

# On the Properties of the Constrained Hansen-Jagannathan Distance

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti\*

October 2015

JEL # C12, C13, G12

\*Gospodinov is from the Federal Reserve Bank of Atlanta. Kan is from the University of Toronto. Robotti is from Imperial College London and Queen Mary University of London. We thank the Editor (Richard T. Baillie), an Associate Editor, Esther Eiling, Wayne Ferson, Jonathan Fletcher, Eric Jondeau, Francisco Peñaranda, B. Ravikumar, Sergei Sarkissian, Enrique Sentana, Jonathan Wright, Chu Zhang, Guofu Zhou, seminar participants at the Chinese University of Hong Kong, EDHEC Business School, Emory University, HEC Lausanne, National University of Singapore, Singapore Management University, University of Montreal, University of South Australia, and participants at the 2010 IFM2 Symposium on Mathematical Finance, the 2010 Meetings of the Society for Nonlinear Dynamics and Econometrics, the 2010 International Symposium on Econometric Theory and Applications, the 2010 Northern Finance Association Meetings, and the 2010 Third Annual SoFiE Conference for helpful discussions and comments. Gospodinov gratefully acknowledges financial support from Fonds de recherche sur la société et la culture (FQRSC), Institut de Finance Mathématique de Montréal (IFM2), and the Social Sciences and Humanities Research Council of Canada. 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. Send correspondence to Cesare Robotti, Imperial College Business School, Imperial College London, Tanaka Building, South Kensington Campus, London SW7 2AZ, United Kingdom; telephone: +44 20 7594 2682. E-mail: [c.robotti@imperial.ac.uk](mailto:c.robotti@imperial.ac.uk).

# On the Properties of the Constrained Hansen-Jagannathan Distance

## ABSTRACT

We provide an in-depth analysis of the theoretical properties of the Hansen-Jagannathan (HJ) distance that incorporates a no-arbitrage constraint. Under a multivariate elliptical distribution assumption, we present explicit expressions for the HJ-distance with a no-arbitrage constraint, the associated Lagrange multipliers, and the stochastic discount factor (SDF) parameters in the case of linear SDFs. This allows us to analyze the benefits and costs of using the HJ-distance with a no-arbitrage constraint to evaluate and rank models. We also study the asymptotic and finite-sample properties of the sample constrained HJ-distance. Finally, we demonstrate the practical relevance of our theoretical findings in an empirical illustration of some popular asset-pricing models.

# 1. Introduction

Since all asset-pricing models can be viewed as approximations to 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 (that is, 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, for example, 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 has received only limited attention in the literature. This can be attributed to some technical difficulties in the analysis of the properties of the population constrained HJ-distance that arise from explicitly incorporating the possibility of arbitrage in the model evaluation process. For example, unlike its unconstrained counterpart, a closed-form analytical expression for the constrained HJ-distance is not currently available even for linear models. As a result, the literature has focused on the sampling behavior (Hansen, Heaton, and Luttmer (1995) and Li, Xu, and Zhang (2010)) and the empirical performance of the constrained HJ-distance for evaluating and comparing asset-pricing models, while a rigorous theoretical analysis of the properties of the underlying population quantity is still largely missing.

Our paper has two main objectives. The first one is to better understand, in the context of linear SDFs, 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 for the constrained HJ-distance under the assumption that the SDF and the returns are conditionally multivariate elliptically 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, we obtain analytical expressions of the SDF parameters that minimize the constrained HJ-distance. When there is an unspanned factor in the SDF, we show that choosing parameters to minimize the constrained HJ-distance instead of the unconstrained HJ-distance will result in a lower probability for the SDF to take negative values, but will lead to a serious deterioration in the ability of the SDF to price the test assets.

Our second objective is to provide an improved analysis of the sample constrained HJ-distance for linear and nonlinear asset-pricing models. 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 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 provide some asymptotic results for the sample constrained HJ-distance and the estimates of the SDF parameters and the Lagrange multipliers that complement the existing literature.

The main findings of our analysis can be summarized as follows. On the positive side, the constrained HJ-distance is an effective tool in detecting arbitrage opportunities in the market. This feature, however, only depends on the properties of the test asset returns and not on the particular choice of a model. Furthermore, the population constrained HJ-distance and its parameters appear

to be well defined in the presence of factors that are uncorrelated with the returns on the test assets, which is not the case for the unconstrained HJ-distance (see, for example, Gospodinov, Kan, and Robotti, 2014). On the downside, the constrained HJ-distance lacks a clear maximum pricing error interpretation and using it to compare and rank competing asset-pricing models can be problematic. Moreover, while the SDF's probability of taking on negative values can be greatly reduced when the parameters are chosen to minimize the constrained HJ-distance, this is generally accompanied by a substantial deterioration in the pricing ability of the model. Intuitively, to reduce the SDF's probability of taking on negative values, the optimization problem based on the constrained HJ-distance forces the candidate SDF to load primarily on the risk-free asset. This, in turn, leads to a lower SDF's variability and to a deterioration in the model's pricing performance. As a result, the differences in the sample constrained HJ-distances of competing asset-pricing models mainly arise from the underlying characteristics of the risk factors (traded versus non-traded) and are often unrelated to their relative pricing performance. Finally, since we are actually not comparing the same sets of SDFs under the unconstrained and constrained HJ-distances, the choice of constrained versus unconstrained HJ-distance should depend on which metric is considered to be more reasonable for model comparison.

The rest of the paper is organized as follows. Section 2 introduces the main setup and notation for the population unconstrained and constrained HJ-distances. Section 3 derives an analytical solution for the constrained HJ-distance under the joint ellipticity assumption on the SDF and the returns on the test assets. Section 4 presents an econometric analysis of the sample constrained HJ-distance. Section 5 illustrates the relevance of our findings with an empirical analysis of popular asset-pricing models. Some concluding remarks are provided in Section 6.

## 2. Setup and Preliminaries

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. 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, so that  $\tilde{r}$  is of dimension  $n = N + 1$ . 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.

We call  $m \in L^2$  an admissible SDF if it prices the test assets correctly, that is,

$$E[\tilde{r}m] = q, \tag{1}$$

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.

Let  $y(\gamma) \in L^2$  be a candidate stochastic discount factor that depends on a  $k$ -vector of unknown parameters  $\gamma \in \Gamma$ , where  $\Gamma$  is the parameter space of  $\gamma$ . If  $y(\gamma)$  prices the  $n$  test assets correctly, then the vector of pricing errors,  $e(\gamma)$ , of the test assets is exactly zero:

$$e(\gamma) = E[\tilde{r}y(\gamma)] - q = 0_n. \tag{2}$$

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

$$\delta^2 = \min_{\gamma \in \Gamma} \min_{m \in \mathcal{M}} E[(y(\gamma) - m)^2] \tag{3}$$

as a misspecification measure of  $y(\gamma)$ . In this paper, we refer to  $\delta^2$  as the squared 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 derivatives on the test assets, for example. To deal with this issue, HJ provide a second model misspecification measure:

$$\delta_+^2 = \min_{\gamma \in \Gamma} \min_{m \in \mathcal{M}^+} E[(y(\gamma) - m)^2]. \tag{4}$$

We refer to  $\delta_+^2$  as the squared constrained HJ-distance. Since  $\mathcal{M}^+$  is a subset of  $\mathcal{M}$ ,  $\delta_+$  cannot be smaller than  $\delta$ .

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

$$\delta^2 = \min_{\gamma \in \Gamma} \max_{\lambda \in \mathfrak{R}^n} E[y(\gamma)^2 - (y(\gamma) - \lambda' \tilde{r})^2] - 2\lambda' q, \quad (5)$$

$$\delta_+^2 = \min_{\gamma \in \Gamma} \max_{\lambda \in \mathfrak{R}^n} E[y(\gamma)^2 - [(y(\gamma) - \lambda' \tilde{r})^+]^2] - 2\lambda' q, \quad (6)$$

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

For a given SDF  $y(\gamma)$ , HJ show that the unconstrained distance is equal to the maximum pricing error of portfolios of  $\tilde{r}$  with unit second moment. Let  $h \in L^2$ , where  $E[h^2] = 1$ . Using Jensen's inequality, we have

$$(E[y(\gamma)h] - E[mh])^2 = (E[(y(\gamma) - m)h])^2 \leq E[(y(\gamma) - m)^2]E[h^2], \quad (7)$$

where  $m \in \mathcal{M}^+$ . It follows that

$$\max_{h \in L^2, E[h^2]=1} |E[y(\gamma)h] - E[mh]| = E[(y(\gamma) - m)^2]^{\frac{1}{2}}. \quad (8)$$

It is easy to verify that the maximum pricing error occurs for  $h = (y(\gamma) - m)/E[(y(\gamma) - m)^2]^{\frac{1}{2}}$ . Unlike the case of the unconstrained HJ-distance, the maximum pricing error expression for the constrained HJ-distance depends on the choice of  $m$  in  $\mathcal{M}^+$ . HJ suggest that we can eliminate this dependence by computing the minimax bound

$$\min_{m \in \mathcal{M}^+} \max_{h \in L^2, E[h^2]=1} |E[y(\gamma)h] - E[mh]| = \min_{m \in \mathcal{M}^+} E[(y(\gamma) - m)^2]^{\frac{1}{2}} = E[(y(\gamma) - m_y^+)^2]^{\frac{1}{2}} = \delta_+, \quad (9)$$

where  $m_y^+$  is the nonnegative admissible SDF that is closest to  $y(\gamma)$ .

It is important to emphasize that  $\delta_+$  generally represents only 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 that the economy uses to price all  $h \in L^2$ . Then, using (8) and (9), we have

$$\max_{h \in L^2, E[h^2]=1} |E[y(\gamma)h] - E[m^*h]| = E[(y(\gamma) - m^*)^2]^{\frac{1}{2}} \geq E[(y(\gamma) - m_y^+)^2]^{\frac{1}{2}} = \delta_+, \quad (10)$$

and the maximum pricing error is generally larger than  $\delta_+$ . The only case in which we can interpret  $\delta_+$  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  $\delta_+$  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 (10), the maximum pricing error of a model is equal to the distance between  $y$  and  $m^*$ . However, a model in  $\mathcal{M}^+$  (that is,  $\delta_+ = 0$ ) can actually be further away from  $m^*$  than a model that is not in  $\mathcal{M}^+$  (that is,  $\delta_+ > 0$ ). This makes it problematic to rank models by  $\delta_+$  because a model with a larger  $\delta_+$  can actually be closer to  $m^*$  and have a smaller maximum pricing error for payoffs in  $L^2$ . In particular, a model with a smaller  $\delta_+$  is not necessarily a better model for pricing derivatives.<sup>1</sup> This is an important point that has been largely ignored in the literature. We further elaborate on this in the next section, where we derive an analytical expression for the constrained HJ-distance in linear models under the assumption that the returns and the SDF are multivariate elliptically distributed.

### 3. Analytical Solution for the Constrained Hansen-Jagannathan Distance in Linear Models

While it is desirable to consider SDFs that are strictly positive, most SDFs used in empirical work are linear. For this reason, in the subsequent analysis we focus on linear SDFs of the form

$$y(\gamma) = \gamma_0 + \gamma_1' f, \quad (11)$$

where  $f$  is a vector of  $K$  systematic factors, and  $\gamma = [\gamma_0, \gamma_1']'$  is a  $k$ -vector of SDF parameters ( $k = K + 1$ ).<sup>2</sup> Define  $\mu_r = E[r]$ ,  $\mu_f = E[f]$ ,  $V_{rf} = \text{Cov}[r, f']$ ,  $V_{rr} = \text{Var}[r]$ ,  $V_{ff} = \text{Var}[f]$ , and  $U = E[\tilde{r}\tilde{r}']$ . Throughout the paper, we assume that the matrices  $V_{rr}$ ,  $V_{ff}$ ,  $U$ , and  $V_{rf}'V_{rr}^{-1}V_{rf}$  are nonsingular. Also, let  $a = \mu_r'V_{rr}^{-1}\mu_r$  be the squared Sharpe ratio of the tangency portfolio of the  $N$  risky assets,  $a_1 = \mu_r'V_{rr}^{-1}V_{rf}(V_{rf}'V_{rr}^{-1}V_{rf})^{-1}V_{rf}'V_{rr}^{-1}\mu_r$  be the squared Sharpe ratio of the tangency portfolio constructed from the  $K$  factor mimicking portfolios, and  $V_{ff \cdot r} = V_{ff} - V_{rf}'V_{rr}^{-1}V_{rf}$  be the covariance matrix of the residuals from projecting the factors onto the returns. Then, the parameter vector  $\gamma = [\gamma_0, \gamma_1']'$  that minimizes the unconstrained HJ-distance is given by

$$\gamma_1 = -\frac{1}{R_0}(V_{rf}'V_{rr}^{-1}V_{rf})^{-1}(V_{rf}'V_{rr}^{-1}\mu_r), \quad \gamma_0 = \frac{1}{R_0} - \gamma_1'\mu_f. \quad (12)$$

---

<sup>1</sup>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.

<sup>2</sup>Results for the case in which the candidate SDF does not depend on parameters are available from the authors upon request.



As a result, the linear SDF that minimizes the unconstrained HJ-distance, the squared unconstrained HJ-distance, and the vector of Lagrange multipliers are given by

$$y = \frac{1}{R_0} + \gamma'_1(f - \mu_f), \quad (13)$$

$$\delta^2 = \frac{a - a_1}{R_0^2}, \quad (14)$$

and

$$\lambda = \begin{bmatrix} -\delta^2 \\ V_{rr}^{-1} \left( V_{rf} \gamma_1 + \frac{\mu_r}{R_0} \right) \end{bmatrix}, \quad (15)$$

respectively.

Unlike the case of the unconstrained HJ-distance, obtaining an analytical solution for  $\delta_+$  requires making a joint distributional assumption on the SDF and the returns on the test assets. More specifically, we assume that the conditional joint distribution of the SDF and the returns is multivariate elliptical (which includes normal, Student  $t$ , Cauchy, Laplace, symmetric stable, and logistic distributions, among others, as special cases). Since an elliptically distributed SDF takes on negative values by construction, it cannot belong to  $\mathcal{M}^+$  and our theoretical analysis of the constrained HJ-distance is conducted under the assumption that the asset-pricing model is misspecified. This assumption seems reasonable in light of the widespread evidence of model misspecification documented in empirical work and our results in Section 5. It is important to emphasize that while we assume that the conditional joint distribution of the SDF and the returns is multivariate elliptical, we do not make any assumption on their time series properties. The mean and the covariance matrix of the SDF and the returns can be time-varying, and many popular time series models like multivariate GARCH with multivariate normal or Student  $t$  errors are allowed under our framework. While there is strong empirical support for the multivariate elliptical distribution assumption for the data and the frequency that we use in our empirical application, we do not argue that this assumption is always a good approximation of the true conditional distribution of the SDF and the returns. Whether ellipticity provides a reasonable approximation or not depends on the problem at hand. Appendix B provides definitions and notation for elliptically distributed random variables, including some new results on the moments of truncated elliptically distributed random variables. These results are of independent interest given the importance of elliptical distributions for portfolio choice theory, equity and option-pricing theory (see Owen and Rabinovitch (1983), Zhou (1993), and Hamada and Valdez (2008), among others).

For the constrained HJ-distance case, the vector of Lagrange multipliers in (6), for a fixed  $\gamma$ , is given by

$$\tilde{\lambda} = \operatorname{argmin}_{\lambda} E[(y(\gamma) - \lambda' \tilde{r})^+{}^2] + 2\lambda' q, \quad (16)$$

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

$$E[\tilde{r}(y(\gamma) - \tilde{\lambda}' \tilde{r})^+] = q. \quad (17)$$

In principle, we can solve the  $n$  nonlinear equations  $E[\tilde{r}(y(\gamma) - \tilde{\lambda}' \tilde{r})^+] = q$  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 solution to the first-order condition in (17) depends on the joint distribution of  $y$  and  $r$  (or equivalently,  $f$  and  $r$ ). Assuming that conditional on  $\mathcal{F}$ ,  $y$  and  $r$  have a multivariate elliptical distribution with finite variance, a linear combination of  $y$  and  $\tilde{r}$ , say  $v$ , also has an elliptical distribution in the same class. We assume that the characteristic function of  $v$  can be expressed as  $\varphi(t) = \exp(it\mu_v)\psi(t^2 s_v^2/2)$  for some function  $\psi(\cdot)$ , where  $\mu_v$  is the mean of  $v$  and  $c^2 s_v^2$  is the variance of  $v$ , with  $c = \sqrt{-\psi'(0)}$ .

