Specification tests of asset pricing models using excess returns

Raymond Kan and Cesare Robotti*

First draft: November 2005
This version: March 2008

*Corresponding author: Cesare Robotti, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street, Atlanta, Georgia, 30309; Tel: (404) 498-8543; Fax: (404) 498-8810; email:cesare.robotti@atl.frb.org. 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. We thank Geert Bekaert (the editor), an anonymous associate editor, two anonymous referees, Pierluigi Balduzzi, Robert Dittmar, Mark Fisher, Nikolay Gospodinov, Chris Kirby, Lukasz Pomorski, Jay Shanken, Tim Simin, Paula Tkac, Kevin Wang, Chu Zhang, Guofu Zhou, seminar participants at Concordia University, Singapore Management University, and participants at the 2006 Topics in Financial Econometrics Conference at the Federal Reserve Bank of Atlanta, the 2006 Northern Finance Meetings, and the 2007 European Finance Association Meetings for helpful discussions and comments. We are also grateful to John Campbell, Kenneth French, Ravi Jagannathan, Yong Wang, and Zhenyu Wang for making their data available to us. Kan gratefully acknowledges financial support from the National Bank Financial of Canada. The views expressed here are the authors' and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Specification tests of asset pricing models using excess returns

Abstract

In this paper, we discuss the impact of different formulations of asset pricing models on the outcome of specification tests that are performed using excess returns. We point out that the popular way of specifying the stochastic discount factor (SDF) as a linear function of the factors is problematic because (1) the specification test statistic is not invariant to an affine transformation of the factors, and (2) the SDFs of competing models can have very different means. In contrast, an alternative specification that defines the SDF as a linear function of the de-meaned factors is free from these two problems and is more appropriate for model comparison. In addition, we suggest that a modification of the traditional Hansen-Jagannathan distance (HJ-distance) is needed when we use the de-meaned factors. The modified HJ-distance uses the inverse of the covariance matrix (instead of the second moment matrix) of excess returns as the weighting matrix to aggregate pricing errors. Asymptotic distributions of the modified HJ-distance and of the traditional HJ-distance based on the de-meaned SDF under correctly specified and misspecified models are provided. Finally, we propose a simple methodology for computing the standard errors of the estimated SDF parameters that are robust to model misspecification. We show that failure to take model misspecification into account is likely to understate the standard errors of the estimates of the SDF parameters and lead us to erroneously conclude that certain factors are priced.

JEL classification: G12

Keywords: Asset pricing models; Specification tests; Modified Hansen-Jagannathan distance; Misspecification robust standard errors; De-meaned stochastic discount factor

1. Introduction

Asset pricing models are, at best, approximations. Although it is of interest to test whether a particular asset pricing model is literally true or not, a more interesting task for empirical researchers is to find out how wrong a model is and to compare the performance of different asset pricing models. The latter task requires a scalar measure of model misspecification. While there are many reasonable measures that can be used, the one introduced by Hansen and Jagannathan (1997) has gained tremendous popularity in the empirical asset pricing literature. Their proposed measure, called the Hansen-Jagannathan distance (HJ-distance), has been used both as a model diagnostic and as a tool for model selection by many researchers. Examples include Jagannathan and Wang (1996), Jagannathan et al. (1998), Campbell and Cochrane (2000), Lettau and Ludvigson (2001), Hodrick and Zhang (2001), Farnsworth et al. (2002), Dittmar (2002), and Jagannathan and Wang (2007), among others.

Many asset pricing models only predict how cross-sectional differences of risk premia are determined. Therefore, empirical performances of these asset pricing models are often judged by how well they price excess returns. The problem is that when only excess returns are used, the mean of the stochastic discount factor (SDF) cannot be identified. As a result, researchers have to choose some normalization of the SDF. It is generally believed that the choice of normalization of the SDF does not matter. In this paper, we show that the normalization of a SDF is only irrelevant when the model is correctly specified. When the model is misspecified, the mean of the SDF can be a very important determinant of the measure of model misspecification. The choice of normalization can also heavily influence the relative rankings of competing asset pricing models.

For the case of linear factor asset pricing models, we show that the standard way of writing the SDF as a linear function of the factors is problematic when only excess returns are used. In particular, the HJ-distance and other specification test statistics are not invariant to an affine transformation of the factors. Under a linear factor asset pricing model, the factors are only unique up to a linear transformation. If one can change the relative rankings of competing models by simply performing an affine transformation of the factors, then it is rather difficult to make sense of the misspecification measure. We suggest that an alternative specification that defines the SDF as a linear function of the de-meaned factors is free from this problem, and should be the preferred specification for linear SDFs. Under the de-meaned linear SDF model, we propose a modified HJ-distance that has a nice economic interpretation and is more appropriate than the traditional HJ-distance. In order to conduct statistical inference, we also provide an asymptotic analysis of the modified HJ-distance and of the traditional HJ-distance based on the de-meaned SDF under both correctly specified and misspecified models.

Besides being interested in specification tests and model comparisons, researchers often ask the question of whether a particular factor in a proposed asset pricing model is "priced" or not. This question is typically addressed by testing whether the SDF parameter associated with the factor is significantly different from zero or not. Without exception, all existing studies perform this test by using a standard error that assumes that the model is correctly specified. In reality, it is hard to justify this assumption when we estimate the SDF parameters for many different models because some (if not all) of the models are bound to be misspecified. In this paper, we propose robust standard errors of the estimates of the SDF parameters that are applicable to both correctly specified and misspecified models. Using popular asset pricing models proposed in the literature as examples, we show that many of the underlying macroeconomic factors are no longer statistically significant, once misspecification robust standard errors are used. Unless we are certain that a model is correct, we should account for potential model misspecification and use the standard errors proposed in this paper.

Although we focus on excess returns in this paper, many of the problems that we discuss are equally applicable to specification tests that use gross returns. Since many of our points are the same for gross returns and for excess returns, we do not repeat our analysis for the case of gross returns. The only problem that does not apply to the case of gross returns is that the misspecification measure is no longer affected by an affine transformation of the factors. Nevertheless, competing models can still have very different means for their SDFs. In addition, testing whether a factor is priced or not is also typically performed by using a standard error that assumes that the model is correctly specified.

The rest of the paper is organized as follows. The next section discusses the population measures of model misspecification and the HJ-distance when only excess returns on test assets are used. We then study the impact of normalization schemes of the linear SDF on the misspecification measures and show how these measures could be affected by affine transformations of the factors. To overcome this problem, we suggest using a de-meaned version of the linear SDF. In addition, we introduce a modified HJ-distance that is more appropriate for the de-meaned version of linear SDFs. Section 3 presents the sample measures of model misspecification and provides their asymptotic distributions under correctly specified and misspecified models. In Section 4, we provide an asymptotic analysis of the estimators of the SDF parameters for a potentially misspecified model. Section 5 provides several empirical applications to illustrate all the issues raised in the paper. The empirical examples show that when the SDF is written as a linear function of the factors and excess returns are used, comparing models with the HJ-distance is problematic. These applications also allow us to demonstrate the differences between the traditional HJ-distance and the modified HJ-distance and to illustrate the potential impact of model misspecification on the standard errors of the estimated SDF parameters. The final section concludes our findings and the Appendix contains proofs of all propositions.

2. Population measures of model misspecification

2.1. Pricing errors and specification tests

Let y be a proposed SDF and r be a vector of the payoffs of N zero-cost portfolios. We define r as the excess returns on the N portfolios. If y correctly prices the N portfolios, we have zero pricing errors on the excess returns on the N portfolios

$$e \equiv E[ry] = 0_N, \tag{1}$$

where 0_N is an N-vector of zeros. However, if y is a misspecified model, then the pricing errors of the model are nonzero. In most cases, the proposed discount factor y involves some unknown parameters λ and it is customary to suggest that $y(\lambda)$ is a misspecified model if for all values of λ, we have

$$e(\lambda) = E[ry(\lambda)] \neq 0_N. \tag{2}$$

When an asset pricing model is misspecified, researchers are often interested in obtaining a scalar measure of the magnitude of the misspecification. For this purpose, we use an aggregate measure of pricing errors, which is often defined as a quadratic form of the pricing errors

$$Q_W = \min_{\lambda} e(\lambda)' W e(\lambda), \tag{3}$$

where W is a positive definite symmetric weighting matrix. Specification tests of asset pricing models are typically sample versions of Q_W. Note that unless the model is correct, Q_W depends on the choice of W. For model comparison it makes sense to use the same W across models. While there are many choices of W that can be used, the one suggested by Hansen and Jagannathan (1997) has emerged as the most popular choice in the literature.

Hansen and Jagannathan (1997) suggest using $W = U^{-1}$ as the weighting matrix, where $U = E[rr']$ is the second moment matrix of the excess returns. The resulting measure of misspecification is commonly known as the HJ-distance, defined as

$$\delta_{HJ} = Q_{U^{-1}}^{\frac{1}{2}} = \left[\min_{\lambda} e(\lambda)' U^{-1} e(\lambda)\right]^{\frac{1}{2}}. \tag{4}$$

Hansen and Jagannathan (1997) provide two nice interpretations of the HJ-distance. The first is that the HJ-distance measures the minimum distance between the proposed SDF and the set of correct SDFs ($\mathcal{M}$),

$$\delta_{HJ} = \min_{m \in \mathcal{M}} \|m - y\|, \tag{5}$$

where $\|X\| = E[X^2]^{\frac{1}{2}}$ is the standard L^2 norm. The second is that it represents the maximum pricing error of a portfolio of r that has a unit second moment. Define ξ as the random payoff of a portfolio. Hansen and Jagannathan (1997) show that

$$\delta_{HJ} = \max_{\|\xi\|=1} |\pi(\xi) - \pi^y(\xi)|, \tag{6}$$

where $\pi(\xi)$ and $\pi^y(\xi)$ are the prices of ξ assigned by the true and the proposed SDF, respectively.

When only excess returns are used to measure model misspecification, one has to be careful with the specification of the proposed SDF. In particular, one cannot specify y in a way such that it can be zero for some values of λ. For example, the popular class of linear factor asset pricing models suggests that y is a linear function of K systematic factors f. However, when only excess returns are used, one cannot specify y as

$$y(\lambda_0, \lambda) = \lambda_0 - f'\lambda. \tag{7}$$

This is because when $\lambda_0 = 0$ and $\lambda = 0_K$, we have $Q_W = 0$ regardless of the validity of the model. When only excess returns are used, it is not possible to identify the mean of the SDF and some normalization of y becomes necessary. It is generally believed that the choice of normalization is entirely one of convenience and that it does not matter which one is used.[1] For linear factor models, a popular choice of normalization is to set $\lambda_0 = 1$ and specify y as[2]

$$y(\lambda) = 1 - f'\lambda. \tag{8}$$

If the model is correct, $Q_W = 0$ for any choice of λ_0. However, when the model is incorrect, the value of Q_W generally depends on the choice of λ_0. Nevertheless, it can be shown that the pricing errors, the p-value of the specification test as well as the relative rankings of competing models do not depend on the choice of λ_0 (as long as the competing linear factor models all use the same λ_0). As a result, researchers often consider the choice of normalization to be rather innocuous.

However, there is a serious problem with imposing $\lambda_0 = 1$ (or any other constant). With such a choice, the misspecification measure Q_W as well as the relative rankings of competing models are sensitive to affine transformations of the factors. This is problematic because under the linear factor asset pricing models, factors are only unique up to an affine transformation. If one can change the relative rankings of competing models by simply performing an affine transformation on some of the factors, then it is rather difficult to make sense of the misspecification measure.

