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

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti*

First draft: May 2009
This draft: December 2011

JEL # G12

*Gospodinov is from Concordia University and CIREQ. Kan is from the University of Toronto. Robotti is from the Federal Reserve Bank of Atlanta and EDHEC Risk Institute. We thank 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, 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. Corresponding author: Raymond Kan, Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada M5S 3E6; Tel: (416) 978-4291; Fax: (416) 978-5433; email:kan@chass.utoronto.ca.

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

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 conditional elliptical distribution assumption, we present explicit expressions for the HJ-distance with a no-arbitrage constraint, the associated Lagrange multipliers, and the 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 rank models. Finally, we illustrate the practical relevance of our theoretical findings in an empirical analysis of some popular asset pricing models.

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 (i.e., the set of SDFs that price a given set of test assets correctly). The second one measures the distance between a proposed SDF and the set of nonnegative admissible SDFs. Since the first measure does not impose the nonnegativity constraint (no-arbitrage condition) on the set of admissible SDFs whereas the second one does, we refer to the first measure as the unconstrained HJ-distance and to the second one as the constrained HJ-distance.

While the unconstrained HJ-distance is analyzed and used in many studies (see, 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 is largely ignored in the literature. The short list of studies that have analyzed and used the constrained HJ-distance includes Hansen, Heaton and Luttmer (1995), Bailey, Li and Zhang (2004), Wang and Zhang (2005), Chen and Ludvigson (2009), Liu, Kuo and Coakley (2009), and Fletcher (2010). In a recent paper, Li, Xu and Zhang (2010) strongly advocate the use of the constrained HJ-distance in empirical work because they find that this metric is more powerful in detecting misspecified models, especially those that are not arbitrage free, and in differentiating between models that have similar pricing errors on a given set of test assets. Their recommendation, however, appears to be driven by empirical evidence and is not based on an analysis of the theoretical properties of the unconstrained and constrained HJ-distances. The lack of a rigorous theoretical analysis of the properties of the constrained HJ-distance is likely due to the fact that an explicit expression for the constrained HJ-distance is not currently available even for linear models.

The main objective of our paper is to better understand the merits and drawbacks of the constrained HJ-distance and the difference between this measure and its unconstrained counterpart. We point out that when the SDF is perfectly correlated with the returns on the test assets, the difference between the squared constrained and unconstrained HJ-distances is the same as the difference between the constrained and unconstrained Hansen-Jagannathan bounds (HJ-bounds, see Hansen and Jagannathan (1991)) constructed from the test assets. This suggests that the difference between the two HJ-distances is identical across all SDFs that are spanned by the returns. Therefore, for two spanned SDFs, testing the equality of unconstrained HJ-distances is the same as testing the equality of constrained HJ-distances. For the more general case in which the SDF is not spanned by the returns on the test assets, we derive an explicit solution of the constrained HJ-distance under the assumption that the SDF and the returns are 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, in the case of linear SDFs, we obtain analytical expressions of the SDF parameters that minimize the constrained HJ-distance. When there is a unspanned factor in the linear 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 linear SDF to take negative values, but will lead to a serious deterioration in the ability of the SDF to price the test assets.

In light of our theoretical findings, we reexamine the empirical performance of the seven asset pricing models considered by Li, Xu and Zhang (2010). We find that their main conclusion that it is easier to differentiate between models when comparing them based on their sample constrained HJ-distances is heavily driven by the very high sample Sharpe ratio of the test assets that they use.

The rest of the paper is organized as follows. Section 1 presents a theoretical analysis of the unconstrained and constrained HJ-distances. Section 2 derives an analytical solution of the constrained HJ-distance under the ellipticity assumption. Section 3 presents our empirical results. Some concluding remarks are provided in Section 4.

1. Unconstrained and Constrained Hansen-Jagannathan Distances

1.1 The setup

Following HJ, let $\mathcal{F}$ be the information that is observed at the date of the asset payoffs. Associated with $\mathcal{F}$ is the space L^2 of all random variables with finite second moments that are in the information set $\mathcal{F}$. This space is used as the collection of hypothetical claims that could be traded. However, for practical reasons, econometricians can only evaluate asset pricing models on a subspace of L^2. Let $\tilde{r} = [R_0,\ r']'$, where R_0 is the gross return on the risk-free asset, and r is a vector of excess returns (in excess of the risk-free rate) on N risky assets.[1] We assume that the payoff space used in an econometric analysis is given by the payoffs of portfolios of $\tilde{r}$:

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

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

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

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

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

[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.

1.2 Pricing errors and Hansen-Jagannathan distances

Let $y \in L^2$ be a candidate stochastic discount factor. If y prices the n test assets correctly, then the vector of pricing errors, e, of the test assets is exactly zero:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where Γ is the parameter space of γ.

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

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

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

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

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

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

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

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

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

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

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

where $m \in \mathcal{M}^+$. It follows that for a given $m \in \mathcal{M}^+$, the maximum pricing error of h with $E[h^2] = 1$ is given by

$$\max_{h \in L^2,\, E[h^2]=1} |E[yh] - E[mh]| = E[(y - m)^2]^{\frac{1}{2}}. \tag{15}$$

It is easy to verify that the maximum pricing error occurs for $h = (y - m)/E[(y - m)^2]^{\frac{1}{2}}$.

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

$$\min_{m \in \mathcal{M}^+} \max_{h \in L^2,\, E[h^2]=1} |E[yh] - E[mh]| = \min_{m \in \mathcal{M}^+} E[(y - m)^2]^{\frac{1}{2}} = E[(y - m_y^+)^2]^{\frac{1}{2}} = \delta_+, \tag{16}$$

where m_y^+ is the nonnegative admissible SDF that is closest to y.

It is important to emphasize that δ_+ 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 (15) and (16), we have

$$\max_{h \in L^2,\, E[h^2]=1} |E[yh] - E[m^*h]| = E[(y - m^*)^2]^{\frac{1}{2}} \geq E[(y - m_y^+)^2]^{\frac{1}{2}} = \delta_+, \tag{17}$$

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

From (17), the maximum pricing error of a model is equal to the distance between y and m^*. However, a model in $\mathcal{M}^+$ (i.e., $\delta_+ = 0$) can actually be further away from m^* than a model that is not in $\mathcal{M}^+$ (i.e., $\delta_+ > 0$). This makes it problematic to rank models by δ_+ because a model with a larger δ_+ can actually be closer to m^* and have a smaller maximum pricing error for payoffs in L^2. In particular, a model with a smaller δ_+ is not necessarily a better model for pricing derivatives.[2]

[2]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.

While it is desirable to consider SDFs that are strictly positive, most SDFs used in empirical work are typically misspecified and some of them can take on negative values. It is often believed that a model with a smaller δ_+ has a smaller probability of taking on negative values because it is closer to $\mathcal{M}^+$. As it turns out, the probability for an SDF to take on negative values has very little to do with the magnitude of δ_+.

To illustrate this point, we adapt an example from Li, Xu and Zhang (2010). We consider an economy with two states (s_1 and s_2) that are equally likely to occur. The only test asset considered by the econometrician is risk-free with gross risk-free rate $R_0 = 1$, so that the payoff space of the test asset ($\mathcal{P}$) can be represented by the dashed line in Figure 1. For an SDF m to be admissible, it has to price the risk-free asset correctly, which implies

$$E[m] = 1 \Rightarrow \frac{1}{2} \times m_1 + \frac{1}{2} \times m_2 = 1 \Rightarrow m_2 = 2 - m_1, \tag{18}$$

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

Figure 1 about here

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

$$\begin{aligned} \delta_+ &= \min_{m \in \mathcal{M}^+} E[(y-m)^2]^{\frac{1}{2}} \\ &= \min_{m \in \mathcal{M}^+} \left[\frac{1}{2} \times (m_1 - y_1)^2 + \frac{1}{2} \times (m_2 - y_2)^2\right]^{\frac{1}{2}} \\ &= \frac{1}{\sqrt{2}} \min_{m \in \mathcal{M}^+} [(m_1 - y_1)^2 + (m_2 - y_2)^2]^{\frac{1}{2}}. \end{aligned} \tag{19}$$

It follows that the shortest distance between $\mathcal{M}^+$ and y is equal to $\sqrt{2}\delta_+$. In Figure 1, we observe that $y^{\mathcal{F}}$ is further away from $\mathcal{M}^+$ than $y^{\mathcal{G}}$. Consequently, we have $\delta_{\mathcal{F},+} > \delta_{\mathcal{G},+}$. Despite having a shorter constrained HJ-distance, $y^{\mathcal{G}}$ takes on negative values in both states. In contrast, $y^{\mathcal{F}}$ always takes on positive values. The important message here is that while $m \in \mathcal{M}^+$ is positive, the distance from $\mathcal{M}^+$ alone tells us little about the probability for an SDF to take on negative values.[3]

We would like to conclude this section with another important observation regarding the concern that some SDFs, especially those that are assumed to be normally distributed, can take on negative values and hence are not arbitrage free. For example, suppose that an SDF y and the returns are jointly normally distributed. It turns out that this SDF can always be converted into a positive SDF without affecting its ability to price the test assets. In particular, consider the following transformation of y:

$$y_+ = \mu_y \exp\left(\frac{y}{\mu_y} - 1 - \frac{\sigma_y^2}{2\mu_y^2}\right), \tag{20}$$

where $\mu_y = E[y]$ and $\sigma_y^2 = \text{Var}[y]$. Note that $y_+ > 0$ when $\mu_y > 0$. Using the properties of the log-normal distribution and Stein's lemma, it is straightforward to show that $E[\tilde{r}y_+] = E[\tilde{r}y]$, so both y and y_+ have the same pricing errors on the test assets. However, this is a mechanical transformation and there is no compelling reason to believe that simply because y_+ is positive, it can price derivatives better than y.

1.3 Hansen-Jagannathan bounds and distances

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

$$\sigma_0^2 = \min_{m \in \mathcal{M}} \text{Var}[m] = \min_{m \in \mathcal{M}} E[m^2] - E[m]^2 = \min_{m \in \mathcal{M}} E[m^2] - \frac{1}{R_0^2}. \tag{21}$$

[3] In Figure 1, the two SDFs are chosen to have the same unconstrained HJ-distance, i.e., $\delta_{\mathcal{F}} = \delta_{\mathcal{G}}$. This is intended to show that even for two models with similar δ's, we still cannot infer that the model with a lower δ_+ has a smaller probability of taking negative values.