We denote the density and cumulative distribution functions of  $\dot{v} = (v - \mu_v)/s_v$  as  $\dot{f}$  and  $\dot{F}$ , respectively. For a given choice of  $\dot{f}$ , we introduce another elliptically distributed random variable,  $w$ , with density function

$$\check{f}(w) = \int_w^\infty cs \dot{f}(cs) ds, \quad (18)$$

and denote the cumulative distribution function of  $w$  by  $\check{F}$ . A more complete discussion of the class of elliptical distributions and an explicit derivation of  $\check{f}$  is provided in Appendix B.

The following proposition presents explicit expressions for the SDF parameters, Lagrange multipliers, and the squared constrained HJ-distance under the multivariate elliptical distribution assumption.

**Proposition 1.** *Let  $\eta$  be the unique solution to*

$$g(u) = \left( a + \alpha' \left[ \frac{1}{\check{F}(u)} I_K - V_{ff}^{-\frac{1}{2}} V_{ff.r} V_{ff}^{-\frac{1}{2}} \right]^{-2} \alpha \right)^{-\frac{1}{2}}, \quad (19)$$

where  $\alpha = V_{ff}^{-1}V_{ff.r}^{\frac{1}{2}}V_{rf}'V_{rr}^{-1}\mu_r$ ,  $0 < \check{F}(u) < 1$  is defined above, and

$$g(u) = \frac{u\check{F}(cu) + \check{f}(u)}{\check{F}(u)}. \quad (20)$$

Then, the vector of SDF parameters that minimizes the constrained HJ-distance is given by  $\tilde{\gamma} = [\tilde{\gamma}_0, \tilde{\gamma}'_1]'$ , where

$$\tilde{\gamma}_1 = -\frac{1}{R_0}[V_{ff} - \check{F}(\eta)V_{ff.r}]^{-1}(V_{rf}'V_{rr}^{-1}\mu_r), \quad \tilde{\gamma}_0 = \frac{1}{R_0} - \tilde{\gamma}'_1\mu_f, \quad (21)$$

and the SDF that minimizes the constrained HJ-distance is

$$\tilde{y} = \frac{1}{R_0} + \tilde{\gamma}'_1(f - \mu_f). \quad (22)$$

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

$$\delta_+^2 = \frac{a + \frac{\eta}{g(\eta)}}{R_0^2\check{F}(\eta)} - \frac{1 + \tilde{a}_1}{R_0^2}, \quad (23)$$

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

$$\tilde{\lambda} = \begin{bmatrix} -\delta_+^2 \\ V_{rr}^{-1}\left(V_{rf}\tilde{\gamma}_1 + \frac{\mu_r}{R_0\check{F}(\eta)}\right) \end{bmatrix}. \quad (24)$$

**Proof.** See Appendix A.

Besides the simplicity of the expressions for  $\tilde{\gamma}$ ,  $\tilde{\lambda}$ , and  $\delta_+^2$ , a few interesting observations emerge from Proposition 1. First, when the factors are spanned by the returns (that is,  $V_{ff.r} = 0_{K \times K}$ ), we have

$$g(\eta) = \frac{1}{\sqrt{a}} \quad (25)$$

and

$$\delta_+^2 - \delta^2 = \frac{a + \sqrt{a}\eta}{R_0^2\check{F}(\eta)} - \frac{1 + a}{R_0^2} = \left[ \frac{\sqrt{a}(\sqrt{a} + \eta)}{R_0^2\check{F}(\eta)} - \frac{1}{R_0^2} \right] - \frac{a}{R_0^2} = \sigma_c^2 - \sigma_0^2, \quad (26)$$

where  $\sigma_c^2 = \min_{m \in \mathcal{M}^+} \text{Var}[m]$  and  $\sigma_0^2 = \min_{m \in \mathcal{M}} \text{Var}[m]$  are the constrained and unconstrained bounds of Hansen and Jagannathan (1991).<sup>3</sup> It should be noted that both HJ-bounds (and their difference) only depend on the choice of the test assets and are model independent. Therefore, when

<sup>3</sup>The equality in (26) is a general result and is not specific to the ellipticity assumption.

the SDF is spanned,  $\delta_+^2 - \delta^2$  is also model independent. This result implies that for two spanned SDFs, the difference between their unconstrained HJ-distances is the same as the difference between their constrained HJ-distances, and the constrained and unconstrained HJ-distances will provide the same ranking of competing asset-pricing models.

Second, when one or more factors are useless, that is, they are uncorrelated with the returns on the test assets, the SDF parameters that minimize the unconstrained HJ-distance are not identified since the matrix  $V_{rf}'V_{rr}^{-1}V_{rf}$  is not of full rank and cannot be inverted. In contrast, 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,  $\delta_+^2$  is equal to the constrained HJ-bound  $\sigma_c^2$ .

Finally, by subtracting  $\delta^2 = (a - a_1)/R_0^2$  from both sides of (23) and rearranging, the difference between the squared constrained and unconstrained HJ-distances is given by

$$\Delta \equiv \delta_+^2 - \delta^2 = \frac{a_1\omega}{R_0^2} + \frac{a + \frac{\eta}{g(\eta)}}{R_0^2\check{F}(\eta)} - \frac{1+a}{R_0^2}, \quad (27)$$

where  $\omega = 1 - \frac{\tilde{a}_1}{a_1}$  can be interpreted as a measure of how unspanned the SDF is. Note that for a spanned SDF ( $V_{ff.r} = 0_{K \times K}$ ), we have  $\omega = 0$  and the expression for  $\Delta$  collapses to the one in equation (26). The following lemma provides the comparative statics for the difference between the constrained and unconstrained HJ-distances with respect to its two main determinants: the squared Sharpe ratio,  $a$ , and  $\omega$ .

**Lemma 1.** *The partial derivative of  $\Delta$  with respect to  $a$  is given by*

$$\frac{\partial \Delta}{\partial a} = \frac{\check{F}(-\eta)}{R_0^2\check{F}(\eta)} > 0. \quad (28)$$

*In addition, if  $V_{ff.r}$  is a positive definite matrix, we have  $\omega > 0$  and*

$$\frac{\partial \Delta}{\partial \omega} = \frac{a_1}{R_0^2} > 0. \quad (29)$$

**Proof.** See Appendix A.

Lemma 1 shows that  $\Delta$  is an increasing function of  $\omega$ , which suggests that  $\sigma_c^2 - \sigma_0^2$  is a lower bound for  $\Delta$ . While a larger  $\omega$  does not affect the SDF's ability to price the test assets (and, hence, the unconstrained HJ-distance), it can affect the SDF's probability of taking on negative values and

hence  $\Delta$  is an increasing function of  $\omega$ . 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  $\omega$  is very different across the two models.

Furthermore, Lemma 1 establishes that  $\Delta$  is an increasing function of the squared Sharpe ratio of the tangency portfolio of the test assets. This result requires some explanation. Consider, for simplicity, the case in which the SDF is spanned. When this happens,  $(\delta_+^2 - \delta^2) \rightarrow (\sigma_c^2 - \sigma_0^2)$  — the difference between the constrained and unconstrained HJ-bounds. Lemma 6 of Kan and Robotti (2015) shows that, under normality,  $(\sigma_c^2 - \sigma_0^2) \rightarrow 0$  when  $a \rightarrow 0$ ,  $(\sigma_c^2 - \sigma_0^2) \rightarrow \infty$  when  $a \rightarrow \infty$ , 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-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 nonnegativity constraint on it has almost no effect.<sup>4</sup>

With the analytical solutions of the linear SDF parameters for the unconstrained and constrained HJ-distances, we can now answer two important questions. The first question is whether the linear SDF  $\tilde{y}$  in (22) results in a lower probability of taking on negative values than the linear SDF  $y$  in (13). 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

$$\tilde{\delta}^2 = \tilde{e}'U^{-1}\tilde{e}, \quad (31)$$

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  $\delta^2$  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  $\delta^2$ , we gain useful insights of the potential

---

<sup>4</sup>The sign of the partial derivative of  $\Delta$  with respect to its third determinant  $R_0$

$$\frac{\partial \Delta}{\partial R_0} = \frac{2}{R_0^3} \left[ 1 - a_1\omega - \frac{a\check{F}(-\eta)}{\check{F}(\eta)} - \frac{\eta}{\check{F}(\eta)g(\eta)} \right] \quad (30)$$

cannot be established unambiguously. It turns out, however, that  $\Delta$  is rather insensitive to changes around the historically observed values of the gross risk-free rate.

cost of using  $\tilde{y}$  instead of  $y$  to price the test assets. The following lemma provides answers to these two questions.

**Lemma 2.** *Let  $y$  and  $\tilde{y}$  be the linear SDFs that minimize the unconstrained and constrained HJ-distances, respectively. Then, we have*

$$P[y < 0] - P[\tilde{y} < 0] = \dot{F} \left( -\frac{c}{R_0 \sqrt{\gamma'_1 V_{ff} \gamma_1}} \right) - \dot{F} \left( -\frac{c}{R_0 \sqrt{\tilde{\gamma}'_1 V_{ff} \tilde{\gamma}_1}} \right) > 0, \quad (32)$$

where  $\dot{F}$  and  $c$  are defined before Proposition 1. In addition, we have

$$\delta^2 \leq \tilde{\delta}^2 \leq \delta_+^2. \quad (33)$$

**Proof.** See Appendix A.

As shown in the proof of Lemma 2, (33) is a general result and does not rely on the linearity of the model and the ellipticity assumption. However, the result that  $P[y < 0] > P[\tilde{y} < 0]$  is specific to linear models under the ellipticity assumption.

Lemma 2 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.

To gain a better understanding of the results in the previous proposition and lemmas, it proves useful to consider the one-factor case. In this setting, we can show that

$$\omega = \frac{\check{F}(-\eta)(1 - \rho^2)}{1 - \check{F}(\eta)(1 - \rho^2)},$$

where  $\eta$  is the unique solution to

$$g(u) = \left[ a + \frac{a_1 \rho^2 (1 - \rho^2) \check{F}(u)^2}{[1 - \check{F}(u)(1 - \rho^2)]^2} \right]^{-\frac{1}{2}}, \quad (34)$$

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. Since there is a monotonically decreasing relation between  $\omega$  and  $\rho^2$  (a

proof of this claim is available upon request), we can express  $P[y < 0] - P[\tilde{y} < 0]$ ,  $\tilde{\delta}^2 - \delta^2$ , and  $\Delta$  in (27) as functions of  $\rho^2$  for different values of the Sharpe ratio of the tangency portfolio of the risky assets. Specifically, the difference in probabilities of taking on negative values for  $y$  and  $\tilde{y}$  is given by

$$P[y < 0] - P[\tilde{y} < 0] = \dot{F}\left(-\frac{c|\rho|}{\sqrt{a_1}}\right) - \dot{F}\left(-\frac{c[1 - \check{F}(\eta)(1 - \rho^2)]}{\sqrt{a_1}|\rho|}\right). \quad (35)$$

In addition, we have

$$\tilde{\delta}^2 - \delta^2 = \frac{a_1}{R_0^2} \left[ \frac{\check{F}(-\eta)(1 - \rho^2)}{1 - \check{F}(\eta)(1 - \rho^2)} \right]^2 \quad (36)$$

and

$$\Delta \equiv \delta_+^2 - \delta^2 = \frac{a_1(1 - \rho^2)\check{F}(-\eta)}{R_0^2[1 - \check{F}(\eta)(1 - \rho^2)]} + \frac{a + \eta/g(\eta)}{R_0^2\dot{F}(\eta)} - \frac{1 + a}{R_0^2}.$$

Assuming that the factor and the excess returns on the test assets are multivariate  $t$ -distributed with six degrees of freedom, Figure 1 plots  $P[y < 0] - P[\tilde{y} < 0]$  as a function of  $\rho^2$  for three different values of the Sharpe ratio of the tangency portfolio ( $\sqrt{a} = 0.25, 0.5, \text{ and } 0.75$ ).<sup>5</sup> In each case, we assume  $a_1 = a/2$ , so that the model explains half of the cross-sectional variation in expected returns.

Figure 1 about here

When  $\rho^2 \rightarrow 0$  ( $y$  is not defined when  $\rho^2 = 0$ ),  $P[y < 0] - P[\tilde{y} < 0] \rightarrow 0.5$ . The reason is that when the unspanned component of the factor increases,  $y$  becomes more volatile (because  $\gamma_1$  does not depend on the unspanned component of the factor) and behaves more like a useless factor. As a result,  $P[y < 0] \rightarrow 0.5$ . In contrast, as  $\rho^2 \rightarrow 0$ ,  $\tilde{y}$  converges to  $1/R_0$  and has almost zero probability of taking on negative values.

When  $\rho^2 \rightarrow 1$ , the SDF behaves more like 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 1 shows that the Sharpe ratio is important in determining  $P[y < 0] - P[\tilde{y} < 0]$ . For a given value of  $\rho^2$ , 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  $y$  assigns relatively

---

<sup>5</sup>Although a Sharpe ratio of 0.75 may seem high, this is in line with the sample Sharpe ratio (0.72) of the tangency portfolio of the 25 Fama-French size and book-to-market portfolios used in the empirical application in Section 5.

little weight to the factor mimicking portfolio and hence  $P[y < 0]$  is small, leaving not much room for  $\tilde{y}$  to improve.

Assuming that  $R_0 = 1.005$  (the plot is not sensitive to other reasonable values of the gross risk-free rate) and using the same parameter values and distributional assumptions as in Figure 1, Figure 2 plots  $\tilde{\delta}^2 - \delta^2$  and  $\Delta \equiv \delta_+^2 - \delta^2$  as functions of  $\rho^2$ .

Figure 2 about here

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  $\tilde{\delta}^2 - \delta^2 \rightarrow 0$ . However, when  $\rho^2 \rightarrow 0$ , 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  $\tilde{\delta}^2 - \delta^2 \rightarrow a_1/R_0^2$ . Similar to Figure 1, 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  $\tilde{y}$ .

In addition, Figure 2 reveals that  $\Delta$  is a decreasing function of  $\rho^2$ . As for  $\tilde{\delta}^2 - \delta^2$ ,  $\Delta$  is also heavily influenced by the Sharpe ratio of the tangency portfolio. When  $\sqrt{a} = 0.25$ , the difference between  $\delta_+^2$  and  $\delta^2$  is indistinguishable from zero. For  $\sqrt{a} = 0.5$ , the difference between  $\delta_+^2$  and  $\delta^2$  is still quite small, even for a relatively small  $\rho^2$ . 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.

In summary, we should expect  $y$  and  $\tilde{y}$  to behave differently if  $a$  is large and  $\rho^2$  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  $\rho^2$  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 even for pricing derivatives since it prices every asset by discounting the future asset payoffs with the gross 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 asymptotic and finite-sample properties. The results in this section no longer rely on the multivariate ellipticity assumption and unlike Section 3, we cover both linear and non-linear models. In addition, 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  $t$  ( $t = 1, \dots, T$ ) by  $x_t$  and the corresponding costs of these  $n$  assets at the end of period  $t - 1$  by  $q_{t-1}$ .<sup>6</sup> Besides being applicable to gross and excess returns, this setup can accommodate payoffs of trading strategies that are based on time-varying information.

