}$, we obtain

$$E[\phi_t(\theta^*)] = \frac{1}{T}\sum_{t=1}^{T}\phi_t(\hat{\theta}) - \frac{1}{T}\sum_{t=1}^{T}\Big(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\Big) + \frac{1}{2}(\hat{\theta}-\theta^*)'\frac{\partial^2 E[\phi_t(\tilde{\theta})]}{\partial\theta\partial\theta'}(\hat{\theta}-\theta^*), \tag{C15}$$

where $\tilde{\theta}$ is an intermediate point between $\hat{\theta}$ and θ^*. A mean value expansion of $\bar{h}_T^*(\theta^*)$ about $\hat{\theta}$ yields

$$0_{n+k} = \sqrt{T}\bar{h}_T^*(\theta^*) = \sqrt{T}\bar{h}_T^*(\hat{\theta}) - \frac{\partial\bar{h}_T^*(\check{\theta})}{\partial\theta}\sqrt{T}(\hat{\theta}-\theta^*), \tag{C16}$$

where $\check{\theta}$ is another intermediate point on the line segment joining $\hat{\theta}$ and θ^*. From Assumption B(iii) and the consistency of $\hat{\theta}$, we have

$$\sqrt{T}(\hat{\theta}-\theta^*) = H^{-1}\sqrt{T}\bar{h}_T^*(\hat{\theta}) + o_p(1). \tag{C17}$$

Using the definition of $\bar{v}_T(\theta)$ in (C7) and the first order condition of $\frac{1}{T}\sum_{t=1}^{T} h_t(\hat{\theta}) = 0_{n+k}$, it follows that

$$\sqrt{T}\bar{v}_T(\hat{\theta}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(h_t(\hat{\theta}) - E[h_t(\hat{\theta})]\Big) = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E[h_t(\hat{\theta})] = -\sqrt{T}\bar{h}_T^*(\hat{\theta}). \tag{C18}$$

This allows us to rewrite $\sqrt{T}\bar{h}_T^*(\hat{\theta})$ as

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\hat{\theta}) = \sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] - \sqrt{T}\bar{v}_T(\theta^*). \tag{C19}$$

By the consistency of $\hat{\theta}$, $P[|\hat{\theta}-\theta^*| > \omega] \to 0$ for any arbitrarily small $\omega > 0$. Then,

$$\sqrt{T}|\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})| \leq \sup_{\theta\in\Theta:|\theta-\theta^*|\leq\omega}\sqrt{T}\,|\bar{v}_T(\theta^*) - \bar{v}_T(\theta)|\,. \tag{C20}$$

From the stochastic equicontinuity of the empirical process $\sqrt{T}\bar{v}_T(\cdot)$,

$$\sup_{\theta\in\Theta:|\theta-\theta^*|\leq\omega}\sqrt{T}\,|\bar{v}_T(\theta^*) - \bar{v}_T(\theta)| \xrightarrow{p} 0. \tag{C21}$$

Therefore, we have $\sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] = o_p(1)$ and

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{C22}$$

Finally, substituting (C22) into (C17) yields

$$\sqrt{T}(\hat{\theta}-\theta^*) = -H^{-1}\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{C23}$$

Thus, from (C23), the consistency of $\hat{\theta}$, and Assumption B(iii), we obtain

$$\begin{aligned}\hat{\delta}_+^2 - \delta_+^2 &= \frac{1}{T}\sum_{t=1}^{T}\Big(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\Big) - \frac{1}{2}(\hat{\theta}-\theta^*)'\frac{\partial^2 E[\phi_t(\tilde{\theta})]}{\partial\theta\partial\theta'}(\hat{\theta}-\theta^*) \\ &= \frac{1}{T}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)]) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right). \qquad \text{(C24)}\end{aligned}$$

This completes the proof.

C.2. Proofs

Proof of Proposition 3: (a) From the definition of H in (C3), we can use the partitioned matrix inverse formula to obtain

$$H^{-1} = \begin{bmatrix} 2G_+ & 2D_+{}' \\ 2D_+ & -2U_+ \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} \tilde{H} & \tilde{H}D_+'U_+^{-1} \\ U_+^{-1}D_+\tilde{H} & -U_+^{-1} + U_+^{-1}D_+\tilde{H}D_+'U_+^{-1} \end{bmatrix}, \qquad \text{(C25)}$$

where $\tilde{H} = (G_+ + D_+'U_+^{-1}D_+)^{-1}$. Under the null hypothesis $H_0 : \delta_+ = 0$, (C14) in Lemma C.1 becomes

$$\hat{\delta}_+^2 = -\frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right) \qquad \text{(C26)}$$

since $\lambda^* = 0_n$ and $\phi_t(\gamma^*, 0_n) = E[\phi_t(\gamma^*, 0_n)] = 0$. Let $\bar{v}_T(\theta^*) = [\bar{v}_{1,T}(\theta^*)', \bar{v}_{2,T}(\theta^*)']'$, where $\bar{v}_{1,T}(\theta^*)$ denotes the first k elements of $\bar{v}_T(\theta^*)$. Under the null, $\bar{v}_{1,T}(\theta^*) = 0_k$, $G_+ = 0_{k\times k}$, $D_+ = D$, and $U_+ = U$.

Then, it follows that

$$\begin{aligned}T\hat{\delta}_+^2 &= -\frac{1}{2}\sqrt{T}\bar{v}_T(\theta^*)'H^{-1}\sqrt{T}\bar{v}_T(\theta^*) + o_p(1) \\ &= \frac{1}{4}\sqrt{T}\bar{v}_{2T}(\theta^*)'[U^{-1} - U^{-1}D(D'U^{-1}D)^{-1}D'U^{-1}]\sqrt{T}\bar{v}_{2,T}(\theta^*) + o_p(1) \\ &= \frac{1}{4}\sqrt{T}\bar{v}_{2T}(\theta^*)'U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}\sqrt{T}\bar{v}_{2,T}(\theta^*) + o_p(1) \qquad \text{(C27)}\end{aligned}$$

by using the fact that $I_n - U^{-\frac{1}{2}}D(D'U^{-1}D)^{-1}D'U^{-\frac{1}{2}} = PP'$. Also, Assumptions A, B, and C ensure that the empirical process $\sqrt{T}\bar{v}_{2,T}(\theta^*)$ obeys the central limit theorem and

$$\sqrt{T}\bar{v}_{2,T}(\theta^*) \overset{A}{\sim} N(0_n, M_{\lambda\lambda}). \qquad \text{(C28)}$$

Thus, using the fact that $M_{\lambda\lambda} = 4S$ under the null, we obtain

$$T\hat{\delta}_+^2 \overset{A}{\sim} z'S^{\frac{1}{2}}U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}S^{\frac{1}{2}}z, \qquad \text{(C29)}$$

where $z \sim N(0_n, I_n)$. Since $S^{\frac{1}{2}}U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}S^{\frac{1}{2}}$ has the same nonzero eigenvalues as $P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$, we have

$$T\hat{\delta}_+^2 \overset{A}{\sim} \sum_{i=1}^{n-k} \xi_i \nu_i, \tag{C30}$$

where the ξ_i's are the eigenvalues of $P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$.

(b) Now consider the case $\delta_+ > 0$. In this situation, the asymptotic behavior of $\sqrt{T}(\hat{\delta}_+^2 - \delta_+^2)$ is determined by $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)])$, which converges weakly to a Gaussian process. Under Assumptions A, B, and C, and since $E[\phi_t(\theta^*)] = \delta_+^2$, we have

$$\sqrt{T}(\hat{\delta}_+^2 - \delta_+^2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)]) + o_p(1) \overset{A}{\sim} N(0, v). \tag{C31}$$

This completes the proof.