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

$$\sigma_c^2 = \min_{m \in \mathcal{M}^+} \text{Var}[m] = \min_{m \in \mathcal{M}^+} E[m^2] - E[m]^2 = \min_{m \in \mathcal{M}^+} E[m^2] - \frac{1}{R_0^2}. \tag{22}$$

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

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

$$\delta^2 = \min_{m \in \mathcal{M}} E[(y - m)^2] = E[y^2] - 2E[ym] + \min_{m \in \mathcal{M}} E[m^2]. \tag{23}$$

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

$$\delta_+^2 = \min_{m \in \mathcal{M}^+} E[(y - m)^2] = E[y^2] - 2E[ym] + \min_{m \in \mathcal{M}^+} E[m^2]. \tag{24}$$

It follows that

$$\delta_+^2 - \delta^2 = \min_{m \in \mathcal{M}^+} E[m^2] - \min_{m \in \mathcal{M}} E[m^2] = \sigma_c^2 - \sigma_0^2. \tag{25}$$

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

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

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

$$\delta_+^2 = \min_{\gamma} \min_{m \in \mathcal{M}^+} E[(y(\gamma) - m)^2] = \min_{\gamma}(E[y(\gamma)^2] - 2E[y(\gamma)m]) + \min_{m \in \mathcal{M}^+} E[m^2]. \tag{27}$$

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

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

$$\text{argmin}_\gamma E[y(\gamma)^2] - 2E[y(\gamma)m], \tag{28}$$

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

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

$$\delta_{\mathcal{F},+}^2 - \delta_{\mathcal{G},+}^2 = \delta_{\mathcal{F}}^2 + (\sigma_c^2 - \sigma_0^2) - \delta_{\mathcal{G}}^2 - (\sigma_c^2 - \sigma_0^2) = \delta_{\mathcal{F}}^2 - \delta_{\mathcal{G}}^2. \tag{29}$$

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

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

$$y = y^* + z, \tag{30}$$

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

$$y^* = \mu_y + V_{ry}' V_{rr}^{-1}(r - \mu_r), \tag{31}$$

with $\mu_y = E[y]$, $\mu_r = E[r]$, $V_{rr} = \text{Var}[r]$, and $V_{ry} = \text{Cov}[r, y]$. It is easy to see that z, the unspanned component, has mean zero and is uncorrelated with r. Since y^* is spanned by the returns, $E[y^*m]$

is constant across $m \in \mathcal{M}$. It follows that the squared unconstrained HJ-distance of y is given by

$$
\begin{aligned}
\delta^2 &= \min_{m\in\mathcal{M}} E[(y-m)^2] \\
&= \min_{m\in\mathcal{M}} E[(y^*+z-m)^2] \\
&= E[y^{*2}] + 2E[y^*(z-m)] + \min_{m\in\mathcal{M}} E[(z-m)^2] \\
&= E[y^{*2}] - 2E[y^*m] + \sigma_0^2 + \frac{1}{R_0^2}. \qquad (32)
\end{aligned}
$$

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

$$\min_{m\in\mathcal{M}} E[(z-m)^2] = \min_{\tilde{m}\in\mathcal{M}} E[\tilde{m}^2] = \sigma_0^2 + \frac{1}{R_0^2}. \qquad (33)$$

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

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

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

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

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

2. Analytical Solution of the Constrained Hansen-Jagannathan Distance

While an explicit solution of the unconstrained HJ-distance is easy to obtain, the constrained HJ-distance problem is much harder and, to the best of our knowledge, an analytical solution is not available in the literature. To a large extent, the lack of an analytical expression has severely hampered our ability to understand the constrained HJ-distance. To overcome this problem, we make a joint distributional assumption on the SDF and the returns on the test assets. Throughout this section, we assume that the 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).[4] 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. Nevertheless, we do not argue that the multivariate elliptical distribution 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. We assume ellipticity solely for the purpose of deriving an analytical solution for δ_+^2. As we demonstrate below, solving for δ_+^2 is nontrivial even under the ellipticity assumption. Appendix A provides some definitions and notation for elliptically distributed random variables. We should also stress that some of the results on moments of censored and truncated elliptically distributed random variables are new and

[4] 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 clearly conducted under the hypothesis that the asset pricing model is misspecified.

of independent interest given the importance of elliptical distributions for portfolio choice theory, asset and option pricing (see Owen and Rabinovitch (1983), Zhou (1993) and Hamada and Valdez (2008), among others.)

2.1 Stochastic discount factors without parameters

We start off with the case in which the SDF y does not depend on unknown parameters. As before, we decompose y into two components y^* (spanned) and z (unspanned) as in (30). For the unconstrained HJ-distance, the vector of pricing errors of $\tilde{r}$ is given by

$$e = E[\tilde{r}y] - q = \begin{bmatrix} R_0\mu_y - 1 \\ V_{ry} + \mu_r\mu_y \end{bmatrix}. \tag{36}$$

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

$$U^{-1} = \begin{bmatrix} R_0^2 & R_0\mu_r' \\ R_0\mu_r & V_{rr} + \mu_r\mu_r' \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1+a}{R_0^2} & -\frac{\mu_r' V_{rr}^{-1}}{R_0} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} & V_{rr}^{-1} \end{bmatrix}, \tag{37}$$

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

$$\lambda = U^{-1}e = \begin{bmatrix} \frac{\mu_y - V_{ry}' V_{rr}^{-1}\mu_r}{R_0} - \frac{1+a}{R_0^2} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0}\right) \end{bmatrix} \tag{38}$$

and the admissible SDF that is closest to y is

$$m_y = y - \lambda'\tilde{r} = z + \frac{1}{R_0} - \frac{\mu_r' V_{rr}^{-1}(r - \mu_r)}{R_0}. \tag{39}$$

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

$$\delta^2 = E[(y - m_y)^2] = \left(\mu_y - \frac{1}{R_0}\right)^2 + \left(V_{ry} + \frac{\mu_r}{R_0}\right)' V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0}\right). \tag{40}$$

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

$$\tilde{\lambda} = \text{argmin}_{\lambda} E[(y - \lambda'\tilde{r})^{+2}] + 2\lambda' q, \tag{41}$$

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

$$E[\tilde{r}(y - \tilde{\lambda}'\tilde{r})^+] = q. \tag{42}$$

In principle, we can solve the n nonlinear equations $E[\tilde{r}(y - \tilde{\lambda}'\tilde{r})^+] = q$ to obtain the vector of Lagrange multipliers $\tilde{\lambda}$, but this can be very complicated. Instead, we simplify the problem so that we only need to solve one nonlinear equation to obtain $\tilde{\lambda}$.

The solution to the first order condition in (42) depends on the joint distribution of y and r. In order to simplify the problem, we assume that conditional on $\mathcal{F}$, y and r have a multivariate elliptical distribution with finite variance. It follows that a linear combination of y and $\tilde{r}$, say v, also has a conditional 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 μ_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 $\tilde{v} = (v - \mu_v)/s_v$ as f and F, respectively. For a given choice of f, we define another elliptically distributed random variable w with the following density function

$$\tilde{f}(w) = \int_w^\infty csf(cs)\mathrm{d}s. \tag{43}$$

and denote the cumulative distribution function of w by $\tilde{F}$. In Appendix A.1, we present a more complete discussion of the class of elliptical distributions and explicitly derive $\tilde{f}$.

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

Proposition 1. *Suppose y and r are jointly elliptically distributed with finite variance, and c, F, $\tilde{f}$ and $\tilde{F}$ are defined as above. Let η be the unique solution of the following equation*

$$g(u) = [a + \sigma_z^2 R_0^2 \tilde{F}(u)^2]^{-\frac{1}{2}}, \tag{44}$$

where a is the squared Sharpe ratio of the tangency portfolio of the N risky assets, σ_z^2 is the variance *of the unspanned component of y, and*

$$g(u) = \frac{uF(cu) + \tilde{f}(u)}{\tilde{F}(u)}. \tag{45}$$

The vector of Lagrange multipliers in the constrained HJ-distance case is given by

$$\tilde{\lambda} = \left[\begin{array}{c} \frac{\mu_y - V_{ry}' V_{rr}^{-1} \mu_r}{R_0} - \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} \\ V_{rr}^{-1} \left(V_{ry} + \frac{\mu_r}{R_0 \tilde{F}(\eta)} \right) \end{array} \right]. \tag{46}$$

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

$$\delta_+^2 = \delta^2 + \sigma_z^2 \tilde{F}(-\eta) + \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} - \frac{1+a}{R_0^2}, \tag{47}$$

where the expression of δ^2 is provided in (40).

Proof. See Appendix B.

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

$$m_y = y - \tilde{\lambda}' \tilde{r} = z + \frac{1}{R_0 \tilde{F}(\eta)} \left[\frac{\eta}{g(\eta)} - \mu_r' V_{rr}^{-1} (r - \mu_r) \right]. \tag{48}$$

Just like m_y in (39) for the unconstrained HJ-distance case, the m_y for the constrained HJ-distance also has two components. The first component, z, is uncorrelated with the returns, and the second component is a linear function of the excess return on the tangency portfolio of the test assets. Let $\Delta = \delta_+^2 - \delta^2$ be the difference between the squared constrained and unconstrained HJ-distances of y. We are interested in the determinants of Δ. Note that η, the solution to (44), depends on σ_z^2, a, and R_0. As a result, Δ also depends on these three parameters. Out of the three parameters, the only one that is related to the SDF is σ_z^2 — the variance of the unspanned component. When $\sigma_z^2 = 0$ (i.e., spanned SDF), we have

$$g(\eta) = \frac{1}{\sqrt{a}} \tag{49}$$

and hence

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

where the last equality is a generalization of Proposition 1 of Kan and Robotti (2008a) to the case of elliptical distributions.[5] This confirms the result in Section 1.3 which suggests that $\delta_+^2 - \delta^2$ for a spanned SDF is equal to the difference between the constrained and unconstrained HJ-bounds.

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

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

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

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

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

Proof. See Appendix B.

Lemma 1 shows that Δ is an increasing function of σ_z^2, which suggests that $\sigma_c^2 - \sigma_0^2$ is a lower bound for Δ. Intuitively, adding an unspanned component z to an SDF does not affect its ability to price the test assets, so the unconstrained HJ-distance (which is a measure of aggregate pricing errors) of a model is unaffected by z. This explains why the expression of δ^2 in (40) is independent of σ_z^2. However, adding z to an SDF can affect its probability of taking on negative values and hence drives the SDF further away from $\mathcal{M}^+$. This explains why δ_+^2 and hence Δ is an increasing function of σ_z^2. Note that Δ does not depend on how good or bad a model is for the test assets. It is only a function of the variance of its unspanned component. This suggests that for two different models, say $\mathcal{F}$ and $\mathcal{G}$, we can expect $\delta_{\mathcal{F},+}^2 - \delta_{\mathcal{G},+}^2$ to differ substantially from $\delta_{\mathcal{F}}^2 - \delta_{\mathcal{G}}^2$ only when the variances of the unspanned components across the two models are very different.

In addition, Lemma 1 shows that Δ is an increasing function of the Sharpe ratio of the tangency portfolio of the test assets. This result requires some explanation. Consider the case in which $\sigma_z^2 \to 0$. When this happens, $(\delta_+^2 - \delta^2) \to (\sigma_c^2 - \sigma_0^2)$ — the difference between the constrained

[5] The proof of this generalization to elliptically distributed random variables is available from the authors upon request.

and unconstrained HJ-bounds. Lemma 6 of Kan and Robotti (2008a) shows that, under normality, $(\sigma_c^2 - \sigma_0^2) \to 0$ when $a \to 0$, $(\sigma_c^2 - \sigma_0^2) \to \infty$ when $a \to \infty$, and $\sigma_c^2 - \sigma_0^2$ is a strictly increasing function of a. Therefore, when a is small, we should not expect large differences between the constrained and unconstrained HJ-bounds and between the constrained and unconstrained HJ-distances. Intuitively, when a is close to zero, the weight of the risk-free asset in the minimum second moment portfolio is close to one, and the gross return on this portfolio has a very small probability of taking on a negative value. Since the minimum variance admissible SDF is proportional to the gross return on this portfolio, imposing the nonnegativity constraint of Hansen and Jagannathan (1991) on it has almost no effect.

To gain some understanding of how σ_z and a affect Δ, Figure 2 plots Δ as a function of σ_z for three different values of the Sharpe ratio of the tangency portfolio ($\sqrt{a} = 0.25$, 0.5, and 0.75) with $R_0 = 1.005$ (the plot is not sensitive to other reasonable values of the gross risk-free rate).[6] In the plot, the SDF and the excess returns on the test assets are assumed to be multivariate t-distributed with six degrees of freedom. As expected, Figure 2 reveals that Δ is an increasing function of σ_z. However, Δ is heavily influenced by the Sharpe ratio of the tangency portfolio. When $\sqrt{a} = 0.25$, the difference between δ_+^2 and δ^2 is indistinguishable from zero. For $\sqrt{a} = 0.5$, the difference between δ_+^2 and δ^2 is still quite small, even for relatively large σ_z. This suggests that for reasonable Sharpe ratio values, we should not expect to find a large difference between the constrained and unconstrained HJ-distances of a model, even if the model contains a large unspanned component.

Figure 2 about here

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

[6] Although a Sharpe ratio of 0.75 may seem high, this is in line with the sample Sharpe ratio (0.71) of the tangency portfolio of the 25 Fama-French size and book-to-market portfolios used in the empirical application in Section 4.

difference could be partly due to sampling variation and partly due to an upward bias of the sample constrained HJ-distance.