Let  $y_t(\gamma)$  and  $m_t$  be the realizations of a candidate SDF and an admissible SDF at time  $t$ , respectively. Throughout this section, it is assumed that  $\{y_t(\gamma)\}_{t=1}^T$  and  $\{x_t\}_{t=1}^T$  are stationary and ergodic processes with finite fourth moments. We adopt the following definitions of correctly specified and misspecified models. An asset-pricing model is correctly specified if there exists a vector of parameters  $\gamma \in \Gamma$  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  $\delta_+ > 0$ .

The sample squared constrained HJ-distance can be obtained as

$$\begin{aligned} \hat{\delta}_+^2 &= \min_{\gamma \in \Gamma} \min_{\{m_t, t=1, \dots, T\}} \frac{1}{T} \sum_{t=1}^T (y_t(\gamma) - m_t)^2, \\ \text{s.t. } &\frac{1}{T} \sum_{t=1}^T m_t x_t = \bar{q}, \\ &m_t \geq 0, \quad t = 1, \dots, T, \end{aligned} \tag{37}$$

where  $\bar{q} = \frac{1}{T} \sum_{t=1}^T q_{t-1}$ . Chernozhukov, Hong, and Tamer (2007) and Chernozhukov, Kocatulum, and Menzel (2015) develop an approach to conducting inference on parameter sets, characterized by a smooth inequality constraint, which can be adapted to the setup in (37). Instead, we base our statistical analysis on the dual (conjugate) problem that gives rise to an unconstrained extremum

---

<sup>6</sup>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, for example, Kan and Robotti (2008)).

estimator.

Let  $\theta = [\gamma', \lambda']'$  and denote by  $\theta^* = [\gamma^{*'}, \lambda^{*'}]'$  the pseudo-true values that solve the dual problem:

$$\theta^* = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \mathbb{R}^n} E[\phi_t(\gamma, \lambda)], \quad (38)$$

where

$$\phi_t(\gamma, \lambda) = y_t(\gamma)^2 - [m_t(\theta)^+]^2 - 2\lambda'q_{t-1} \quad (39)$$

and

$$m_t(\theta) = y_t(\gamma) - \lambda'x_t. \quad (40)$$

When  $y_t(\gamma^*) \in \mathcal{M}^+$ , we have  $\lambda^* = 0_n$  and we refer to  $\gamma^*$  as the true value. The estimator of  $\theta^*$  can then be obtained as

$$\hat{\theta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\lambda} \end{bmatrix} = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \mathbb{R}^n} \frac{1}{T} \sum_{t=1}^T \phi_t(\gamma, \lambda), \quad (41)$$

and the sample squared constrained HJ-distance can be expressed as

$$\hat{\delta}_+^2 = \frac{1}{T} \sum_{t=1}^T \phi_t(\hat{\gamma}, \hat{\lambda}). \quad (42)$$

Within this framework, there are two possible ways to proceed with the estimation of the unknown parameters. The first approach is parametric and hinges upon the linearity of the SDF and the multivariate elliptical assumption introduced in Section 3. Under this method, the estimates of  $\gamma$  and  $\lambda$  are obtained by substituting, for a particular member of the class of multivariate elliptical distributions, the corresponding sample quantities in the analytical expressions presented in Proposition 1. In what follows, we pursue a semi-parametric approach, which is directly based on the saddle-point optimization problem described above and does not require any distributional assumptions on the data or linearity of  $y(\gamma)$ .

## 4.1 Asymptotic Results

Let

$$S_+ = \sum_{j=-\infty}^{\infty} E [(x_t m_t(\theta^*)^+ - q_{t-1})(x_{t+j} m_{t+j}(\theta^*)^+ - q_{t+j-1})'], \quad (43)$$

$$G_+ = E \left[ (y_t(\gamma^*) - m_t(\theta^*)^+) \frac{\partial^2 y_t(\gamma^*)}{\partial \gamma \partial \gamma'} + \frac{\partial y_t(\gamma^*)}{\partial \gamma} \frac{\partial y_t(\gamma^*)}{\partial \gamma'} I_{\{m_t(\theta^*) \leq 0\}} \right], \quad (44)$$

$$D_+ = E \left[ x_t \frac{\partial y_t(\gamma^*)}{\partial \gamma'} I_{\{m_t(\theta^*) > 0\}} \right], \quad (45)$$

and

$$U_+ = E \left[ x_t x_t' I_{\{m_t(\theta^*) > 0\}} \right], \quad (46)$$

where  $I_{\{\cdot\}}$  denotes the indicator function. 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  $S_+ = S \equiv \sum_{j=-\infty}^{\infty} E[(x_t y_t(\gamma^*) - q_{t-1})(x_{t+j} y_{t+j}(\gamma^*) - q_{t+j-1})']$ . In addition, let  $Z = (Z_1, \dots, Z_s)'$  be a vector of  $s$  independent standard normal random variables, and let  $\xi = (\xi_1, \dots, \xi_s)'$  be a vector of  $s$  real numbers. Then,  $F_s(\xi) = \sum_{i=1}^s \xi_i Z_i^2$  denotes a random variable which is distributed as a weighted sum of chi-squares with parameters  $(s, \xi)$ . Finally, let  $\overset{A}{\sim}$  stand for ‘‘asymptotically distributed as.’’

The regularity conditions for establishing the subsequent limiting results are stated in Gospodinov, Kan, and Robotti (2013). More specifically, in addition to Assumptions A, B, and C in Gospodinov, Kan, and Robotti (2013), we assume that  $P[y_t(\gamma) = \lambda' x_t] = 0$  almost surely in a small neighborhood of  $\theta^*$ . The asymptotic distributions of the sample squared constrained HJ-distance, SDF parameter estimates, and sample Lagrange multipliers are provided in the following lemma.

**Lemma 3.** *Under the above assumptions,*

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

$$T \hat{\delta}_+^2 \overset{A}{\sim} F_{n-k}(\xi), \quad (47)$$

where the  $\xi_i$ 's are the eigenvalues of

$$A = P' U^{-\frac{1}{2}} S U^{-\frac{1}{2}} P, \quad (48)$$

with  $P$  being an  $n \times (n - k)$  orthonormal matrix whose columns are orthogonal to  $U^{-\frac{1}{2}} D$ .

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

$$\sqrt{T}(\hat{\delta}_+^2 - \delta_+^2) \overset{A}{\sim} N(0, v_{\hat{\delta}_+}), \quad (49)$$

where  $v_{\hat{\delta}_+} = \sum_{j=-\infty}^{\infty} E[(\phi_t(\theta^*) - \delta_+^2)(\phi_{t+j}(\theta^*) - \delta_+^2)]$ , and

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, \Sigma_{\hat{\theta}}), \quad (50)$$

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

$$l_{1t} = (G_+ + D'_+ U_+^{-1} D_+)^{-1} \left[ (y_t(\gamma^*) - m_t(\theta^*)^+) \frac{\partial y_t(\gamma^*)}{\partial \gamma} + D'_+ U_+^{-1} [x_t m_t(\theta^*)^+ - q_{t-1}] \right] \quad (51)$$

$$l_{2t} = U_+^{-1} [D_+ l_{1t} - x_t m_t(\theta^*)^+ + q_{t-1}]. \quad (52)$$

**Proof.** See Appendix A.

Interestingly, Lemma 3(a) shows the equivalence of the asymptotic distributions of the sample constrained and unconstrained HJ-distance tests (for the asymptotic distribution of the unconstrained HJ-distance test, see, for example, Parker and Julliard (2005)). Hence, the existing specification test developed for the sample unconstrained HJ-distance is also applicable to the sample constrained HJ-distance. The intuition behind this result is that, under the null of a correctly specified model, the constraints are not binding and the two tests should be asymptotically equivalent.<sup>7</sup> A similar result holds for the limiting distributions of the parameter estimates under correctly specified models which coincide with those for the unconstrained HJ-distance (for their explicit expressions, see Gospodinov, Kan, and Robotti (2013)). Lemma 3(b) provides the asymptotic distribution of the estimates of the SDF parameters and the Lagrange multipliers for the case of misspecified models. Finally, to conduct inference, the matrix in (48), the variance  $v_{\hat{\delta}_+}$ , and the variance matrix  $\Sigma_{\hat{\theta}}$  should be replaced with consistent estimators. Simulation results and extensions to pairwise model comparison tests are available from the authors upon request.

## 4.2 Finite-Sample Results

In this section, we discuss some issues related to the finite-sample properties of the constrained HJ-distance. In particular, 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

---

<sup>7</sup>In the case of a linear SDF,  $\delta_+^2 = -\lambda^* E[q_{t-1}]$  and  $\lambda^* = 0_n$  implies  $\delta_+^2 = 0$ . Therefore, testing for correct model specification can be performed using the asymptotically pivotal Lagrange multiplier (LM) test proposed by Gospodinov, Kan, and Robotti (2013). In addition to its excellent finite-sample properties, the LM statistic allows us to test directly the hypothesis of zero pricing errors, which is of primary interest to financial economists. It should be emphasized that using the LM test as a specification test is valid only in the case of linear models since, in general,  $\lambda^* = 0_n$  does not always imply that  $\delta_+^2 = 0$ .

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 (2015) prove that this occurs with nonzero probability in any finite sample. What these 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  $\{m_t, t = 1, \dots, T\}$  that satisfies the constraints in (37) 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  $\{x_t, t = 1, \dots, T\}$  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  $m_t \geq 0, t = 1, \dots, T$ , such that

$$\frac{1}{T} \sum_{t=1}^T x_t m_t = \bar{q}. \quad (53)$$

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

$$w' x_t \geq 0, \quad t = 1, \dots, T. \quad (54)$$

The equivalence of these two statements suggests that if one finds a portfolio  $w$  such that it has negative average cost (that is,  $w' \bar{q} < 0$ ) but nonnegative payoffs in every period of the sample, then  $\hat{\mathcal{M}}^+$  is an empty set. When such an event (that is, existence of an in-sample arbitrage portfolio) occurs, we show that  $\hat{\delta}_+ = \infty$  for all models. Suppose  $w' \bar{q} = -\varepsilon$ , where  $\varepsilon > 0$ . Consider  $\lambda = \vartheta w$ , where  $\vartheta$  is a nonnegative scalar. Since  $\vartheta w$  for  $\vartheta \geq 0$  is only a subset of  $\mathfrak{R}^n$  and using the minimax

inequality, we have

$$\begin{aligned}
\hat{\delta}_+^2 &= \min_{\gamma \in \Gamma} \max_{\lambda \in \mathbb{R}^n} \frac{1}{T} \sum_{t=1}^T [y_t(\gamma)^2 - [(y_t(\gamma) - \lambda'x_t)^+]^2] - 2\lambda'\bar{q} \\
&\geq \max_{\lambda \in \mathbb{R}^n} \min_{\gamma \in \Gamma} \frac{1}{T} \sum_{t=1}^T [y_t(\gamma)^2 - [(y_t(\gamma) - \lambda'x_t)^+]^2] - 2\lambda'\bar{q} \\
&\geq \max_{\vartheta \geq 0} \min_{\gamma \in \Gamma} \frac{1}{T} \sum_{t=1}^T [y_t(\gamma)^2 - [(y_t(\gamma) - \vartheta w'x_t)^+]^2] + 2\varepsilon\vartheta.
\end{aligned} \tag{55}$$

Note that since  $\vartheta w'x_t \geq 0$ , we have  $(y_t(\gamma) - \vartheta w'x_t)^+ \leq |y_t(\gamma)|$  and  $y_t(\gamma)^2 - [(y_t(\gamma) - \vartheta w'x_t)^+]^2 \geq 0$  for all  $\gamma \in \Gamma$ . Hence, it follows that

$$\hat{\delta}_+^2 \geq \max_{\vartheta \geq 0} 2\varepsilon\vartheta = \infty. \tag{56}$$

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  $(x_t, q_{t-1})$ ,  $t = 1, \dots, T$ . 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, that is,  $q = [1, 0'_N]'$ . When the excess returns are i.i.d. multivariate elliptically 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 (the proof of this result is available upon request).

Figure 3 about here

In Figure 3, we assume that the excess returns on the  $N$  risky assets are i.i.d. multivariate  $t$ -distributed with six degrees of freedom, and 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 3,  $P[\hat{\delta}_+ = \infty]$  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 Sharpe ratio, it is easier to observe an in-sample arbitrage opportunity.

The result  $P[\hat{\delta}_+ = \infty] > 0$  has some implications for the use of the constrained HJ-distance for model evaluation and multiple model comparison. The positive aspect of this result is that a value of infinity for the constrained HJ-distance suggests that the model admits arbitrage and should be rejected. In this respect, the constrained HJ-distance has the added advantage of detecting arbitrage. However, it should be emphasized that the probability for the sample constrained HJ-distance to take the value of infinity does not depend on the particular model but only on the properties of the returns on the test assets. Therefore, in a multiple model comparison, the sample constrained HJ-distances of all models will be infinity and we cannot use this distance to rank models. Furthermore,  $P[\hat{\delta}_+ = \infty] > 0$  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 (42) to obtain  $\hat{\delta}_+$ , we should first find out if the primal problem in (37) 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 (42). We can then use standard linear programming routines (`linprog` in MATLAB, for example) to determine whether such a problem is feasible.<sup>8</sup>

## 5. Empirical Application

In this section, we illustrate the implications of our theoretical results using actual data. 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.

### 5.1 Data and Asset-Pricing Models

The return data are from Kenneth French's website and consist of the quarterly value-weighted returns on the 25 Fama-French size and book-to-market ranked portfolios plus the one-month T-

---

<sup>8</sup>We implement the numerical procedure for estimating  $\delta_+$  and other associated parameters in a set of Matlab programs. The Matlab programs for this paper are available upon request.

bill. The data span the period from the third quarter of 1953 until the third quarter of 2007, the onset of the most recent financial crisis (217 quarterly observations). The beginning date of our sample period is dictated by bond yield data availability.

We analyze seven asset-pricing models that have yielded encouraging results in cross-sectional asset pricing, starting with a conditional version of the consumption CAPM (CC-CAY) due to Lettau and Ludvigson (2001). The SDF specification for this model is

$$y_t^{CC-CAY}(\gamma) = \gamma_0 + \gamma_{cay}cay_{t-1} + \gamma_{c_{nd}}c_{nd,t} + \gamma_{c_{nd} \cdot cay}c_{nd,t} \cdot cay_{t-1}, \quad (57)$$

where  $c_{nd}$  is the growth rate in real per capita nondurable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis and  $cay$  is the consumption-aggregate wealth ratio of Lettau and Ludvigson (2001) from Martin Lettau's website. The SDF specification in Equation (57) is obtained by scaling the constant term and the  $c_{nd}$  factor by a constant and  $cay$ .

The second model is another conditional version of the consumption CAPM (CC-MY) proposed by Lustig and Van Nieuwerburgh (LV, 2005) with the housing collateral ratio ( $my$ , constructed as in LV),  $c_{nd}$ , and the interaction term  $c_{nd} \cdot my$  as risk factors. The SDF of this model is given by

$$y_t^{CC-MY}(\gamma) = \gamma_0 + \gamma_{my}my_{t-1} + \gamma_{c_{nd}}c_{nd,t} + \gamma_{c_{nd} \cdot my}c_{nd,t} \cdot my_{t-1}. \quad (58)$$

The SDF specification in Equation (58) is obtained by scaling the constant term and the  $c_{nd}$  factor by a constant and  $my$ .

Next, we consider a conditional version of the CAPM (C-SW) due to Santos and Veronesi (SV, 2006) with an SDF specification

$$y_t^{C-SW}(\gamma) = \gamma_0 + \gamma_{r_{mkt}}r_{mkt,t} + \gamma_{r_{mkt} \cdot s^w}r_{mkt,t} \cdot s_{t-1}^w, \quad (59)$$

where  $r_{mkt}$  is the excess return (in excess of the one-month T-bill rate) on the value-weighted stock market index (NYSE-AMEX-NASDAQ) from Kenneth French's website and  $s^w$  is the labor income-consumption ratio (constructed as in SV).