Proof of Proposition 4: For $\delta_+ > 0$ and under Assumptions A, B, and C,

$$\sqrt{T}\bar{v}_T(\theta^*) \overset{A}{\sim} N(0_{n+k}, M). \tag{C32}$$

Then, combining (C23) and (C32), we obtain

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, H^{-1}MH^{-1}). \tag{C33}$$

To derive an explicit expression for the asymptotic covariance matrix of $\hat{\theta}$, we write

$$H^{-1}MH^{-1} = \sum_{j=-\infty}^{\infty} E[\tilde{h}_t \tilde{h}'_{t+j}], \tag{C34}$$

where

$$\tilde{h}_t \equiv \begin{bmatrix} \tilde{h}_{1t} \\ \tilde{h}_{2t} \end{bmatrix} = H^{-1}\frac{\partial \phi_t(\theta^*)}{\partial \theta}. \tag{C35}$$

Using (C25), (C8), and (C9), we can express $\tilde{h}_{1t}$ and $\tilde{h}_{2t}$ as

$$\begin{aligned} \tilde{h}_{1t} &= (G_+ + D'_+U_+^{-1}D_+)^{-1}\left[(y_t(\gamma^*) - m_t(\theta^*)^+)\frac{\partial y_t(\gamma^*)}{\partial \gamma} + D'_+U_+^{-1}[x_t m_t(\theta^*)^+ - q_{t-1}]\right], \\ \tilde{h}_{2t} &= U_+^{-1}[D_+\tilde{h}_{1t} - x_t m_t(\theta^*)^+ + q_{t-1}]. \end{aligned}$$

This completes the proof.

Proof of Lemma 4: We only provide the proof for $E[q_{t-1}]'\lambda^* = -\delta_+^2$. The proof for $\bar{q}'\hat{\lambda} = -\hat{\delta}_+^2$ can be similarly obtained by replacing the expectation operator with $\frac{1}{T}\sum_{t=1}^{T}$. From the dual problem, the squared constrained HJ-distance is given by

$$\delta_+^2 = E[y_t(\gamma^*)^2 - (m_t(\theta^*)^+)^2 - 2\lambda^{*\prime}q_{t-1}]. \tag{C36}$$

Using the fact that $m_t(\theta^*)^+ = (y_t(\gamma^*) - \lambda^{*\prime}x_t)^+ \in \mathcal{M}^+$, we have

$$E[m_t(\theta^*)^+(\lambda^{*\prime}x_t)] = \lambda^{*\prime}E[q_{t-1}] \tag{C37}$$

and we can write

$$E[(m_t(\theta^*)^+)^2] = E[m_t(\theta^*)^+(y_t(\gamma^*) - \lambda^{*\prime}x_t)] = E[m_t(\theta^*)^+y_t(\gamma^*)] - \lambda^{*\prime}E[q_{t-1}]. \tag{C38}$$

Therefore,

$$\delta_+^2 = E[y_t(\gamma^*)^2] - E[m_t(\theta^*)^+y_t(\gamma^*)] - \lambda^{*\prime}E[q_{t-1}] = E[y_t(\gamma^*)(y_t(\gamma^*) - m_t(\theta^*)^+)] - \lambda^{*\prime}E[q_{t-1}]. \tag{C39}$$

When $y_t(\gamma^*) = \gamma^{*\prime}\tilde{f}_t$, we have the following first order condition:

$$E[\tilde{f}_t\tilde{f}_t']\gamma^* - E[\tilde{f}_t m_t(\theta^*)^+] = 0_k. \tag{C40}$$

Premultiplying both sides of (C40) by $\gamma^{*\prime}$, we obtain

$$E[y_t(\gamma^*)^2] - E[y_t(\gamma^*)m_t(\theta^*)^+] = E[y_t(\gamma^*)(y_t(\gamma^*) - m_t(\theta^*)^+)] = 0. \tag{C41}$$

Then, it follows that $\delta_+^2 = -\lambda^{*\prime}E[q_{t-1}]$. This completes the proof.

Example of two strictly non-nested models that are both correctly specified: Let R be the gross returns on N risky assets and R_0 be the gross return on the risk-free asset. Suppose that R_p is the gross return on the tangency portfolio of the N risky assets. Simple mean-variance mathematics gives

$$E[R] = R_0 1_N + \mathrm{Cov}[R, R_p]\gamma^*, \tag{C42}$$

where $\gamma^* = (\mu_p - R_0)/\sigma_p^2$, with $\mu_p = E[R_p]$ and $\sigma_p^2 = \mathrm{Var}[R_p]$. In addition, assume that R is multivariate normally distributed. Consider the following SDF

$$y^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{1}{R_0}\exp\left(a_0 - \gamma_{\mathcal{F}}R_p\right), \tag{C43}$$

where

$$a_0 = -\frac{\gamma^{*2}\sigma_p^2}{2} + \gamma^*\mu_p. \tag{C44}$$

Using Stein's lemma as in the proof of Lemma 2, we can easily establish that

$$E[R_0 y^{\mathcal{F}}(\gamma^*)] = 1, \qquad E[R y^{\mathcal{F}}(\gamma^*)] = 1_N, \tag{C45}$$

so that $y^{\mathcal{F}}(\gamma^*) \in \mathcal{M}^+$.

Now consider a factor $f = R_p + \epsilon$, where ϵ is a normal mean-zero measurement error independent of the returns. It follows that $\mu_f = E[f] = \mu_p$ and $\sigma_f^2 = \text{Var}[f] > \sigma_p^2$. Consider an alternative SDF

$$y^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{1}{R_0}\exp\left(\tilde{a}_0 - \gamma_{\mathcal{G}} f\right), \tag{C46}$$

where

$$\tilde{a}_0 = -\frac{\gamma^{*2}\sigma_f^2}{2} + \gamma^*\mu_p. \tag{C47}$$

Using Stein's lemma, we obtain

$$E[R_0 y^{\mathcal{G}}(\gamma^*)] = 1, \qquad E[R y^{\mathcal{G}}(\gamma^*)] = 1_N, \tag{C48}$$

and $y^{\mathcal{G}}(\gamma^*)$ is also a correctly specified model. Note that $y^{\mathcal{F}}(\gamma_{\mathcal{F}})$ and $y^{\mathcal{G}}(\gamma_{\mathcal{G}})$ are two strictly non-nested models because there are no choices of $\gamma_{\mathcal{F}}$ and $\gamma_{\mathcal{G}}$ that can make these two SDFs identical. This example shows that we can have two strictly non-nested positive SDFs that are both correctly specified.

Proof of Proposition 5: From Lemma C.1 and under the null $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$, we can use (C27) to obtain

$$\begin{aligned} & T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \\ = \; & \frac{1}{4}\begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}) \end{bmatrix}' \begin{bmatrix} U^{-\frac{1}{2}} P_{\mathcal{F}} P'_{\mathcal{F}} U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}} P_{\mathcal{G}} P'_{\mathcal{G}} U^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}) \end{bmatrix} + o_p(1). \end{aligned} \tag{C49}$$

From Assumptions A, B, and C, we have

$$\begin{bmatrix} \sqrt{T}\bar{v}^{\mathcal{F}}_{2,T}(\theta^*_{\mathcal{F}}) \\ \sqrt{T}\bar{v}^{\mathcal{G}}_{2,T}(\theta^*_{\mathcal{G}}) \end{bmatrix} \overset{A}{\sim} N\left(0_{2n}, 4\mathcal{S}\right). \tag{C50}$$