2.2 Linear stochastic discount factors

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

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

where f is a vector of K systematic factors, and $\gamma = [\gamma_0,\ \gamma_1']'$ is the vector of SDF parameters. In addition to facilitating the derivation of γ, linear SDFs deserve a thorough investigation because of their popularity in the literature. Throughout this subsection, we assume that f and r are jointly elliptically distributed (which implies that $y(\gamma)$ and r are jointly elliptically distributed).

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

$$\gamma_1 = -\frac{1}{R_0}(V_{rf}'V_{rr}^{-1}V_{rf})^{-1}(V_{rf}'V_{rr}^{-1}\mu_r), \qquad \gamma_0 = \frac{1}{R_0} - \gamma_1'\mu_f. \tag{55}$$

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

$$y = \frac{1}{R_0} + \gamma_1'(f - \mu_f). \tag{56}$$

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

$$\delta^2 = \frac{a - a_1}{R_0^2} \tag{57}$$

and

$$\lambda = \begin{bmatrix} -\delta^2 \\ V_{rr}^{-1}\left(V_{rf}\gamma_1 + \frac{\mu_r}{R_0}\right) \end{bmatrix}, \tag{58}$$

respectively.

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

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

$$g(u) = \left(a + \alpha'\left[\frac{1}{\tilde{F}(u)}I_K - V_{ff}^{-\frac{1}{2}}V_{ff\cdot r}V_{ff}^{-\frac{1}{2}}\right]^{-2}\alpha\right)^{-\frac{1}{2}}, \tag{59}$$

where $\alpha = V_{ff}^{-1}V_{ff\cdot r}^{\frac{1}{2}}V_{rf}'V_{rr}^{-1}\mu_r$, $g(u)$ is defined in (45), and $\tilde{F}$ is defined before Proposition 1. Then, the vector of SDF parameters that minimizes the constrained HJ-distance is given by $\tilde{\gamma} = [\tilde{\gamma}_0,\ \tilde{\gamma}_1']'$, where

$$\tilde{\gamma}_1 = -\frac{1}{R_0}[V_{ff} - \tilde{F}(\eta)V_{ff\cdot r}]^{-1}(V_{rf}'V_{rr}^{-1}\mu_r), \qquad \tilde{\gamma}_0 = \frac{1}{R_0} - \tilde{\gamma}_1'\mu_f, \tag{60}$$

and the SDF that minimizes the constrained HJ-distance is

$$\tilde{y} = \frac{1}{R_0} + \tilde{\gamma}_1'(f - \mu_f). \tag{61}$$

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

$$\delta_+^2 = \frac{a + \frac{\eta}{g(\eta)}}{R_0^2\tilde{F}(\eta)} - \frac{1 + \tilde{a}_1}{R_0^2}, \tag{62}$$

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

$$\tilde{\lambda} = \begin{bmatrix} -\delta_+^2 \\ V_{rr}^{-1}\left(V_{rf}\tilde{\gamma}_1 + \frac{\mu_r}{R_0\tilde{F}(\eta)}\right) \end{bmatrix}. \tag{63}$$

Proof. See Appendix B.

Besides the simplicity of the expressions for $\tilde{\gamma}$, $\tilde{\lambda}$, and δ_+^2, a few interesting observations emerge from Proposition 2. First, the Lagrange multiplier on the risk-free asset is equal to $-\delta_+^2$ (expression (58) shows that a similar result holds for the unconstrained HJ-distance). Second, in contrast to the SDF case without parameters, η does not depend on R_0 since $\tilde{y}$ prices the risk-free asset correctly. Third, when the factors are spanned by the returns (i.e., $V_{ff\cdot r} = 0_{K\times K}$), it can be readily shown that the difference between the squared constrained and unconstrained HJ-distances coincides with the difference between the constrained and unconstrained HJ-bounds. This result confirms our earlier findings for spanned SDFs in Section 1.3. Finally, when one or more factors are useless, i.e., they are uncorrelated with the returns, 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. However, the SDF parameters that minimize the constrained HJ-distance are still well defined. For example, when all factors are useless, we have $\tilde{\gamma} = [1/R_0,\ 0_K']'$ and $\tilde{y} = 1/R_0$. In this case, δ_+^2 is equal to the constrained HJ-bound σ_c^2.

With the analytical solutions of the linear SDF parameters for the unconstrained and constrained HJ-distances, we can now answer two interesting questions. The first question is whether the linear SDF $\tilde{y}$ in (61) results in a lower probability of taking on negative values than the linear SDF y in (56). If this is the case, one can think of this as a potential benefit of using the constrained HJ-distance. The second question is whether there is a trade-off between getting the linear SDF closer to $\mathcal{M}^+$ and the ability of the SDF to price the test assets. For this purpose, we introduce an aggregate measure of pricing errors of $\tilde{y}$ as

$$\tilde{\delta}^2 = \tilde{e}'U^{-1}\tilde{e}, \tag{64}$$

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

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

$$P[y < 0] - P[\tilde{y} < 0] = F\left(-\frac{c}{R_0\sqrt{\gamma_1' V_{ff}\gamma_1}}\right) - F\left(-\frac{c}{R_0\sqrt{\tilde{\gamma}_1' V_{ff}\tilde{\gamma}_1}}\right) > 0, \tag{65}$$

where F and c are defined before Proposition 1. In addition, we have

$$\delta^2 \le \tilde{\delta}^2 \le \delta_+^2. \tag{66}$$

Proof. See Appendix B.

As shown in the proof of Lemma 2, (66) is a general result. It is not specific to the linear model and our proof does not rely on the ellipticity assumption. However, we can only establish $P[y < 0] > P[\tilde{y} < 0]$ for the case of linear models and under the ellipticity assumption. Whether this inequality continues to hold for nonlinear models or without the ellipticity assumption is an open question.[7]

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. For the one-factor case, we can show that

$$P[y < 0] - P[\tilde{y} < 0] = F\left(-\frac{c|\rho|}{\sqrt{a_1}}\right) - F\left(-\frac{c[1 - \tilde{F}(\eta)(1-\rho^2)]}{\sqrt{a_1}|\rho|}\right), \tag{67}$$

where η is the unique solution of

$$g(u) = \left[a + \frac{a_1\rho^2(1-\rho^2)\tilde{F}(u)^2}{[1 - \tilde{F}(u)(1-\rho^2)]^2}\right]^{-\frac{1}{2}}, \tag{68}$$

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

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

$$\tilde{\delta}^2 - \delta^2 = \frac{a_1}{R_0^2} \left[\frac{\tilde{F}(-\eta)(1-\rho^2)}{1-\tilde{F}(\eta)(1-\rho^2)} \right]^2 . \tag{69}$$

Note that both (67) and (69) depend on a, a_1 and ρ^2. In these expressions, a is the squared Sharpe ratio of the tangency portfolio of the test assets, which is a measure of the cross-sectional difference in expected excess returns across the test assets; a_1 measures how good the model is in explaining the expected returns on the test assets (recall that $\delta^2 = (a - a_1)/R_0^2$); and, finally, ρ^2 measures how well the factor is spanned by the returns.

Assuming that the factor and the excess returns on the test are multivariate t-distributed with six degrees of freedom, Figure 3 plots $P[y < 0] - P[\tilde{y} < 0]$ as a function of ρ^2 for three different values of the Sharpe ratio of the tangency portfolio ($\sqrt{a} = 0.25$, 0.5, and 0.75). In each case, we assume $a_1 = a/2$, so that the model explains half of the cross-sectional variation in expected returns.

Figure 3 about here

From Figure 3, we can see that when $\rho^2 \to 0$ (y is not defined when $\rho^2 = 0$), $P[y < 0] - P[\tilde{y} < 0] \to 0.5$. The reason is that when the unspanned component of the factor increases, y becomes more volatile (because γ_1 does not depend on the unspanned component of the factor) and behaves more like a useless factor. As a result, $P[y < 0] \to 0.5$. However, as $\rho^2 \to 0$, $\tilde{y}$ converges to $1/R_0$ and has almost zero probability of taking on negative values. In contrast, when $\rho^2 \to 1$, the SDF behaves more like a spanned SDF. For a spanned SDF, the SDF parameters and hence the probabilities of taking on negative values are the same for y and $\tilde{y}$. Finally, Figure 3 shows that the Sharpe ratio is important in determining $P[y < 0] - P[\tilde{y} < 0]$. For a given value of ρ^2, we can see that the difference between the two probabilities is an increasing function of a. The reason is that the spanned component of the SDF y is a linear function of the return on the factor mimicking portfolio. When a is small, a_1 is also small, so y puts relatively little weight on the factor mimicking portfolio and hence $P[y < 0]$ is small, leaving not much room for $\tilde{y}$ to improve.

Using the same parameters and distributional assumption as in Figure 3, Figure 4 plots $\tilde{\delta}^2 - \delta^2$ as a function of ρ^2 for $R_0 = 1.005$.

Figure 4 about here

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

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

3. Empirical Analysis

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

3.1 Data and asset pricing models

For ease of comparison, we focus on the same asset pricing models considered by Li, Xu and Zhang (2010) and use their data to perform our empirical analysis.[8] The return data consist of quarterly gross returns on the three-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios. The data are from 1952:2 to 2000:4 (195 quarterly observations). The seven models that are considered are:

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

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

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

LV: a version of the conditional consumption CAPM of Lustig and Van Nieuwerburgh (2005)

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

where my is the housing collateral ratio;

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

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

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

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

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

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

[8] We thank Haitao Li, Yuewu Xu and Xiaoyan Zhang for making their data available to us and refer to their paper for a more detailed description of the data.

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

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

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

YOGO: the durable consumption CAPM of Yogo (2006)

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

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

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

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

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

3.2 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. The table also reports the probability that the estimated SDF takes on negative values in the sample, the standard deviation of the estimated SDF, and 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 $\hat{\tilde{\delta}}$, the sample estimate of $\tilde{\delta}$ defined in (64), which is a measure of the maximum pricing error on the test assets for the SDF that minimizes the constrained HJ-distance.

Table 1 about here

While some of the results in Table 1 are already reported and discussed in Li, Xu and Zhang (2010), we would like to emphasize and reinterpret several important findings that naturally emerge from the predictions of our theoretical analyses in Sections 1 and 2. First, Table 1 clearly shows that the largest increases in the sample constrained HJ-distance over its unconstrained counterpart occur for models with high probabilities of taking on negative values (such as LV, LVX1, and LVX2). For these models, the probability for their SDF to take on negative values can be greatly reduced when the parameters are chosen to minimize the constrained HJ-distance. However, as discussed in Section 2.2, this reduction in probability generally comes at the cost of higher pricing errors on the test assets (a higher $\tilde{\delta}$). For models with large differences between $\hat{\delta}_+$ and $\hat{\delta}$ (such as LV, LVX1, and LVX2), we also see a significant difference between $\hat{\tilde{\delta}}$ and $\hat{\delta}$, indicating a substantial deterioration in the ability of the SDF to price the test assets when its parameters are chosen to minimize the constrained HJ-distance instead of the unconstrained HJ-distance. The deterioration in the pricing ability of LV, LVX1, and LVX2 is also reflected in the standard deviations of their SDFs, which significantly drop from 0.817, 1.229, and 1.478 in Panel A to 0.294, 0.318, and 0.350 in Panel B. This implies that it would be even harder for these models to satisfy the sample HJ-bounds if their parameters were chosen to minimize the constrained HJ-distance.

Second, the variation in the differences between the sample unconstrained and constrained HJ-distances across models deserves some remarks. As expected from our theoretical analysis, the differences are relatively small for SDFs that are close to being spanned by the returns on the test assets. For example, the percentage difference between the sample constrained and unconstrained HJ-distances of FF3 is only 4.2% since this model has a very high R^2 of 0.983. While we do not report the parameter estimates of the various models to preserve space, the pattern of the differences in the parameter estimates provides further support to our theoretical predictions: the largest differences in parameter estimates arise in models with non-traded factors and almost no differences arise in models with traded factors.[9] As a result, the different effects of imposing the no-arbitrage constraint across models are driven by the underlying structure of the problem and

[9]The full set of parameter estimates is available upon request.

characteristics of the factors (traded versus non-traded).