To prepare for our analysis of this problem, we define $Y = [f', r']'$ and its mean and covariance matrix as

$$\mu = E[Y] \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \tag{9}$$

$$V = \mathrm{Var}[Y] \equiv \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{10}$$

Under the linear SDF of $y = 1 - f'\lambda$, the pricing errors of the N assets are given by

$$e(\lambda) = E[ry] = E[r(1 - f'\lambda)] = \mu_2 - B\lambda, \tag{11}$$

where $B = E[rf'] = V_{21} + \mu_2\mu_1'$. It follows that

$$Q_W = \min_{\lambda}(\mu_2 - B\lambda)'W(\mu_2 - B\lambda) = \mu_2'W\mu_2 - \mu_2'WB(B'WB)^{-1}B'W\mu_2. \tag{12}$$

Throughout the paper, we assume that V_{21} is of full column rank (which implies that B is also of full column rank). Hence, there exists a unique λ that minimizes $e(\lambda)'We(\lambda)$, which we denote by

$$\lambda_W = (B'WB)^{-1}(B'W\mu_2). \tag{13}$$

[1] See Cochrane (2005, pp. 256–9) for a discussion of this view.

[2] See Jagannathan and Wang (1996), Kan and Zhou (1999, 2001), Kan and Zhang (1999a, 1999b), Cochrane (2005, pp. 256–9), and Jagannathan and Wang (2002) among others.

Note that unless the model is correctly specified, λ_W depends on the choice of W.

The following proposition shows that when the asset pricing model is misspecified, Q_W depends on the mean of the factors (μ_1). As a consequence, one can easily alter the relative ranking of a specific model by performing an affine transformation of the factors.

Proposition 1. *Suppose the model is misspecified, i.e., μ_2 is not in the span of the column space of B. The μ_1 that maximizes Q_W is $\mu_1 = -V_{12}W\mu_2(\mu_2'W\mu_2)^{-1}$ and the μ_1 that minimizes Q_W is $\mu_1 \to \pm\infty$. In addition, we have*

$$\max_{\mu_1} Q_W = \mu_2'W\mu_2, \tag{14}$$

$$\inf_{\mu_1} Q_W = 0. \tag{15}$$

Intuitively, if we choose μ_1 to make $W^{\frac{1}{2}}B$ orthogonal to $W^{\frac{1}{2}}\mu_2$ (i.e., $B'W\mu_2 = 0_K$), we get the maximum possible Q_W and the model appears to be very poor. On the other hand, if μ_1 is very large in absolute value, then $B = V_{21} + \mu_2\mu_1'$ is dominated by the term $\mu_2\mu_1'$, and B can explain the expected excess returns (μ_2) very well regardless of how poor the covariances (V_{21}) or betas are in explaining the expected excess returns.

Proposition 1 has serious implications because it suggests that when using the linear SDF in (8), one can manipulate the outcome of a specification test by simply adding a constant to the original factors. Although the factors are suggested by theory in most empirical applications, they are only unique up to a linear transformation. Therefore, one could easily justify using any rescaling of the proposed factors. For example, under the CAPM, one can choose to write the SDF as a linear function of excess returns, raw returns, or gross returns on the market portfolio, and there is no strong reason to believe that one particular choice is superior to the others. However, the misspecification measure Q_W will be different across these three plausible specifications of the CAPM. While Proposition 1 is stated in terms of population moments, a similar result holds for the sample version of Q_W and it holds even for a correctly specified model. This is because the sample pricing errors will not be exactly zero even for a correctly specified model. As a result, one can always manipulate the sample version of Q_W by rescaling the original factors, regardless of whether the model is correctly specified or misspecified.[3]

[3]It is worth pointing out that the optimal GMM specification test is also in general not invariant to an affine

2.2. An alternative specification of the linear stochastic discount factor

For a misspecification measure to make sense, we would like it to be invariant to an affine transformation of the factors. For the case of a linear SDF, there indeed exists an alternative specification that has this nice property. Under this alternative specification, we write the SDF as a linear function of the de-meaned factors[4]

$$y(\lambda) = 1 - (f - E[f])'\lambda. \tag{16}$$

By comparing (16) with (7), one may think that this alternative specification is simply the original linear SDF model with a normalization of $\lambda_0 = 1 + E[f]'\lambda$. One might further conjecture that since the choice of λ_0 does not affect the pricing errors or the p-value of the specification test, using this alternative specification of the linear SDF model would not make any real difference. However, there are some subtle differences between the two specifications that could lead to very different results. The first difference is that in the de-meaned version of the linear SDF, λ_0 is not a fixed constant but a function of λ. As it turns out, when the model is misspecified, the λ that minimizes the quadratic form of the pricing errors is not the same across the raw and de-meaned specifications of the linear SDF model. Therefore, the pricing errors and the p-values of the specification tests are not identical under these two specifications. Another difference is that when it comes to model comparison, the de-meaned specification imposes the constraint that $E[y] = 1$ across models whereas the original specification does not.[5] The advantage of this alternative specification is that the pricing errors and Q_W are invariant to affine transformations of the factors. To see this, write the pricing errors of the N assets under the de-meaned version of the linear SDF as

$$e(\lambda) = E[ry] = E[r] - E[r(f - \mu_1)']\lambda = \mu_2 - V_{21}\lambda. \tag{17}$$

transformation of the factors. Both the two-step and the iterative GMM specification tests are affected by an affine transformation of the factors. The only optimal GMM specification test that is not plagued by this problem is the one that uses the continuous-updating estimator of Hansen et al. (1996). Surprisingly, the specification test statistic of the continuous updating GMM does not depend on whether we use the linear factor SDF here or the linear de-meaned factor SDF in the next subsection. Proof of this result is available upon request.

[4] See, for example, Balduzzi and Kallal (1997), Kirby (1998), Cochrane (2005, p.257), Balduzzi and Robotti (2008), and Burnside (2007). Note that even when the model is correctly specified, the value of λ in (16) is not the same as the value of λ in (8) unless $\mu_1 = 0_K$.

[5] Another way to normalize the SDF to have unit mean is to define $y^* = y/E[y]$. For the case of the linear SDF, this alternative specification gives us a normalized SDF $y^*(\lambda) = (1 - f'\lambda)/(1 - \mu_1'\lambda)$. It can be shown that this alternative normalization gives us the same misspecification measure as the de-meaned factor specification.

It follows that the misspecification measure of the model is given by

$$Q_W = \min_{\lambda}(\mu_2 - V_{21}\lambda)'W(\mu_2 - V_{21}\lambda) = \mu_2'W\mu_2 - \mu_2'WV_{21}(V_{12}WV_{21})^{-1}V_{12}W\mu_2, \tag{18}$$

which is independent of the choice of μ_1.

Note that in (16), we restrict the candidate SDFs to have unit mean. Such an assumption is innocuous. We can instead restrict the candidate SDFs to have mean c, where c is an arbitrary nonzero constant. The only effect this has is that the resulting Q_W will be c^2 times the Q_W with unit mean. It will not change any statistical inference or the relative rankings of competing models.

Despite the nice property of being invariant to affine transformations of the factors, the de-meaned version of the linear SDF has not been very popular in the empirical literature. We suspect that researchers stay away from the de-meaned version because it requires the estimation of the mean of the factors which, in turn, adds some complications to statistical inference. One notable exception is Burnside (2007). Burnside studies the power of GMM specification tests in rejecting misspecified models with factors that have very low correlations with returns. He finds that the GMM tests have better power under the de-meaned SDF specification than under the raw SDF specification. Although Burnside's paper compares the same two specifications of the linear SDFs, the focus of his paper is very different from ours. His paper focuses on the relative power of specification tests under the two SDF specifications, whereas our paper focuses on the invariance issue of specification tests as well as on deriving misspecification robust standard errors for the parameter estimates.

2.3. The modified HJ-distance

Although so far we have focused on linear SDFs, the problem that we discuss also applies to nonlinear SDFs. Namely, the misspecification measure and relative rankings of models are generally not invariant to the rescaling of the factors. When only zero-cost portfolios are used as test assets, it makes sense to restrict all the SDFs to have a unit (or a common constant) mean in order to have a fair comparison between models. This constraint amounts to requiring all competing SDFs to assign the same price to the risk-free asset, so that we only compare their performances based on their pricing errors on excess returns. An added benefit of this restriction is that it allows us to interpret the pricing errors as expected return errors. This is because when y has unit mean, the pricing errors are given by

$$E[ry] = E[r(1 + y - E[y])] = E[r] + \text{Cov}[r, y], \tag{19}$$

and we can interpret $-\text{Cov}[r, y]$ as the expected excess returns based on the proposed asset pricing model y.

Once we restrict the candidate SDF, y, to have unit mean, we should no longer use the traditional HJ-distance to measure model misspecification since the set of admissible SDFs, $\mathcal{M}$, contains many SDFs that do not have unit mean. Therefore, we need to modify the definition of the HJ-distance. Our proposed modified HJ-distance is defined as

$$\delta_m = \min_{m \in \mathcal{M}, E[m]=1} \|m - y\|, \tag{20}$$

and it is a measure of how far y is from an admissible SDF that has unit mean.

For the original HJ-distance δ_{HJ}, δ_{HJ}^2 can be interpreted as a misspecification measure Q_W with $W = U^{-1}$, where $U = E[rr']$ is the second moment matrix of the excess returns. For our modified HJ-distance, we have a similar interpretation, but we need to replace U^{-1} by V_{22}^{-1}, where V_{22} is the covariance matrix of the excess returns. Such a modification was first suggested by Balduzzi and Yao (2007) and the following proposition generalizes this result.

Proposition 2. *Let x be a vector of payoffs on N test assets at the end of the period and q be a vector of prices of the N assets at the beginning of the period. When we restrict a candidate SDF y to have a mean of $E[y] = c$, the modified HJ-distance of y is given by*

$$\delta_m = \min_{m \in \mathcal{M}, E[m]=c} ||m - y|| = (e'\text{Var}[x]^{-1}e)^{\frac{1}{2}}, \tag{21}$$

where $\text{Var}[x]$ *is the variance-covariance matrix of x and $e = E[xy] - q$ is the vector of pricing errors.*

In many cases, the proposed SDF, y, involves some unknown parameters λ. For these cases, it makes sense to define the modified HJ-distance as

$$\delta_m = \left[\min_{\lambda} e(\lambda)'\text{Var}[x]^{-1}e(\lambda)\right]^{\frac{1}{2}}, \tag{22}$$

where $e(\lambda) = E[xy(\lambda)] - q$.

Note that the result in Proposition 2 is quite general: It works for both linear and nonlinear models, and for gross returns as well as excess returns. When we choose x to be a vector of excess returns and use the de-meaned version of the linear SDF in (16), the pricing errors are given by $e(\lambda) = \mu_2 - V_{21}\lambda$ and the squared modified HJ-distance can be expressed as

$$\delta_m^2 = \min_\lambda (\mu_2 - V_{21}\lambda)' V_{22}^{-1} (\mu_2 - V_{21}\lambda) = \mu_2' V_{22}^{-1} \mu_2 - \mu_2' V_{22}^{-1} V_{21} (V_{12} V_{22}^{-1} V_{21})^{-1} V_{12} V_{22}^{-1} \mu_2. \quad (23)$$

It is interesting to note that for a linear SDF, using a nonsingular $W = (V_{22} + V_{21} C V_{12})^{-1}$ as the weighting matrix (where C is a $K \times K$ matrix) produces the same results as using V_{22}^{-1} as the weighting matrix.[6] For example, we can use Σ^{-1} as the weighting matrix, where $\Sigma = V_{22} - V_{21} V_{11}^{-1} V_{12}$ is the covariance matrix of the residuals from regressing r on $[1,\ f']'$. However, we cannot use U^{-1} as the weighting matrix, where $U = E[rr'] = V_{22} + \mu_2\mu_2'$ is the second moment matrix of the excess returns. Unless the model is correct, μ_2 is not spanned by the column space of V_{21}. Hence, using U^{-1} as the weighting matrix will give a δ_{HJ} that is different from the δ_m that uses V_{22}^{-1} or Σ^{-1} as the weighting matrix. Since $U - V_{22} = \mu_2\mu_2'$ is a positive semidefinite matrix, δ_{HJ} is in general smaller than δ_m. Although δ_{HJ} is not the same as δ_m, the following lemma shows that δ_{HJ} is just a monotonic transformation of δ_m for the case of de-meaned version of linear SDF.