Hence,

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} z'\mathcal{S}^{\frac{1}{2}} \begin{bmatrix} U^{-\frac{1}{2}} P_{\mathcal{F}} P'_{\mathcal{F}} U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}} P_{\mathcal{G}} P'_{\mathcal{G}} U^{-\frac{1}{2}} \end{bmatrix} \mathcal{S}^{\frac{1}{2}} z, \tag{C51}$$

where $z \sim N(0_{2n}, I_{2n})$. Furthermore, the nonzero eigenvalues of

$$\mathcal{S}^{\frac{1}{2}} \begin{bmatrix} U^{-\frac{1}{2}} P_{\mathcal{F}} P_{\mathcal{F}}' U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}} P_{\mathcal{G}} P_{\mathcal{G}}' U^{-\frac{1}{2}} \end{bmatrix} \mathcal{S}^{\frac{1}{2}} \tag{C52}$$

are the same as the eigenvalues of the matrix

$$\begin{bmatrix} P_{\mathcal{F}}' U^{-\frac{1}{2}} & 0_{(n-k_1)\times n} \\ 0_{(n-k_2)\times n} & P_{\mathcal{G}}' U^{-\frac{1}{2}} \end{bmatrix} \mathcal{S} \begin{bmatrix} U^{-\frac{1}{2}} P_{\mathcal{F}} & 0_{n\times(n-k_2)} \\ 0_{n\times(n-k_1)} & -U^{-\frac{1}{2}} P_{\mathcal{G}} \end{bmatrix}. \tag{C53}$$

Then, it follows that

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} \sum_{i=1}^{2n-k_1-k_2} \xi_i \nu_i, \tag{C54}$$

where the ξ_i's are the eigenvalues of the matrix

$$\begin{bmatrix} P_{\mathcal{F}}' U^{-\frac{1}{2}} S_{\mathcal{F}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P_{\mathcal{F}}' U^{-\frac{1}{2}} S_{\mathcal{F}\mathcal{G}} U^{-\frac{1}{2}} P_{\mathcal{G}} \\ P_{\mathcal{G}}' U^{-\frac{1}{2}} S_{\mathcal{G}\mathcal{F}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P_{\mathcal{G}}' U^{-\frac{1}{2}} S_{\mathcal{G}} U^{-\frac{1}{2}} P_{\mathcal{G}} \end{bmatrix}. \tag{C55}$$

This completes the proof.

Proof of Proposition 6: Since $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ under the null, it follows that $\lambda_{\mathcal{F}}^* = \lambda_{\mathcal{G}}^*$ and $m_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = m_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$ which implies that $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$. Using these identities, we have

$$\frac{\partial \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)}{\partial \lambda_{\mathcal{F}}} = 2[x_t m_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)^+ - q_{t-1}] = 2[x_t m_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)^+ - q_{t-1}] = \frac{\partial \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)}{\partial \lambda_{\mathcal{G}}}, \tag{C56}$$

and

$$\bar{v}_{2,T}^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \bar{v}_{2,T}^{\mathcal{G}}(\theta_{\mathcal{G}}^*). \tag{C57}$$

It is convenient to express the null hypothesis $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$ as a functional dependence

$$H_0 : \gamma_{\mathcal{G}}^* = g(\gamma_{\mathcal{F}}^*), \tag{C58}$$

where $g(\cdot)$ is a twice continuously differentiable function from $\Gamma_{\mathcal{F}}$ to $\Gamma_{\mathcal{G}}$ (see Gallant, 1987; Vuong, 1989).[28] Denote by

$$G(\gamma_{\mathcal{F}}) = \frac{\partial g(\gamma_{\mathcal{F}})}{\partial \gamma_{\mathcal{F}}'} \tag{C59}$$

the $k_2 \times k_1$ matrix of derivatives of $g(\gamma_{\mathcal{F}})$ with respect to $\gamma_{\mathcal{F}}$. Gallant (1987, p. 241) shows that

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) G(\gamma_{\mathcal{F}}^*) = \Psi^{\mathcal{G}}(g(\gamma_{\mathcal{F}}^*)) G(\gamma_{\mathcal{F}}^*) = 0_{(k_2-k_1)\times k_1}. \tag{C60}$$

[28] Gallant (1987, Section 3.6) provides a discussion of these two alternative representations of the null hypothesis.

Define the matrices

$$\mathbb{S} = [\Psi_*^{\mathcal{G}},\ 0_{(k_2-k_1)\times n}], \qquad \mathbb{Q} = \begin{bmatrix} G(\gamma_{\mathcal{F}}^*) & 0_{k_2\times n} \\ 0_{n\times k_1} & I_n \end{bmatrix} \tag{C61}$$

and note that $\mathbb{S}\mathbb{Q} = 0_{(k_2-k_1)\times(n+k_1)}$. Then, using (C57) and (C58), it follows that (see Lemma B in Vuong, 1989)

$$\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \mathbb{Q}'\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \tag{C62}$$

and

$$H_{\mathcal{F}} = \mathbb{Q}' H_{\mathcal{G}} \mathbb{Q}. \tag{C63}$$

By Lemma C.1 and the fact that $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$ under the null, we obtain

$$\begin{aligned} & T(\hat{\delta}_{\mathcal{F},+}^2 - \hat{\delta}_{\mathcal{G},+}^2) \\ = & -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)' H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\ = & -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'\mathbb{Q}(\mathbb{Q}' H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\ = & \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-\frac{1}{2}}[I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\mathbb{Q}(\mathbb{Q}' H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}' H_{\mathcal{G}}^{\frac{1}{2}}] H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \end{aligned} \tag{C64}$$

Using $\mathbb{S}\mathbb{Q} = 0_{(k_2-k_1)\times(n+k_1)}$, it can be shown that (see Gallant, 1987, pp. 241–242)

$$I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\mathbb{Q}(\mathbb{Q}' H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}' H_{\mathcal{G}}^{\frac{1}{2}} = H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}. \tag{C65}$$

Substituting (C65) into (C64) yields

$$\begin{aligned} T(\hat{\delta}_{\mathcal{F},+}^2 - \hat{\delta}_{\mathcal{G},+}^2) &= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-\frac{1}{2}}[H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}] H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\ &= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \end{aligned} \tag{C66}$$

Furthermore, invoking

$$\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \overset{A}{\sim} N\left(0_{n+k_2}, M_{\mathcal{G}}\right), \tag{C67}$$

we have

$$T(\hat{\delta}_{\mathcal{F},+}^2 - \hat{\delta}_{\mathcal{G},+}^2) \overset{A}{\sim} \frac{1}{2} z' \left[M_{\mathcal{G}}^{\frac{1}{2}} H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1} M_{\mathcal{G}}^{\frac{1}{2}} \right] z, \tag{C68}$$

where $z \sim N(0_{n+k_2}, I_{n+k_2})$. Since the eigenvalues of the matrix $\frac{1}{2}M_{\mathcal{G}}^{\frac{1}{2}} H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1} M_{\mathcal{G}}^{\frac{1}{2}}$ are the same as the eigenvalues of the matrix

$$\frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1} M_{\mathcal{G}} H_{\mathcal{G}}^{-1}\mathbb{S}' = \frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}\Sigma_{\hat{\theta}_{\mathcal{G}}}\mathbb{S}' = (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}, \tag{C69}$$

we conclude that

$$T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \overset{A}{\sim} \sum_{i=1}^{k_2-k_1} \xi_i \nu_i, \tag{C70}$$

where the ν_i's are independent chi-squared random variables with one degree of freedom and the ξ_i's are the eigenvalues of the matrix in (C69). Since $A = \Psi_*^{\mathcal{G}} \tilde{H}_{\mathcal{G}} \Psi_*^{\mathcal{G}\prime}$ and $B = \Psi_*^{\mathcal{G}} \Sigma_{\hat{\gamma}_{\mathcal{G}}} \Psi_*^{\mathcal{G}\prime}$ are two symmetric positive definite matrices, $A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ is also symmetric positive definite with positive eigenvalues. Furthermore, because $A^{-1}B$ and $A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ share the same eigenvalues, the eigenvalues of $A^{-1}B$ are also positive. This completes the proof.