Our theoretical results suggest that we can expect some meaningful differences between the unconstrained and constrained HJ-distances only when the Sharpe ratio of the tangency portfolio of the risky assets is very high. As it turns out, the tangency portfolio of the 25 Fama-French size and book-to-market portfolios used by Li, Xu and Zhang (2010) has a relatively high sample Sharpe ratio (0.71). To understand whether this is an important reason for the difference in results between the two HJ-distances, we consider another set of test assets with a smaller sample Sharpe ratio. Specifically, we use quarterly gross returns on the three-month T-bill and 10 size and 12 industry portfolios from Kenneth French's website.[10] The sample Sharpe ratio of the tangency portfolio of this new set of risky assets is 0.53. 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 2.

Table 2 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 unconstrained HJ-distance of each model is now substantially smaller. As an example, consider LVX2. For this model, the difference between $\hat{\delta}_+$ and $\hat{\delta}$ is only 3.9%, a much smaller number than the 25.2% difference reported in Table 1.

4. Conclusion

This paper examines the population 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

[10] Considering size and industry portfolios in addition to or instead of the 25 Fama-French size and book-to-market portfolios is consistent with one of the prescriptions of Lewellen, Nagel and Shanken (2010).

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.

Appendix A: Definitions and Preliminary Lemmas

A.1 Elliptical distributions: Definitions and notation

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

$$\varphi(t) = \exp(it'\mu)\psi\left(\frac{t'St}{2}\right) \tag{A1}$$

for some

$$\mu = \begin{bmatrix} \mu_u \\ \mu_v \end{bmatrix}, \qquad S = \begin{bmatrix} s_u^2 & s_{uv} \\ s_{vu} & s_v^2 \end{bmatrix}, \tag{A2}$$

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

$$f_v(v) = \frac{c_1}{s_v} h\left(\frac{(v-\mu_v)^2}{2s_v^2}\right), \tag{A3}$$

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

$$c_1 = \frac{1}{\sqrt{2}}\left[\int_0^\infty x^{-\frac{1}{2}} h(x)\mathrm{d}x\right]^{-1} \tag{A4}$$

is a normalization constant.

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

$$\psi(s) = e^{-s}, \quad (A5)$$

$$h(t) = e^{-t}, \quad (A6)$$

$$c_1 = \frac{1}{\sqrt{2\pi}}, \quad (A7)$$

$$f_v(v) = \frac{1}{\sqrt{2\pi}s_v} e^{-\frac{(v-\mu_v)^2}{2s_v^2}}. \quad (A8)$$

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 ν degrees of freedom, we have

$$\psi(s) = \frac{K_{\nu/2}(\sqrt{2\nu s})\left(\frac{\nu s}{2}\right)^{\frac{\nu}{4}}}{2\Gamma\left(\frac{\nu}{2}\right)}, \quad (A9)$$

$$h(t) = \left(1 + \frac{2t}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (A10)$$

$$c_1 = \frac{1}{B\left(\frac{1}{2}, \frac{\nu}{2}\right)\sqrt{\nu}}, \quad (A11)$$

$$f_v(v) = \frac{1}{B\left(\frac{1}{2}, \frac{\nu}{2}\right)\sqrt{\nu}s_v}\left[1 + \frac{(v-\mu_v)^2}{\nu s_v^2}\right]^{-\frac{\nu+1}{2}}, \quad (A12)$$

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

$$\psi'(s) = -\frac{\nu K_{(\nu-2)/2}(\sqrt{2\nu s})\left(\frac{\nu s}{2}\right)^{\frac{\nu-2}{4}}}{\Gamma\left(\frac{\nu}{2}\right)}, \quad (A13)$$

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 μ_v and s_v^2, we define

$$\tilde{v} = \frac{v - \mu_v}{s_v}. \quad (A14)$$

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

$$f(\tilde{v}) = c_1 h\left(\frac{\tilde{v}^2}{2}\right). \quad (A15)$$

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

For every spherical random variable $\tilde{v}$ with finite variance, Landsman and Valdez (2003) show that a random variable w with the following density function[11]

$$\tilde{f}(w) = \int_w^\infty c\tilde{v} f(c\tilde{v}) \mathrm{d}\tilde{v} = \frac{1}{c} \int_{cw}^\infty s f(s) \mathrm{d}s \tag{A16}$$

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

$$\tilde{f}(w) = \frac{c_1}{c} H\left(\frac{c^2 w^2}{2}\right), \tag{A17}$$

where

$$H(x) = \int_x^\infty h(t) \mathrm{d}t. \tag{A18}$$

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

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

$$H(x) = \int_x^\infty e^{-t} \mathrm{d}t = e^{-x}. \tag{A19}$$

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

$$\tilde{f}(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \tag{A20}$$

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

When $\tilde{v} \sim t_\nu$ for $\nu > 2$, we use (A10) to obtain

$$H(x) = \int_x^\infty \left(1 + \frac{2t}{\nu}\right)^{-\frac{\nu+1}{2}} \mathrm{d}t = \frac{\nu}{\nu - 1} \left(1 + \frac{2x}{\nu}\right)^{-\frac{\nu-1}{2}}. \tag{A21}$$

[11] Instead of mapping $\tilde{v}$ to w, Landsman and Valdez (2003) define a slightly different mapping from $\tilde{v}$ to $Z^* = cw$.

Then using (A11) and $c = [\nu/(\nu-2)]^{\frac{1}{2}}$, we obtain

$$\begin{aligned}\tilde{f}(w) &= \frac{1}{\sqrt{\nu}B\left(\frac{1}{2},\frac{\nu}{2}\right)}\frac{\sqrt{\nu-2}}{\sqrt{\nu}}\frac{\nu}{\nu-1}\left(1+\frac{w^2}{\nu-2}\right)^{-\frac{\nu-1}{2}} \\ &= \frac{1}{\sqrt{\nu-2}B\left(\frac{1}{2},\frac{\nu-2}{2}\right)}\left(1+\frac{w^2}{\nu-2}\right)^{-\frac{\nu-1}{2}}, \qquad \text{(A22)}\end{aligned}$$

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

A.2 Preliminary lemmas

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

$$E[v^+] = \mu_v F(c\eta) + \sigma_v \tilde{f}(\eta) = \tilde{F}(\eta)\sigma_v g(\eta), \qquad \text{(A23)}$$

$$\begin{aligned}E[uv^+] &= \sigma_{uv}\tilde{F}(\eta) + \mu_u[\mu_v F(c\eta) + \sigma_v \tilde{f}(\eta)] \\ &= \tilde{F}(\eta)\left(E[uv] + \mu_u\sigma_v[g(\eta)-\eta]\right), \qquad \text{(A24)}\end{aligned}$$

where F is the cumulative distribution function of $\tilde{v} = (v-\mu_v)/s_v$, $c = \sigma_{\tilde{v}}$, $\tilde{f}$ and $\tilde{F}$ are the density and cumulative distribution functions of another spherical random variable w that is associated with $\tilde{v}$ as defined in (A16), and

$$g(\eta) = \frac{\eta F(c\eta) + \tilde{f}(\eta)}{\tilde{F}(\eta)}. \qquad \text{(A25)}$$

Proof of Lemma A.1. For a given f, we define the functions

$$\tilde{h}(x) = \int_{-x}^{\infty} \tilde{v} f(\tilde{v})\mathrm{d}\tilde{v}, \qquad \text{(A26)}$$

$$\tilde{H}(x) = \int_{-x}^{\infty} \tilde{v}^2 f(\tilde{v})\mathrm{d}\tilde{v}. \qquad \text{(A27)}$$

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

$$E[v^+] = \int_0^{\infty} v f_v(v)\mathrm{d}v = \int_{-c\eta}^{\infty}(\mu_v + s_v\tilde{v})f(\tilde{v})\mathrm{d}\tilde{v} = \mu_v F(c\eta) + s_v\tilde{h}(c\eta) = \mu_v F(c\eta) + \sigma_v\tilde{f}(\eta), \quad \text{(A28)}$$

where the last equality follows from (A16) and the fact that $\tilde{h}(c\eta) = c\tilde{f}(-\eta) = c\tilde{f}(\eta)$.

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

$$E[v^{+2}] = \int_0^\infty v^2 f_v(v)\mathrm{d}v = \int_{-c\eta}^\infty (\mu_v + s_v\tilde{v})^2 f(\tilde{v})\mathrm{d}\tilde{v} = \mu_v^2 F(c\eta) + 2\mu_v s_v \tilde{h}(c\eta) + s_v^2 \tilde{H}(c\eta). \quad \text{(A29)}$$

Since

$$\frac{\mathrm{d}\tilde{h}(\eta)}{\mathrm{d}\eta} = \eta f(-\eta) = \eta f(\eta), \quad \text{(A30)}$$

we can use integration by parts to obtain

$$\begin{aligned}
\tilde{H}(c\eta) &= \int_{-c\eta}^\infty \tilde{v}^2 f(\tilde{v})\mathrm{d}\tilde{v} \\
&= \tilde{v}\tilde{h}(\tilde{v})\Big|_{-c\eta}^\infty - \int_{-c\eta}^\infty \tilde{h}(\tilde{v})\mathrm{d}\tilde{v} \\
&= d - c\eta\tilde{h}(c\eta) - c\int_{-c\eta}^\infty \tilde{f}\left(\frac{\tilde{v}}{c}\right)\mathrm{d}\tilde{v} \\
&= d - c^2\eta\tilde{f}(\eta) - c^2\int_{-\eta}^\infty \tilde{f}(s)\mathrm{d}s \\
&= d - c^2\eta\tilde{f}(\eta) + c^2\tilde{F}(\eta), \quad \text{(A31)}
\end{aligned}$$

where $d \equiv \lim_{\tilde{v}\to\infty} \tilde{v}\tilde{h}(\tilde{v})$. We now show that $d = 0$ when $c < \infty$. Since w is a symmetric random variable, $\tilde{F}(0) = 1/2$ and it follows that

$$\tilde{H}(0) = d + c^2\tilde{F}(0) = d + \frac{c^2}{2}. \quad \text{(A32)}$$

However, we know that

$$\tilde{H}(0) = \int_0^\infty \tilde{v}^2 f(\tilde{v})\mathrm{d}\tilde{v} = \frac{c^2}{2}, \quad \text{(A33)}$$

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

$$\tilde{H}(c\eta) = -c^2\eta\tilde{f}(c\eta) + c^2\tilde{F}(\eta). \quad \text{(A34)}$$

Using (A29) and $\tilde{h}(c\eta) = c\tilde{f}(\eta)$, we have

$$\begin{aligned}
E[v^{+2}] &= \mu_v^2 F(c\eta) + 2\mu_v s_v \tilde{h}(c\eta) - s_v^2 c^2 \eta\tilde{f}(\eta) + s_v^2 c^2 \tilde{F}(\eta) \\
&= \mu_v^2 F(c\eta) + \mu_v\sigma_v\tilde{f}(\eta) + \sigma_v^2\tilde{F}(\eta). \quad \text{(A35)}
\end{aligned}$$

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

$$E[u|v] = \mu_u + \frac{\sigma_{uv}}{\sigma_v^2}(v - \mu_v). \tag{A36}$$

It then follows that

$$\begin{aligned} E[uv^+] &= E[E[u|v]v^+] \\ &= E\left[\left(\mu_u + \frac{\sigma_{uv}}{\sigma_v^2}(v-\mu_v)\right)v^+\right] \\ &= \left(\mu_u - \frac{\sigma_{uv}}{\sigma_v^2}\mu_v\right)E[v^+] + \frac{\sigma_{uv}}{\sigma_v^2}E[v^{+2}] \\ &= \left(\mu_u - \frac{\sigma_{uv}}{\sigma_v^2}\mu_v\right)[\mu_v F(c\eta) + \sigma_v \tilde{f}(\eta)] + \frac{\sigma_{uv}}{\sigma_v^2}[\mu_v^2 F(c\eta) + \sigma_v^2 \tilde{F}(\eta) + \mu_v\sigma_v\tilde{f}(\eta)] \\ &= \sigma_{uv}\tilde{F}(\eta) + \mu_u[\mu_v F(c\eta) + \sigma_v\tilde{f}(\eta)]. \end{aligned} \tag{A37}$$

This completes the proof.