Lemma 1. *For the de-meaned version of the linear SDF, $y(\lambda) = 1 - (f - E[f])'\lambda$, the squared HJ-distance and the squared modified HJ-distance are monotonic transformations of each other, and the relations are given by $\delta_{HJ}^2 = \frac{\delta_m^2}{1+\delta_m^2}$ and $\delta_m^2 = \frac{\delta_{HJ}^2}{1-\delta_{HJ}^2}$. In addition, we have $\lambda_{HJ} = \frac{\lambda_m}{1+\delta_m^2}$ and $\lambda_m = \frac{\lambda_{HJ}}{1-\delta_{HJ}^2}$, where $\lambda_{HJ} = \mathrm{argmin}_\lambda (\mu_2 - V_{21}\lambda)' U^{-1} (\mu_2 - V_{21}\lambda) = (V_{12} U^{-1} V_{21})^{-1} (V_{12} U^{-1} \mu_2)$ and $\lambda_m = \mathrm{argmin}_\lambda (\mu_2 - V_{21}\lambda)' V_{22}^{-1} (\mu_2 - V_{21}\lambda) = (V_{12} V_{22}^{-1} V_{21})^{-1} (V_{12} V_{22}^{-1} \mu_2)$.*[7]

Note that Lemma 1 also holds for the sample counterparts of δ_{HJ}^2, δ_m^2, λ_{HJ} and λ_m. Therefore, ranking models by $\hat{\delta}_{HJ}$ is the same as ranking models by $\hat{\delta}_m$. In addition, once the asymptotic distribution of $\hat{\delta}_m^2$ is established, we can use Lemma 1 and the delta method to obtain the asymptotic distribution of $\hat{\delta}_{HJ}^2$. Another point to note is that Lemma 1 suggests that δ_{HJ}^2 (and also $\hat{\delta}_{HJ}^2$) is bounded above by one. In computing the p-value of $\hat{\delta}_{HJ}^2$ under the correctly specified model, one often uses the asymptotic distribution of a linear combination of χ_1^2 random variables. The fact

[6] This result can be proved using the matrix identities in the Appendix of Kan and Zhou (2004).

[7] Besides being the quantity that minimizes δ_m^2, λ_m also has a nice economic interpretation. Let $f^\star = V_{12} V_{22}^{-1} r$ be the mimicking portfolios of the K factors. Then, we can write $\lambda_m = \mathrm{Var}[f^\star]^{-1} E[f^\star]$. For the case of $K = 1$, λ_m is simply the risk premium of the factor mimicking portfolio over its variance.

that $\hat{\delta}^2_{HJ}$ has a bounded distribution suggests that the asymptotic distribution may have problems with approximating the right tail of the distribution of $\hat{\delta}^2_{HJ}$.

Since using V_{22}^{-1} or Σ^{-1} as the weighting matrix does not affect the modified HJ-distance for a linear factor model, we can provide an alternative expression of the squared modified HJ-distance for the de-meaned linear SDF model as

$$\delta_m^2 = \mu_2'[\Sigma^{-1} - \Sigma^{-1}\beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}]\mu_2 = \min_{\gamma}(\mu_2 - \beta\gamma)'\Sigma^{-1}(\mu_2 - \beta\gamma), \tag{24}$$

where $\beta = V_{21}V_{11}^{-1}$ are the regression slope coefficients from regressing r on f (and an intercept). Note that the last expression is analogous to the cross-sectional regression test of Shanken (1985), which is simply an aggregate measure of the pricing errors from the GLS cross-sectional regression of μ_2 on β.[8]

2.4. An alternative interpretation of the modified HJ-distance

When a risk-free asset is available, Farnsworth et al. (2002) suggest that we should add the risk-free asset to the set of test assets to improve the performance of asset pricing models. It turns out that, once we augment the excess returns on the N portfolios with the gross return on the risk-free asset, the traditional HJ-distance on these augmented returns is closely related to our modified HJ-distance on excess returns. To understand this relation, we define R_0 as the gross return on the risk-free asset. Suppose that we want to evaluate a linear asset pricing model using both R_0 and r.[9] Since we have a positive investment asset, we can now write the SDF as in (7). Then, the pricing errors of the $N+1$ assets are given by

$$e(\lambda_0, \lambda) = E\begin{bmatrix} R_0 y(\lambda_0,\lambda) - 1 \\ r y(\lambda_0,\lambda) \end{bmatrix} = \begin{bmatrix} R_0 & -R_0\mu_1' \\ \mu_2 & -(V_{21} + \mu_2\mu_1') \end{bmatrix}\begin{bmatrix} \lambda_0 \\ \lambda \end{bmatrix} - \begin{bmatrix} 1 \\ 0_N \end{bmatrix}. \tag{25}$$

The traditional squared HJ-distance computed using the $N+1$ assets is given by

$$\tilde{\delta}^2_{HJ} = \min_{\lambda_0,\lambda} e(\lambda_0,\lambda)'\tilde{U}^{-1}e(\lambda_0,\lambda), \tag{26}$$

[8] It can be shown that the test of Gibbons et al. (1989) is a special version of the squared modified HJ-distance that is computed using both the test assets and the benchmark assets as well as imposing the restriction that the benchmark assets are priced without errors (i.e., $\lambda = V_{11}^{-1}\mu_1$).

[9] We can also evaluate the asset pricing model using all gross returns, i.e., R_0 and $r + R_0 1_N$. The results are identical.

where $\tilde{U}$ is the second moment matrix of $[R_0,\ r']'$ and is given by

$$\tilde{U} = \begin{bmatrix} R_0^2 & R_0\mu_2' \\ R_0\mu_2 & V_{22} + \mu_2\mu_2' \end{bmatrix}. \tag{27}$$

The following lemma shows that there is a one-to-one correspondence between the traditional HJ-distance $\tilde{\delta}_{HJ}$ that is computed using the $N+1$ assets and the modified HJ-distance δ_m that is computed using just the excess returns on the N risky assets.

Lemma 2. *The traditional HJ-distance ($\tilde{\delta}_{HJ}$) based on the excess returns on N risky portfolios and the gross return on the risk-free asset using the model $y(\lambda_0, \lambda) = \lambda_0 - f'\lambda$ is related to the modified HJ-distance (δ_m) based on just the excess returns on the N risky portfolios using the model $y(\lambda) = 1 - (f - E[f])'\lambda$ as follows*

$$\tilde{\delta}_{HJ} = \frac{\delta_m}{R_0}, \tag{28}$$

where R_0 is the gross risk-free rate.

Lemma 2 suggests that once the risk-free asset is available, there is no difference in ranking models by $\tilde{\delta}_{HJ}$ or by δ_m. To force δ_m to be identical to $\tilde{\delta}_{HJ}$, we just need to write the SDF for the excess returns as

$$y(\lambda) = \frac{1}{R_0} - (f - E[f])'\lambda. \tag{29}$$

This way, we have $E[y] = 1/R_0$ and the SDF will price the risk-free asset correctly.

As a population measure, we can use either $\tilde{\delta}_{HJ}$ or δ_m to rank models. However, the risk-free rate is not constant over time and, as a result, the sample counterparts of $\tilde{\delta}_{HJ}$ and δ_m do not always give us the same rankings across models. In reality, the risk-free rate is typically much less volatile than the excess returns on the risky assets. For statistical reasons, we tend to choose the parameters of the SDF to price the risk-free asset well. Therefore, once the risk-free asset is included as a test asset, the means of the competing SDFs tend to be very close to each other and model comparison using the traditional HJ-distance becomes more meaningful.

Some asset pricing models, such as the zero-beta CAPM of Black (1972), are not designed to price the risk-free asset correctly. This is because the return on a risk-free asset, like the T-bill, is considered to be the risk-free lending rate and should be below the zero-beta rate. For those models, it would be unreasonable to force their SDFs to price the risk-free asset correctly. Instead of defining r as the returns on the test assets in excess of the risk-free rate, it makes sense to follow Ferson et al. (1993) and define r as the returns on the test assets in excess of the return on a benchmark asset, where the benchmark asset can be any other risky asset. When excess returns are constructed in this fashion, our modified HJ-distance only imposes the constraint that all competing models have the same zero-beta rate. Therefore, it is still applicable even when the asset pricing model does not hold for the risk-free asset.

3. Sample measures of model misspecification

3.1. Asymptotic analysis under correctly specified models

In practice, the population misspecification measure Q_W of a model is unobservable and has to be estimated. In this subsection, we discuss the asymptotic distribution of the sample measure of Q_W for the case of linear factor models. We assume that the SDF at time t is a linear function of f_t, which is a vector of K systematic factors. There are two ways to write the linear SDF, one being $y_t = 1 - f_t'\lambda$ and the other one being the de-meaned version $y_t = 1 - (f_t - E[f_t])'\lambda$. Since the first specification is not invariant to an affine transformation of the factors, we will focus our discussion on the sample misspecification measure based on the second specification.

We assume that the model is estimated using excess returns on N ($N > K$) test assets. Let $Y_t = [f_t',\ r_t']'$, where r_t is a vector of excess returns on N test assets at time t. Suppose that we have T observations of Y_t and denote the sample moments of Y_t by

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} Y_t, \tag{30}$$

$$\hat{V} = \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\mu})(Y_t - \hat{\mu})'. \tag{31}$$

Let W_T be a symmetric positive definite weighting matrix on the pricing errors with $W_T \xrightarrow{\text{a.s.}} W$, where W is a symmetric positive definite matrix. The sample version of the model misspecification measure in (18) is given by

$$\hat{Q}_W = \min_{\lambda}(\hat{\mu}_2 - \hat{V}_{21}\lambda)'W_T(\hat{\mu}_2 - \hat{V}_{21}\lambda) = \hat{\mu}_2'W_T\hat{\mu}_2 - \hat{\mu}_2'W_T\hat{V}_{21}(\hat{V}_{12}W_T\hat{V}_{21})^{-1}\hat{V}_{12}W_T\hat{\mu}_2. \tag{32}$$

In this subsection, we present the asymptotic distribution of $\hat{Q}_W$ under the correctly specified model. In order to obtain the distribution of $\hat{Q}_W$, we employ the Generalized Method of Moments (GMM) of Hansen (1982). Under the correctly specified model, we have the following population moment conditions

$$E[g_t(\theta)] = E\begin{bmatrix} f_t - \mu_1 \\ r_t[1 - (f_t - \mu_1)'\lambda] \end{bmatrix} = 0_{N+K}, \tag{33}$$

where $\theta = [\mu_1',\ \lambda']'$. The sample moment conditions are then given by

$$\bar{g}_T(\theta) = \begin{bmatrix} \bar{g}_{1T}(\mu_1) \\ \bar{g}_{2T}(\mu_1, \lambda) \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^T (f_t - \mu_1) \\ \frac{1}{T}\sum_{t=1}^T r_t[1 - (f_t - \mu_1)'\lambda] \end{bmatrix}. \tag{34}$$