Proof of Proposition 7: Define the following matrices

$$\mathbb{S}_{\mathcal{F}} = \left[\Psi_*^{\mathcal{F}},\ 0_{(k_1-k_3)\times n}\right], \qquad \mathbb{S}_{\mathcal{G}} = \left[\Psi_*^{\mathcal{G}},\ 0_{(k_2-k_3)\times n}\right]. \tag{C71}$$

Since $\mathcal{H} \subset \mathcal{F}$ and $\mathcal{H} \subset \mathcal{G}$, we can use the results from the proof of Proposition 6 to obtain

$$T(\hat{\delta}^2_{\mathcal{H},+} - \hat{\delta}^2_{\mathcal{F},+}) = \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)' H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}' (\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}')^{-1} \mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + o_p(1) \tag{C72}$$

and

$$T(\hat{\delta}^2_{\mathcal{H},+} - \hat{\delta}^2_{\mathcal{G},+}) = \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}' (\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}')^{-1} \mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \tag{C73}$$

Taking the difference yields

$$\begin{aligned} T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \;=\; & -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)' H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}' (\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}')^{-1} \mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ & + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)' H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}' (\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}')^{-1} \mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \end{aligned} \tag{C74}$$

From Assumptions A, B, and C, the joint empirical process $\sqrt{T}[\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)',\ \bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)']'$ converges to a Gaussian process:

$$\begin{bmatrix} \sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ \sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \end{bmatrix} \overset{A}{\sim} N\left(0_{2n+k_1+k_2}, \mathbb{M}\right), \tag{C75}$$

where

$$\mathbb{M} = \lim_{T\to\infty} \mathrm{Var} \begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)}{\partial \theta_{\mathcal{F}}} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)}{\partial \theta_{\mathcal{G}}} \end{bmatrix}. \tag{C76}$$

Hence,

$$\begin{aligned} & T(\hat{\delta}^2_{\mathcal{F},+} - \hat{\delta}^2_{\mathcal{G},+}) \\ & \overset{A}{\sim} z' \left[\frac{1}{2} \mathbb{M}^{\frac{1}{2}} \begin{pmatrix} -H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}' (\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}')^{-1} \mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}' (\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}')^{-1} \mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \end{pmatrix} \mathbb{M}^{\frac{1}{2}} \right] z, \end{aligned} \tag{C77}$$

where $z \sim N(0_{2n+k_1+k_2}, I_{2n+k_1+k_2})$. Then, using the fact that AB and BA share the same nonzero eigenvalues, the matrix in the square brackets in (C77) has the same nonzero eigenvalues as the matrix

$$\frac{1}{2}\begin{bmatrix} -(\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}')^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}')^{-1} \end{bmatrix} \times \begin{bmatrix} \mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} & 0_{(k_1-k_3)\times(n+k_2)} \\ 0_{(k_2-k_3)\times(n+k_1)} & \mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \end{bmatrix} \mathbb{M} \begin{bmatrix} H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}' & 0_{(n+k_1)\times(k_2-k_3)} \\ 0_{(n+k_2)\times(k_1-k_3)} & H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}' \end{bmatrix}. \quad \text{(C78)}$$

Using the fact that $\mathbb{S}_{\mathcal{F}} H_{\mathcal{F}}^{-1} \mathbb{S}_{\mathcal{F}}' = \frac{1}{2}\Psi_*^{\mathcal{F}} \tilde{H}_{\mathcal{F}} \Psi_*^{\mathcal{F}\prime}$, $\mathbb{S}_{\mathcal{G}} H_{\mathcal{G}}^{-1} \mathbb{S}_{\mathcal{G}}' = \frac{1}{2}\Psi_*^{\mathcal{G}} \tilde{H}_{\mathcal{G}} \Psi_*^{\mathcal{G}\prime}$, and

$$\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}} = \begin{bmatrix} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1} \end{bmatrix} \mathbb{M} \begin{bmatrix} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1} \end{bmatrix} \quad \text{(C79)}$$

is the asymptotic covariance matrix of $[\hat{\theta}_{\mathcal{F}}', \ \hat{\theta}_{\mathcal{G}}']'$, the matrix in (C78) can be written as

$$\begin{bmatrix} -(\Psi_*^{\mathcal{F}} \tilde{H}_{\mathcal{F}} \Psi_*^{\mathcal{F}\prime})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi_*^{\mathcal{G}} \tilde{H}_{\mathcal{G}} \Psi_*^{\mathcal{G}\prime})^{-1} \end{bmatrix} \Psi_*^{\mathcal{F}\mathcal{G}} \Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}} \Psi_*^{\mathcal{F}\mathcal{G}\prime}. \quad \text{(C80)}$$

Therefore,

$$T(\hat{\delta}_{\mathcal{F},+}^2 - \hat{\delta}_{\mathcal{G},+}^2) \overset{A}{\sim} \sum_{i=1}^{k_1+k_2-2k_3} \xi_i \nu_i, \quad \text{(C81)}$$

where the ξ_i's are the eigenvalues of the matrix in (C80). This completes the proof.

References

Almeida, C., Garcia, R., 2009. Assessing misspecified asset pricing models with empirical likelihood estimators. Unpublished working paper. EDHEC Business School.

Andrews, D. W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D. W. K., 1993. An introduction to econometric applications of empirical process theory for dependent random variables. Econometric Reviews 12, 183–216.

Andrews, D. W. K., 1994. Empirical process methods in econometrics. In R. F. Engle and D. L. McFadden (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, 2247–2294.

Andrews, D. W. K., Pollard, D., 1994. An introduction to functional central limit theorems for dependent stochastic processes. International Statistical Review 62, 119–132.

Bailey, W., Li, H., Zhang, X., 2004. Hedge fund performance evaluation: A stochastic discount factor approach. Unpublished working paper. Cornell University.

Bansal, R., Hsieh, D. A., Viswanathan, S., 1993. A new approach to international arbitrage pricing. Journal of Finance 48, 1719–1747.

Boyle, P., Feng, S., Tian, W., Wang, T., 2008. Robust stochastic discount factors. Review of Financial Studies 21, 1077–1122.

Burnside, A. C., 1994. Hansen-Jagannathan bounds as classical tests of asset pricing models. Journal of Business and Economic Statistics 12, 57–79.

Burnside, A. C., 2007. Empirical asset pricing and statistical power in the presence of weak risk factors. Unpublished working paper. Duke University.

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

Chen, X., Ludvigson, S. C., 2009. Land of addicts? An empirical investigation of habit-based asset pricing models. Journal of Applied Econometrics 24, 1057–1093.

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

Fama, E. F., French, K. R., 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, 473–503.

Fletcher, J., 2009. Arbitrage and the evaluation of linear factor models in U.K. stock returns. Financial Review, forthcoming.

Gallant, R. A., 1987. Nonlinear Statistical Models. Wiley, New York.

Golden, R. M., 2003. Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. Psychometrica 68, 229–249.

Hall, A. R., Pelletier, D., 2008. Non-nested testing in models estimated via generalized method of moments. Unpublished working paper. North Carolina State University.

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

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

Hansen, L., 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.

Inoue, A., Kilian, L., 2002. Bootstrapping autoregressive processes with possible unit roots. Econometrica 70, 377–391.

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.

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

Kan, R., Robotti, C., 2008. The exact distribution of the Hansen-Jagannathan bound. Unpublished working paper. University of Toronto.

Kan, R., Robotti, C., 2008. Specification tests of asset pricing models using excess returns. Journal of Empirical Finance 15, 816–838.

Kan, R., Robotti, C., 2009. Model comparison using the Hansen-Jagannathan distance. Review of Financial Studies 22, 3449–3490.

Kan, R., Robotti, C., Shanken, J., 2009. Pricing model performance and the two-pass cross-sectional regression methodology. Unpublished working paper. Federal Reserve Bank of Atlanta.

Kan, R., Zhang, C., 1999a. GMM tests of stochastic discount factor models with useless factors. Journal of Financial Economics 54, 103–127.

Kan, R., Zhang, C., 1999b. Two-pass tests of asset pricing models with useless factors. Journal of Finance 54, 204–235.

Kan, R., Zhou, G., 2003. Hansen-Jagannathan distance: Geometry and exact distribution. Unpublished working paper. University of Toronto.

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.

Lewellen, J. W., Nagel, S., Shanken, J., 2009. A skeptical appraisal of asset-pricing tests. Journal of Financial Economics, forthcoming.