Our fourth specification is the investment growth model (IGM) of Li, Vassalou, and Xing (2006). Its SDF is given by

$$y_t^{IGM}(\gamma) = \gamma_0 + \gamma_{i_{hh}}i_{hh,t} + \gamma_{i_{corp}}i_{corp,t} + \gamma_{i_{ncorp}}i_{ncorp,t} + \gamma_{i_{fcorp}}i_{fcorp,t} + \gamma_{i_{fm}}i_{fm,t}, \quad (60)$$



where  $i_{hh}$ ,  $i_{corp}$ ,  $i_{ncorp}$ ,  $i_{fcorp}$ , and  $i_{fm}$  represent the log (gross fixed) investment growth rates for households, non-financial corporations, non-corporate sector, financial corporations, and farm sector, respectively. All the data for these five series are from the Flow of Funds of the Board of Governors of the Federal Reserve System.

We also consider an empirical implementation of Merton's (1973) intertemporal extension of the CAPM (ICAPM) based on Campbell (1996), who argues that innovations in state variables that forecast future investment opportunities should serve as factors. The five-factor specification proposed by Petkova (2006) is

$$y_t^{ICAPM}(\gamma) = \gamma_0 + \gamma_{r_{mkt}} r_{mkt,t} + \gamma_{term} term_t + \gamma_{def} def_t + \gamma_{div} div_t + \gamma_{rf} rf_t, \quad (61)$$

where  $term$  is the difference between the yields of ten-year and one-year government bonds (from the Board of Governors of the Federal Reserve System),  $def$  is the difference between the yields of long-term corporate Baa bonds (from the Board of Governors of the Federal Reserve System) and long-term government bonds (from Ibbotson Associates),  $div$  is the dividend yield on the Center for Research in Security Prices (CRSP) value-weighted stock market portfolio, and  $rf$  is the one-month T-bill yield (from CRSP, Fama Risk Free Rates). The actual factors for  $term$ ,  $def$ ,  $div$ , and  $rf$  are their innovations from a VAR(4) system of seven state variables that also includes the market, size, and value factors of Fama and French (1993).

The sixth model (D-CCAPM) is due to Yogo (2006) and highlights the cyclical role of durable consumption in asset pricing. The specification is

$$y_t^{D-CCAPM}(\gamma) = \gamma_0 + \gamma_{r_{mkt}} r_{mkt,t} + \gamma_{c_{nd}} c_{nd,t} + \gamma_{c_d} c_{d,t}, \quad (62)$$

where  $c_d$  is the growth rate in real per capita durable consumption (seasonally adjusted at annual rates) from the Bureau of Economic Analysis.

Finally, we consider the three-factor model (FF3) of Fama and French (1993). This model extends the CAPM by including the two empirically motivated factors  $r_{smb}$  and  $r_{hml}$ , where  $r_{smb}$  is the return difference between portfolios of stocks with small and large market capitalizations, and  $r_{hml}$  is the return difference between portfolios of stocks with high and low book-to-market ratios ("value" and "growth" stocks, respectively) from Kenneth French's website. The SDF specification is

$$y_t^{FF3}(\gamma) = \gamma_0 + \gamma_{r_{mkt}} r_{mkt,t} + \gamma_{r_{smb}} r_{smb,t} + \gamma_{r_{hml}} r_{hml,t}. \quad (63)$$

## 5.2 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.<sup>9</sup> 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. In addition to the standard deviations of the estimated SDF, 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 (31), is a measure of the maximum pricing error on the test assets.

Table 1 about here

In the following, we would like to emphasize and interpret several important findings that naturally emerge from the predictions of our theoretical analysis in Section 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 CC-MY, IGM, and ICAPM). 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, 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 CC-MY, IGM, and ICAPM), 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 CC-MY, IGM, and ICAPM is also reflected in the standard deviations of their SDFs, which significantly drop from 0.920, 1.409, and 1.060 in Panel A to 0.394, 0.491, and 0.622 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.

---

<sup>9</sup>For the unconstrained HJ-distance and related tests, we refer the readers to Kan and Robotti (2009).

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 4.3% since this model has a very high  $R^2$  of 0.975. The pattern of the differences in the parameter estimates reported in Table 2 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.<sup>10</sup>

Table 2 about here

For example, consider the estimated SDF parameters on the interaction variable  $c_{nd} \cdot my$  in CC-MY. The coefficient estimate goes from 404.75 to 27.49 when using the constrained HJ-distance measure. In contrast, when the factors are traded, as is the case in FF3, there is no noticeable difference in the magnitude of the parameter estimates across metrics. As a result, the different effects of imposing the no-arbitrage constraint across models are driven by the underlying structure of the problem and characteristics of the factors (traded versus non-traded).

Third, 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  $\hat{\delta}_+$  is not always smaller than the standard error of  $\hat{\delta}$ . In fact, we find that for four 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. In unreported empirical experiments (available from the authors upon request), we performed pair-

---

<sup>10</sup>Given the strong evidence of model misspecification documented in Table 1, in Table 2 we report misspecification-robust standard errors of the SDF parameter estimates based on (50) and (51). Interestingly, Table 1 shows that IGM and ICAPM are not rejected at conventional significance levels when using the unconstrained HJ-distance, while they are strongly rejected when using the constrained HJ-distance. To understand why IGM and ICAPM are not rejected by the test based on the unconstrained HJ-distance, it is important to realize that these models are not properly identified since they contain unspanned factors that exhibit very low correlations with the returns on the test assets (for example, using the reduced rank test of Kleibergen and Paap (2006), we cannot reject the null hypothesis of lack of identification for these models at any conventional level). For such models, Gospodinov, Kan, and Robotti (2014) show that the unconstrained HJ-distance test has low power in detecting misspecification, and that this test is inconsistent under the alternative of model misspecification in the extreme case in which one or more factors are useless. In contrast, the test based on the constrained HJ-distance seems to have higher power in rejecting misspecified and unidentified asset-pricing models.

wise model comparison tests using the two distance measures. For the unconstrained HJ-distance, out of 21 pairwise model comparison tests, no model outperforms its competitors at the 5% confidence level. For the constrained HJ-distance, FF3 outperforms C-SW and D-CCAPM.<sup>11</sup> The main reason for our finding is that in this dataset the differences in sample constrained HJ-distances across models are slightly larger than the corresponding differences in sample unconstrained HJ-distances. However, trying to determine whether the constrained HJ-distance is more powerful than the unconstrained HJ-distance in distinguishing models is not a very meaningful exercise. The reason is that the SDF parameters for a given model are in general different under the two metrics. As a result, we are actually not comparing the same sets of SDFs under the unconstrained and constrained HJ-distances. Ultimately, the choice of constrained versus unconstrained HJ-distance should depend on which metric is considered to be more reasonable for model comparison.

Finally, 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 here has a relatively high sample Sharpe ratio (0.72). To understand whether this is an important reason for the difference in results between the two HJ-distances, we consider another set of test assets with a smaller sample Sharpe ratio. Specifically, we use quarterly gross returns on the one-month T-bill and 10 size and 12 industry portfolios from Kenneth French’s website.<sup>12</sup> The sample Sharpe ratio of the tangency portfolio of this new set of risky assets is 0.55. We then perform the same analysis as in Table 1, leaving the sample period and the models unchanged. The results of this exercise are reported in Table 3.

Table 3 about here

Consistent with our theoretical results, we find that a decrease in the Sharpe ratio of the tangency portfolio of the risky assets causes the unconstrained and constrained HJ-distances to behave similarly across different models. In addition, the difference between the constrained and

---

<sup>11</sup>It is interesting to note that, while FF3 dominates C-SW and D-CCAPM statistically, ICAPM has the lowest sample constrained HJ-distance and yet does not outperform any other model in formal pairwise model comparison tests. Precision appears to play a role in this. The standard error of ICAPM’s  $\hat{\delta}_+$  is the highest of all the models. The importance of taking information about sampling variation into account is evident.

<sup>12</sup>Analyzing the performance of these different models on an alternative set of test assets should be considered as purely illustrative of our methodological findings, and not as a critique of the use of the 25 Fama-French size and book-to-market portfolios in empirical work.

unconstrained HJ-distances of each model is now substantially smaller. As an example, consider IGM. For this model, the difference between  $\hat{\delta}_+$  and  $\hat{\delta}$  is only 4.1%, a much smaller number than the 23.9% difference reported in Table 1. When using formal pairwise model comparison tests as before, we find that only D-CCAPM outperforms CC-CAY at the 5% confidence level for the unconstrained as well as the constrained HJ-distance measures.

## 6. Conclusion

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 general, the constrained HJ-distance is only a measure of the 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 an ellipticity assumption on the conditional joint distribution of 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, which is applicable to linear as well to nonlinear models, 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. Finally, we provide a limiting theory for estimation and testing of SDFs using the constrained HJ-distance measure.

## Appendix A

**Proof of Proposition 1.** 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

$$\delta_+^2 = E[\tilde{y}^2] - E[m_{\tilde{y}}^{+2}] - 2\tilde{\lambda}'q \quad (\text{A.1})$$

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

$$C\tilde{\gamma} - E[\tilde{f}m_{\tilde{y}}^+] = 0_{K+1}, \quad (\text{A.2})$$

$$E[\tilde{r}m_{\tilde{y}}^+] = q. \quad (\text{A.3})$$

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

$$E[\tilde{f}m_{\tilde{y}}^+] = \check{F}(\eta)(C\tilde{\gamma} - D'\tilde{\lambda} + \sigma_m[g(\eta) - \eta]E[\tilde{f}]), \quad (\text{A.4})$$

$$E[\tilde{r}m_{\tilde{y}}^+] = \check{F}(\eta)(D\tilde{\gamma} - U\tilde{\lambda} + \sigma_m[g(\eta) - \eta]E[\tilde{r}]), \quad (\text{A.5})$$

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

$$\begin{bmatrix} \check{F}(-\eta)C & \check{F}(\eta)D' \\ \check{F}(\eta)D & -\check{F}(\eta)U \end{bmatrix} \begin{bmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \check{F}(\eta)\sigma_m[g(\eta) - \eta]E[\tilde{f}] \\ q - \check{F}(\eta)\sigma_m[g(\eta) - \eta]E[\tilde{r}] \end{bmatrix}. \quad (\text{A.6})$$

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

$$\begin{bmatrix} \check{F}(-\eta)C & \check{F}(\eta)D' \\ \check{F}(\eta)D & -\check{F}(\eta)U \end{bmatrix}^{-1} = \begin{bmatrix} H & HD'U^{-1} \\ U^{-1}DH & -\frac{1}{\check{F}(\eta)}U^{-1} + U^{-1}DHD'U^{-1} \end{bmatrix} \quad (\text{A.7})$$

and

$$U^{-1} = \begin{bmatrix} R_0^2 & R_0\mu_r' \\ R_0\mu_r & V_{rr} + \mu_r\mu_r' \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1+a}{R_0^2} & -\frac{\mu_r'V_{rr}^{-1}}{R_0} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} & V_{rr}^{-1} \end{bmatrix}. \quad (\text{A.8})$$

Using (A.8), we can easily verify that  $U^{-1}E[\tilde{r}] = q/R_0$  and hence

$$D'U^{-1}E[\tilde{r}] = \frac{1}{R_0}D'q = \frac{1}{R_0}E[R_0\tilde{f}] = E[\tilde{f}]. \quad (\text{A.9})$$

Using this identity, we can then show that

$$\tilde{\gamma} = HD'U^{-1}q. \quad (\text{A.10})$$

From the partitioned matrix inverse formula and after some algebra, we obtain

$$H = \begin{bmatrix} 1 + \mu'_f Q \mu_f & -\mu'_f Q \\ -Q \mu_f & Q \end{bmatrix}, \quad (\text{A.11})$$

where  $Q = [V_{ff} - \check{F}(\eta)V_{ff.r}]^{-1}$ . Using this expression and (A.8), we can rewrite (A.10) as

$$\tilde{\gamma} = \begin{bmatrix} 1 + \mu'_f Q \mu_f & -\mu'_f Q \\ -Q \mu_f & Q \end{bmatrix} \begin{bmatrix} R_0 & \mu'_r \\ \mu_f R_0 & V'_{rf} + \mu_f \mu'_r \end{bmatrix} \begin{bmatrix} \frac{(1+a)}{R_0^2} \\ -\frac{V_{rr}^{-1} \mu_r}{R_0} \end{bmatrix}. \quad (\text{A.12})$$

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

$$\tilde{\gamma}_1 = -\frac{1}{R_0} Q V'_{rf} V_{rr}^{-1} \mu_r, \quad \tilde{\gamma}_0 = \frac{1}{R_0} - \tilde{\gamma}'_1 \mu_f. \quad (\text{A.13})$$

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

When the candidate SDF  $y$  does not depend on parameters, it can be shown that

$$\tilde{\lambda} = \begin{bmatrix} \frac{\mu_y - V'_{ry} V_{rr}^{-1} \mu_r}{R_0} - \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \check{F}(\eta)} \\ V_{rr}^{-1} \left( V_{ry} + \frac{\mu_r}{R_0 \check{F}(\eta)} \right) \end{bmatrix}, \quad (\text{A.14})$$

where  $\mu_y = E[y]$  and  $V_{ry} = \text{Cov}[r, y]$ . Since the above expression for  $\tilde{\lambda}$  also works for  $\tilde{y}$ , we can use  $\mu_{\tilde{y}} = E[\tilde{y}] = 1/R_0$  and  $V'_{r\tilde{y}} V_{rr}^{-1} \mu_r = \text{Cov}[r, \tilde{y}]' V_{rr}^{-1} \mu_r = \tilde{\gamma}'_1 V'_{rf} V_{rr}^{-1} \mu_r = -\tilde{a}_1/R_0$  to obtain

$$\tilde{\lambda} = \begin{bmatrix} \frac{1 + \tilde{a}_1}{R_0^2} - \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \check{F}(\eta)} \\ V_{rr}^{-1} \left( V_{rf} \tilde{\gamma}_1 + \frac{\mu_r}{R_0 \check{F}(\eta)} \right) \end{bmatrix}. \quad (\text{A.15})$$

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

$$m_{\tilde{y}} = \tilde{\gamma}'_1 \tilde{f} - \tilde{\lambda}'_1 \tilde{r} = \tilde{\gamma}'_1 \epsilon - \frac{\mu'_r V_{rr}^{-1} (r - \mu_r)}{R_0 \check{F}(\eta)} + \frac{\eta}{R_0 \check{F}(\eta) g(\eta)}. \quad (\text{A.16})$$

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

$$\sigma_m^2 = \tilde{\gamma}'_1 V_{ff.r} \tilde{\gamma}_1 + \frac{a}{R_0^2 \check{F}(\eta)^2}. \quad (\text{A.17})$$

Since  $m_{\tilde{y}}^+$  prices the risk-free asset correctly, it follows that

$$E[m_{\tilde{y}}^+] = \check{F}(\eta) \sigma_m g(\eta) = \frac{1}{R_0}. \quad (\text{A.18})$$



Then, plugging the expression of  $\sigma_m^2$  from (A.17) into (A.18), we obtain

$$g(\eta) = \left[ a + \tilde{\gamma}'_1 V_{ff \cdot r} \tilde{\gamma}_1 R_0^2 \check{F}(\eta)^2 \right]^{-\frac{1}{2}}. \quad (\text{A.19})$$

Using the expression for  $\tilde{\gamma}_1$  in (A.13) and rearranging terms, we can see that  $\eta$  satisfies the following equation:

$$g(u) = \left( a + \alpha' \left[ \frac{1}{\check{F}(u)} I_K - V_{ff}^{-\frac{1}{2}} V_{ff \cdot r} V_{ff}^{-\frac{1}{2}} \right]^{-2} \alpha \right)^{-\frac{1}{2}}. \quad (\text{A.20})$$

For establishing the uniqueness of the solution to equation (A.20), we need to show that (i)  $g(u) > 0$ , (ii)  $g(\infty) = \infty$ , (iii)  $g(-\infty) = 0$ , and (iv)  $g'(u) > 0$ . Condition (i) follows from equation (B.23) in Lemma B.1 of Appendix B. Condition (ii) follows from the definition of  $g$ . For condition (iii), it is convenient to write

$$g(u) = \frac{u\dot{F}(cu) + \check{f}(u)}{\check{F}(u)} = \frac{g_1(u)}{g_2(u)}. \quad (\text{A.21})$$

Then,

$$g'_1(u) = \dot{F}(cu), \quad (\text{A.22})$$

$$g'_2(u) = \check{f}(u), \quad (\text{A.23})$$

$$g''_1(u) = c\dot{f}(cu), \quad (\text{A.24})$$

$$g''_2(u) = -cu\dot{f}(cu). \quad (\text{A.25})$$

Using L'Hôpital's rule twice, we have

$$\lim_{u \rightarrow -\infty} g(u) = \lim_{u \rightarrow -\infty} \frac{g'_1(u)}{g'_2(u)} = \lim_{u \rightarrow -\infty} \frac{c\dot{f}(cu)}{-cu\dot{f}(cu)} = \lim_{u \rightarrow -\infty} -\frac{1}{u} = 0. \quad (\text{A.26})$$

For (iv), taking the derivative of  $g(u)$ , we have

$$g'(u) = \frac{\dot{F}(cu)}{\check{F}(u)^2} \left[ \check{F}(u) - u\check{f}(u) - \frac{\check{f}(u)^2}{\dot{F}(cu)} \right] > 0, \quad (\text{A.27})$$

where the inequality is obtained by using Lemma B.2 in Appendix B and the fact that  $\text{Var}[x] > 0$ . Since the left hand side of (A.20) is positive and increasing in  $u$  and the right hand side of (A.20) is positive and decreasing in  $u$  (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), (A.20) has a unique solution.