It is straightforward to verify that

$$\hat{Q}_W = \bar{g}_{2T}(\hat{\theta})' W_T \bar{g}_{2T}(\hat{\theta}), \tag{35}$$

where $\hat{\theta} = [\hat{\mu}_1',\ \hat{\lambda}']'$ with

$$\hat{\lambda} = (\hat{V}_{12} W_T \hat{V}_{21})^{-1} (\hat{V}_{12} W_T \hat{\mu}_2). \tag{36}$$

Cochrane (2005) suggests that $\hat{\theta}$ can be written as the solution to the following conditions

$$A_T \bar{g}_T(\theta) = 0_{2K}, \tag{37}$$

where

$$A_T = \begin{bmatrix} I_K & 0_{K\times N} \\ 0_{K\times K} & \hat{V}_{12} W_T \end{bmatrix} \stackrel{\text{a.s.}}{\longrightarrow} \begin{bmatrix} I_K & 0_{K\times N} \\ 0_{K\times K} & V_{12} W \end{bmatrix} \equiv A. \tag{38}$$

We define the derivative of the sample moment conditions with respect to the parameters as

$$D_T(\theta) = \frac{\partial \bar{g}_T(\theta)}{\partial \theta'} = \begin{bmatrix} -I_K & 0_{K\times K} \\ \hat{\mu}_2 \lambda' & -\frac{1}{T}\sum_{t=1}^T r_t (f_t - \mu_1)' \end{bmatrix} \stackrel{\text{a.s.}}{\longrightarrow} \begin{bmatrix} -I_K & 0_{K\times K} \\ \mu_2 \lambda' & -V_{21} \end{bmatrix} \equiv D. \tag{39}$$

Note that under the correctly specified model, $\mu_2 = V_{21}\lambda$, and D can be simplified to

$$D = \begin{bmatrix} -I_K & 0_{K\times K} \\ V_{21}\lambda\lambda' & -V_{21} \end{bmatrix}. \tag{40}$$

Under joint stationarity and ergodicity assumptions on Y_t and assuming that its fourth moments exist, the asymptotic distribution of $\hat{\theta}$ is then given by

$$\sqrt{T}(\hat{\theta} - \theta) \stackrel{A}{\sim} N(0_{2K}, (AD)^{-1} ASA' (D'A')^{-1}), \tag{41}$$

where

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E[g_t(\theta)g_{t+j}(\theta)'], \tag{42}$$

and the asymptotic distribution of $\bar{g}_T(\hat{\theta})$ is given by

$$\sqrt{T}\bar{g}_T(\hat{\theta}) \stackrel{A}{\sim} N(0_{N+K}, [I_{N+K} - D(AD)^{-1}A]S[I_{N+K} - D(AD)^{-1}A]'). \tag{43}$$

Under the correctly specified model, we have

$$\sqrt{T}\bar{g}_{2T}(\hat{\theta}) \stackrel{A}{\sim} N(0_N, [I_N - V_{21}(V_{12}WV_{21})^{-1}V_{12}W]S_{22}[I_N - V_{21}(V_{12}WV_{21})^{-1}V_{12}W]'). \tag{44}$$

Therefore, the asymptotic distribution of $T\hat{Q}_W$ under the correctly specified model is a linear combination of $N-K$ independent chi-squared random variables with one degree of freedom

$$T\hat{Q}_W \stackrel{A}{\sim} \sum_{i=1}^{N-K} \xi_i \chi_1^2, \tag{45}$$

where ξ_i are the $N-K$ nonzero eigenvalues of

$$[I_N - W^{\frac{1}{2}}V_{21}(V_{12}WV_{21})^{-1}V_{12}W^{\frac{1}{2}}]W^{\frac{1}{2}}S_{22}W^{\frac{1}{2}}[I_N - W^{\frac{1}{2}}V_{21}(V_{12}WV_{21})^{-1}V_{12}W^{\frac{1}{2}}], \tag{46}$$

or equivalently the eigenvalues of $P'W^{\frac{1}{2}}S_{22}W^{\frac{1}{2}}P$, where P is an $N \times (N-K)$ orthonormal matrix with its columns orthogonal to $W^{\frac{1}{2}}V_{21}$. Note that when the model is correctly specified, the asymptotic distribution of $\hat{Q}_W$ only depends on S_{22}, and not on S_{11} and S_{12}. This implies that the asymptotic distribution of $\hat{Q}_W$ does not depend on whether we know μ_1 or not. In addition, when the model is correctly specified, the asymptotic distribution of $\hat{Q}_W$ does not depend on whether we use W or its consistent estimate W_T as the weighting matrix.

For the de-meaned version of the linear SDF, the traditional sample HJ-distance is defined as

$$\hat{\delta}_{HJ} = \left[\hat{\mu}_2'\hat{U}^{-1}\hat{\mu}_2 - \hat{\mu}_2'\hat{U}^{-1}\hat{V}_{21}(\hat{V}_{12}\hat{U}^{-1}\hat{V}_{21})^{-1}\hat{V}_{12}\hat{U}^{-1}\hat{\mu}_2\right]^{\frac{1}{2}}, \tag{47}$$

where $\hat{U} = \hat{V}_{22} + \hat{\mu}_2\hat{\mu}_2'$ is the sample second moment matrix of the excess returns. Similarly, we define the sample modified HJ-distance as the sample counterpart of (23)

$$\hat{\delta}_m = \left[\hat{\mu}_2'\hat{V}_{22}^{-1}\hat{\mu}_2 - \hat{\mu}_2'\hat{V}_{22}^{-1}\hat{V}_{21}(\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}\hat{V}_{12}\hat{V}_{22}^{-1}\hat{\mu}_2\right]^{\frac{1}{2}}. \tag{48}$$

The squared modified HJ-distance is simply $\hat{Q}_W$ with $W_T = \hat{V}_{22}^{-1}$. It follows that $T\hat{\delta}_m^2$ has a similar asymptotic distribution

$$T\hat{\delta}_m^2 \stackrel{A}{\sim} \sum_{i=1}^{N-K} \xi_i \chi_1^2, \tag{49}$$

where ξ_i are the eigenvalues of $P' V_{22}^{-\frac{1}{2}} S_{22} V_{22}^{-\frac{1}{2}} P$, and P is an $N \times (N-K)$ orthonormal matrix with its columns orthogonal to $V_{22}^{-\frac{1}{2}} V_{21}$. The asymptotic distribution of $\hat{\delta}_{HJ}^2$ under the correctly specified model can be similarly obtained by setting $W = U^{-1}$.

3.2. Asymptotic analysis under misspecified models

To gain a good understanding of the behavior of a sample misspecification measure, we also need to obtain the asymptotic distribution of $\hat{\delta}_m$ and $\hat{\delta}_{HJ}$ under misspecified models. Our approach to solving this problem is the delta method. We note that $\hat{\delta}_m^2$ is just a complicated but smooth function of $\hat{\mu}$ and $\hat{V}$. Therefore, once we have the asymptotic distribution of $\hat{\mu}$ and $\hat{V}$, we can use the delta method to obtain the asymptotic distributions of $\hat{\delta}_m^2$ and $\hat{\delta}_m$. Let

$$\phi = \begin{bmatrix} \mu \\ \mathrm{vec}(V) \end{bmatrix}, \qquad \hat{\phi} = \begin{bmatrix} \hat{\mu} \\ \mathrm{vec}(\hat{V}) \end{bmatrix}. \tag{50}$$

Under some standard regularity conditions, we can assume[10]

$$\sqrt{T}(\hat{\phi} - \phi) \overset{A}{\sim} N(0_{(N+K)\times(N+K+1)}, S_0). \tag{51}$$

Then using the delta method, the asymptotic distributions of $\hat{\delta}_m^2$ and $\hat{\delta}_m$ under the misspecified model are given by

$$\sqrt{T}(\hat{\delta}_m^2 - \delta_m^2) \overset{A}{\sim} N(0, d' S_0 d), \tag{52}$$

$$\sqrt{T}(\hat{\delta}_m - \delta_m) \overset{A}{\sim} N\left(0, \frac{d' S_0 d}{4\delta_m^2}\right), \tag{53}$$

where $d = \partial \delta_m^2 / \partial \phi$. In addition, since there is a monotonic transformation between δ_{HJ} and δ_m as given in Lemma 1, we can also use the delta method to obtain the asymptotic distribution of $\hat{\delta}_{HJ}^2$ and $\hat{\delta}_{HJ}$ under the misspecified model as

$$\sqrt{T}(\hat{\delta}_{HJ}^2 - \delta_{HJ}^2) \overset{A}{\sim} N(0, (1-\delta_{HJ}^2)^4 d' S_0 d), \tag{54}$$

$$\sqrt{T}(\hat{\delta}_{HJ} - \delta_{HJ}) \overset{A}{\sim} N\left(0, \frac{(1-\delta_{HJ}^2)^4 d' S_0 d}{4\delta_{HJ}^2}\right). \tag{55}$$

One may think that Proposition 2.2 of Hansen et al. (1995) (see also equation (44) of Hansen and Jagannathan (1997)) has already presented the asymptotic distribution of $\hat{\delta}_{HJ}^2$ under misspecified

[10] Note that $\hat{V}$ is a symmetric matrix. Therefore, $\hat{\phi}$ contains some redundant elements and S_0 is a singular matrix. We could have written $\hat{\phi}$ as $[\hat{\mu}',\ \mathrm{vech}(\hat{V})']'$, but the results are the same under both specifications.

models and that our equation (54) is simply a restatement of their results. Our results are actually different from the results of Hansen et al. (1995) because their results are only applicable when the mean of the SDF is unconstrained. When the competing SDFs are restricted to have the same mean as in our case, we need to take into account this constraint in deriving the asymptotic standard error of $\hat{\delta}^2_{HJ}$. Simply using the results of Hansen et al. (1995) would give us the wrong asymptotic standard error for $\hat{\delta}^2_{HJ}$.

In order to apply the delta method, we need to obtain the analytical expression of the derivative vector d. We derive and present this expression in the following lemma.

Lemma 3. *Let* $\lambda_m = (V_{12}V_{22}^{-1}V_{21})^{-1}V_{12}V_{22}^{-1}\mu_2$ *and* $e_m = \mu_2 - V_{21}\lambda_m$*, we have*

$$d = \frac{\partial \delta_m^2}{\partial \phi} = \begin{bmatrix} 2 \\ -2\lambda_m \\ -V_{22}^{-1}e_m \end{bmatrix} \otimes \begin{bmatrix} 0_K \\ V_{22}^{-1}e_m \end{bmatrix}. \tag{56}$$

With the analytical expression of d available, we proceed to simplify the asymptotic variance of $\hat{\delta}^2_m$. We first note that $\hat{\mu}$ and $\hat{V}$ can be written as the GMM estimator that uses the moment conditions $E[h_t(\phi)] = 0_{(N+K)(N+K+1)}$, where

$$h_t(\phi) = \begin{bmatrix} Y_t - \mu \\ \text{vec}((Y_t - \mu)(Y_t - \mu)' - V) \end{bmatrix}. \tag{57}$$

Since this is an exactly identified system of moment conditions, it is straightforward to verify that the asymptotic variance of $\hat{\phi}$ is given by

$$S_0 = \sum_{j=-\infty}^{\infty} E[h_t(\phi)h_{t+j}(\phi)']. \tag{58}$$

It follows that

$$\text{Avar}[\hat{\delta}^2_m] = d'S_0 d = \sum_{j=-\infty}^{\infty} E[q_t(\phi)q_{t+j}(\phi)], \tag{59}$$

where

$$q_t(\phi) = d'h_t(\phi) = 2u_t y_t - u_t^2 + \delta_m^2, \tag{60}$$

with $u_t = e_m' V_{22}^{-1}(r_t - \mu_2)$ and $y_t = 1 - \lambda_m'(f_t - \mu_1)$.[11]

[11] Note that equation (60) is for a de-meaned linear SDF. For a general nonlinear SDF y_t that is normalized to have unit mean, q_t in (60) has to be replaced by $q_t = y_t^2 - (y_t - \lambda_1' r_t - \lambda_2)^2 - 2\lambda_2 - \delta_m^2$, where $\lambda_1 = V_{22}^{-1}E[r_t y_t]$ and $\lambda_2 = -\lambda_1'\mu_2$. Proof of this result is available upon request.