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

Li, H., Xu, Y., Zhang, X., 2009. Evaluating asset pricing models using the second Hansen-Jagannathan distance. Journal of Financial Economics, forthcoming.

Lien, D., Vuong, Q. H., 1987. Selecting the best linear regression model: A classical approach. Journal of Econometrics 35, 3–23.

Liu, X., Kuo, J. M., Coakley, J., 2009. A pricing kernel approach to valuing interest rate options. Unpublished working paper. University of Essex.

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

Marcellino, M., Rossi, B., 2008. Model selection for nested and overlapping nonlinear, dynamic and possibly mis-specified models. Oxford Bulletin of Economics and Statistics 70, 867–893.

Newey, W. K., West, K. D., 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 55, 703–708.

Newey, W. K., West, K. D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.

Parker, J. A., Julliard, C., 2005. Consumption risk and the cross section of expected returns. Journal of Political Economy 113, 185–222.

Pollard, D., 1990. Empirical processes: Theory and applications. NSF-CBMS Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics, Hayward, California.

Rivers, D., Vuong, Q. H., 2002. Model selection tests for nonlinear dynamic models. Econometrics Journal 5, 1–39.

Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. Review of Financial Studies 19, 1–44.

Sims, C. A., Stock, J. H., Watson, M. W., 1990. Inference in linear time series with some unit roots. Econometrica 58, 113–144.

Stock, J. H., Wright, J. H., 2000. GMM with weak identification. Econometrica 68, 1055–1096.

Vuong, Q. H., 1989. Likelihood ration tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Wang, Z., Zhang, X., 2005. Empirical evaluation of asset pricing models: Arbitrage and pricing errors over contingent claims. Unpublished working paper. Cornell University.

Yogo, M., 2006. A consumption-based explanation of expected stock returns. Journal of Finance 61, 539–580.

Table 1
Summary of the models

The table presents the sample unconstrained and constrained HJ-distances ($\hat{\delta}$ and $\hat{\delta}_+$, respectively) of seven linear asset pricing models. The models include the conditional consumption CAPM (LL) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (LV) of Lustig and Van Nieuwerburgh (2004), the conditional CAPM (SV) of Santos and Veronesi (2006), the simple and extended sector investment models (LVX1 and LVX2, respectively) of Li, Vassalou, and Xing (2006), the durable consumption CAPM (YOGO) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The models are estimated using quarterly gross returns on the three-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios. The data are from 1952:2 to 2000:4 (195 observations). $p(\delta = 0)$ is the p-value for the test of $H_0 : \delta = 0$. $p_n(\delta_+ = 0)$ and $p_a(\delta_+ = 0)$ are the p-values for the test of $H_0 : \delta_+ = 0$ imposing the null and not imposing the null, respectively. $\text{se}(\hat{\delta})$ ($\text{se}(\hat{\delta}_+)$) is the standard error of $\hat{\delta}$ ($\hat{\delta}_+$) under the assumption that $\delta > 0$ ($\delta_+ > 0$). We use the Newey and West (1994) HAC estimator with Bartlett kernel and automatic lag selection to account for serial correlation in the data. $P[\hat{y} < 0]$ is the probability for the estimated SDF to take on negative values during the sample period. $\sigma_{\hat{y}}$ is the standard deviation of the estimated SDF. $\sigma_{\hat{\phi}}$ is the standard deviation of the estimated ϕ function. ρ_c^2 is the centered R^2 from the linear regression of the estimated SDF on the returns on the test assets. $(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$ is the percentage difference between the sample constrained and unconstrained HJ-distances. $\hat{\tilde{\delta}}$ is a sample measure of the maximum pricing error on the test assets for the SDF that minimizes the constrained HJ-distance.

Panel A: Unconstrained Hansen-Jagannathan distance

Model	**LL**	**LV**	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
$\hat{\delta}$	0.643	0.643	0.642	0.580	0.546	0.651	0.582
$p(\delta = 0)$	0.004	0.003	0.000	0.320	0.494	0.000	0.015
$\text{se}(\hat{\delta})$	0.090	0.080	0.089	0.108	0.110	0.087	0.092
$P[\hat{y} < 0]$	0.021	0.103	0.010	0.138	0.154	0.000	0.015
$\sigma_{\hat{y}}$	0.592	0.817	0.335	1.229	1.478	0.273	0.389
$\sigma_{\hat{\phi}}$	1.667	1.660	1.495	1.862	1.846	1.481	1.365
ρ_c^2	0.234	0.112	0.687	0.105	0.088	0.845	0.983

Panel B: Constrained Hansen-Jagannathan distance

Model	**LL**	**LV**	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
$\hat{\delta}_+$	0.685	0.700	0.667	0.691	0.684	0.673	0.607
$p_n(\delta_+ = 0)$	0.000	0.038	0.000	0.000	0.000	0.000	0.000
$p_a(\delta_+ = 0)$	0.014	0.028	0.007	0.009	0.004	0.005	0.028
$\text{se}(\hat{\delta}_+)$	0.094	0.089	0.096	0.094	0.099	0.097	0.099
$P[\hat{y} < 0]$	0.000	0.000	0.000	0.005	0.010	0.000	0.015
$\sigma_{\hat{y}}$	0.392	0.294	0.296	0.318	0.350	0.260	0.389
$\sigma_{\hat{\phi}}$	1.870	1.777	1.702	1.828	1.828	1.696	1.576
ρ_c^2	0.239	0.138	0.795	0.129	0.136	0.925	0.983
$(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$	6.4%	8.8%	3.9%	19.2%	25.2%	3.4%	4.2%
$\hat{\tilde{\delta}}$	0.651	0.666	0.643	0.646	0.638	0.651	0.582

Table 2

Model selection tests

The table presents the p-values for a sequential test based on the squared unconstrained and constrained HJ-distances of seven linear asset pricing models. The models include the conditional consumption CAPM (LL) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (LV) of Lustig and Van Nieuwerburgh (2004), the conditional CAPM (SV) of Santos and Veronesi (2006), the simple and extended sector investment models (LVX1 and LVX2, respectively) of Li, Vassalou, and Xing (2006), the durable consumption CAPM (YOGO) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The models are estimated using quarterly gross returns on the three-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios. The data are from 1952:2 to 2000:4 (195 observations). Panel A presents the p-values for the test of $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$ given in Propositions 6 (for nested models) and 7 (for overlapping models). For those overlapping models where the null hypothesis is rejected at the 5% level, Panel B reports the p-values for the test of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ (unconstrained HJ-distance) and $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$ (constrained HJ-distance) derived in Proposition 5. Finally, for the overlapping models where the null hypothesis in Panel B is rejected at the 5% level, Panel C presents the p-values for the test of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} \neq 0$ (unconstrained HJ-distance) and $H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} \neq 0$ (constrained HJ-distance) based on the normal test in equation (100). All p-values are computed using the Newey and West (1994) HAC estimator with Bartlett kernel and automatic lag selection. Bold entries denote statistical significance at the 5% level.