Using Lemma B.1 in Appendix B, it is straightforward to obtain

$$E[\tilde{y}m_{\tilde{y}}^+] = \text{Cov}[\tilde{y}, m_{\tilde{y}}]\check{F}(\eta) + \frac{1}{R_0^2} = \left[ \tilde{\gamma}'_1 V_{ff.r} \tilde{\gamma}_1 + \frac{\tilde{a}_1}{R_0^2 \check{F}(\eta)} \right] \check{F}(\eta) + \frac{1}{R_0^2}, \quad (\text{A.28})$$

$$E[m_{\tilde{y}}^{+2}] = E[m_{\tilde{y}}m_{\tilde{y}}^+] = \sigma_m^2 \check{F}(\eta) + \frac{\mu_m}{R_0}. \quad (\text{A.29})$$

The squared constrained HJ-distance is then given by

$$\begin{aligned} \delta_+^2 &= E[(\tilde{y} - m_{\tilde{y}}^+)^2] \\ &= E[\tilde{y}^2] - 2E[\tilde{y}m_{\tilde{y}}^+] + E[m_{\tilde{y}}^{+2}] \\ &= \frac{1}{R_0^2} + \tilde{\gamma}'_1 V_{ff} \tilde{\gamma}_1 - 2 \left[ \tilde{\gamma}'_1 V_{ff.r} \tilde{\gamma}_1 + \frac{\tilde{a}_1}{R_0^2 \check{F}(\eta)} \right] \check{F}(\eta) - \frac{2}{R_0^2} + \sigma_m^2 \check{F}(\eta) + \frac{\mu_m}{R_0} \\ &= \tilde{\gamma}'_1 V_{ff} \tilde{\gamma}_1 - 2\tilde{\gamma}'_1 V_{ff.r} \tilde{\gamma}_1 \check{F}(\eta) - \frac{1 + 2\tilde{a}_1}{R_0^2} + \left[ \tilde{\gamma}'_1 V_{ff.r} \tilde{\gamma}_1 + \frac{a}{R_0^2 \check{F}(\eta)^2} \right] \check{F}(\eta) + \frac{\eta \sigma_m}{R_0} \\ &= \tilde{\gamma}'_1 [V_{ff} - \check{F}(\eta) V_{ff.r}] \tilde{\gamma}_1 - \frac{1 + 2\tilde{a}_1}{R_0^2} + \frac{a}{R_0^2 \check{F}(\eta)} + \frac{\eta}{R_0^2 \check{F}(\eta) g(\eta)} \\ &= \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \check{F}(\eta)} - \frac{1 + \tilde{a}_1}{R_0^2}, \end{aligned} \quad (\text{A.30})$$

where the second last equality is obtained by using  $\sigma_m = 1/[R_0 \check{F}(\eta) g(\eta)]$  from (A.18). Finally, we can easily see that the first element of  $\tilde{\lambda}$  in (A.15) is equal to  $-\delta_+^2$ . This completes the proof.

**Proof of Lemma 1.** Let  $b = \left( \alpha' \left[ I_K - V_{ff}^{-\frac{1}{2}} \check{F}(\eta) V_{ff.r} V_{ff}^{-\frac{1}{2}} \right]^{-2} \alpha \right)$ . We first show that  $\partial\Delta/\partial\eta = 0$ . More specifically, the derivative of  $\tilde{a}_1$  with respect to  $\eta$  is

$$\frac{\partial \tilde{a}_1}{\partial \eta} = \check{f}(\eta) \mu_r' V_{rr}^{-1} V_{rf} [V_{ff} - \check{F}(\eta) V_{ff.r}]^{-1} V_{ff.r} [V_{ff} - \check{F}(\eta) V_{ff.r}]^{-1} V_{rf}' V_{rr}^{-1} \mu_r = b \check{f}(\eta) \quad (\text{A.31})$$

as in the proof of Proposition 1. Furthermore,

$$\begin{aligned} \frac{\partial \Delta}{\partial \eta} &= -\frac{b \check{f}(\eta)}{R_0^2} - \frac{a \check{f}(\eta)}{R_0^2 \check{F}(\eta)^2} + \frac{R_0^2 \check{F}(\eta) g(\eta) - \eta [R_0^2 \check{F}(\eta) g'(\eta) + R_0^2 \check{f}(\eta) g(\eta)]}{R_0^4 \check{F}(\eta)^2 g(\eta)^2} \\ &= -\frac{b \check{f}(\eta)}{R_0^2} - \frac{a \check{f}(\eta)}{R_0^2 \check{F}(\eta)^2} + \frac{1}{R_0^2 g(\eta) \check{F}(\eta)} - \frac{\eta g'(\eta)}{R_0^2 g(\eta)^2 \check{F}(\eta)} - \frac{\eta \check{f}(\eta)}{R_0^2 g(\eta) \check{F}(\eta)^2} \\ &= -\frac{\check{f}(\eta)}{R_0^2 \check{F}(\eta)^2} (a + b \check{F}(\eta)^2) + \frac{1}{R_0^2 g(\eta) \check{F}(\eta)} - \frac{\eta g'(\eta)}{R_0^2 g(\eta)^2 \check{F}(\eta)} - \frac{\eta \check{f}(\eta)}{R_0^2 g(\eta) \check{F}(\eta)^2} \\ &= -\frac{\check{f}(\eta)}{R_0^2 \check{F}(\eta)^2 g(\eta)^2} + \frac{g(\eta) \check{F}(\eta) - \eta \dot{F}(c\eta)}{R_0^2 \check{F}(\eta)^2 g(\eta)^2} \\ &= -\frac{\check{f}(\eta)}{R_0^2 \check{F}(\eta)^2 g(\eta)^2} + \frac{\check{f}(\eta)}{R_0^2 \check{F}(\eta)^2 g(\eta)^2} \\ &= 0. \end{aligned} \quad (\text{A.32})$$

The fourth equality follows because

$$a + b\check{F}(\eta)^2 = \left( a + \alpha' \left[ \frac{1}{\check{F}(\eta)} I_K - V_{ff}^{-\frac{1}{2}} V_{ff.r} V_{ff}^{-\frac{1}{2}} \right]^{-2} \alpha \right) = \frac{1}{g(\eta)^2} \quad (\text{A.33})$$

and

$$g'(\eta) = \frac{\dot{F}(c\eta) - g(\eta)\check{f}(\eta)}{\check{F}(\eta)}, \quad (\text{A.34})$$

which can be easily verified by using  $g(\eta) = \frac{\eta\dot{F}(c\eta) + \check{f}(\eta)}{\check{F}(\eta)}$  (see Lemma B.1 in Appendix B) and (A.27).

This suggests that the partial derivatives of  $\Delta$  with respect to the remaining parameters can treat  $\eta$  as a constant. Thus, it follows that

$$\frac{\partial \Delta}{\partial a} = \frac{1}{R_0^2 \check{F}(\eta)} - \frac{1}{R_0^2} = \frac{\check{F}(-\eta)}{R_0^2 \check{F}(\eta)} > 0. \quad (\text{A.35})$$

Next, we show that  $\omega > 0$  if  $V_{ff.r}$  is a positive definite matrix. After substituting for  $V_{ff.r} = V_{ff} - V_{rf}' V_{rr}^{-1} V_{rf}$  and rearranging, we have

$$\begin{aligned} \omega &= \frac{\mu_r' V_{rr}^{-1} V_{rf} \left( (V_{rf}' V_{rr}^{-1} V_{rf})^{-1} - [V_{ff} - \check{F}(\eta) V_{ff.r}]^{-1} \right) V_{rf}' V_{rr}^{-1} \mu_r}{\mu_r' V_{rr}^{-1} V_{rf} (V_{rf}' V_{rr}^{-1} V_{rf})^{-1} V_{rf}' V_{rr}^{-1} \mu_r} \\ &= \frac{\mu_r' V_{rr}^{-1} V_{rf} \left( (V_{rf}' V_{rr}^{-1} V_{rf})^{-1} - [\check{F}(\eta) V_{rf}' V_{rr}^{-1} V_{rf} + (1 - \check{F}(\eta)) V_{ff}]^{-1} \right) V_{rf}' V_{rr}^{-1} \mu_r}{\mu_r' V_{rr}^{-1} V_{rf} (V_{rf}' V_{rr}^{-1} V_{rf})^{-1} V_{rf}' V_{rr}^{-1} \mu_r}. \end{aligned} \quad (\text{A.36})$$

Note that the matrix

$$[\check{F}(\eta) V_{rf}' V_{rr}^{-1} V_{rf} + (1 - \check{F}(\eta)) V_{ff}] - V_{rf}' V_{rr}^{-1} V_{rf} \quad (\text{A.37})$$

is positive definite since  $V_{ff} - V_{rf}' V_{rr}^{-1} V_{rf}$  is positive definite and  $0 < \check{F}(\eta) < 1$ . Then, it follows that

$$(V_{rf}' V_{rr}^{-1} V_{rf})^{-1} - [\check{F}(\eta) V_{rf}' V_{rr}^{-1} V_{rf} + (1 - \check{F}(\eta)) V_{ff}]^{-1} \quad (\text{A.38})$$

is also positive definite.<sup>13</sup> Therefore,  $0 < \omega < 1$  and

$$\frac{\partial \Delta}{\partial \omega} = \frac{a_1}{R_0^2} > 0. \quad (\text{A.39})$$

This completes the proof.

---

<sup>13</sup>From the properties of positive definite matrices, if  $A - B$  is positive definite, then  $(B^{-1} - A^{-1})$  is also positive definite.

**Proof of Lemma 2.** The probability for  $y$  to take on negative values is equal to

$$P[y < 0] = \dot{F}\left(-\frac{c\mu_y}{\sigma_y}\right) = \dot{F}\left(-\frac{c}{R_0\sigma_y}\right) = \dot{F}\left(-\frac{c}{R_0\sqrt{\gamma'_1 V_{ff}\gamma_1}}\right), \quad (\text{A.40})$$

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

$$P[\tilde{y} < 0] = \dot{F}\left(-\frac{c\mu_{\tilde{y}}}{\sigma_{\tilde{y}}}\right) = \dot{F}\left(-\frac{c}{R_0\sigma_{\tilde{y}}}\right) = \dot{F}\left(-\frac{c}{R_0\sqrt{\tilde{\gamma}'_1 V_{ff}\tilde{\gamma}_1}}\right), \quad (\text{A.41})$$

where  $\sigma_{\tilde{y}}$  is the standard deviation of the SDF  $\tilde{y}$ . The inequality holds because

$$\begin{aligned} R_0^2 \tilde{\gamma}'_1 V_{ff} \tilde{\gamma}_1 &= \mu'_r V_{rr}^{-1} V_{rf} (V_{ff} - \check{F}(\eta) V_{ff.r})^{-1} V_{ff} (V_{ff} - \check{F}(\eta) V_{ff.r})^{-1} V'_{rf} V_{rr}^{-1} \mu_r \\ &= \mu'_r V_{rr}^{-1} V_{rf} V_{ff}^{\frac{1}{2}} \left( I_K - \check{F}(\eta) V_{ff}^{-\frac{1}{2}} V_{ff.r} V_{ff}^{-\frac{1}{2}} \right)^{-2} V_{ff}^{\frac{1}{2}} V'_{rf} V_{rr}^{-1} \mu_r \\ &\leq \mu'_r V_{rr}^{-1} V_{rf} V_{ff}^{\frac{1}{2}} (I_K - V_{ff}^{-\frac{1}{2}} V_{ff.r} V_{ff}^{-\frac{1}{2}})^{-2} V_{ff}^{\frac{1}{2}} V'_{rf} V_{rr}^{-1} \mu_r \\ &= \mu'_r V_{rr}^{-1} V_{rf} (V'_{rf} V_{rr}^{-1} V_{rf})^{-1} V_{ff} (V'_{rf} V_{rr}^{-1} V_{rf})^{-1} V'_{rf} V_{rr}^{-1} \mu_r \\ &= R_0^2 \gamma'_1 V_{ff} \gamma_1. \end{aligned} \quad (\text{A.42})$$

For (33), the first inequality,  $\delta^2 \leq \tilde{\delta}^2$ , is obvious since  $\gamma$  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

$$\min_{m \in \mathcal{M}^+} (E[\tilde{y}h] - E[mh])^2 \leq \delta_+^2. \quad (\text{A.43})$$

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

$$\varpi^* = \frac{U^{-1}\tilde{e}}{(\tilde{e}'U^{-1}\tilde{e})^{\frac{1}{2}}}. \quad (\text{A.44})$$

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

$$\delta_+^2 \geq \inf_{m \in \mathcal{M}^+} (E[\tilde{y}h] - E[mh])^2 = (E[\tilde{y}h] - E[mh])^2 = (\varpi^{*'}(E[\tilde{y}\tilde{r}] - q))^2 = (\varpi^{*'}\tilde{e})^2 = \tilde{\delta}^2. \quad (\text{A.45})$$

This completes the proof.