In conducting statistical tests, we need a consistent estimate of $\text{Avar}[\hat{\delta}_m^2]$. This can be accomplished by replacing $q_t(\phi)$ with

$$q_t(\hat{\phi}) = 2\hat{u}_t\hat{y}_t - \hat{u}_t^2 + \hat{\delta}_m^2, \tag{61}$$

where $\hat{u}_t = \hat{e}_m'\hat{V}_{22}^{-1}(r_t - \hat{\mu}_2)$, $\hat{y}_t = 1 - \hat{\lambda}_m'(f_t - \hat{\mu}_1)$, $\hat{\lambda}_m = (\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}\hat{V}_{12}\hat{V}_{22}^{-1}\hat{\mu}_2$, and $\hat{e}_m = \hat{\mu}_2 - \hat{V}_{21}\hat{\lambda}_m$. For example, if $q_t(\phi)$ is uncorrelated over time, then we have $\text{Avar}[\hat{\delta}_m^2] = E[q_t^2(\phi)]$, and its consistent estimator is given by[12]

$$\widehat{\text{Avar}}[\hat{\delta}_m^2] = \frac{1}{T}\sum_{t=1}^{T} q_t^2(\hat{\phi}), \tag{62}$$

which is extremely convenient to compute.

4. Asymptotic analysis of the estimates of the stochastic discount factor parameters under potentially misspecified models

In many empirical studies, there is substantial interest in the point estimates of λ. A significant $\hat{\lambda}$ associated with a given factor is often interpreted as evidence that the factor is priced. However, in computing the standard error of $\hat{\lambda}$, researchers typically rely on the asymptotic distribution under the assumption that the model is correctly specified. This practice is somewhat difficult to justify, especially when the model is rejected by the data. In those cases, it is hard to interpret the reported t-ratios and p-values for $\hat{\lambda}$. In order to deal with this problem, we present an analysis of the asymptotic distribution of $\hat{\lambda}$ under potentially misspecified models. A similar asymptotic analysis was presented in Hall and Inoue (2003), Hou and Kimmel (2006), and Shanken and Zhou (2007). Hou and Kimmel (2006) and Shanken and Zhou (2007) derive misspecification robust standard errors for the two-pass cross-sectional regressions estimators under multivariate normality assumptions. Hall and Inoue (2003) derive misspecification robust standard errors for GMM estimators under fairly general assumptions. Our methodology is similar to the one proposed by Hall and Inoue (2003) in the sense that it is free of distributional assumptions. However, given that the GMM for our problem is a recursive one, Theorem 2 of Hall and Inoue (2003) is not directly applicable to our setup and needs to be generalized. In addition, we go beyond Hall and Inoue (2003) by providing a consistent estimator of the asymptotic variance of $\hat{\lambda}$. Finally, we provide an explicit expression of

[12]When $q_t(\phi)$ is autocorrelated, one can use the Newey and West's (1987) method to obtain a consistent estimator of $\text{Avar}[\hat{\delta}_m^2]$.

the asymptotic variance of $\hat{\lambda}$ for the multivariate elliptical case which allows us to show that when one uses the linear de-meaned SDFs and V_{22}^{-1} as the weighting matrix, the misspecification robust standard errors are always bigger than the traditional standard errors that are derived under the correctly specified model. We also explain in detail what determines this difference and conduct an empirical analysis to illustrate the importance of our results.

It is important to emphasize that when a model is misspecified (i.e., μ_2 is not in the span of the column space of V_{21}), λ is no longer unique but is determined by the choice of the weighting matrix. We denote the sample estimators of λ_m and λ_{HJ} defined in Lemma 1 by

$$\hat{\lambda}_m = (\hat{V}_{12}\hat{V}_{22}^{-1}\hat{V}_{21})^{-1}(\hat{V}_{12}\hat{V}_{22}^{-1}\hat{\mu}_2), \quad (63)$$

$$\hat{\lambda}_{HJ} = (\hat{V}_{12}\hat{U}^{-1}\hat{V}_{21})^{-1}(\hat{V}_{12}\hat{U}^{-1}\hat{\mu}_2). \quad (64)$$

Because $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ are just functions of $\hat{\mu}$ and $\hat{V}$, we can use the delta method to obtain

$$\sqrt{T}(\hat{\lambda}_m - \lambda_m) \overset{A}{\sim} N\left(0_K, \left[\frac{\partial \lambda_m}{\partial \phi'}\right] S_0 \left[\frac{\partial \lambda_m}{\partial \phi'}\right]'\right), \quad (65)$$

$$\sqrt{T}(\hat{\lambda}_{HJ} - \lambda_{HJ}) \overset{A}{\sim} N\left(0_K, \left[\frac{\partial \lambda_{HJ}}{\partial \phi'}\right] S_0 \left[\frac{\partial \lambda_{HJ}}{\partial \phi'}\right]'\right), \quad (66)$$

where $S_0 = \text{Avar}[\hat{\phi}]$ and $\hat{\phi} = [\hat{\mu}', \ \text{vec}(\hat{V})']'$. We present the partial derivatives in the following lemma.

Lemma 4. *Let $e_m = \mu_2 - V_{21}\lambda_m$ and $H = (V_{12}V_{22}^{-1}V_{21})^{-1}$, we have*

$$\frac{\partial \lambda_m}{\partial \phi'} = [1, \ -\lambda_m', \ 0_N'] \otimes [0_{K\times K}, \ HV_{12}V_{22}^{-1}] + [0_{K+1}', \ e_m'V_{22}^{-1}] \otimes [H, \ -HV_{12}V_{22}^{-1}], \quad (67)$$

$$\frac{\partial \lambda_{HJ}}{\partial \phi'} = \frac{\frac{\partial \lambda_m}{\partial \phi'} - \lambda_{HJ}\frac{\partial \delta_m^2}{\partial \phi'}}{1+\delta_m^2} = (1-\delta_{HJ}^2)\left(\frac{\partial \lambda_m}{\partial \phi'} - \lambda_{HJ}\frac{\partial \delta_m^2}{\partial \phi'}\right), \quad (68)$$

where $\partial \delta_m^2/\partial \phi'$ is given in Lemma 3.

With this lemma and (58), we can simplify the asymptotic variance of $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ to

$$\text{Avar}[\hat{\lambda}_m] = \sum_{j=-\infty}^{\infty} E[q_t^m(\phi)q_{t+j}^m(\phi)'], \quad (69)$$

$$\text{Avar}[\hat{\lambda}_{HJ}] = \sum_{j=-\infty}^{\infty} E[q_t^{HJ}(\phi)q_{t+j}^{HJ}(\phi)'], \quad (70)$$

where

$$
\begin{aligned}
q_t^m(\phi) &= \frac{\partial \lambda_m}{\partial \phi'} h_t(\phi) \\
&= HV_{12}V_{22}^{-1}(r_t-\mu_2)y_t + H[(f_t-\mu_1)-V_{12}V_{22}^{-1}(r_t-\mu_2)]u_t + \lambda_m, \qquad (71)\\
q_t^{HJ}(\phi) &= \frac{\partial \lambda_{HJ}}{\partial \phi'} h_t(\phi) \\
&= (1-\delta_{HJ}^2)\left[HV_{12}V_{22}^{-1}(r_t-\mu_2)y_t + H[(f_t-\mu_1)-V_{12}V_{22}^{-1}(r_t-\mu_2)]u_t \right.\\
&\quad \left. + \lambda_{HJ}(1-2u_ty_t+u_t^2)\right], \qquad (72)
\end{aligned}
$$

with $u_t = e_m' V_{22}^{-1}(r_t-\mu_2)$ and $y_t = 1-\lambda_m'(f_t-\mu_1)$. Note that when the model is correctly specified, we have $\lambda_m = \lambda_{HJ} = \lambda$, $\delta_{HJ}^2 = 0$, $e_m = 0_N$ and $u_t = 0$. In this case, we have

$$
q_t^m(\phi) = q_t^{HJ}(\phi) = HV_{12}V_{22}^{-1}(r_t-\mu_2)y_t + \lambda \qquad (73)
$$

and both $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ have the same asymptotic distribution. However, when the model is misspecified, the asymptotic distributions of $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ are not the same. When estimating the standard errors of $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$, it is advisable to use the sample counterparts of (71) and (72) instead of the sample counterpart of (73). This is because the latter is only valid when the model is correctly specified whereas the former are valid for both correctly specified and misspecified models.

Note that the impact of the misspecification adjustments on the asymptotic variances of $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ can be relatively unimportant when either u_t is small or $(f_t-\mu_1)-V_{12}V_{22}^{-1}(r_t-\mu_2)$ is small. Since $E[u_t^2] = \delta_m^2$, u_t will be small when the model is close to being correctly specified. For the second case, we note that the term $(f_t-\mu_1)-V_{12}V_{22}^{-1}(r_t-\mu_2)$ is the residual from regressing the factors on the returns. Its magnitude will be small when the factors are well mimicked by the returns.

In order to gain more intuition about the determinants of the misspecification adjustment, we make additional assumptions to further simplify the expressions of $\text{Avar}[\hat{\lambda}_m]$ and $\text{Avar}[\hat{\lambda}_{HJ}]$. In Proposition 3, we present the asymptotic variances of $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ when Y_t is i.i.d. multivariate elliptically distributed.

Proposition 3. *When $Y_t = [f_t',\ r_t']'$ is i.i.d. multivariate elliptically distributed with finite fourth moments, we have*

$$
\text{Avar}[\hat{\lambda}_m] = [1+(1+\kappa)\lambda_m'V_{11}\lambda_m]H + (1+2\kappa)\lambda_m\lambda_m' + (1+\kappa)\delta_m^2 H(V_{11}-V_{12}V_{22}^{-1}V_{21})H, \qquad (74)
$$

$$\text{Avar}[\hat{\lambda}_{HJ}] = (1-\delta_{HJ}^2)^2(\text{Avar}[\hat{\lambda}_m] + \text{Avar}[\hat{\delta}_m^2]\lambda_{HJ}\lambda_{HJ}') - 2(2+3\kappa)\delta_{HJ}^2\lambda_{HJ}\lambda_{HJ}' + 2(1+\kappa)\delta_{HJ}^2(HV_{11}\lambda_{HJ}\lambda_{HJ}' + \lambda_{HJ}\lambda_{HJ}'V_{11}H), \quad (75)$$

where $H = (V_{12}V_{22}^{-1}V_{21})^{-1}$, κ *is the kurtosis parameter of* Y_t *and* $\text{Avar}[\hat{\delta}_m^2] = 4[1+(1+\kappa)\lambda_m'V_{11}\lambda_m]\delta_m^2 + (2+3\kappa)\delta_m^4$.