Panel A: Test of $H_0 : y^{\mathcal{F}} = y^{\mathcal{G}}$

	Unconstrained Hansen-Jagannathan distance					
	LV	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
LL	0.981	0.265	0.390	0.644	0.975	**0.001**
LV		0.282	0.368	0.608	0.884	**0.001**
SV			0.721	0.978	0.140	**0.003**
LVX1				0.512	0.393	0.310
LVX2					0.682	0.321
YOGO						**0.001**

	Constrained Hansen-Jagannathan distance					
	LV	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
LL	0.531	**0.044**	0.984	0.961	0.444	**0.000**
LV		**0.021**	0.501	0.535	0.226	**0.000**
SV			**0.035**	**0.023**	0.088	**0.003**
LVX1				0.389	0.517	**0.000**
LVX2					0.186	**0.001**
YOGO						**0.001**

Table 2
Model selection tests (continued)

Panel B: Test of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} = 0$ ($H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} = 0$)

	Unconstrained Hansen-Jagannathan distance					
	LV	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
LL						0.259
LV						0.466
SV						**0.029**
LVX1						
LVX2						
YOGO						**0.003**

	Constrained Hansen-Jagannathan distance					
	LV	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
LL		0.584				**0.023**
LV		0.052				**0.001**
SV			0.263	0.263		**0.007**
LVX1						**0.001**
LVX2						**0.001**
YOGO						**0.001**

Panel C: Test of $H_0 : \delta^2_{\mathcal{F}} = \delta^2_{\mathcal{G}} \neq 0$ ($H_0 : \delta^2_{\mathcal{F},+} = \delta^2_{\mathcal{G},+} \neq 0$)

	Unconstrained Hansen-Jagannathan distance					
	LV	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
LL						
LV						
SV						0.077
LVX1						
LVX2						
YOGO						**0.042**

	Constrained Hansen-Jagannathan distance					
	LV	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
LL						0.077
LV						**0.033**
SV						0.059
LVX1						0.070
LVX2						0.089
YOGO						0.059

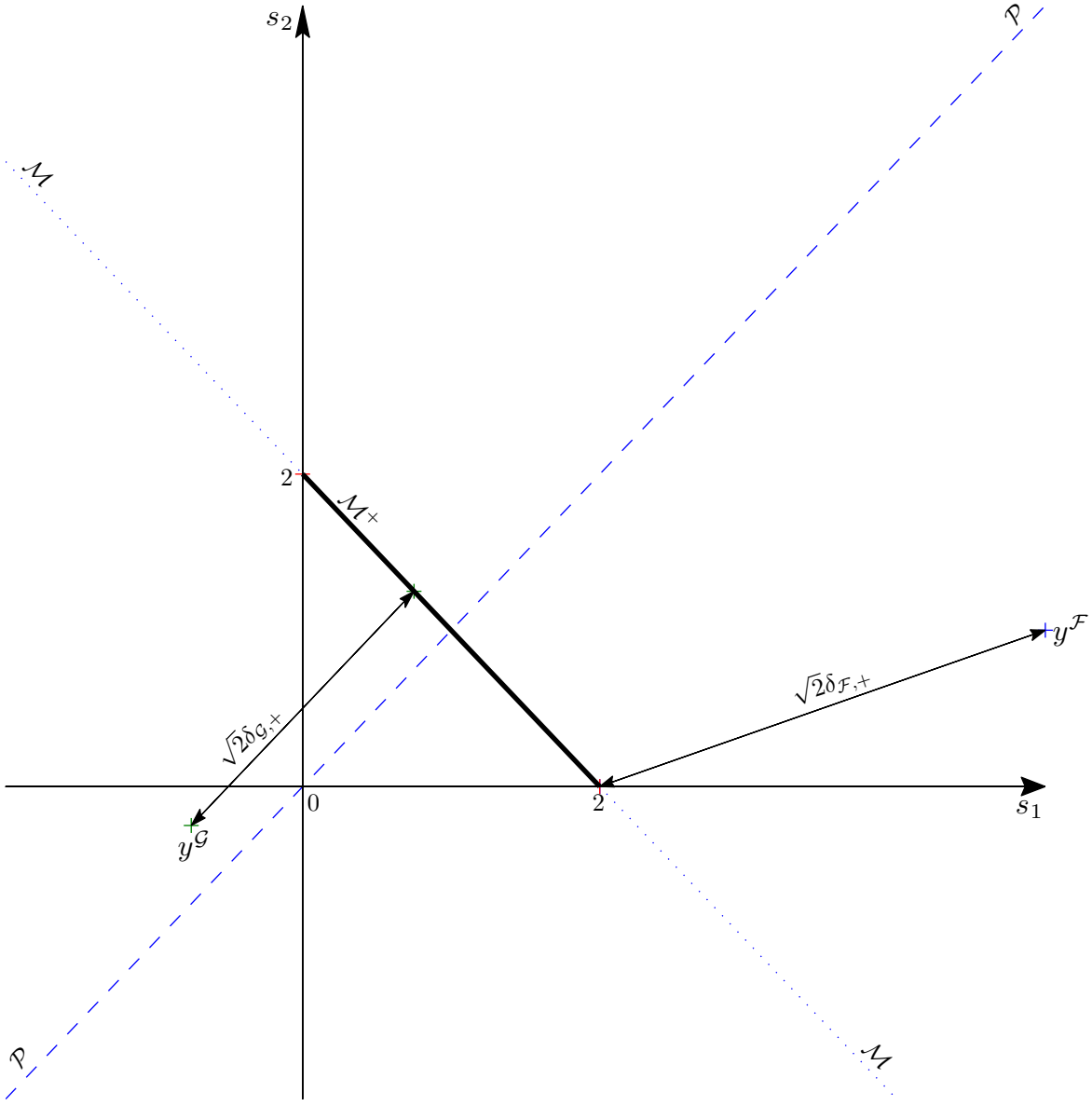

Figure 1
Graphical representation of the constrained Hansen-Jagannathan distance. The figure presents two SDFs ($y^{\mathcal{F}}$ and $y^{\mathcal{G}}$) in an economy with two states (s_1 and s_2) and one risk-free asset. The dashed line represents the payoff space of the risk-free asset ($\mathcal{P}$). The dotted line represents the admissible set of SDFs ($\mathcal{M}$) and the thick solid line represents the set of nonnegative admissible SDFs ($\mathcal{M}^+$). The shortest distance between $y^{\mathcal{F}}$ ($y^{\mathcal{G}}$) and $\mathcal{M}^+$ is proportional to its constrained HJ-distance, labeled as $\delta_{\mathcal{F},+}$ ($\delta_{\mathcal{G},+}$).