**Proof of Lemma 3.** (a) Let

$$M \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 E[\phi_t(\theta^*)]}{\partial \theta \partial \theta'} = \begin{bmatrix} 2G_+ & 2D_+' \\ 2D_+ & -2U_+ \end{bmatrix}. \quad (\text{A.46})$$

From the definition of  $M$  in (A.46), we can use the partitioned matrix inverse formula to obtain

$$M^{-1} = \begin{bmatrix} 2G_+ & 2D_+' \\ 2D_+ & -2U_+ \end{bmatrix}^{-1} = \frac{1}{2} \begin{bmatrix} M & \tilde{M}D_+'U_+^{-1} \\ U_+^{-1}D_+\tilde{M} & -U_+^{-1} + U_+^{-1}D_+\tilde{M}D_+'U_+^{-1} \end{bmatrix}, \quad (\text{A.47})$$

where  $\tilde{M} = (G_+ + D_+'U_+^{-1}D_+)^{-1}$ . Following similar arguments as in Lemma A.1 in Gospodinov, Kan, and Robotti (2013), it can be shown that

$$\hat{\delta}_+^2 - \delta_+^2 = \frac{1}{T} \sum_{t=1}^T (\phi_t(\theta^*) - E[\phi_t(\theta^*)]) - \frac{1}{2} \bar{v}_T(\theta^*)' M^{-1} \bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right), \quad (\text{A.48})$$

where  $\bar{v}_T(\theta) = \frac{1}{T} \sum_{t=1}^T (\partial \phi_t(\theta) / \partial \theta - E[\partial \phi_t(\theta) / \partial \theta])$ . Then, under the null hypothesis  $H_0 : \delta_+ = 0$ , (A.48) becomes

$$\hat{\delta}_+^2 = -\frac{1}{2} \bar{v}_T(\theta^*)' M^{-1} \bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right) \quad (\text{A.49})$$

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

$$\begin{aligned} T\hat{\delta}_+^2 &= -\frac{1}{2} \sqrt{T} \bar{v}_T(\theta^*)' M^{-1} \sqrt{T} \bar{v}_T(\theta^*) + o_p(1) \\ &= \frac{1}{4} \sqrt{T} \bar{v}_{2,T}(\theta^*)' U^{-\frac{1}{2}} P P' U^{-\frac{1}{2}} \sqrt{T} \bar{v}_{2,T}(\theta^*) + o_p(1) \end{aligned} \quad (\text{A.50})$$

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

$$\sqrt{T} \bar{v}_{2,T}(\theta^*) \overset{A}{\rightsquigarrow} N(0_n, 4S) \quad (\text{A.51})$$

under the null  $H_0 : \delta_+ = 0$ . Then, we obtain

$$T\hat{\delta}_+^2 \overset{A}{\rightsquigarrow} z' S^{\frac{1}{2}} U^{-\frac{1}{2}} P P' U^{-\frac{1}{2}} S^{\frac{1}{2}} z, \quad (\text{A.52})$$

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

$$T\hat{\delta}_+^2 \overset{A}{\rightsquigarrow} F_{n-k}(\xi), \quad (\text{A.53})$$

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

(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 the stated assumptions and since  $E[\phi_t(\theta^*)] = \delta_+^2$ , we have

$$\sqrt{T}(\hat{\delta}_+^2 - \delta_+^2) = \frac{1}{\sqrt{T}} \sum_{t=1}^T (\phi_t(\theta^*) - E[\phi_t(\theta^*)]) + o_p(1) \stackrel{A}{\sim} N(0, v_{\hat{\delta}_+}). \quad (\text{A.54})$$

Deriving the asymptotic distribution of  $\sqrt{T}(\hat{\theta} - \theta^*)$  closely follows Gospodinov, Kan, and Robotti (2013) and is omitted to conserve space. This completes the proof.

## Appendix B

### B.1 Elliptical Distributions: Definitions and Notation

In this section, we introduce the definitions and notation for the class of multivariate elliptical distributions, following closely Landsman and Valdez (2003). We say that two random variables  $(u, v)$  have a bivariate elliptical distribution, written as  $\mathcal{E}(\mu, S, \psi)$ , if their characteristic function can be expressed as

$$\varphi(t) = \exp(it' \mu) \psi \left( \frac{t' S t}{2} \right) \quad (\text{B.1})$$

for some

$$\mu = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix}, \quad S = \begin{bmatrix} s_u^2 & s_{uv} \\ s_{vu} & s_v^2 \end{bmatrix}, \quad (\text{B.2})$$

and  $\psi(\cdot)$ , which is called the *characteristic generator*. When the mean of  $[u, v]'$  exists, we have  $E[u] = \mu_u$  and  $E[v] = \mu_v$ . When the variance of  $[u, v]'$  exists, we have  $\sigma_u^2 = \text{Var}[u] = c^2 s_u^2$ ,  $\sigma_v^2 = \text{Var}[v] = c^2 s_v^2$ , and  $\sigma_{uv} = \text{Cov}[u, v] = c^2 s_{uv}$ , where  $c = \sqrt{-\psi'(0)}$ . It is important to remember that  $S$  is not the covariance matrix of  $[u, v]'$  in general.

We assume that the density functions of  $u$  and  $v$  exist. The density function of  $v$  (the density function of  $u$  is similarly defined) is given by

$$f_v(v) = \frac{c_1}{s_v} h \left( \frac{(v - \mu_v)^2}{2s_v^2} \right), \quad (\text{B.3})$$

where  $h(\cdot)$  is a nonnegative function (called the *density generator*) and

$$c_1 = \frac{1}{\sqrt{2}} \left[ \int_0^\infty x^{-\frac{1}{2}} h(x) dx \right]^{-1} \quad (\text{B.4})$$

is a normalization constant.

We provide two examples of elliptical distributions: normal and Student  $t$ . For the normal distribution, we have

$$\psi(s) = e^{-s}, \quad (\text{B.5})$$

$$h(t) = e^{-t}, \quad (\text{B.6})$$

$$c_1 = \frac{1}{\sqrt{2\pi}}, \quad (\text{B.7})$$

$$f_v(v) = \frac{1}{\sqrt{2\pi} s_v} e^{-\frac{(v - \mu_v)^2}{2s_v^2}}. \quad (\text{B.8})$$

It follows that  $\psi'(s) = -e^{-s}$ ,  $c = \sqrt{-\psi'(0)} = 1$ , and  $\sigma_v^2 = s_v^2$ .

For the Student  $t$  distribution with  $\nu$  degrees of freedom, we have

$$\psi(s) = \frac{K_{\nu/2}(\sqrt{2\nu s}) \left(\frac{\nu s}{2}\right)^{\frac{\nu}{4}}}{2\Gamma\left(\frac{\nu}{2}\right)}, \quad (\text{B.9})$$

$$h(t) = \left(1 + \frac{2t}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (\text{B.10})$$

$$c_1 = \frac{1}{\text{B}\left(\frac{1}{2}, \frac{\nu}{2}\right) \sqrt{\nu}}, \quad (\text{B.11})$$

$$f_v(v) = \frac{1}{\text{B}\left(\frac{1}{2}, \frac{\nu}{2}\right) \sqrt{\nu} s_v} \left[1 + \frac{(v - \mu_v)^2}{\nu s_v^2}\right]^{-\frac{\nu+1}{2}}, \quad (\text{B.12})$$

where  $K_\nu(x)$  is the modified Bessel function of the second kind,  $\Gamma(a)$  is the gamma function, and  $\text{B}(a, b)$  is the beta function. It is straightforward to show that

$$\psi'(s) = -\frac{\nu K_{(\nu-2)/2}(\sqrt{2\nu s}) \left(\frac{\nu s}{2}\right)^{\frac{\nu-2}{4}}}{\Gamma\left(\frac{\nu}{2}\right)}, \quad (\text{B.13})$$

and  $c = \sqrt{-\psi'(0)} = [\nu/(\nu - 2)]^{\frac{1}{2}}$ . In addition, when  $\nu > 1$ , the mean of  $v$  exists and when  $\nu > 2$ , the variance of  $v$  exists and  $\sigma_v^2 = \nu s_v^2/(\nu - 2)$ .

For a given elliptical random variable  $v$  with parameters  $\mu_v$  and  $s_v^2$ , we define

$$\dot{v} = \frac{v - \mu_v}{s_v}. \quad (\text{B.14})$$

The random variable  $\dot{v}$  has a spherical distribution (that is, an elliptical distribution with parameters  $\mu_{\dot{v}} = 0$  and  $s_{\dot{v}} = 1$ ). We denote its density and cumulative distribution functions by  $\dot{f}(\dot{v})$  and  $\dot{F}(\dot{v})$ , respectively. Note that

$$\dot{f}(\dot{v}) = c_1 h\left(\frac{\dot{v}^2}{2}\right). \quad (\text{B.15})$$

By symmetry, we have  $\dot{f}(-\dot{v}) = \dot{f}(\dot{v})$  and  $1 - \dot{F}(-\dot{v}) = \dot{F}(\dot{v})$ . In addition, we have  $\sigma_{\dot{v}}^2 = c^2$  when the variance of  $\dot{v}$  exists.

For every spherical random variable  $\dot{v}$  with finite variance, Landsman and Valdez (2003) show that a random variable  $w$  with density function<sup>14</sup>

$$\check{f}(w) = \int_w^\infty c \dot{v} \dot{f}(c\dot{v}) d\dot{v} = \frac{1}{c} \int_{cw}^\infty s \dot{f}(s) ds \quad (\text{B.16})$$

<sup>14</sup>Instead of mapping  $\dot{v}$  into  $w$ , Landsman and Valdez (2003) define a slightly different mapping from  $\dot{v}$  into  $Z^* = cw$ .



is also a spherical random variable. The density function of  $w$  can alternatively be written as

$$\check{f}(w) = \frac{c_1}{c} H\left(\frac{c^2 w^2}{2}\right), \quad (\text{B.17})$$

where

$$H(x) = \int_x^\infty h(t) dt. \quad (\text{B.18})$$

From this expression, we can easily see that the density function of  $w$  depends only on  $w^2$ , which implies that  $w$  has a spherical distribution. The distribution of  $w$  is crucial for us to obtain the tail conditional expectation of  $\dot{v}$ .

For a given spherical random variable  $\dot{v}$ , the above definitions allow us to quickly obtain the density function of the associated spherical random variable  $w$ . For example, when  $\dot{v} \sim N(0, 1)$ , we have  $h(t) = e^{-t}$  and

$$H(x) = \int_x^\infty e^{-t} dt = e^{-x}. \quad (\text{B.19})$$

Therefore, using  $c = 1$  and  $c_1 = 1/\sqrt{2\pi}$ , we obtain

$$\check{f}(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \quad (\text{B.20})$$

and  $w \sim N(0, 1)$ .

When  $\dot{v} \sim t_\nu$  for  $\nu > 2$ , we use (B.10) to obtain

$$H(x) = \int_x^\infty \left(1 + \frac{2t}{\nu}\right)^{-\frac{\nu+1}{2}} dt = \frac{\nu}{\nu-1} \left(1 + \frac{2x}{\nu}\right)^{-\frac{\nu-1}{2}}. \quad (\text{B.21})$$

Then, using (B.11) and  $c = [\nu/(\nu-2)]^{\frac{1}{2}}$ , we obtain

$$\begin{aligned} \check{f}(w) &= \frac{1}{\sqrt{\nu} \text{B}\left(\frac{1}{2}, \frac{\nu}{2}\right)} \frac{\sqrt{\nu-2}}{\sqrt{\nu}} \frac{\nu}{\nu-1} \left(1 + \frac{w^2}{\nu-2}\right)^{-\frac{\nu-1}{2}} \\ &= \frac{1}{\sqrt{\nu-2} \text{B}\left(\frac{1}{2}, \frac{\nu-2}{2}\right)} \left(1 + \frac{w^2}{\nu-2}\right)^{-\frac{\nu-1}{2}}, \end{aligned} \quad (\text{B.22})$$

and  $w \sim t_{\nu-2}$ .

## B.2 Lemmas

**Lemma B.1.** *Suppose  $[u, v]^t$  is bivariate elliptically distributed with finite variance. Let  $\eta = \mu_v/\sigma_v$ , where  $\mu_v$  and  $\sigma_v$  are the mean and standard deviation of  $v$ , respectively. We have*

$$E[v^+] = \mu_v \dot{F}(c\eta) + \sigma_v \check{f}(\eta) = \check{F}(\eta) \sigma_v g(\eta), \quad (\text{B.23})$$

$$\begin{aligned} E[uv^+] &= \sigma_{uv} \check{F}(\eta) + \mu_u [\mu_v \dot{F}(c\eta) + \sigma_v \check{f}(\eta)] \\ &= \check{F}(\eta) (E[uv] + \mu_u \sigma_v [g(\eta) - \eta]), \end{aligned} \quad (\text{B.24})$$

where  $\dot{F}$  is the cumulative distribution function of  $\dot{v} = (v - \mu_v)/s_v$ ,  $c = \sigma_v$ ,  $\check{f}$  and  $\check{F}$  are the density and cumulative distribution functions of another spherical random variable  $w$  that is associated with  $\dot{v}$  as defined in (B.16), and

$$g(\eta) = \frac{\eta \dot{F}(c\eta) + \check{f}(\eta)}{\check{F}(\eta)}. \quad (\text{B.25})$$

**Proof of Lemma B.1.** For a given  $\check{f}$ , we define the functions

$$\dot{h}(x) = \int_{-x}^{\infty} \dot{v} \check{f}(\dot{v}) d\dot{v}, \quad (\text{B.26})$$

$$\dot{H}(x) = \int_{-x}^{\infty} \dot{v}^2 \check{f}(\dot{v}) d\dot{v}. \quad (\text{B.27})$$

We are interested in obtaining  $E[v^+]$ , which is given by

$$E[v^+] = \int_0^{\infty} v f_v(v) dv = \int_{-c\eta}^{\infty} (\mu_v + s_v \dot{v}) \dot{f}(\dot{v}) d\dot{v} = \mu_v \dot{F}(c\eta) + s_v \dot{h}(c\eta) = \mu_v \dot{F}(c\eta) + \sigma_v \check{f}(\eta), \quad (\text{B.28})$$

where the last equality follows from (B.16) and the fact that  $\dot{h}(c\eta) = c\check{f}(-\eta) = c\check{f}(\eta)$ .

In order to obtain  $E[uv^+]$ , we need to first derive  $E[v^{+2}]$ , which is given by

$$E[v^{+2}] = \int_0^{\infty} v^2 f_v(v) dv = \int_{-c\eta}^{\infty} (\mu_v + s_v \dot{v})^2 \dot{f}(\dot{v}) d\dot{v} = \mu_v^2 \dot{F}(c\eta) + 2\mu_v s_v \dot{h}(c\eta) + s_v^2 \dot{H}(c\eta). \quad (\text{B.29})$$

Since

$$\frac{d\dot{h}(\eta)}{d\eta} = -\eta \dot{f}(-\eta) = -\eta \dot{f}(\eta), \quad (\text{B.30})$$

we can use integration by parts to obtain

$$\begin{aligned}
\dot{H}(c\eta) &= \int_{-c\eta}^{\infty} \dot{v}^2 f(\dot{v}) d\dot{v} \\
&= -\dot{v} \dot{h}(\dot{v}) \Big|_{-c\eta}^{\infty} + \int_{-c\eta}^{\infty} \dot{h}(\dot{v}) d\dot{v} \\
&= d - c\eta \dot{h}(-c\eta) + c \int_{-c\eta}^{\infty} \check{f}\left(\frac{\dot{v}}{c}\right) d\dot{v} \\
&= d - c^2 \eta \check{f}(\eta) + c^2 \int_{-\eta}^{\infty} \check{f}(s) ds \\
&= d - c^2 \eta \check{f}(\eta) + c^2 \check{F}(\eta),
\end{aligned} \tag{B.31}$$

where  $d \equiv \lim_{\dot{v} \rightarrow \infty} -\dot{v} \dot{h}(\dot{v})$  and the fourth equality follows from the fact that  $\dot{h}(-c\eta) = \dot{h}(c\eta) = c \check{f}(\eta)$ . We now show that  $d = 0$  when  $c < \infty$ . Since  $w$  is a symmetric random variable,  $\check{F}(0) = 1/2$  and it follows that

$$\dot{H}(0) = d + c^2 \check{F}(0) = d + \frac{c^2}{2}. \tag{B.32}$$

However, we know that

$$\dot{H}(0) = \int_0^{\infty} \dot{v}^2 f(\dot{v}) d\dot{v} = \frac{c^2}{2}, \tag{B.33}$$

and hence  $d = 0$  when  $c$  is finite. Therefore, we have

$$\dot{H}(c\eta) = -c^2 \eta \check{f}(\eta) + c^2 \check{F}(\eta). \tag{B.34}$$

Using (B.29) and  $\dot{h}(c\eta) = c \check{f}(\eta)$ , we have

$$\begin{aligned}
E[v^{+2}] &= \mu_v^2 \dot{F}(c\eta) + 2\mu_v s_v \dot{h}(c\eta) - s_v^2 c^2 \eta \check{f}(\eta) + s_v^2 c^2 \check{F}(\eta) \\
&= \mu_v^2 \dot{F}(c\eta) + \mu_v \sigma_v \check{f}(\eta) + \sigma_v^2 \check{F}(\eta).
\end{aligned} \tag{B.35}$$

Under the bivariate elliptical assumption on  $u$  and  $v$ , we have

$$E[u|v] = \mu_u + \frac{\sigma_{uv}}{\sigma_v^2} (v - \mu_v). \tag{B.36}$$

It then follows that

$$\begin{aligned}
E[uv^+] &= E[E[u|v]v^+] \\
&= E\left[\left(\mu_u + \frac{\sigma_{uv}}{\sigma_v^2} (v - \mu_v)\right) v^+\right] \\
&= \left(\mu_u - \frac{\sigma_{uv}}{\sigma_v^2} \mu_v\right) E[v^+] + \frac{\sigma_{uv}}{\sigma_v^2} E[v^{+2}] \\
&= \left(\mu_u - \frac{\sigma_{uv}}{\sigma_v^2} \mu_v\right) [\mu_v \dot{F}(c\eta) + \sigma_v \check{f}(\eta)] + \frac{\sigma_{uv}}{\sigma_v^2} [\mu_v^2 \dot{F}(c\eta) + \sigma_v^2 \check{F}(\eta) + \mu_v \sigma_v \check{f}(\eta)] \\
&= \sigma_{uv} \check{F}(\eta) + \mu_u [\mu_v \dot{F}(c\eta) + \sigma_v \check{f}(\eta)].
\end{aligned} \tag{B.37}$$

This completes the proof.