When the model is correctly specified, we have $\delta_m^2 = \delta_{HJ}^2 = 0$ and $\lambda_m = \lambda_{HJ} = \lambda$, and $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$ have the same asymptotic variance

$$\text{Avar}[\hat{\lambda}_m] = \text{Avar}[\hat{\lambda}_{HJ}] = [1+(1+\kappa)\lambda'V_{11}\lambda]H + (1+2\kappa)\lambda\lambda'. \quad (76)$$

By comparing (74) with (76), we can see that our asymptotic variance of $\hat{\lambda}_m$ is larger than the traditional one by the following positive definite matrix $(1+\kappa)\delta_m^2 H(V_{11} - V_{12}V_{22}^{-1}V_{21})H$, and we call this term the misspecification adjustment. Note that this adjustment is determined by δ_m^2, κ, H and $V_{11} - V_{12}V_{22}^{-1}V_{21}$. As expected, the adjustment is positively related to the squared modified HJ-distance δ_m^2. Hence, the degree of model misspecification plays an important role in determining the magnitude of this adjustment. The adjustment is also positively related to κ which suggests that the fatter the tails of the returns, the larger the adjustment. The final determinant of the adjustment is related to H and $V_{11} - V_{12}V_{22}^{-1}V_{21}$. To understand what these two matrices are about, consider a projection of the factors on the returns (and a constant term) and denote the factor mimicking portfolio by $f^\star = V_{12}V_{22}^{-1}r$. It follows that $H = (V_{12}V_{22}^{-1}V_{21})^{-1} = \text{Var}[f^\star]^{-1}$ and $V_{11} - V_{12}V_{22}^{-1}V_{21} = \text{Var}[f] - \text{Var}[f^\star]$. Hence, these two matrices are both measures of how well the factors can be explained by the excess returns. When the factors are portfolio returns, we can expect these two terms to be small and the misspecification adjustment to be relatively unimportant. However, when the factors are macroeconomic factors, they may have very low correlations with excess returns and $\text{Var}[f^\star]$ may be very small. In those cases, the magnitude of this bias can be huge and model misspecification can have a serious impact on the standard errors of $\hat{\lambda}_m$. Ignoring model misspecification and using the traditional way of computing standard errors (i.e., assuming the model is correct), one can mistakenly conclude that a factor is priced. An extreme case of this is the useless factor model studied by Kan and Zhang (1999a, 1999b), where they find that when using the traditional method of computing standard errors, a useless factor is priced with probability one as T goes to infinity. This is because in the useless factor case, we have

$V_{12} = 0_{K\times N}$. The matrix $H = (V_{12}V_{22}^{-1}V_{21})^{-1}$ explodes and $\text{Avar}[\hat{\lambda}_m] \to \infty$. As a result, $\hat{\lambda}_m$ does not converge to a constant value.

The impact of misspecification on the asymptotic variance of $\hat{\lambda}_{HJ}$ is less clear. The difference between the two matrices in (75) and (76) is not a positive definite matrix. Consequently, it is possible that for some elements of $\hat{\lambda}_{HJ}$ the asymptotic variance increases with misspecification, whereas for other elements of $\hat{\lambda}_{HJ}$ the asymptotic variance decreases with misspecification.

Before moving on, it is important to compare our results with those in Shanken (1985) and Shanken and Zhou (2007) and highlight the additional contribution that we make in Sections 3 and 4. On the modified HJ-distance, we show in (24) that $\hat{\delta}_m^2$ is analogous to the cross-sectional regression test statistic of Shanken (1985). While Shanken provides the asymptotic distribution of this test statistic under the correctly specified model, his derivations are based on the normality assumption. Shanken and Zhou (2007) evaluate the power of this test statistic under some misspecified models using simulations but the asymptotic distribution of the test statistic under misspecified models is not known. We go beyond Shanken (1985) and Shanken and Zhou (2007) by providing the asymptotic distributions of $\hat{\delta}_m^2$ under correctly specified as well as misspecified models and under general distributional assumptions. On the parameter estimates, Shanken and Zhou (2007) provide the asymptotic distribution of the risk premium estimates in the two-pass GLS cross-sectional regression framework under misspecified models but their results are derived under the normality assumption. In contrast, our results on the asymptotic distribution of the SDF parameter estimates are obtained under general distributional assumptions. In addition, we also show that under the elliptical distributional assumption, the standard error of the parameter estimate under misspecification is always larger than the one under the correctly specified model, a result that Shanken and Zhou (2007, p.46) are unable to prove.

5. Empirical examples

The first empirical example is based on Jagannathan and Wang (1996, JW hereafter). JW (1996) propose a conditional CAPM that helps to explain the size and the book-to-market effect. In their Table V, they test their model using the following moment conditions

$$E[r_t(1 - \lambda_{vw}R_t^{vw} - \lambda_{prem}R_{t-1}^{prem} - \lambda_{labor}R_t^{labor})] = 0_N, \tag{77}$$

where r_t are the excess returns on 100 size and beta ranked portfolios, R_t^{vw} is the return on the value-weighted market portfolio, R_{t-1}^{prem} is the yield spread between high and low grade corporate bonds, and R_t^{labor} is the growth rate of per capita income. In Panel A of Table 1, we present the λ estimates and the sample HJ-distances using the same data. The results are largely identical to the ones reported in Table V of JW. In particular, we find that the sample HJ-distance of the JW model has a very low value of 0.1442 and a p-value of 0.965. In the same table, JW also test the Fama-French (1993, FF3 hereafter) three-factor model using the following moment conditions

$$E[r_t(1 - \lambda_{vw} r_t^{vw} - \lambda_{smb} r_t^{smb} - \lambda_{hml} r_t^{hml})] = 0_N, \tag{78}$$

where r_t^{vw} is the excess return on the value-weighted market portfolio, r_t^{smb} is the return difference between portfolios of small and large stocks, and r_t^{hml} is the return difference between portfolios of high and low book-to-market ratios. In Panel A of Table 1, we also present the estimation results of (78). The results are qualitatively similar to the ones reported in JW's Table V. In particular, we find that the FF3 model has a high sample HJ-distance of 0.5494 and a p-value of 0.264. Based on the HJ-distance alone, one would obviously prefer the JW model.

Table 1 about here

The huge difference in HJ-distances between the JW and FF3 models is at odds with other evidence in JW, which for the most part shows similar performance between the two models. As it turns out, the huge difference in HJ-distances is due to the fact that the SDFs in the two models have very different means. When the means of the factors are nonzero, imposing the same intercept on the linear SDFs across models actually forces the means of the SDFs to be very different. In Panel A of Table 1, we report the estimated mean of the SDF for the two models, computed using the sample mean of the factors and the estimated λs. We can now clearly see that the SDF of the JW model has an estimated mean of 0.1228 whereas the SDF of the FF3 model has an estimated mean of 0.9478. As we showed in Proposition 1, the results of the linear SDF can be manipulated by adding and subtracting a constant to the factors. Suppose that we subtract 0.017 from R_t^{vw}, 0.0121 from R_{t-1}^{prem} and 0.059 from R_t^{lab} in the JW model, and we add one to the three factors of the FF3 model. In Panel B, we report the estimation results of the two models using these transformed factors. We now see a dramatic reversal of the HJ-distance comparison. The sample HJ-distance of the JW model is now 0.5832 and none of the three factors are statistically significant. In contrast, the sample HJ-distance of the FF3 model is only 0.0105 with all factors significantly priced. Of course, this does not mean that the FF3 model performs better than the JW model. These results, just like the results in Panel A, are simply unreliable because the means of the SDFs across the two models are vastly different. Another point to note is that the p-value for testing $H_0 : \delta_{HJ} = 0$ is also not invariant to affine transformations of the factors and is subject to manipulation just like the sample HJ-distance.

Before we proceed, we should emphasize that the objective of the paper is not to discredit JW's results or to suggest that their conclusions are invalid. JW's conclusions are not just based on their results in Table V but on many other careful analyses. Our objective is to point out that the results based on linear SDFs are unreliable when only excess returns are used to estimate the models. More importantly, we propose solutions to help researchers deal with this issue.

Knowing that the results of the linear SDF are not invariant to affine transformations of the factors, we now present the estimation results that use the de-meaned version of the linear SDF. In Panel A of Table 2, we present the estimation results of the two models, JW and FF3, using the de-meaned version of the linear SDF. These results have the advantage of being invariant to affine transformations of the factors and both SDFs have the same unit mean. In computing the standard error of $\hat{\lambda}$ and the p-value of $\hat{\delta}_{HJ}$, we rely on the asymptotic results in (41) and (45) with $W = U^{-1}$, where $U = E[rr']$. In our implementation, we replace the population parameters by their sample estimates, and in computing the consistent estimate of S, we assume $g_t(\theta)$ is uncorrelated over time. Using the de-meaned version of the SDF, the sample HJ-distances of the two models no longer differ by a large amount as in Panel A of Table 1. While the JW model still has a slightly smaller sample HJ-distance (0.5624 vs. 0.5726) than the FF3 model, the t-ratios of the coefficient estimates of the model are dramatically lower than the corresponding ones in Panel A of Table 1. This is also true for the FF3 model. Except for $\hat{\lambda}_{prem}$, none of the coefficients of the two models are significantly different from zero. Overall, there is no strong evidence that suggests that the JW model significantly outperforms the FF3 model once we require the SDFs of the two models to have the same mean.

Table 2 about here

The traditional HJ-distance is a measure of how far away the candidate SDF is from the set of all admissible SDFs. As we argue in Section 2.3, when we use the de-meaned factors and force the linear SDFs to have unit mean, it makes sense to also restrict the set of admissible SDFs to have unit mean. The resulting distance measure that we derive, the modified HJ-distance, uses the inverse of the covariance matrix (instead of the second moment matrix) of excess returns as the weighting matrix. In Panel B of Table 2, we report the estimation results of the two models using the de-meaned version of the linear SDF and the inverse of the covariance matrix of excess returns as the weighting matrix. By construction, the modified HJ-distance is larger than the traditional HJ-distance, which is what we observe when we compare $\hat{\delta}_{HJ}$ in Panel A with $\hat{\delta}_m$ in Panel B. Similar to Panel A, the sample modified HJ-distances for the JW and the FF3 models are very close, suggesting that the two models have similar performance. However, the p-values of $\hat{\delta}_m$ show that both the JW and FF3 models are rejected by the data, contrary to the results in Panel A that rely on the traditional sample HJ-distance. Since $\hat{\delta}_{HJ}$ and $\hat{\delta}_m$ are just monotonic transformations of each other, an exact test should give us the same p-value regardless of whether we use $\hat{\delta}_{HJ}$ or $\hat{\delta}_m$ to test the model. The fact that we obtain vastly different test outcomes is an indication of serious problems with using the asymptotic tests.[13]

Although asymptotically both $\hat{\delta}^2_{HJ}$ and $\hat{\delta}^2_m$ have the same distribution under the correctly specified model, the fact that $\hat{\delta}_{HJ} < \hat{\delta}_m$ suggests that in finite samples, the test that uses $\hat{\delta}_m$ will favor rejection whereas the test that uses $\hat{\delta}_{HJ}$ will favor acceptance. A similar problem also exists in $\hat{\lambda}_m$ and $\hat{\lambda}_{HJ}$, but in the opposite direction. Using the same proof as in Lemma 1, we can easily establish that $\hat{\lambda}_{HJ} = \hat{\lambda}_m/(1+\hat{\delta}^2_m)$, which suggests that $\hat{\lambda}_{HJ}$ and $\hat{\lambda}_m$ must have the same sign but the absolute value of $\hat{\lambda}_{HJ}$ is always smaller than the absolute value of $\hat{\lambda}_m$. Although under the correctly specified model, both $\hat{\lambda}_{HJ}$ and $\hat{\lambda}_m$ have the same asymptotic distribution, the fact that $\hat{\lambda}_m$ dominates $\hat{\lambda}_{HJ}$ in every sample suggests that we are more likely to find $\hat{\lambda}_m$ to have greater statistical significance than $\hat{\lambda}_{HJ}$. When we compare the t-ratios of $\hat{\lambda}_{HJ}$ in Panel A with the t-ratios of $\hat{\lambda}_m$ in Panel B, we find exactly this relation. This is an indication that either the models are incorrect or that the asymptotic distributions that we use to compute the standard errors of $\hat{\lambda}$s are inappropriate.