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

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

$$\text{Var}[x] = \frac{c^2}{F(cu)}\left[\tilde{F}(u) - u\tilde{f}(u) - \frac{\tilde{f}(u)^2}{F(cu)}\right], \tag{A38}$$

where $c = \sigma_{\tilde{v}}$, and $\tilde{f}$ and $\tilde{F}$ are the density and cumulative distribution functions of another elliptical random variable w that is associated with $\tilde{v}$ as defined in (A16).

Proof of Lemma A.2. Using the fact that

$$\frac{\partial c\tilde{f}(x/c)}{\partial x} = -xf(x), \tag{A39}$$

we can easily obtain

$$E[x] = \frac{1}{F(cu)}\int_{-\infty}^{cu} xf(x)\mathrm{d}q = -\frac{c}{F(cu)}\,\tilde{f}\left(\frac{x}{c}\right)\Big|_{-\infty}^{cu} = -\frac{c\tilde{f}(u)}{F(cu)}. \tag{A40}$$

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

$$\begin{aligned} E[x^2] &= \frac{1}{F(cu)}\int_{-\infty}^{cu} x^2 f(x)\mathrm{d}x \\ &= \frac{1}{F(cu)}\left[-cx\tilde{f}\left(\frac{x}{c}\right)\Big|_{-\infty}^{cu} + \int_{-\infty}^{cu} c\tilde{f}\left(\frac{x}{c}\right)\mathrm{d}q\right] \\ &= \frac{1}{F(cu)}[-c^2u\tilde{f}(u) + c^2\tilde{F}(u)]. \end{aligned} \tag{A41}$$

Then, the variance of x is given by (A38). This completes the proof.

Appendix B: Proofs of Main Results

Proof of Proposition 1. Since $m_y = y - \tilde{\lambda}'\tilde{r}$ follows an elliptical distribution with mean μ_m and variance σ_m^2, we can invoke Lemma A.1 to obtain

$$E[\tilde{r}m_y^+] = \tilde{F}(\eta)(E[\tilde{r}y] - U\tilde{\lambda} + E[\tilde{r}]\sigma_m[g(\eta) - \eta]), \tag{B1}$$

where $\eta = \mu_m/\sigma_m$.

Using the first order condition $E[\tilde{r}m_y^+] = q$ and the expression of U^{-1} in (37), we obtain $U^{-1}E[\tilde{r}] = q/R_0$ and

$$\begin{aligned} \tilde{\lambda} &= U^{-1}E[\tilde{r}y] + \sigma_m[g(\eta) - \eta]U^{-1}E[\tilde{r}] - \frac{1}{\tilde{F}(\eta)}U^{-1}q \\ &= \begin{bmatrix} \frac{\mu_y - V_{ry}'V_{rr}^{-1}\mu_r}{R_0} \\ V_{rr}^{-1}V_{ry} \end{bmatrix} + \begin{bmatrix} \frac{\sigma_m[g(\eta)-\eta]}{R_0} \\ 0_N \end{bmatrix} - \frac{1}{\tilde{F}(\eta)}\begin{bmatrix} \frac{1+a}{R_0^2} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} \end{bmatrix}. \end{aligned} \tag{B2}$$

Using (B2) and after simplification, we have

$$m_y = y - \tilde{\lambda}'\tilde{r} = z + \frac{1}{\tilde{F}(\eta)}\left[\frac{1}{R_0} - \frac{\mu_r'V_{rr}^{-1}(r - \mu_r)}{R_0}\right] - \sigma_m[g(\eta) - \eta], \tag{B3}$$

and the variance of m_y is given by

$$\sigma_m^2 = \sigma_z^2 + \frac{a}{R_0^2\tilde{F}(\eta)^2}. \tag{B4}$$

With (B2) and (B4), we can see that once η is identified, $\tilde{\lambda}$ will be uniquely determined and there is no need to solve n nonlinear equations to obtain $\tilde{\lambda}$.

Now, using the fact that m_y^+ prices the risk-free asset correctly, we have $E[m_y^+] = 1/R_0$. Then using (A23) to express $E[m_y^+] = \tilde{F}(\eta)\sigma_m g(\eta)$, we obtain

$$g(\eta) = \frac{1}{\sigma_m R_0 \tilde{F}(\eta)}. \tag{B5}$$

Substituting σ_m from (B4) into this expression, we can see that η satisfies the following first order condition:

$$g(u) = [a + \sigma_z^2 R_0^2 \tilde{F}(u)^2]^{-\frac{1}{2}}. \tag{B6}$$

For establishing the uniqueness of the solution to equation (B6), 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 (A23). Condition (ii) follows from the definition of g. For condition (iii), it is convenient to write

$$g(u) = \frac{uF(cu) + \tilde{f}(u)}{\tilde{F}(u)} = \frac{g_1(u)}{g_2(u)}. \tag{B7}$$

Then,

$$g_1'(u) = F(cu), \tag{B8}$$

$$g_2'(u) = \tilde{f}(u), \tag{B9}$$

$$g_1''(u) = cf(cu), \tag{B10}$$

$$g_2''(u) = -cuf(cu). \tag{B11}$$

Using L'Hôpital's rule twice, we have

$$\lim_{u\to-\infty} g(u) = \lim_{u\to-\infty} \frac{g_1''(u)}{g_2''(u)} = \lim_{u\to-\infty} \frac{cf(cu)}{-cuf(cu)} = \lim_{u\to-\infty} -\frac{1}{u} = 0. \tag{B12}$$

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

$$g'(u) = \frac{F(cu)}{\tilde{F}(u)^2}\left[\tilde{F}(u) - u\tilde{f}(u) - \frac{\tilde{f}(u)^2}{F(cu)}\right] > 0, \tag{B13}$$

where the inequality is obtained by using (A38) and the fact that $\text{Var}[x] > 0$. Since the left hand side of (B6) is positive and increasing in η, and the right hand side of (B6) is positive and decreasing in η, the solution to (B6) is unique.

Using (B5), we can express the vector of Lagrange multipliers as

$$\tilde{\lambda} = \begin{bmatrix} \frac{\mu_y - V_{ry}' V_{rr}^{-1}\mu_r}{R_0} + \frac{1-\frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} - \frac{1+a}{R_0^2 \tilde{F}(\eta)} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0 \tilde{F}(\eta)}\right) \end{bmatrix} = \begin{bmatrix} \frac{\mu_y - V_{ry}' V_{rr}^{-1}\mu_r}{R_0} - \frac{a+\frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} \\ V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0 \tilde{F}(\eta)}\right) \end{bmatrix}. \tag{B14}$$

The nonnegative admissible SDF that is closest to y is m_y^+, where m_y is defined in (B3). It follows that the squared constrained HJ-distance of y is given by

$$\delta_+^2 = E[(y - m_y^+)^2] = E[y^2] - 2E[y m_y^+] + E[m_y^{+2}]. \tag{B15}$$

It is straightforward to show that

$$E[y^2] = \sigma_z^2 + \mu_y^2 + V_{ry}' V_{rr}^{-1} V_{ry}. \tag{B16}$$

Using (A24) and the fact that $E[m_y^+] = 1/R_0$, we obtain

$$E[y m_y^+] = \left(\sigma_z^2 - \frac{V_{ry}' V_{rr}^{-1}\mu_r}{R_0 \tilde{F}(\eta)}\right)\tilde{F}(\eta) + \frac{\mu_y}{R_0}, \tag{B17}$$

$$E[m_y^{+2}] = E[m_y m_y^+] = \sigma_m^2 \tilde{F}(\eta) + \frac{\mu_m}{R_0}. \tag{B18}$$

With these expressions, we obtain

$$\begin{aligned} \delta_+^2 &= \sigma_z^2 + \mu_y^2 + V_{ry}' V_{rr}^{-1} V_{ry} - 2\left(\sigma_z^2 \tilde{F}(\eta) - \frac{V_{ry}' V_{rr}^{-1}\mu_r}{R_0}\right) - \frac{2\mu_y}{R_0} + \sigma_m^2 \tilde{F}(\eta) + \frac{\mu_m}{R_0} \\ &= \left(\mu_y - \frac{1}{R_0}\right)^2 + \left(V_{ry} + \frac{\mu_r}{R_0}\right)' V_{rr}^{-1}\left(V_{ry} + \frac{\mu_r}{R_0}\right) \\ &\quad + \sigma_z^2 - 2\sigma_z^2 \tilde{F}(\eta) + \sigma_m^2 \tilde{F}(\eta) + \frac{\mu_m}{R_0} - \frac{1+a}{R_0^2} \\ &= \delta^2 + \sigma_z^2[1 - \tilde{F}(\eta)] + (\sigma_m^2 - \sigma_z^2)\tilde{F}(\eta) + \frac{\eta\sigma_m}{R_0} - \frac{1+a}{R_0^2} \\ &= \delta^2 + \sigma_z^2 \tilde{F}(-\eta) + \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} - \frac{1+a}{R_0^2}, \end{aligned} \tag{B19}$$

where the last equality follows from (B4) and (B5). This completes the proof.

Proof of Lemma 1. We first show that $\partial\Delta/\partial\eta = 0$.

$$\begin{aligned}
\frac{\partial\Delta}{\partial\eta} &= -\sigma_z^2\tilde{f}(-\eta) - \frac{a\tilde{f}(\eta)}{R_0^2\tilde{F}(\eta)^2} + \frac{R_0^2\tilde{F}(\eta)g(\eta) - \eta[R_0^2\tilde{F}(\eta)g'(\eta) + R_0^2\tilde{f}(\eta)g(\eta)]}{R_0^4\tilde{F}(\eta)^2g(\eta)^2} \\
&= -\sigma_z^2\tilde{f}(\eta) - \frac{a\tilde{f}(\eta)}{R_0^2\tilde{F}(\eta)^2} + \frac{1}{R_0^2g(\eta)\tilde{F}(\eta)} - \frac{\eta g'(\eta)}{R_0^2g(\eta)^2\tilde{F}(\eta)} - \frac{\eta\tilde{f}(\eta)}{R_0^2g(\eta)\tilde{F}(\eta)^2} \\
&= -\frac{\tilde{f}(\eta)}{R_0^2\tilde{F}(\eta)^2}[a + \sigma_z^2R_0^2\tilde{F}(\eta)^2] + \frac{1}{R_0^2g(\eta)\tilde{F}(\eta)} - \frac{\eta g'(\eta)}{R_0^2g(\eta)^2\tilde{F}(\eta)} - \frac{\eta\tilde{f}(\eta)}{R_0^2g(\eta)\tilde{F}(\eta)^2} \\
&= -\frac{\tilde{f}(\eta)}{R_0^2\tilde{F}(\eta)^2g(\eta)^2} + \frac{g(\eta)\tilde{F}(\eta) - \eta F(c\eta)}{R_0^2\tilde{F}(\eta)^2g(\eta)^2} \\
&= -\frac{\tilde{f}(\eta)}{R_0^2\tilde{F}(\eta)^2g(\eta)^2} + \frac{\tilde{f}(\eta)}{R_0^2\tilde{F}(\eta)^2g(\eta)^2} \\
&= 0. \qquad \text{(B20)}
\end{aligned}$$

The fourth equality follows because

$$a + \sigma_z^2R_0^2\tilde{F}(\eta)^2 = \frac{1}{g(\eta)^2} \qquad \text{(B21)}$$

and

$$g'(\eta) = \frac{F(c\eta) - g(\eta)\tilde{f}(\eta)}{\tilde{F}(\eta)}, \qquad \text{(B22)}$$

which can be easily verified by using (A25) and (B13). This suggests that as far as the partial derivatives of Δ with respect to $(\sigma_z^2,\ a,\ R_0)$ are concerned, we can treat η as a constant. Thus, it follows that

$$\begin{aligned}
\frac{\partial\Delta}{\partial\sigma_z^2} &= \tilde{F}(-\eta) > 0, \qquad \text{(B23)} \\
\frac{\partial\Delta}{\partial a} &= \frac{1}{R_0^2\tilde{F}(\eta)} - \frac{1}{R_0^2} = \frac{\tilde{F}(-\eta)}{R_0^2\tilde{F}(\eta)} > 0, \qquad \text{(B24)} \\
\frac{\partial\Delta}{\partial R_0} &= \frac{2}{R_0^3}\left[1 - \frac{a\tilde{F}(-\eta)}{\tilde{F}(\eta)} - \frac{\eta}{\tilde{F}(\eta)g(\eta)}\right]. \qquad \text{(B25)}
\end{aligned}$$

This completes the proof.