The following lemma is used in proving the uniqueness of the solution to the equation in Proposition 1 for elliptically distributed random variables.

**Lemma B.2.** *Let  $\dot{f}$  and  $\dot{F}$  be the density and cumulative distribution functions of a spherical random variable  $\dot{v}$  with finite variance. By truncating  $\dot{v}$  from above at the value of  $cu$ , we define a truncated random variable  $x$  with density function  $\dot{f}(x)/\dot{F}(cu)$  for  $-\infty < x < cu$ . The variance of  $x$  is given by*

$$\text{Var}[x] = \frac{c^2}{\dot{F}(cu)} \left[ \check{F}(u) - u\check{f}(u) - \frac{\check{f}(u)^2}{\dot{F}(cu)} \right], \quad (\text{B.38})$$

where  $c = \sigma_{\dot{v}}$ , and  $\check{f}$  and  $\check{F}$  are the density and cumulative distribution functions of another elliptical random variable  $w$  that is associated with  $\dot{v}$  as defined in equation (B.16).

**Proof of Lemma B.2.** Using the fact that

$$\frac{\partial c\check{f}(x/c)}{\partial x} = -x\check{f}(x), \quad (\text{B.39})$$

we can easily obtain

$$E[x] = \frac{1}{\dot{F}(cu)} \int_{-\infty}^{cu} x\dot{f}(x)dx = -\frac{c}{\dot{F}(cu)} \check{f}\left(\frac{x}{c}\right)\Big|_{-\infty}^{cu} = -\frac{c\check{f}(u)}{\dot{F}(cu)}. \quad (\text{B.40})$$

Then, using integration by parts, we obtain the second moment of  $x$  as

$$\begin{aligned} E[x^2] &= \frac{1}{\dot{F}(cu)} \int_{-\infty}^{cu} x^2\dot{f}(x)dx \\ &= \frac{1}{\dot{F}(cu)} \left[ -cx\check{f}\left(\frac{x}{c}\right)\Big|_{-\infty}^{cu} + \int_{-\infty}^{cu} c\check{f}\left(\frac{x}{c}\right) dx \right] \\ &= \frac{1}{\dot{F}(cu)} [-c^2u\check{f}(u) + c^2\check{F}(u)]. \end{aligned} \quad (\text{B.41})$$

It follows that the variance of  $x$  is given by (B.38). This completes the proof.

## References

- Bansal, Ravi, David A. Hsieh, and S. “Vish” Viswanathan, 1993, A new approach to international arbitrage pricing, *Journal of Finance* 48, 1719–1747.
- Boyle, Phelim P., Shui Feng, Weidong Tian, and Tan Wang, 2008, Robust stochastic discount factors, *Review of Financial Studies* 21, 1077–1122.
- Burnside, Craig A., 1994, Hansen-Jagannathan bounds as classical tests of asset pricing models, *Journal of Business and Economic Statistics* 12, 57–79.
- Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy* 104, 298–345.
- Campbell, John, and John H. Cochrane, 2000, Explaining the poor performance of consumption-based asset pricing models, *Journal of Finance* 55, 2863–2878.
- Chernozhukov, Victor, Han Hong, and Elie Tamer, 2007, Estimation and confidence regions for parameter sets in econometric models, *Econometrica* 75, 1243–1284.
- Chernozhukov, Victor, Emre Kocatulum, and Konrad Menzel, 2015, Inference on sets in finance, *Quantitative Economics* 6, 309–358.
- Dittmar, Robert F., 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross-section of equity returns, *Journal of Finance* 57, 369–403.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Farnsworth, Heber, Wayne E. Ferson, David Jackson, and Steven Todd, 2002, Performance evaluation with stochastic discount factors, *Journal of Business* 75, 473–503.
- Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2013, Chi-squared tests for evaluation and comparison of asset pricing models, *Journal of Econometrics* 173, 108–125.
- Gospodinov, Nikolay, Raymond Kan, and Cesare Robotti, 2014, Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors, *Review of Financial Studies* 27, 2139–2170.

- Hamada, Mahmoud, and Emiliano A. Valdez, 2008, CAPM and option pricing with elliptically contoured distributions, *Journal of Risk and Insurance* 75, 387–409.
- Hansen, Lars Peter, John C. Heaton, and Erzo G. J. Luttmer, 1995, Econometric evaluation of asset pricing models, *Review of Financial Studies* 8, 237–274.
- Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.
- Hansen, Lars Peter, and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.
- Hodrick, Robert J., and Xiaoyan Zhang, 2001, Evaluating the specification errors of asset pricing models, *Journal of Financial Economics* 62, 327–376.
- Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara, 1998, Relationship between labor-income risk and average return: Empirical evidence from the Japanese stock market, *Journal of Business* 71, 319–348.
- Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.
- Kan, Raymond, and Cesare Robotti, 2008, Specification tests of asset pricing models using excess returns, *Journal of Empirical Finance* 15, 816–838.
- Kan, Raymond, and Cesare Robotti, 2009, Model comparison using the Hansen-Jagannathan distance, *Review of Financial Studies* 22, 3449–3490.
- Kan, Raymond, and Cesare Robotti, 2015, The exact distribution of the Hansen-Jagannathan bound, *Management Science*, forthcoming.
- Kan, Raymond, and Guofu Zhou, 2003, Hansen-Jagannathan distance: Geometry and exact distribution, Working Paper, University of Toronto.
- Kleibergen, Frank, and Richard Paap, 2006, Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics* 133, 97–126.

- Landsman, Zinovity M., and Emiliano A. Valdez, 2003, Tail conditional expectations for elliptical distributions, *North American Actuarial Journal* 7, 55–71.
- Lettau, Martin, and Sydney C. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy* 109, 1238–1287.
- Li, Quing, Maria Vassalou, and Yuhang Xing, 2006, Sector investment growth rates and the cross-section of equity returns, *Journal of Business* 79, 1637–1665.
- Li, Haitao, Yuewu Xu, and Xiaoyan Zhang, 2010, Evaluating asset pricing models using the second Hansen-Jagannathan distance, *Journal of Financial Economics* 97, 279–301.
- Lustig, Hanno, and Stijn Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia, *Journal of Finance* 60, 1167–1221.
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Owen, Joel, and Ramon Rabinovitch, 1983, On the class of elliptical distributions and their applications to the theory of portfolio choice, *Journal of Finance* 38, 745–752.
- Parker, Jonathan A., and Christian Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy* 113, 185–222.
- Petkova, Ralitsa, 2006, Do the Fama-French factors proxy for innovations in predictive variables?, *Journal of Finance* 61, 581–612.
- Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.
- Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance* 61, 539–580.
- Zhou, Guofu, 1993, Asset-pricing tests under alternative distributions, *Journal of Finance* 48, 1927–1942.

**Table 1**

**Summary of the models using quarterly gross returns on the one-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios**

Panel A: Unconstrained HJ-distance

Model	CC-CAY	CC-MY	C-SW	IGM	ICAPM	D-CCAPM	FF3
$\hat{\delta}$	0.623	0.635	0.650	0.541	0.527	0.650	0.582
$p(\delta = 0)$	0.001	0.003	0.000	0.356	0.133	0.000	0.000
$se(\hat{\delta})$	0.097	0.094	0.073	0.125	0.109	0.081	0.079
$P[\hat{y} < 0]$	0.115	0.111	0.014	0.207	0.115	0.023	0.014
$\sigma_{\hat{y}}$	0.772	0.920	0.390	1.409	1.060	0.475	0.407
$\rho_c^2$	0.193	0.138	0.512	0.123	0.199	0.350	0.975

Panel B: Constrained HJ-distance

Model	CC-CAY	CC-MY	C-SW	IGM	ICAPM	D-CCAPM	FF3
$\hat{\delta}_+$	0.666	0.692	0.681	0.670	0.595	0.679	0.607
$p(\delta_+ = 0)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$se(\hat{\delta}_+)$	0.099	0.091	0.085	0.096	0.107	0.087	0.088
$P[\hat{y} < 0]$	0.009	0.018	0.005	0.018	0.023	0.000	0.014
$\sigma_{\hat{y}}$	0.510	0.394	0.283	0.491	0.622	0.316	0.405
$\rho_c^2$	0.201	0.186	0.766	0.165	0.328	0.634	0.975
$(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$	6.9%	8.9%	4.8%	23.9%	12.9%	4.6%	4.3%
$\hat{\hat{\delta}}$	0.634	0.660	0.654	0.618	0.563	0.655	0.582

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 (CC-CAY) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (CC-MY) of Lustig and Van Nieuwerburgh (2005), the conditional CAPM (C-SW) of Santos and Veronesi (2006), the investment growth model (IGM) of Li, Vassalou, and Xing (2006), the intertemporal CAPM (ICAPM) of Petkova (2006), the durable consumption CAPM (D-CCAPM) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The data are from 1953:3 to 2007:3 (217 observations).  $p(\delta = 0)$  is the  $p$ -value for the test of  $H_0 : \delta = 0$ .  $p(\delta_+ = 0)$  is the  $p$ -value for the test of  $H_0 : \delta_+ = 0$ .  $se(\hat{\delta})$  ( $se(\hat{\delta}_+)$ ) is the standard error of  $\hat{\delta}$  ( $\hat{\delta}_+$ ) under the assumption that  $\delta > 0$  ( $\delta_+ > 0$ ).  $P[\hat{y} < 0]$  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.  $\rho_c^2$  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.  $\hat{\hat{\delta}}$  is a sample measure of the maximum pricing error on the test assets for the SDF that minimizes the constrained HJ-distance.



**Table 2**

**Estimates and  $t$ -ratios of parameters for various models using quarterly gross returns on the one-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios**

CC-CAY											
Unconstrained HJ-distance						Constrained HJ-distance					
	$\hat{\gamma}_0$	$\hat{\gamma}_{cay}$	$\hat{\gamma}_{c_{nd}}$	$\hat{\gamma}_{c_{nd}\cdot cay}$		$\hat{\gamma}_0$	$\hat{\gamma}_{cay}$	$\hat{\gamma}_{c_{nd}}$	$\hat{\gamma}_{c_{nd}\cdot cay}$		
$\hat{\gamma}$	1.28	-40.11	-54.14	-666.16		1.18	-28.04	-36.38	489.24		
$t(\hat{\gamma})$	7.46	-1.50	-1.13	-0.17		12.44	-1.95	-1.74	0.25		

CC-MY											
Unconstrained HJ-distance						Constrained HJ-distance					
	$\hat{\gamma}_0$	$\hat{\gamma}_{my}$	$\hat{\gamma}_{c_{nd}}$	$\hat{\gamma}_{c_{nd}\cdot my}$		$\hat{\gamma}_0$	$\hat{\gamma}_{my}$	$\hat{\gamma}_{c_{nd}}$	$\hat{\gamma}_{c_{nd}\cdot my}$		
$\hat{\gamma}$	1.32	2.70	-90.08	404.75		1.15	2.15	-41.47	27.49		
$t(\hat{\gamma})$	6.71	0.81	-1.84	0.78		16.53	1.10	-2.56	0.21		

C-SW											
Unconstrained HJ-distance						Constrained HJ-distance					
	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{r_{mkt}\cdot sw}$			$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{r_{mkt}\cdot sw}$			
$\hat{\gamma}$	1.01	117.47	-91.11			1.03	57.61	-45.80			
$t(\hat{\gamma})$	23.20	1.24	-1.27			26.66	1.06	-1.12			

IGM													
Unconstrained HJ-distance						Constrained HJ-distance							
	$\hat{\gamma}_0$	$\hat{\gamma}_{i_{hh}}$	$\hat{\gamma}_{i_{corp}}$	$\hat{\gamma}_{i_{incorp}}$	$\hat{\gamma}_{i_{fcorp}}$	$\hat{\gamma}_{i_{fm}}$		$\hat{\gamma}_0$	$\hat{\gamma}_{i_{hh}}$	$\hat{\gamma}_{i_{corp}}$	$\hat{\gamma}_{i_{incorp}}$	$\hat{\gamma}_{i_{fcorp}}$	$\hat{\gamma}_{i_{fm}}$
$\hat{\gamma}$	0.68	-17.10	9.19	-10.09	20.47	10.54		0.99	-8.90	1.31	-4.20	5.95	4.25
$t(\hat{\gamma})$	1.52	-1.42	1.08	-1.59	1.32	0.61		7.07	-2.06	0.93	-3.50	1.28	1.02

**Table 2 (continued)**

**Estimates and  $t$ -ratios of parameters for various models using quarterly gross returns on the one-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios**

ICAPM												
Unconstrained HJ-distance						Constrained HJ-distance						
	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{term}$	$\hat{\gamma}_{def}$	$\hat{\gamma}_{div}$	$\hat{\gamma}_{rf}$
$\hat{\gamma}$	0.93	3.12	-178.07	162.45	179.53	-9.95	0.99	-0.42	-98.73	123.22	71.14	-13.90
$t(\hat{\gamma})$	6.98	0.43	-1.33	1.41	0.69	-0.15	12.15	-0.10	-2.11	2.08	0.49	-0.58

D-CCAPM								
Unconstrained HJ-distance				Constrained HJ-distance				
	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{c_{nd}}$	$\hat{\gamma}_{c_d}$	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{c_{nd}}$	$\hat{\gamma}_{c_d}$
$\hat{\gamma}$	1.22	-1.98	-59.16	1.86	1.14	-2.54	-24.47	-1.97
$t(\hat{\gamma})$	7.75	-1.50	-1.10	0.23	14.36	-2.34	-1.01	-0.38

FF3								
Unconstrained HJ-distance				Constrained HJ-distance				
	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{r_{smb}}$	$\hat{\gamma}_{r_{hml}}$	$\hat{\gamma}_0$	$\hat{\gamma}_{r_{mkt}}$	$\hat{\gamma}_{r_{smb}}$	$\hat{\gamma}_{r_{hml}}$
$\hat{\gamma}$	1.15	-4.57	-0.26	-6.49	1.15	-4.53	-0.32	-6.45
$t(\hat{\gamma})$	20.61	-3.86	-0.18	-4.66	20.80	-3.88	-0.23	-4.65

The table presents the estimation results of seven linear asset-pricing models. The models include the conditional consumption CAPM (CC-CAY) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (CC-MY) of Lustig and Van Nieuwerburgh (2005), the conditional CAPM (C-SW) of Santos and Veronesi (2006), the investment growth model (IGM) of Li, Vassalou, and Xing (2006), the intertemporal CAPM (ICAPM) of Petkova (2006), the durable consumption CAPM (D-CCAPM) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The data are from 1953:3 to 2007:3 (217 observations). We report SDF parameter estimates  $\hat{\gamma}$  (with  $\hat{\gamma}_0$  being the estimated intercept of the SDF) and associated  $t$ -statistics ( $t(\hat{\gamma})$ ) based on (50) and (51).