[13] In the statistics literature, it is quite common to find that equivalent asymptotic tests can lead to vastly different outcomes. For example, in testing the uniform linear hypothesis in multivariate regressions, Berndt and Savin (1977) show that the Wald test statistic must be greater than the likelihood ratio test statistic, which in turn is greater than the Lagrange multiplier test statistic, even though all three tests have the same asymptotic distribution.

As in all the existing studies, the standard errors (and p-values) of the $\hat{\lambda}$s in Tables 1 and 2 are computed under the assumption that the model is correctly specified. This assumption is probably hard to justify and using these standard errors to test whether a particular factor is priced can be misleading. Having derived the asymptotic distribution of $\hat{\lambda}$ under a potentially misspecified model in Section 4, it is of interest to see how the inferences are altered with our method of computing standard errors. In Table 3, we report the same estimation results of Table 2, except that the standard errors for $\hat{\lambda}$ are robust to model misspecification. In Section 4, we suggest that misspecification increases the asymptotic variance of $\hat{\lambda}_m$. Therefore, when we account for potential misspecification in the model, the standard error of $\hat{\lambda}_m$ should go up and its t-ratio should be smaller. This is exactly what we observe when we compare the results in Panel B of Tables 2 and 3. Going from Table 2 to Table 3, we see uniformly smaller t-ratios for the two models. For example, the t-ratio of $\hat{\lambda}_{lab}$ goes down from 2.29 to 1.53, suggesting that the growth rate of per capita income is no longer significantly priced.

Table 3 about here

Unlike the standard errors of $\hat{\lambda}_m$, the standard errors of $\hat{\lambda}_{HJ}$ do not uniformly go up after we account for model misspecification. Some t-ratios in Panel A end up being higher while some being lower. Interestingly, once we take into account potential misspecification, both Panels in Table 3 produce roughly the same t-ratios for $\hat{\lambda}$, regardless of whether we use the inverse of the second moment matrix or the covariance matrix as the weighting matrix. This is in sharp contrast with the results in Table 2 which show quite a bit of difference in the t-ratios of $\hat{\lambda}_{HJ}$ and $\hat{\lambda}_m$. This suggests that computing the t-ratios under the assumption that the model is correctly specified is probably the reason why we have far less robust results in Table 2.

In Table 3, we also report the t-ratios for $\hat{\delta}_{HJ}$ and $\hat{\delta}_m$, which are computed using asymptotic standard errors that are valid under misspecified models. Note that since these standard errors are invalid under the correctly specified model, we cannot use the t-ratios of $\hat{\delta}_{HJ}$ and $\hat{\delta}_m$ to test the validity of the model (i.e., the null hypothesis of $H_0 : \delta = 0$). Instead, it is more appropriate to use the standard errors to construct confidence intervals for δ_{HJ} and δ_m.[14] The 95% confidence intervals for δ_{HJ} and δ_m are far away from zero for both models. In addition, the confidence

[14]See Kan and Robotti (2008) for details on how to construct a confidence interval for the HJ-distance.

intervals for δ_{HJ} (or δ_m) for the two models significantly overlap with each other. Therefore, after accounting for sampling variability, we cannot find material difference in terms of the performance of the two models as measured by δ_{HJ} or δ_m.

Although in our first empirical example the use of the de-meaned linear SDF does not alter the ranking of the JW and FF3 models, this is not always the case as we demonstrate using a second empirical example based on Jagannathan and Wang (2007, JW2 hereafter). JW2 advocate an empirical specification of the consumption CAPM (CCAPM) in which annual consumption growth is computed based on the consumption data in the fourth quarter. In their Table V, JW2 compare the performance of their CCAPM with the FF3 model (both SDFs are written as linear functions of the raw factors) using the excess returns on the 25 Fama and French size and book-to-market ranked portfolios. Using the same data, we replicate their results in Panel A of Table 4 below. Note that the quarter four to quarter four consumption growth factor is significantly priced with a t-ratio of 13.36 and the sample HJ-distance of the CCAPM (0.2888) is much smaller than the HJ-distance of the FF3 model (0.6316). In addition, both the CCAPM and the FF3 model are not rejected by the data because the p-values of the sample HJ-distances are 0.70 and 0.11, respectively. Based on these results, JW2 conclude that there is fairly strong empirical support for their CCAPM.

Table 4 about here

As we discussed in the previous example, this kind of comparison can be very misleading because the means of the competing SDFs can be very different. To find out if this is the case, we also report in Panel A the sample means of the SDFs for the CCAPM (0.1948) and the FF3 model (0.6912), which are indeed quite different from each other. In order to overcome this problem, we advocate the use of the de-meaned linear SDF specification as well as the use of the modified HJ-distance. Panel B shows that the results dramatically change when using the de-meaned SDF specification and the inverse of the covariance matrix of excess returns as the weighting matrix. In addition, we report the estimation results under this alternative specification together with a misspecification robust standard error for $\hat{\lambda}$. The results in Panel B show that the modified HJ-distance of the CCAPM is 1.4239, which is higher than the modified HJ-distance of the FF3 model of 1.4048. This completely reverses the JW2's ranking of the two models. In addition, an inspection of the p-values of the modified HJ-distances shows that both models are now significantly rejected by the data, and the t-ratio of the SDF parameter associated with the consumption factor drops from 13.36 to 2.48. In summary, this second empirical example clearly shows that using the de-meaned linear SDF specification and the modified HJ-distance can make a qualitative difference in terms of specification test and in terms of ranking of models.

Our last empirical example is chosen to illustrate the importance of reporting misspecification robust standard errors of the parameter estimates instead of relying on the standard errors of the estimates based on the correctly specified model. We consider six popular asset pricing models in the literature. These are the same models that were considered by Hodrick and Zhang (2001).[15] The SDF of these models are in the form of $y_t = 1 - \lambda'(f_t - E[f_t])$, and they differ in terms of the specification of f_t. These models are estimated using monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate. The returns on the size and book-to-market ranked portfolios are from Kenneth French's web site and the one-month T-bill rate is from Ibbotson Associates (SBBI module). For most of our time series, the data are from 1952/1 to 2006/12 (660 monthly observations).

The first model that we consider is the CAPM which assumes the f_t is r_t^{vw}, where r_t^{vw} is the excess return on the value-weighted combined NYSE-AMEX-NASDAQ index. The second model is a linearized consumption CAPM (CCAPM) which assumes that the f_t is r_t^{cg}, where r_t^{cg} is the monthly growth rate in real nondurables consumption. For the CCAPM, we only have monthly data starting in 1959/2 (575 monthly observations). The third model (JW model) is the conditional CAPM of Jagannathan and Wang (1996) which assumes that $f_t = [r_t^{jvw},\ r_{t-1}^{prem},\ r_t^{lab}]$, where r_t^{jvw} is the return on the valued-weighted combined NYSE-AMEX-NASDAQ index, r_{t-1}^{prem} is the lagged yield spread between BAA and AAA rated corporate bonds, and r_t^{lab} is the growth rate in per capita labor income. The fourth model (CAMP model) is a linearized version of Campbell's (1996) intertemporal capital asset pricing model which assumes that $f_t = [r_t^{rvw},\ r_t^{clab},\ r_{t-1}^{div},\ r_{t-1}^{rtb},\ r_{t-1}^{trm}]$, where r_t^{rvw} is the real return on the CRSP value-weighted index, r_t^{clab} is the monthly growth rate in real labor income (constructed differently from the JW labor series), r_{t-1}^{div} is the dividend yield on the value-weighted market portfolio, r_{t-1}^{rtb} is the difference between the one-month T-bill rate and its one-year backward moving average, and r_{t-1}^{trm} is the yield spread between long and short-term government bonds. For the CAMP model, the data are directly obtained from Campbell and we

[15] See Hodrick and Zhang (2001) for a detailed description of the risk factors of each model.

only have monthly data covering the period 1952/2 to 1990/12 (467 monthly observations). The fifth model (FF3 model) is the Fama-French (1993) three-factor model which assumes that the SDF is $f_t = [r_t^{vw},\ r_t^{smb},\ r_t^{hml}]$, where r_t^{smb} is the return difference between portfolios of small and large stocks and r_t^{hml} is the return difference between portfolios of high and low book-to-market ratios. The Fama-French factors are from Kenneth French's web site. The sixth model (FF5 model) is the Fama-French (1993) five-factor model which assumes that $f_t = [r_t^{vw},\ r_t^{smb},\ r_t^{hml},\ r_t^{term},\ r_t^{def}]$, where r_t^{term} is the return spread between a 30-year Treasury bond and the one-month T-bill and r_t^{def} is the return spread between long-term corporate and long-term government bonds, and both factors are obtained from Ibbotson Associates.

In Table 5, we report parameter estimates $\hat{\lambda}$ and their t-ratios under correctly specified models (t-ratio$_{cs}$) as well as under potentially misspecified models (t-ratio$_m$) for the six models. The parameter estimates $\hat{\lambda}$ in Table 5 are computed using (63), the standard errors under the correctly specified model are computed using (69) and (73), and the misspecification robust standard errors are computed using (69) and (71). Consistent with our theoretical results and our first two empirical examples, the t-ratios under correctly specified and potentially misspecified models are about the same for factors that are portfolio returns. For example, for the CAPM, the t-ratio on the excess market return under the correctly specified model is 3.33, very close to the t-ratio of 3.32 under potential misspecified models. The same type of conclusion holds when considering the FF3 model. However, for models with macroeconomic factors such as the CCAPM, the conclusion of whether consumption growth is significantly priced or not crucially depends on the type of standard error used. When using a misspecification robust standard error, consumption growth is not priced at the conventional 5% level since t-ratio$_m = 1.91$. On the contrary, using the traditional way of computing standard errors would lead us to the conclusion that consumption growth is a priced factor (t-ratio$_{cs} = 2.96$). Similarly, once we use misspecification robust standard errors, the labor factor of the JW model and the labor and *rtb* factors of the CAMP model go from being significantly priced at the 5% level to being insignificant. For macroeconomic factors, the evidence that the standard errors under potentially misspecified models are bigger than the standard errors under correctly specified models is overwhelming. Consequently, we believe it is important to report misspecification robust standard errors when estimating asset pricing models with macroeconomic factors.

Table 5 about here

6. Conclusion

This paper studies specification tests of asset pricing models that are performed using excess returns. We find that the popular specification that writes the SDF as a linear function of the factors is problematic because the outcome of the specification test can be affected by an affine transformation of the factors. In contrast, a less popular version of the linear SDF which writes the SDF as a linear function of the de-meaned factors is free from this problem. We also point out that the traditional HJ-distance is inappropriate when we impose a constraint on the mean of the candidate SDFs, and we propose a modified HJ-distance that is more suited for this purpose. The only difference between the modified HJ-distance and the traditional HJ-distance is that we use the inverse of the covariance matrix rather than the second moment matrix of the excess returns as the weighting matrix. These two HJ-distances have the same asymptotic distribution when the model is correctly specified, but their asymptotic distributions are not the same under misspecified models.

For statistical inference, we provide the asymptotic distributions for both the modified HJ-distance and the traditional HJ-distance based on the de-meaned SDF as well as for the estimates of the SDF parameters. We derive the asymptotic distributions not just for the case of correctly specified models, but also for the case of misspecified models. Another contribution of the paper is to propose a simple method to compute standard errors on the estimates of the SDF parameters that are robust to model misspecification.