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

$$\delta_+^2 = E[\tilde{y}^2] - E[m_{\tilde{y}}^{+2}] - 2\tilde{\lambda}'q \qquad \text{(B26)}$$

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

$$C\tilde{\gamma} - E[\tilde{f} m_{\tilde{y}}^{+}] = 0_{K+1}, \tag{B27}$$

$$E[\tilde{r} m_{\tilde{y}}^{+}] = q. \tag{B28}$$

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

$$E[\tilde{f} m_{\tilde{y}}^{+}] = \tilde{F}(\eta)(C\tilde{\gamma} - D'\tilde{\lambda} + \sigma_m[g(\eta) - \eta]E[\tilde{f}]), \tag{B29}$$

$$E[\tilde{r} m_{\tilde{y}}^{+}] = \tilde{F}(\eta)(D\tilde{\gamma} - U\tilde{\lambda} + \sigma_m[g(\eta) - \eta]E[\tilde{r}]), \tag{B30}$$

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

$$\begin{bmatrix} \tilde{F}(-\eta)C & \tilde{F}(\eta)D' \\ \tilde{F}(\eta)D & -\tilde{F}(\eta)U \end{bmatrix} \begin{bmatrix} \tilde{\gamma} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} \tilde{F}(\eta)\sigma_m[g(\eta) - \eta]E[\tilde{f}] \\ q - \tilde{F}(\eta)\sigma_m[g(\eta) - \eta]E[\tilde{r}] \end{bmatrix}. \tag{B31}$$

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

$$\begin{bmatrix} \tilde{F}(-\eta)C & \tilde{F}(\eta)D' \\ \tilde{F}(\eta)D & -\tilde{F}(\eta)U \end{bmatrix}^{-1} = \begin{bmatrix} H & HD'U^{-1} \\ U^{-1}DH & -\frac{1}{\tilde{F}(\eta)}U^{-1} + U^{-1}DHD'U^{-1} \end{bmatrix}. \tag{B32}$$

Using (37), we can easily verify that $U^{-1}E[\tilde{r}] = q/R_0$ and hence

$$D'U^{-1}E[\tilde{r}] = \frac{1}{R_0}D'q = \frac{1}{R_0}E[R_0\tilde{f}] = E[\tilde{f}]. \tag{B33}$$

Using this identity, we can then show that

$$\tilde{\gamma} = HD'U^{-1}q. \tag{B34}$$

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

$$H = \begin{bmatrix} 1 + \mu_f' P \mu_f & -\mu_f' P \\ -P\mu_f & P \end{bmatrix}, \tag{B35}$$

where $P = [V_{ff} - \tilde{F}(\eta)V_{ff\cdot r}]^{-1}$. Using this expression and (37), we can then rewrite (B34) as

$$\tilde{\gamma} = \begin{bmatrix} 1 + \mu_f' P \mu_f & -\mu_f' P \\ -P\mu_f & P \end{bmatrix} \begin{bmatrix} R_0 & \mu_r' \\ \mu_f R_0 & V_{rf}' + \mu_f \mu_r' \end{bmatrix} \begin{bmatrix} \frac{(1+a)}{R_0^2} \\ -\frac{V_{rr}^{-1}\mu_r}{R_0} \end{bmatrix}. \tag{B36}$$

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

$$\tilde{\gamma}_1 = -\frac{1}{R_0} P V_{rf}' V_{rr}^{-1} \mu_r, \qquad \tilde{\gamma}_0 = \frac{1}{R_0} - \tilde{\gamma}_1' \mu_f. \tag{B37}$$

As a result, we can write $\tilde{y} = \frac{1}{R_0} + \tilde{\gamma}_1'(f - \mu_f)$. Since the expression of $\tilde{\lambda}$ in (46) also works for $\tilde{y}$, we can use $\mu_{\tilde{y}} = 1/R_0$, and $V_{r\tilde{y}}' V_{rr}^{-1} \mu_r = \tilde{\gamma}_1' V_{rf}' V_{rr}^{-1} \mu_r = -\tilde{a}_1/R_0$ to obtain

$$\tilde{\lambda} = \begin{bmatrix} \frac{1+\tilde{a}_1}{R_0^2} - \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} \\ V_{rr}^{-1} \left(V_{rf} \tilde{\gamma}_1 + \frac{\mu_r}{R_0 \tilde{F}(\eta)} \right) \end{bmatrix}. \tag{B38}$$

Note that we only need to solve for η 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

$$m_{\tilde{y}} = \tilde{\gamma}' \tilde{f} - \tilde{\lambda}' \tilde{r} = \tilde{\gamma}_1' \epsilon - \frac{\mu_r' V_{rr}^{-1}(r - \mu_r)}{R_0 \tilde{F}(\eta)} + \frac{\eta}{R_0 \tilde{F}(\eta) g(\eta)}. \tag{B39}$$

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

$$\sigma_m^2 = \mathrm{Var}[m_{\tilde{y}}] = \tilde{\gamma}_1' V_{ff\cdot r} \tilde{\gamma}_1 + \frac{a}{R_0^2 \tilde{F}(\eta)^2}. \tag{B40}$$

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

$$E[m_{\tilde{y}}^+] = \tilde{F}(\eta) \sigma_m g(\eta) = \frac{1}{R_0}. \tag{B41}$$

Then, plugging the expression of σ_m^2 from (B40) into (B41), we obtain

$$g(\eta) = \left[a + \tilde{\gamma}_1' V_{ff\cdot r} \tilde{\gamma}_1 R_0^2 \tilde{F}(\eta)^2 \right]^{-\frac{1}{2}}. \tag{B42}$$

Using the expression for $\tilde{\gamma}_1$ in (B37) and rearranging terms, we can see that η satisfies the following equation:

$$g(u) = \left(a + \alpha' \left[\frac{1}{\tilde{F}(u)} I_K - V_{ff}^{-\frac{1}{2}} V_{ff\cdot r} V_{ff}^{-\frac{1}{2}} \right]^{-2} \alpha \right)^{-\frac{1}{2}}, \tag{B43}$$

where $\alpha = V_{ff}^{-1} V_{ff\cdot r}^{\frac{1}{2}} V_{rf}' V_{rr}^{-1} \mu_r$. Since the left hand side is positive and increasing in u and the right hand side 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), (B43) has a unique solution. Using (A24), it is straightforward to obtain

$$E[\tilde{y} m_{\tilde{y}}^+] = \mathrm{Cov}[\tilde{y}, m_{\tilde{y}}] \tilde{F}(\eta) + \frac{1}{R_0^2} = \left[\tilde{\gamma}_1' V_{ff\cdot r} \tilde{\gamma}_1 + \frac{\tilde{a}_1}{R_0^2 \tilde{F}(\eta)} \right] \tilde{F}(\eta) + \frac{1}{R_0^2}, \tag{B44}$$

$$E[m_{\tilde{y}}^{+2}] = E[m_{\tilde{y}} m_{\tilde{y}}^+] = \sigma_m^2 \tilde{F}(\eta) + \frac{\mu_m}{R_0}. \tag{B45}$$

The squared constrained HJ-distance is then given by

$$
\begin{aligned}
\delta_+^2 &= E[(\tilde{y} - m_{\tilde{y}}^+)^2] \\
&= E[\tilde{y}^2] - 2E[\tilde{y} m_{\tilde{y}}^+] + E[m_{\tilde{y}}^{+2}] \\
&= \frac{1}{R_0^2} + \tilde{\gamma}_1' V_{ff} \tilde{\gamma}_1 - 2\left[\tilde{\gamma}_1' V_{ff\cdot r} \tilde{\gamma}_1 + \frac{\tilde{a}_1}{R_0^2 \tilde{F}(\eta)}\right] \tilde{F}(\eta) - \frac{2}{R_0^2} + \sigma_m^2 \tilde{F}(\eta) + \frac{\mu_m}{R_0} \\
&= \tilde{\gamma}_1' V_{ff} \tilde{\gamma}_1 - 2\tilde{\gamma}_1' V_{ff\cdot r} \tilde{\gamma}_1 \tilde{F}(\eta) - \frac{1 + 2\tilde{a}_1}{R_0^2} + \left[\tilde{\gamma}_1' V_{ff\cdot r} \tilde{\gamma}_1 + \frac{a}{R_0^2 \tilde{F}(\eta)^2}\right] \tilde{F}(\eta) + \frac{\eta \sigma_m}{R_0} \\
&= \tilde{\gamma}_1' [V_{ff} - \tilde{F}(\eta) V_{ff\cdot r}] \tilde{\gamma}_1 - \frac{1 + 2\tilde{a}_1}{R_0^2} + \frac{a}{R_0^2 \tilde{F}(\eta)} + \frac{\eta}{R_0^2 \tilde{F}(\eta) g(\eta)} \\
&= \frac{a + \frac{\eta}{g(\eta)}}{R_0^2 \tilde{F}(\eta)} - \frac{1 + \tilde{a}_1}{R_0^2},
\end{aligned}
\tag{B46}
$$

where the second last equality is obtained by using $\sigma_m = 1/[R_0 \tilde{F}(\eta) g(\eta)]$ from (B41). Finally, we can easily see that the first element of $\tilde{\lambda}$ in (B38) is equal to $-\delta_+^2$. This completes the proof.

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

$$
P[y < 0] = F\left(-\frac{c\mu_y}{\sigma_y}\right) = F\left(-\frac{c}{R_0 \sigma_y}\right) = F\left(-\frac{c}{R_0 \sqrt{\gamma_1' V_{ff} \gamma_1}}\right), \tag{B47}
$$

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

$$
P[\tilde{y} < 0] = F\left(-\frac{c\mu_{\tilde{y}}}{\sigma_{\tilde{y}}}\right) = F\left(-\frac{c}{R_0 \sigma_{\tilde{y}}}\right) = F\left(-\frac{c}{R_0 \sqrt{\tilde{\gamma}_1' V_{ff} \tilde{\gamma}_1}}\right), \tag{B48}
$$

where $\mu_{\tilde{y}}$ and $\sigma_{\tilde{y}}$ are the mean and standard deviation of the SDF $\tilde{y}$, respectively. The inequality holds because

$$
\begin{aligned}
R_0^2 \tilde{\gamma}_1' V_{ff} \tilde{\gamma}_1 &= \mu_r' V_{rr}^{-1} V_{rf} (V_{ff} - \tilde{F}(\eta) V_{ff\cdot r})^{-1} V_{ff} (V_{ff} - \tilde{F}(\eta) V_{ff\cdot r})^{-1} V_{rf}' V_{rr}^{-1} \mu_r \\
&= \mu_r' V_{rr}^{-1} V_{rf} V_{ff}^{\frac{1}{2}} \left(I_K - \tilde{F}(\eta) V_{ff}^{-\frac{1}{2}} V_{ff\cdot r} V_{ff}^{-\frac{1}{2}}\right)^{-2} V_{ff}^{\frac{1}{2}} V_{rf}' V_{rr}^{-1} \mu_r \\
&\leq \mu_r' V_{rr}^{-1} V_{rf} V_{ff}^{\frac{1}{2}} (I_K - V_{ff}^{-\frac{1}{2}} V_{ff\cdot r} V_{ff}^{-\frac{1}{2}})^{-2} V_{ff}^{\frac{1}{2}} V_{rf}' V_{rr}^{-1} \mu_r \\
&= \mu_r' V_{rr}^{-1} V_{rf} (V_{rf}' V_{rr}^{-1} V_{rf})^{-1} V_{ff} (V_{rf}' V_{rr}^{-1} V_{rf})^{-1} V_{rf}' V_{rr}^{-1} \mu_r \\
&= R_0^2 \gamma_1' V_{ff} \gamma_1.
\end{aligned}
\tag{B49}
$$

For (66), the first inequality, $\delta^2 \le \tilde{\delta}^2$, is obvious since γ is chosen to minimize $\delta^2 = e'U^{-1}e$ but $\tilde{\gamma}$ is not. For the second inequality, $\tilde{\delta}^2 \le \delta_+^2$, note that for every $h \in L^2$ with $E[h^2] = 1$, we have

$$\min_{m \in \mathcal{M}^+} (E[\tilde{y}h] - E[mh])^2 \le \delta_+^2. \tag{B50}$$

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

$$w^* = \frac{U^{-1}\tilde{e}}{(\tilde{e}'U^{-1}\tilde{e})^{\frac{1}{2}}}. \tag{B51}$$

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

$$\delta_+^2 \ge \inf_{m \in \mathcal{M}^+} (E[\tilde{y}h] - E[mh])^2 = (E[\tilde{y}h] - E[mh])^2 = (w^{*\prime}(E[\tilde{y}\tilde{r}] - q))^2 = (w^{*\prime}\tilde{e})^2 = \tilde{\delta}^2. \tag{B52}$$

This completes the proof.