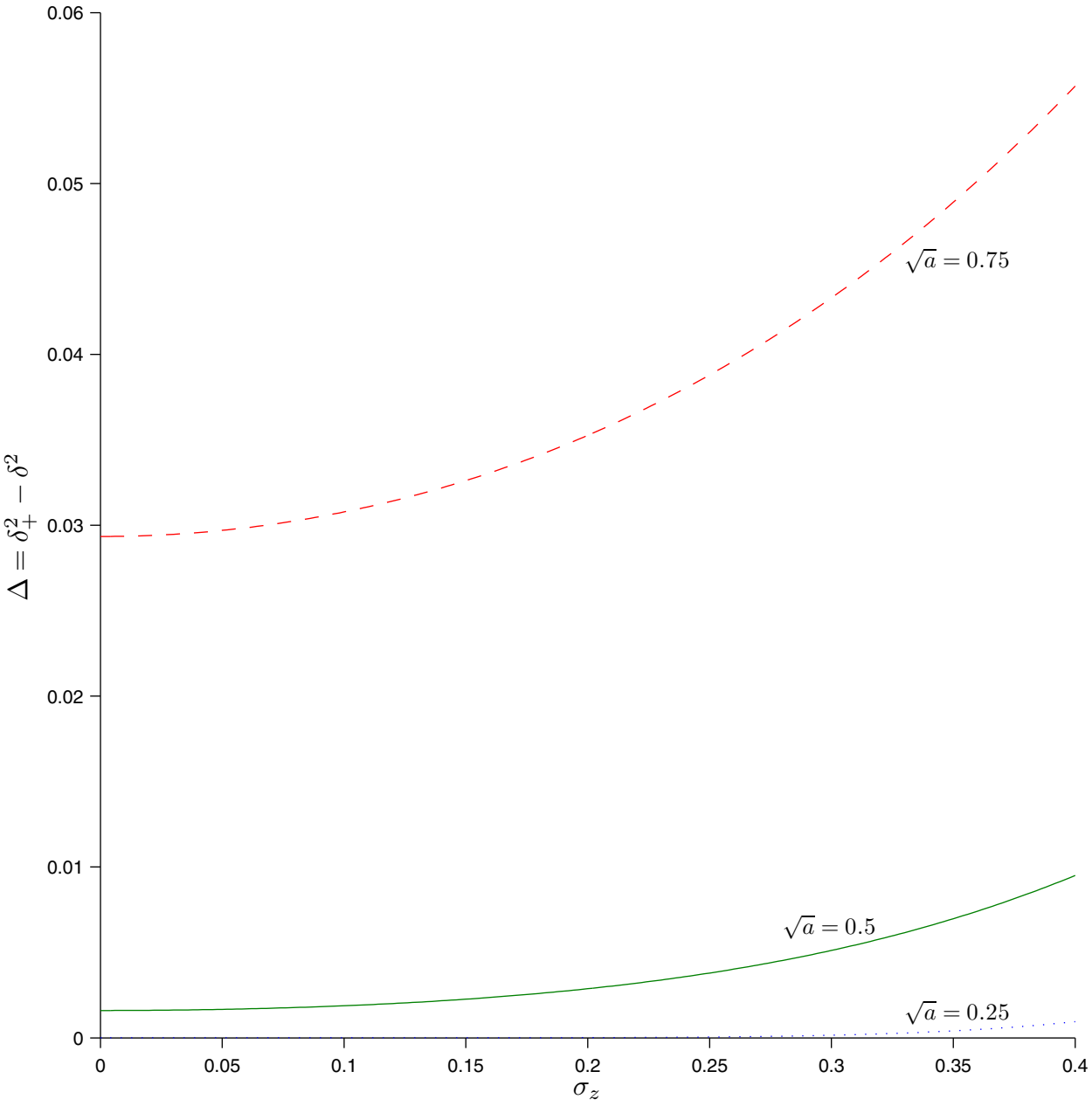

Figure 2
Difference between squared constrained and unconstrained Hansen-Jagannathan distances. The figure plots the difference between squared constrained and unconstrained HJ-distances (Δ) as a function of the standard deviation of the unspanned component (σ_z) of the candidate SDF. The gross risk-free rate is assumed to be 1.005. The dotted line represents the case in which the Sharpe ratio of the tangency portfolio ($\sqrt{a}$) is 0.25. The solid line is for $\sqrt{a} = 0.5$, and the dashed line is for $\sqrt{a} = 0.75$.

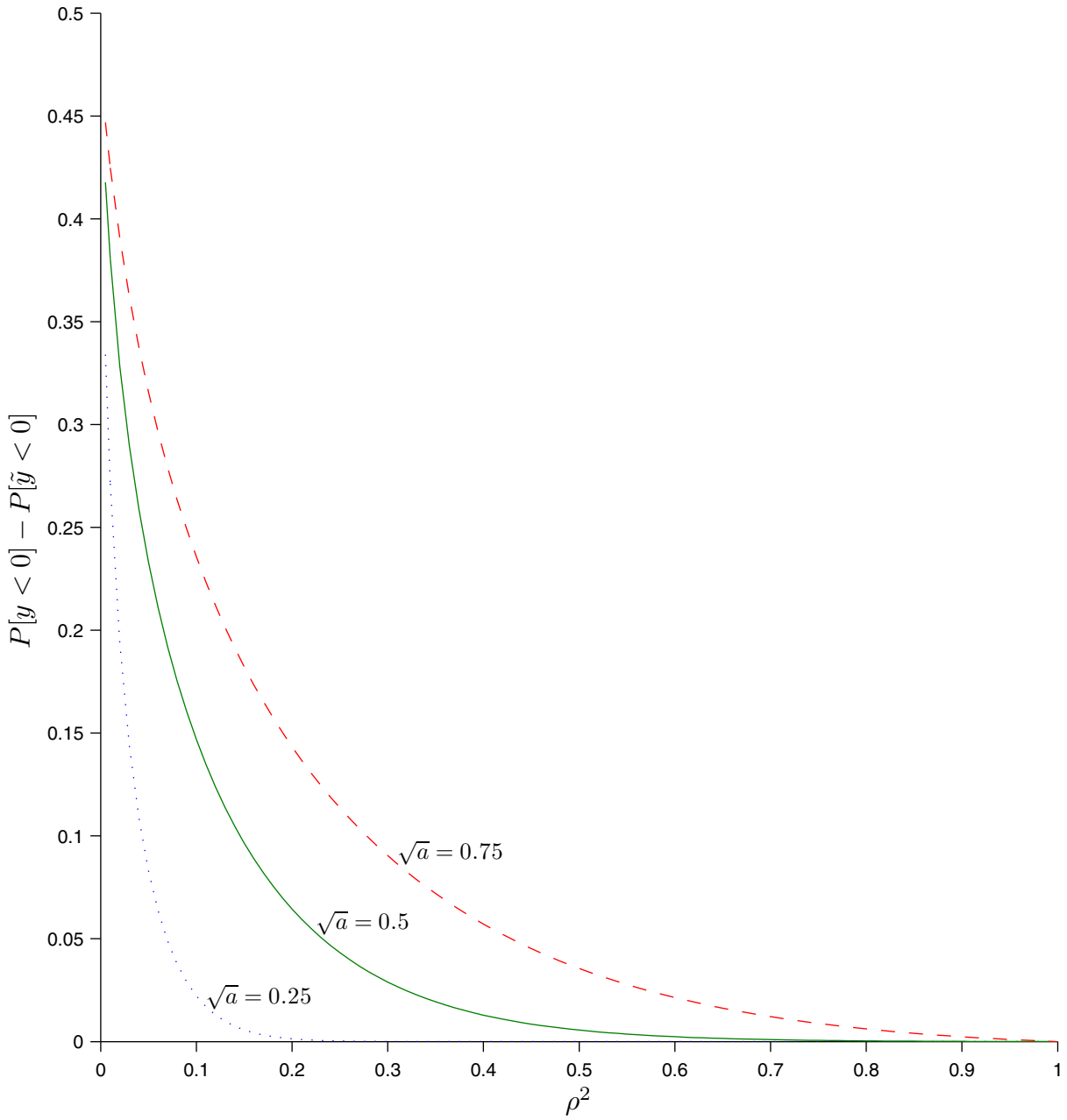

Figure 3
Difference in the probabilities of taking on negative values for two linear SDFs. The figure plots $P[y<0]-P[\tilde{y}<0]$ as a function of ρ^2 in a 1-factor setting, where y and $\tilde{y}$ are the linear SDFs chosen to minimize the unconstrained and constrained HJ-distances, respectively. ρ^2 is the proportion of variability of the factor that is explained by the returns. The dotted line represents the case in which the Sharpe ratio of the tangency portfolio ($\sqrt{a}$) is 0.25. The solid line is for $\sqrt{a}=0.5$, and the dashed line is for $\sqrt{a}=0.75$. In each case, we assume that the squared Sharpe ratio of the factor mimicking portfolio (a_1) is half of the value of a.