**Table 3**

**Summary of the models using quarterly gross returns on the one-month T-bill and 10 size and 12 industry portfolios**

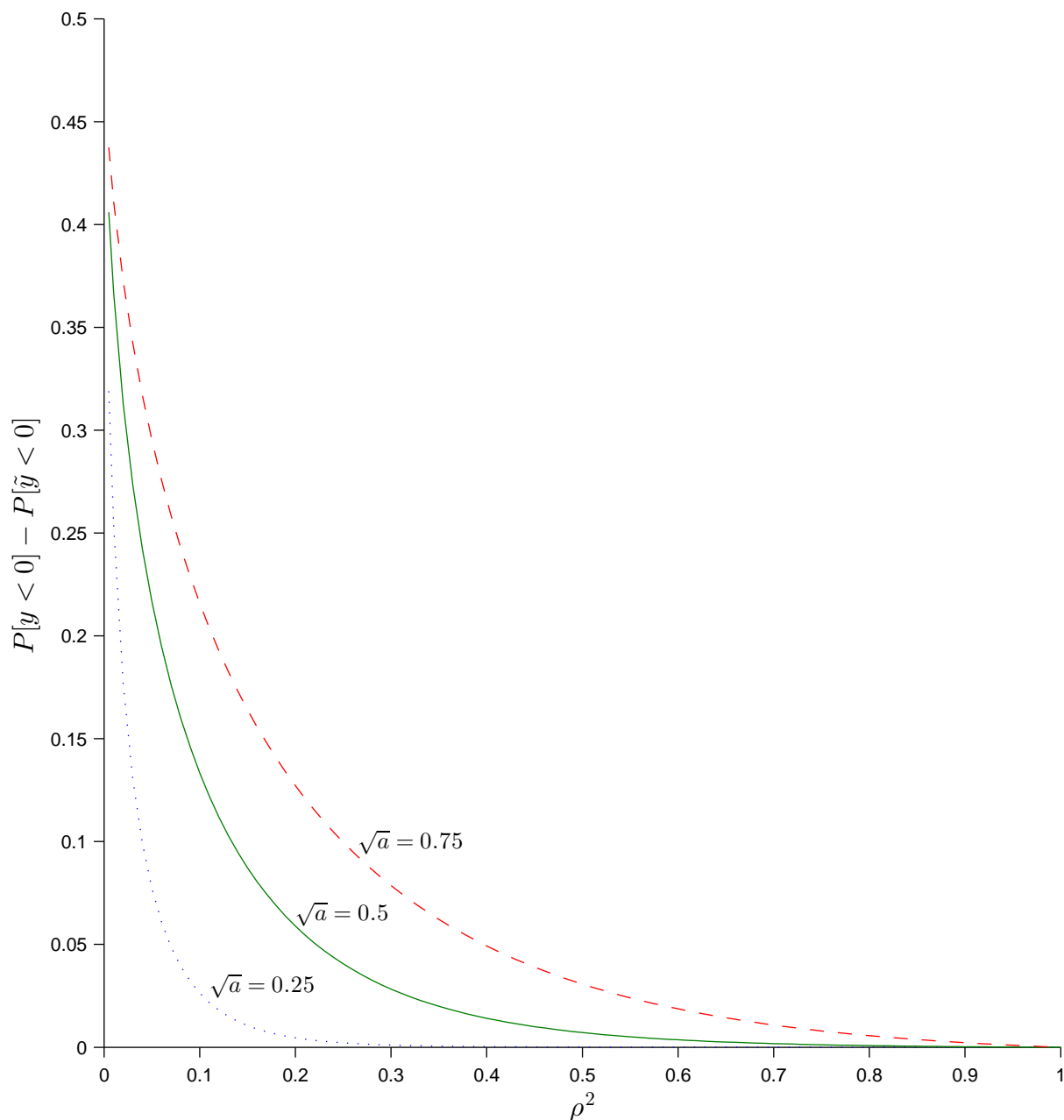
Panel A: Unconstrained HJ-distance

Model	CC-CAY	CC-MY	C-SW	IGM	ICAPM	D-CCAPM	FF3
$\hat{\delta}$	0.541	0.489	0.488	0.497	0.491	0.439	0.491
$p(\delta = 0)$	0.000	0.020	0.000	0.003	0.000	0.038	0.000
$se(\hat{\delta})$	0.077	0.081	0.075	0.084	0.073	0.078	0.075
$P[\hat{y} < 0]$	0.005	0.083	0.005	0.069	0.000	0.046	0.000
$\sigma_{\hat{y}}$	0.196	0.674	0.270	0.673	0.265	0.549	0.241
$\rho_c^2$	0.164	0.224	0.829	0.125	0.825	0.362	0.996

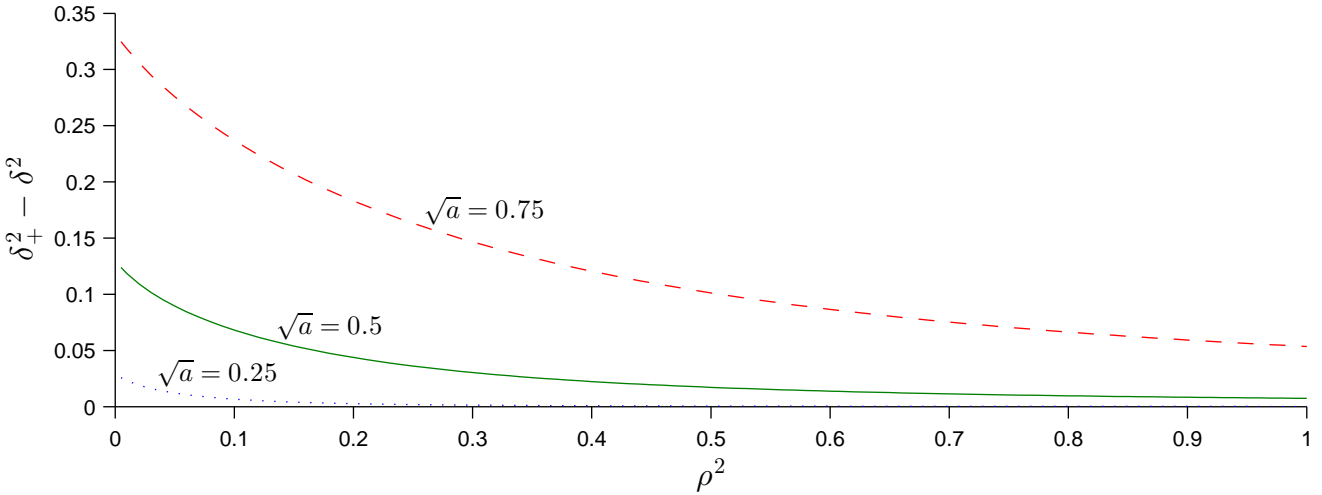
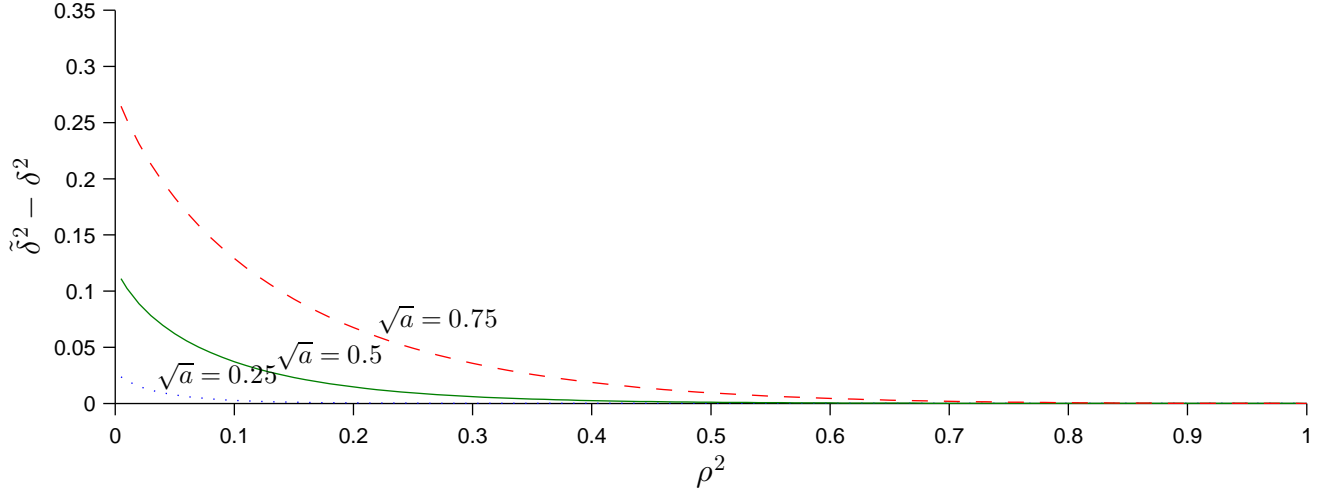
Panel B: Constrained HJ-distance

Model	CC-CAY	CC-MY	C-SW	IGM	ICAPM	D-CCAPM	FF3
$\hat{\delta}_+$	0.547	0.510	0.490	0.517	0.496	0.461	0.495
$p(\delta_+ = 0)$	0.000	0.000	0.000	0.000	0.000	0.005	0.000
$se(\hat{\delta}_+)$	0.081	0.084	0.078	0.086	0.078	0.081	0.079
$P[\hat{y} < 0]$	0.000	0.014	0.005	0.018	0.000	0.023	0.000
$\sigma_{\hat{y}}$	0.135	0.411	0.277	0.395	0.255	0.398	0.241
$\rho_c^2$	0.163	0.202	0.803	0.124	0.889	0.490	0.997
$(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$	1.2%	4.3%	0.4%	4.1%	0.9%	5.1%	0.9%
$\hat{\hat{\delta}}$	0.542	0.497	0.488	0.506	0.492	0.446	0.491

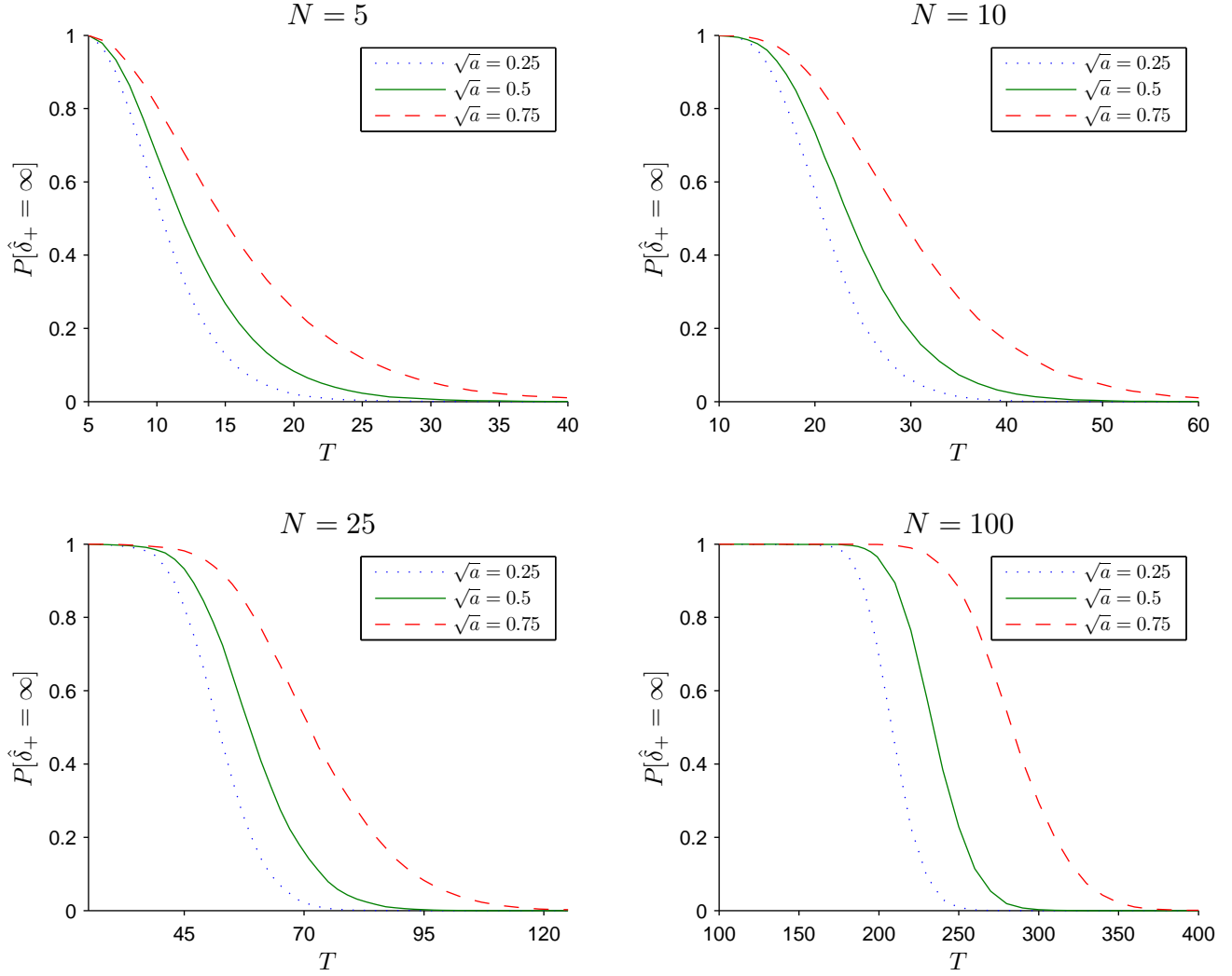
The table presents the sample unconstrained and constrained HJ-distances ( $\hat{\delta}$  and  $\hat{\delta}_+$ , respectively) of seven linear asset-pricing models. The models include the conditional consumption CAPM (CC-CAY) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (CC-MY) of Lustig and Van Nieuwerburgh (2005), the conditional CAPM (C-SW) of Santos and Veronesi (2006), the investment growth model (IGM) of Li, Vassalou, and Xing (2006), the intertemporal CAPM (ICAPM) of Petkova (2006), the durable consumption CAPM (D-CCAPM) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The data are from 1953:3 to 2007:3 (217 observations).  $p(\delta = 0)$  is the  $p$ -value for the test of  $H_0 : \delta = 0$ .  $p(\delta_+ = 0)$  is the  $p$ -value for the test of  $H_0 : \delta_+ = 0$ .  $se(\hat{\delta})$  ( $se(\hat{\delta}_+)$ ) is the standard error of  $\hat{\delta}$  ( $\hat{\delta}_+$ ) under the assumption that  $\delta > 0$  ( $\delta_+ > 0$ ).  $P[\hat{y} < 0]$  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.  $\rho_c^2$  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.  $\hat{\hat{\delta}}$  is a sample measure of the maximum pricing error on the test assets for the SDF that minimizes the constrained HJ-distance.



**Figure 1. 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  $\rho^2$  in a one-factor setting, where  $y$  and  $\tilde{y}$  are the linear SDFs chosen to minimize the unconstrained and constrained HJ-distances, respectively.  $\rho^2$  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 factor and the excess returns on the test assets are assumed to be multivariate  $t$ -distributed with six degrees of freedom.



**Figure 2. Differences in aggregate measures of pricing errors and in constrained and unconstrained HJ-distances.** The figure plots  $\tilde{\delta}^2 - \delta^2$  and  $\Delta = \delta_+^2 - \delta^2$  as functions of  $\rho^2$  in a one-factor setting.  $\rho^2$  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. The factor and the excess returns on the test assets are assumed to be multivariate  $t$ -distributed with six degrees of freedom.



**Figure 3. Probability for the sample constrained HJ-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  $t$ -distributed with six degrees of freedom and their unconditional variances are set equal to one. 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.