References

Bailey, Warren, Haitao Li, and Xiaoyan Zhang, 2004, Hedge fund performance evaluation: A stochastic discount factor approach, Working paper, Cornell University.

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.

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

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

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.

Fletcher, Jonathan, 2010, Arbitrage and the evaluation of linear factor models in U.K. stock returns, *Financial Review* 45, 449–468.

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, 2008a, The exact distribution of the Hansen-Jagannathan bound, Working paper, University of Toronto.

Kan, Raymond, and Cesare Robotti, 2008b, 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 Guofu Zhou, 2003, Hansen-Jagannathan distance: Geometry and exact distribution, Working paper, University of Toronto.

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.

Lewellen, Jonathan W., Stefan Nagel, and Jay A. Shanken, 2010, A skeptical appraisal of asset-pricing tests, *Journal of Financial Economics* 96, 175–194.

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.

Liu, Xiaoquan, Jing-Ming Kuo, and Jerry Coakley, 2009, A pricing kernel approach to valuing interest rate options, Working paper, University of Essex.

Lustig, Hanno, and Stijn Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia, *Journal of Finance* 60, 1167–1221.

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.

Santos, Tano, and Pietro Veronesi, 2006, Labor income and predictable stock returns, *Review of Financial Studies* 19, 1–44.

Wang, Zhenyu, and Xiaoyan Zhang, 2005, Empirical evaluation of asset pricing models: Arbitrage and pricing errors over contingent claims, Working paper, Cornell University.

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 three-month T-bill and the 25 Fama-French size and book-to-market ranked portfolios

Panel A: Unconstrained Hansen-Jagannathan distance

Model	**LL**	**LV**	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
$\hat{\delta}$	0.643	0.643	0.642	0.580	0.546	0.651	0.582
$P[\hat{y} < 0]$	0.021	0.103	0.010	0.138	0.154	0.000	0.015
$\sigma_{\hat{y}}$	0.592	0.817	0.335	1.229	1.478	0.273	0.389
ρ_c^2	0.234	0.112	0.687	0.105	0.088	0.845	0.983

Panel B: Constrained Hansen-Jagannathan distance

Model	**LL**	**LV**	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
$\hat{\delta}_+$	0.685	0.700	0.667	0.691	0.684	0.673	0.607
$P[\hat{y} < 0]$	0.000	0.000	0.000	0.005	0.010	0.000	0.015
$\sigma_{\hat{y}}$	0.392	0.294	0.296	0.318	0.350	0.260	0.389
ρ_c^2	0.239	0.138	0.795	0.129	0.136	0.925	0.983
$(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$	6.4%	8.8%	3.9%	19.2%	25.2%	3.4%	4.2%
$\hat{\hat{\delta}}$	0.651	0.666	0.643	0.646	0.638	0.651	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 (LL) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (LV) of Lustig and Van Nieuwerburgh (2004), the conditional CAPM (SV) of Santos and Veronesi (2006), the simple and extended sector investment models (LVX1 and LVX2, respectively) of Li, Vassalou and Xing (2006), the durable consumption CAPM (YOGO) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The data are from 1952:2 to 2000:4 (195 observations). $P[\hat{y} < 0]$ 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. ρ_c^2 is the centered R^2 from the linear regression of the estimated SDF on the returns on the test assets. $(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$ is the percentage difference between the sample constrained and unconstrained HJ-distances. $\hat{\hat{\delta}}$ is a sample measure of the maximum pricing error on the test assets for the SDF that minimizes the constrained HJ-distance.

Table 2
Summary of the models using quarterly gross returns on the three-month T-bill and 10 size and 12 industry portfolios

Panel A: Unconstrained Hansen-Jagannathan distance

Model	**LL**	**LV**	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
$\hat{\delta}$	0.489	0.487	0.462	0.495	0.464	0.443	0.463
$P[\hat{y} < 0]$	0.010	0.026	0.000	0.036	0.041	0.010	0.000
$\sigma_{\hat{y}}$	0.437	0.569	0.264	0.419	0.636	0.473	0.249
ρ_c^2	0.217	0.200	0.910	0.178	0.153	0.368	0.996

Panel B: Constrained Hansen-Jagannathan distance

Model	**LL**	**LV**	**SV**	**LVX1**	**LVX2**	**YOGO**	**FF3**
$\hat{\delta}_+$	0.506	0.498	0.464	0.502	0.482	0.451	0.465
$P[\hat{y} < 0]$	0.010	0.010	0.000	0.005	0.010	0.000	0.000
$\sigma_{\hat{y}}$	0.275	0.416	0.255	0.291	0.432	0.358	0.250
ρ_c^2	0.216	0.199	0.958	0.180	0.154	0.548	0.995
$(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$	3.3%	2.3%	0.5%	1.4%	3.9%	1.7%	0.4%
$\hat{\bar{\delta}}$	0.496	0.489	0.462	0.498	0.471	0.446	0.463

The table presents the sample unconstrained and constrained HJ-distances ($\hat{\delta}$ and $\hat{\delta}_+$, respectively) of seven linear asset pricing models. The models include the conditional consumption CAPM (LL) of Lettau and Ludvigson (2001), a version of the conditional consumption CAPM (LV) of Lustig and Van Nieuwerburgh (2004), the conditional CAPM (SV) of Santos and Veronesi (2006), the simple and extended sector investment models (LVX1 and LVX2, respectively) of Li, Vassalou and Xing (2006), the durable consumption CAPM (YOGO) of Yogo (2006), and the three-factor model (FF3) of Fama and French (1993). The data are from 1952:2 to 2000:4 (195 observations). $P[\hat{y} < 0]$ 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. ρ_c^2 is the centered R^2 from the linear regression of the estimated SDF on the returns on the test assets. $(\hat{\delta}_+ - \hat{\delta})/\hat{\delta}$ is the percentage difference between the sample constrained and unconstrained HJ-distances. $\hat{\bar{\delta}}$ is a sample measure of the maximum pricing error on the test assets for the SDF that minimizes the constrained HJ-distance.

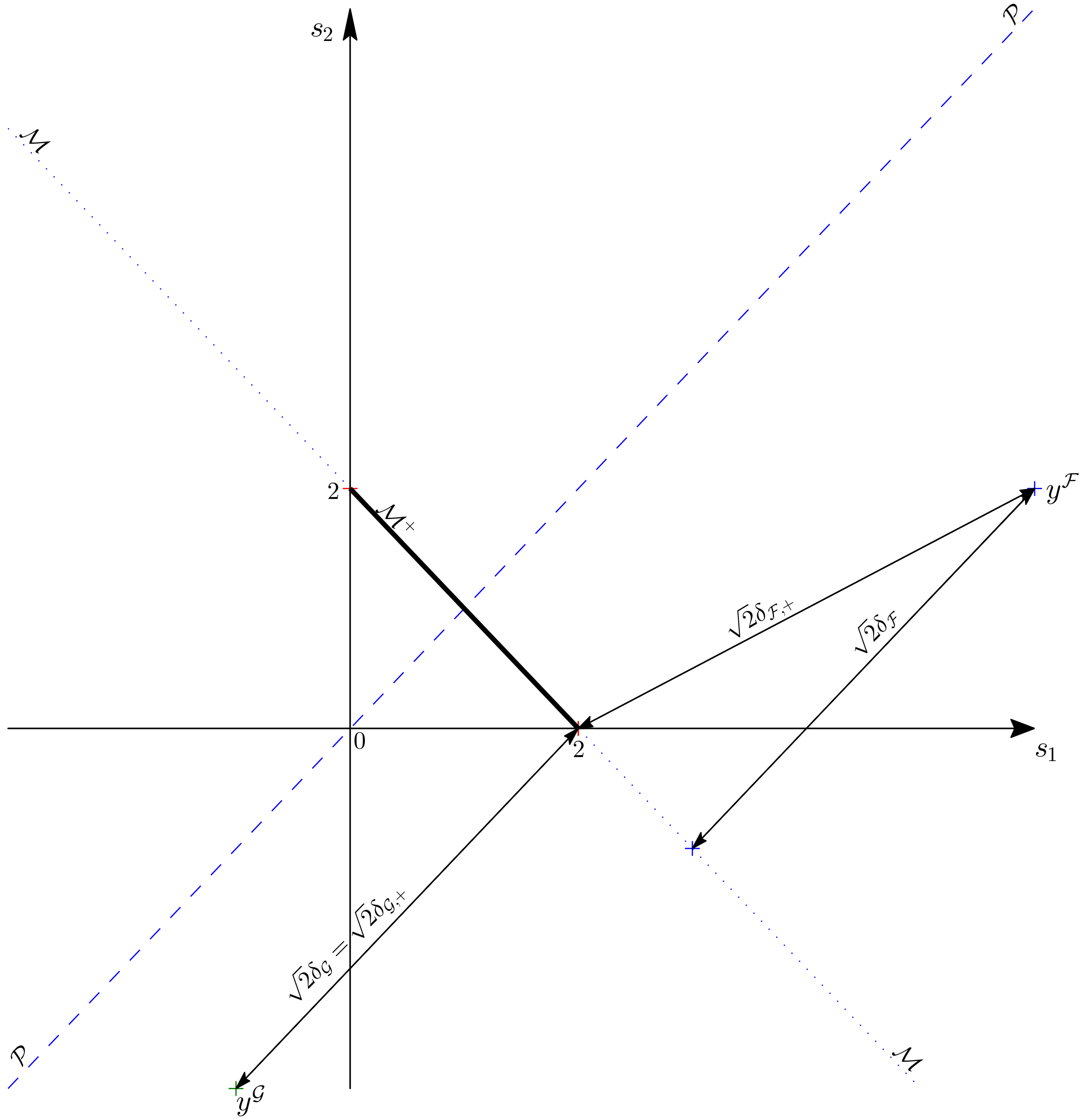

Figure 1. Graphical representation of the constrained Hansen-Jagannathan distance. The figure presents two SDFs ($y^{\mathcal{F}}$ and $y^{\mathcal{G}}$) in an economy with two states (s_1 and s_2) and one risk-free asset. The dashed line represents the payoff space of the risk-free asset ($\mathcal{P}$).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