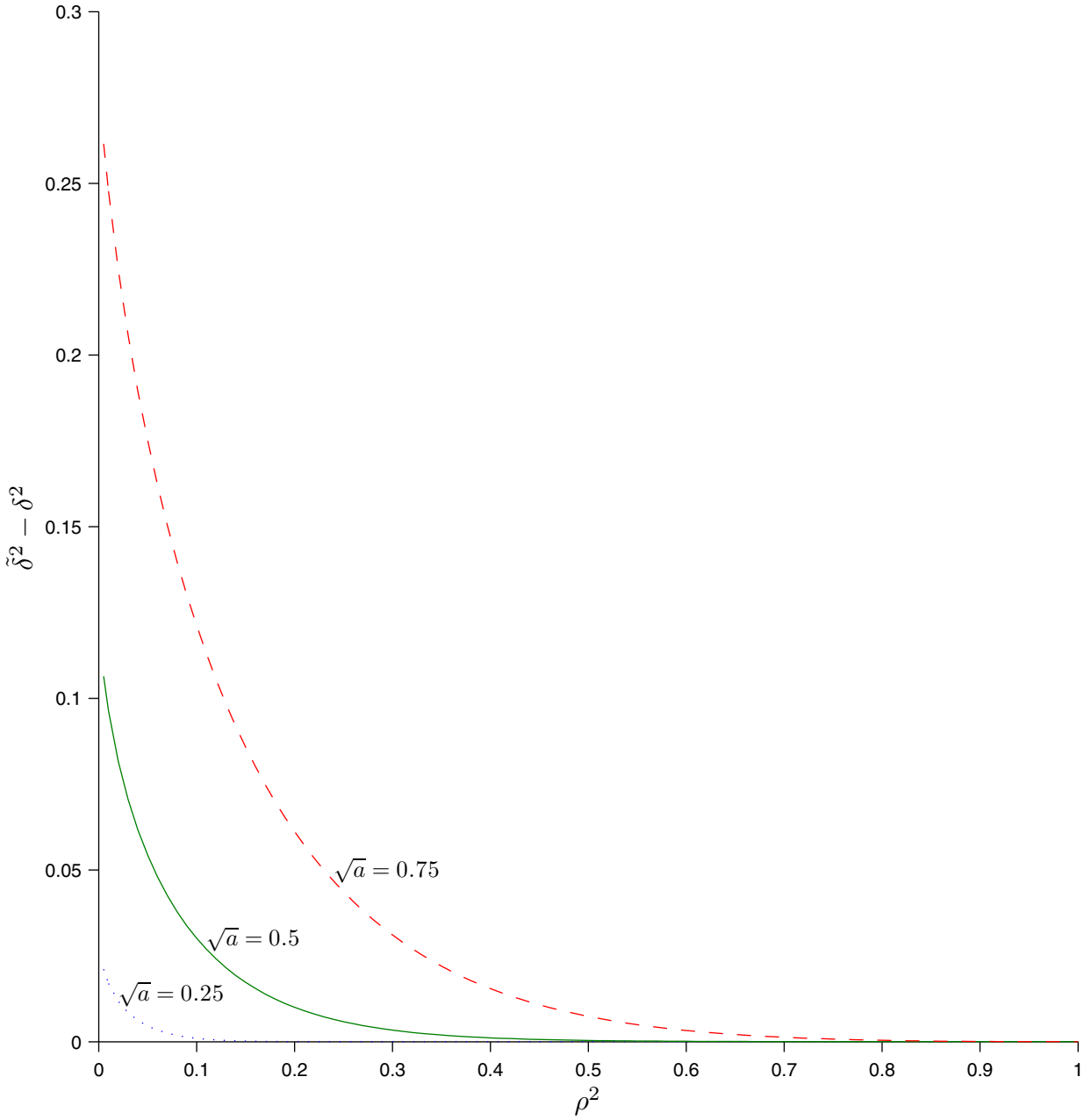

Figure 4
Difference in the aggregate measures of pricing errors of two linear SDFs. The figure plots $\tilde{\delta}^2 - \delta^2$ as a function of ρ^2 in a one-factor setting, where δ^2 and $\tilde{\delta}^2$ are the aggregate measures of pricing errors of the test assets when the linear SDF is chosen to minimize the unconstrained and constrained HJ-distances, respectively. ρ^2 is the proportion of variability of the factor that is explained by the returns. The dotted line represents the case in which the Sharpe ratio of the tangency portfolio ($\sqrt{a}$) is 0.25. The solid line is for $\sqrt{a} = 0.5$, and the dashed line is for $\sqrt{a} = 0.75$. In each case, we assume that the squared Sharpe ratio of the factor mimicking portfolio (a_1) is half of the value of a. The gross risk-free rate is assumed to be 1.005.

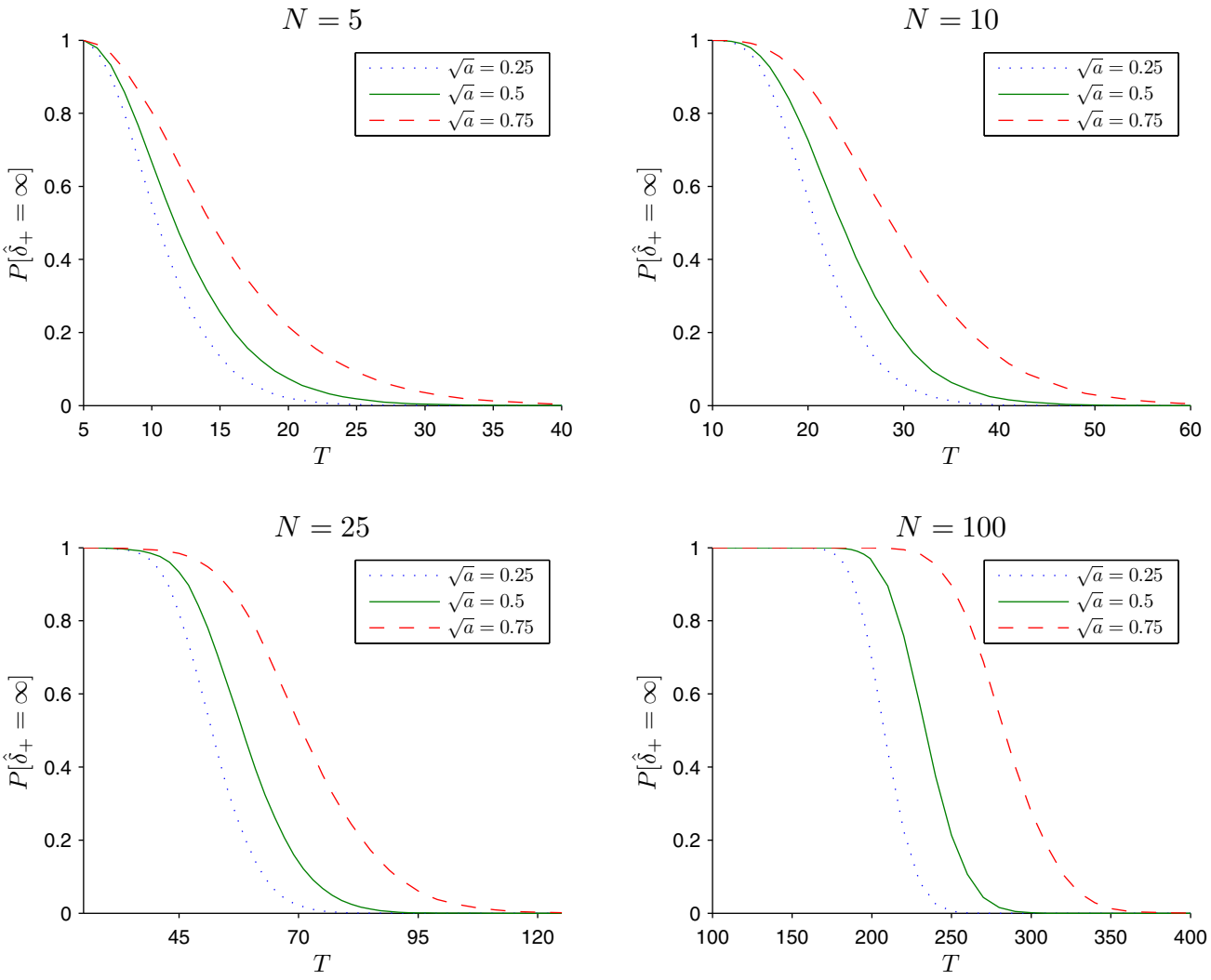

Figure 5
Probability for the sample constrained Hansen-Jagannathan distance to take on the value of infinity. The figure plots $P[\hat{\delta}_+ = \infty]$ as a function of the length of the time series (T). The payoffs of the test assets consist of a constant risk-free rate and the excess returns on N risky assets. The excess returns are assumed to be i.i.d. multivariate normally distributed. The four graphs in the figure present $P[\hat{\delta}_+ = \infty]$ for $N = 5$, 10, 25, 100. Within each graph, the dotted line represents the case in which the Sharpe ratio of the tangency portfolio of the risky assets ($\sqrt{a}$) is 0.25. The solid line is for $\sqrt{a} = 0.5$, and the dashed line is for $\sqrt{a} = 0.75$. The probabilities are computed based on 10,000 simulated series.