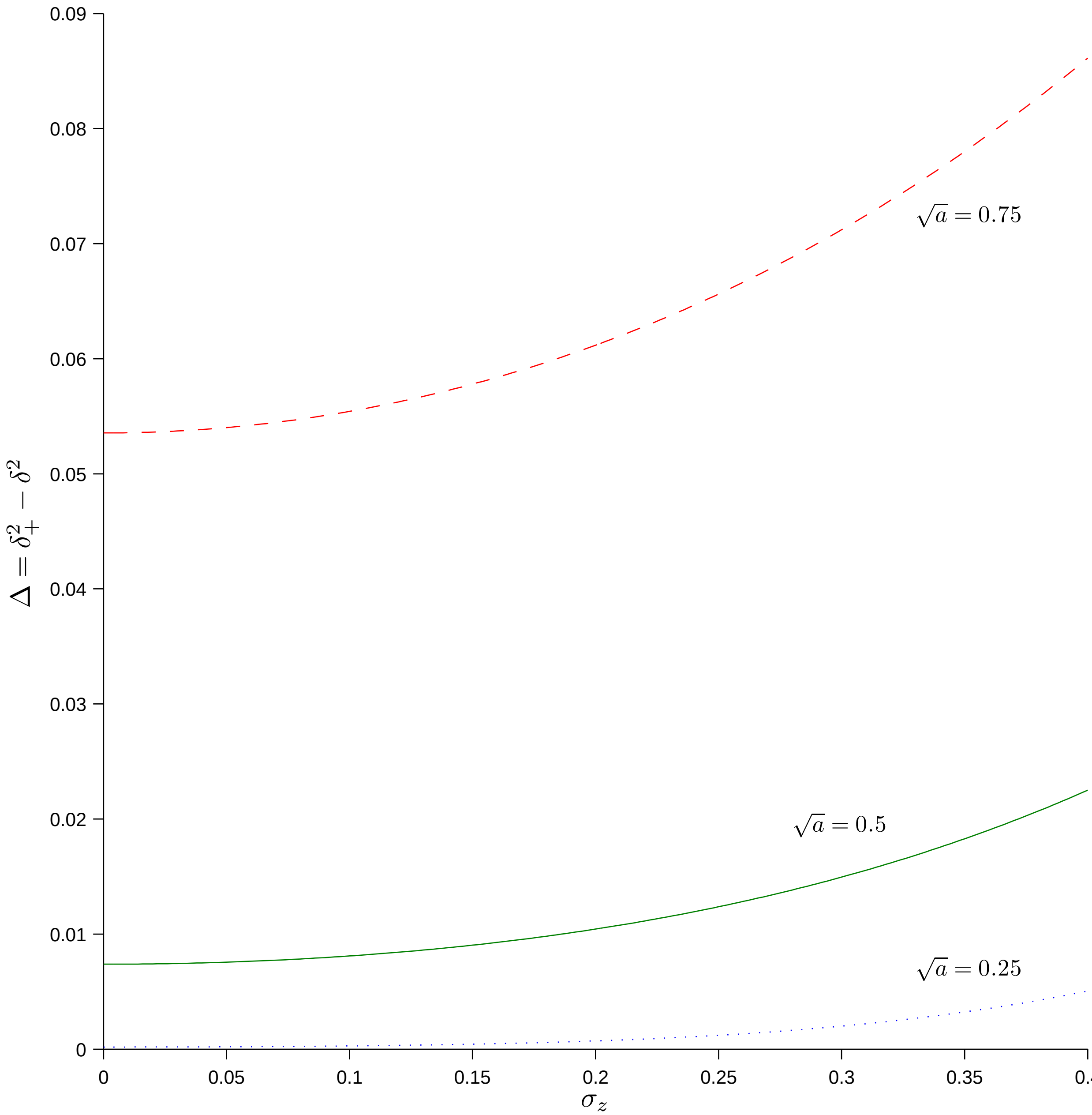

Figure 2. Difference between squared constrained and unconstrained Hansen-Jagannathan distances. The figure plots the difference between squared constrained and unconstrained HJ-distances (Δ) as a function of the standard deviation of the unspanned component (σ_z) of the candidate SDF. The gross risk-free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

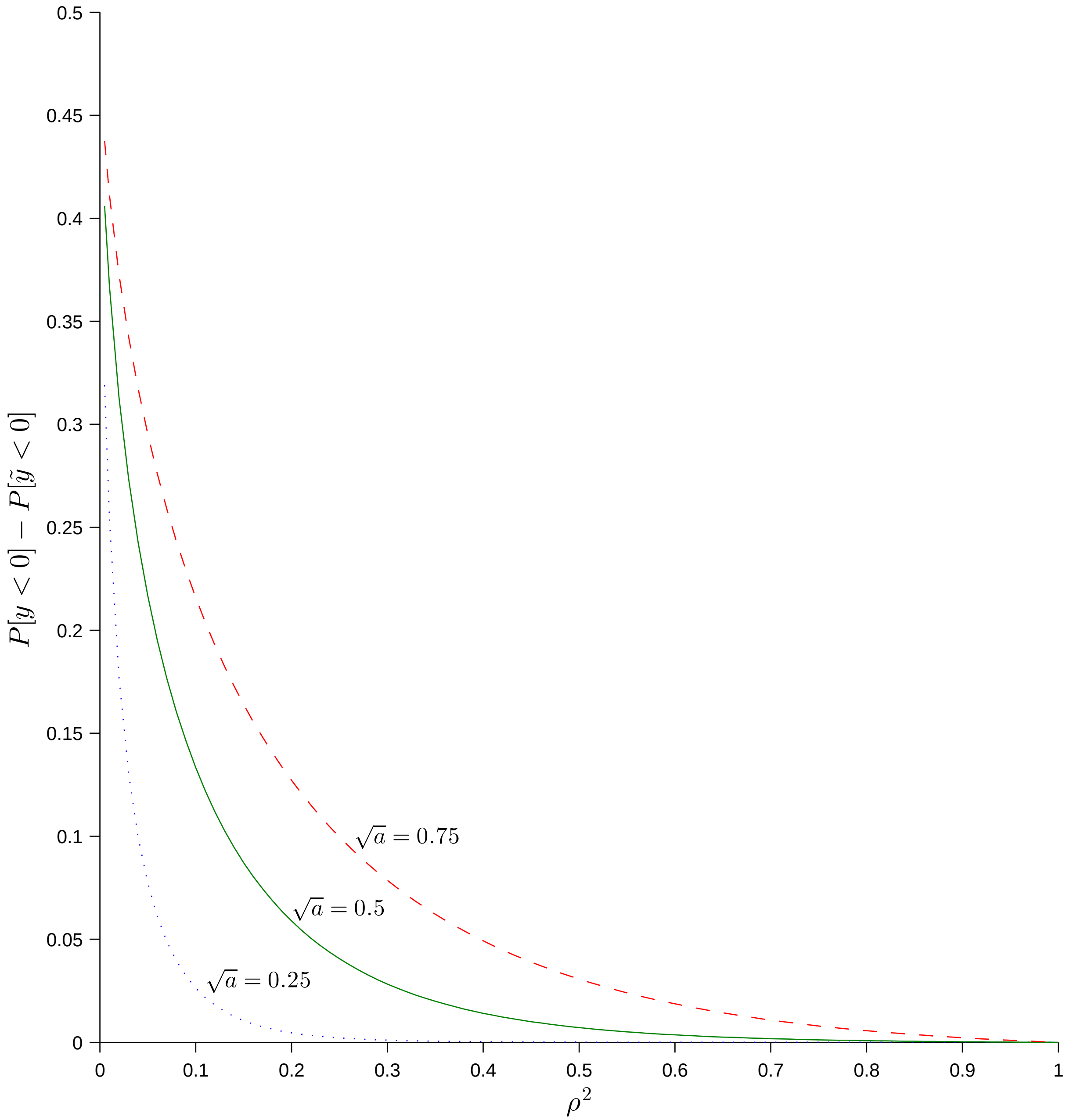

Figure 3. Difference in the probabilities of taking on negative values for two linear SDFs. The figure plots $P[y < 0] - P[\tilde{y} < 0]$ as a function of ρ^2 in a 1-factor setting, where y and $\tilde{y}$ are the linear SDFs chosen to minimize the unconstrained and constrained HJ-distances, respectively.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

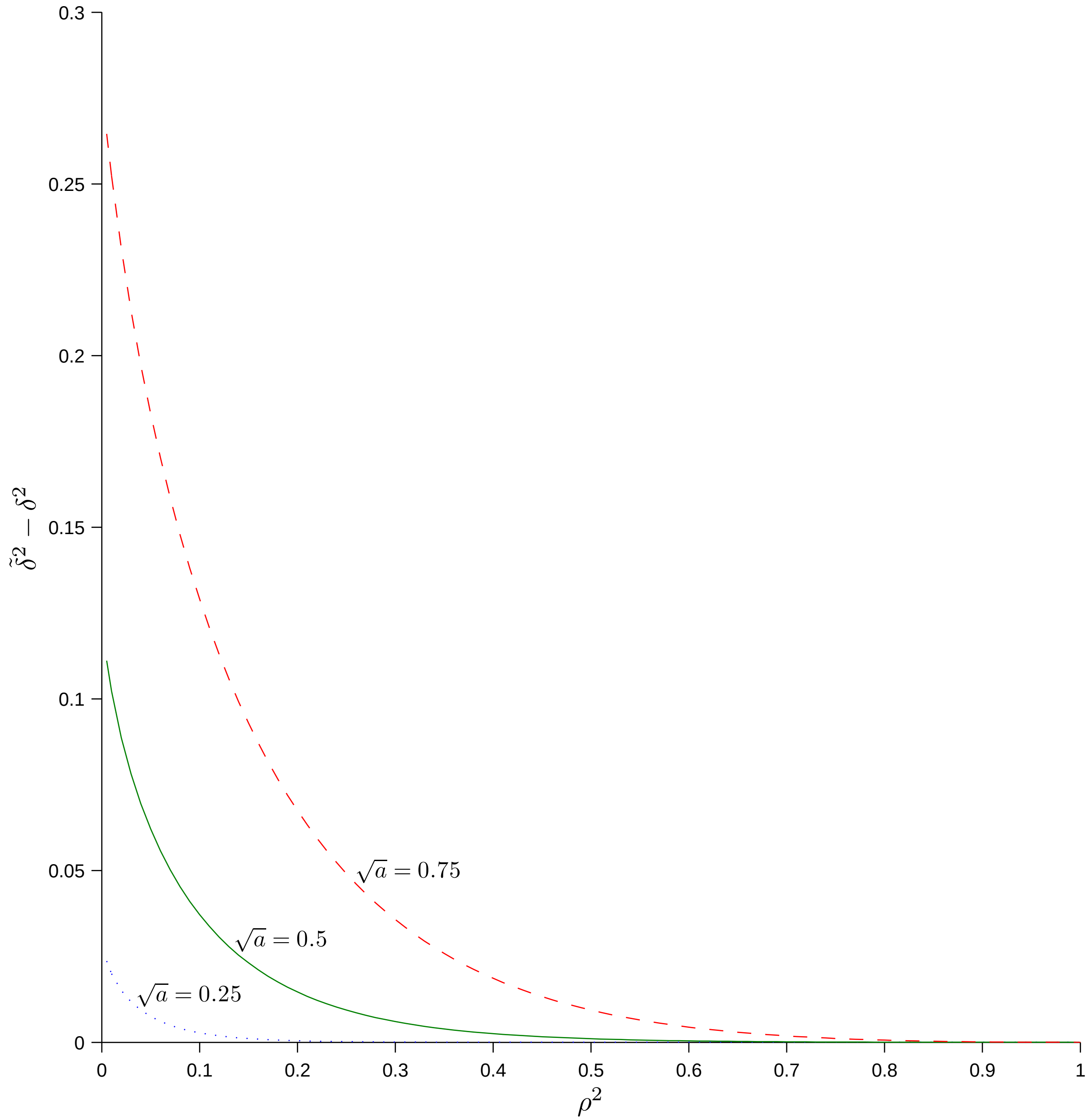

Figure 4. Difference in the aggregate measures of pricing errors of two linear SDFs. The figure plots $\tilde{\delta}^2 - \delta^2$ as a function of ρ^2 in a one-factor setting, where δ^2 and $\tilde{\delta}^2$ are the aggregate measures of pricing errors of the test assets when the linear SDF is chosen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