Using Jagannathan and Wang (1996) and Jagannathan and Wang (2007) as examples, we illustrate the importance of using the de-meaned version of the linear SDF and demonstrate the substantial differences that one can get by using the modified HJ-distance instead of the traditional HJ-distance. We also show that the misspecification adjustment term in the standard error of the estimate can make a substantial difference in determining whether a macroeconomic factor is priced or not. Failure to take model misspecification into account can understate the standard errors of the estimates of the SDF parameters and lead us to erroneously conclude that certain factors are priced.

Appendix

Proof of Proposition 1. To simplify this problem, we define

$$a = V_{12}W\mu_2(\mu_2'W\mu_2)^{-1}, \tag{A1}$$

$$E = V_{21} - \mu_2 a' = V_{21} - \mu_2(\mu_2'W\mu_2)^{-1}\mu_2'WV_{21}. \tag{A2}$$

Note that when the model is misspecified, E is of full column rank because μ_2 is not in the span of the column space of V_{21}. It is straightforward to show that

$$E'W\mu_2 = (V_{21} - \mu_2 a')'W\mu_2 = V_{12}W\mu_2 - V_{12}W\mu_2 = 0_K. \tag{A3}$$

Therefore, we have

$$B'W\mu_2 = [E + \mu_2(a+\mu_1)']'W\mu_2 = (\mu_2'W\mu_2)(a+\mu_1), \tag{A4}$$

$$B'WB = [E + \mu_2(a+\mu_1)']'W[E + \mu_2(a+\mu_1)'] = E'WE + (\mu_2'W\mu_2)(a+\mu_1)(a+\mu_1)'. \tag{A5}$$

Writing $F = E'WE$, $b = a + \mu_1$, $\eta = \mu_2'W\mu_2$ and using the identity

$$(F + \eta bb')^{-1} = F^{-1} - \frac{F^{-1}bb'F^{-1}}{b'F^{-1}b + \eta^{-1}}, \tag{A6}$$

the objective function (12) can be written as

$$Q_W = \eta - \eta^2 b'\left(F^{-1} - \frac{F^{-1}bb'F^{-1}}{b'F^{-1}b + \eta^{-1}}\right)b = \eta - \eta^2\left[\frac{b'F^{-1}b}{1 + \eta(b'F^{-1}b)}\right] = \frac{\eta}{1 + \eta(b'F^{-1}b)}. \tag{A7}$$

Therefore, maximizing/minimizing Q_W by choosing μ_1 is the same as minimizing/maximizing $b'F^{-1}b$ by choosing b. This is accomplished by choosing $b = 0_K$ (i.e., $\mu_1 = -a$) and $b \to \pm\infty$ (i.e., $\mu_1 \to \pm\infty$), respectively. This completes the proof.

Proof of Proposition 2. The optimization problem in (21) is

$$\begin{aligned} \delta_m^2 &= \min_m E[(y-m)^2] \\ \text{s.t.} \quad E[xm] &= q, \\ E[m] &= c. \end{aligned}$$

Define λ_1 and λ_2 as the Lagrange multipliers of the two equality constraints. We investigate the saddle point problem

$$\begin{aligned} \delta_m^2 &= \min_m \sup_{\lambda_1,\lambda_2} E[(y-m)^2] + 2\lambda_1'E[xm - q] + 2\lambda_2(E[m] - c) \\ &= \max_{\lambda_1,\lambda_2} \min_m E[(y-m)^2] + 2\lambda_1'E[xm - q] + 2\lambda_2(E[m] - c). \end{aligned} \tag{A8}$$

Using the fact that $E[y] = c$, we can write

$$\begin{aligned} & E[(y-m)^2] + 2\lambda_1' E[xm - q] + 2\lambda_2(E[m] - c) \\ = \; & E[(y - \lambda_1' x - \lambda_2 - m)^2] + 2\lambda_1' E[xy - q] + 2\lambda_2 E[y] - \lambda_1' E[xx']\lambda_1 - 2\lambda_1' E[x]\lambda_2 - \lambda_2^2 - 2\lambda_2 c \\ = \; & E[(y - \lambda_1' x - \lambda_2 - m)^2] + 2\lambda_1' E[xy - q] - \lambda_1' E[xx']\lambda_1 - 2\lambda_1' E[x]\lambda_2 - \lambda_2^2 \qquad \text{(A9)} \end{aligned}$$

and only the first term in this expression involves m. Hence, for any λ_1 and λ_2, the inner minimization problem can be solved by choosing

$$m^* = y - \lambda_1' x - \lambda_2, \qquad \text{(A10)}$$

and we are left with a simple maximization problem of

$$\delta_m^2 = \max_{\lambda_1, \lambda_2} 2\lambda_1' E[xy - q] - \lambda_1' E[xx']\lambda_1 - 2\lambda_1' E[x]\lambda_2 - \lambda_2^2. \qquad \text{(A11)}$$

The first order conditions of the maximization problem are

$$\begin{aligned} E[xy - q] - E[xx']\lambda_1 - E[x]\lambda_2 &= 0_N, \qquad \text{(A12)} \\ E[x]'\lambda_1 + \lambda_2 &= 0. \qquad \text{(A13)} \end{aligned}$$

It follows that $\lambda_2 = -E[x]'\lambda_1$ and $\lambda_1 = \text{Var}[x]^{-1}(E[xy] - q) = \text{Var}[x]^{-1}e$. With these optimal λ_1 and λ_2, the optimal choice of m is

$$m^* = y - e'\text{Var}[x]^{-1}(x - E[x]). \qquad \text{(A14)}$$

The squared modified HJ-distance is therefore

$$\delta_m^2 = E[(y - m^*)^2] = E[e'\text{Var}[x]^{-1}(x - E[x])(x - E[x])'\text{Var}[x]^{-1}e] = e'\text{Var}[x]^{-1}e. \qquad \text{(A15)}$$

This completes the proof.

Proof of Lemma 1. Using the identity (A6), we have

$$U^{-1} = (V_{22} + \mu_2\mu_2')^{-1} = V_{22}^{-1} - \frac{V_{22}^{-1}\mu_2\mu_2' V_{22}^{-1}}{1 + \mu_2' V_{22}^{-1}\mu_2}. \qquad \text{(A16)}$$

Using (A16) and denoting $c = \mu_2' V_{22}^{-1}\mu_2$ and $H = (V_{12}V_{22}^{-1}V_{21})^{-1}$, we have

$$\begin{aligned} \mu_2' U^{-1}\mu_2 &= \frac{c}{1+c}, \qquad \text{(A17)} \\ V_{12}U^{-1}\mu_2 &= \frac{V_{12}V_{22}^{-1}\mu_2}{1+c}, \qquad \text{(A18)} \\ (V_{12}U^{-1}V_{21})^{-1} &= H + \frac{H(V_{12}V_{22}^{-1}\mu_2)(\mu_2' V_{22}^{-1}V_{21})H}{1 + \delta_m^2}. \qquad \text{(A19)} \end{aligned}$$

It follows that

$$
\begin{aligned}
\delta_{HJ}^2 &= \mu_2' U^{-1}\mu_2 - \mu_2' U^{-1} V_{21}(V_{12}U^{-1}V_{21})^{-1}V_{12}U^{-1}\mu_2 \\
&= \frac{c}{1+c} - \frac{\mu_2' V_{22}^{-1} V_{21}\left[H + \frac{H(V_{12}V_{22}^{-1}\mu_2)(\mu_2' V_{22}^{-1}V_{21})H}{1+\delta_m^2}\right] V_{12}V_{22}^{-1}\mu_2}{(1+c)^2} \\
&= \frac{c}{1+c} - \frac{(c-\delta_m^2) + \frac{(c-\delta_m^2)^2}{1+\delta_m^2}}{(1+c)^2} \\
&= \frac{c}{1+c} - \frac{(c-\delta_m^2)\frac{1+c}{1+\delta_m^2}}{(1+c)^2} \\
&= \frac{\delta_m^2}{1+\delta_m^2}, \qquad \text{(A20)}
\end{aligned}
$$

where the third equality uses the fact that $\delta_m^2 = c - \mu_2' V_{22}^{-1} V_{21} H V_{12} V_{22}^{-1}\mu_2$. Similarly

$$
\begin{aligned}
\lambda_{HJ} &= \left[H + \frac{H(V_{12}V_{22}^{-1}\mu_2)(\mu_2' V_{22}^{-1}V_{21})H}{1+\delta_m^2}\right]\frac{V_{12}V_{22}^{-1}\mu_2}{1+c} \\
&= \frac{\lambda_m}{1+c} + \frac{\lambda_m(c-\delta_m^2)}{(1+\delta_m^2)(1+c)} \\
&= \frac{\lambda_m}{1+\delta_m^2}. \qquad \text{(A21)}
\end{aligned}
$$

This completes the proof.

Proof of Lemma 2. Defining $e_1 = [1,\ 0_K']'$ and

$$
A = \begin{bmatrix} R_0 & -R_0\mu_1' \\ \mu_2 & -(V_{21} + \mu_2\mu_1') \end{bmatrix}, \qquad \text{(A22)}
$$

we can write the squared HJ-distance based on the $N+1$ assets as

$$
\tilde{\delta}_{HJ}^2 = e_1'[\tilde{U}^{-1} - \tilde{U}^{-1}A(A'\tilde{U}^{-1}A)^{-1}A'\tilde{U}^{-1}]e_1. \qquad \text{(A23)}
$$

Using the partitioned matrix inverse formula, we have

$$
\tilde{U}^{-1} = \begin{bmatrix} \frac{1+\mu_2' V_{22}^{-1}\mu_2}{R_0^2} & -\frac{\mu_2' V_{22}^{-1}}{R_0} \\ -\frac{V_{22}^{-1}\mu_2}{R_0} & V_{22}^{-1} \end{bmatrix} \qquad \text{(A24)}
$$

and it follows that

$$
A'\tilde{U}^{-1}A = \begin{bmatrix} 1 & -\mu_1' \\ -\mu_1 & V_{12}V_{22}^{-1}V_{21} + \mu_1\mu_1' \end{bmatrix}. \qquad \text{(A25)}
$$

Using the partitioned matrix inverse formula again, we have

$$
(A'\tilde{U}^{-1}A)^{-1} = \begin{bmatrix} 1 + \mu_1' H \mu_1 & \mu_1' H \\ H\mu_1 & H \end{bmatrix}, \qquad \text{(A26)}
$$

where $H=(V_{12}V_{22}^{-1}V_{21})^{-1}$. Since $e_1'\tilde{U}^{-1}A=[1/R_0,\ (\mu_2'V_{22}^{-1}V_{21}-\mu_1')/R_0]$, we have

$$\begin{aligned}&\tilde{\delta}^2_{HJ}\\ =\ & \frac{1+\mu_2'V_{22}^{-1}\mu_2}{R_0^2}-\frac{1+\mu_1'H\mu_1+2\mu_1'H(V_{12}V_{22}^{-1}\mu_2-\mu_1)+(V_{12}V_{22}^{-1}\mu_2-\mu_1)'H(V_{12}V_{22}^{-1}\mu_2-\mu_1)}{R_0^2}\\ =\ & \frac{\mu_2'V_{22}^{-1}\mu_2-\mu_2'V_{22}^{-1}V_{21}HV_{12}V_{22}^{-1}\mu_2}{R_0^2}=\frac{\delta_m^2}{R_0^2}. \qquad \text{(A27)}\end{aligned}$$

This completes the proof.

Proof of Lemma 3. It is straightforward to show that

$$\frac{\partial\delta_m^2}{\partial\mu_1} = 0_K, \qquad \text{(A28)}$$

$$\frac{\partial\delta_m^2}{\partial\mu_2} = 2[V_{22}^{-1}-V_{22}^{-1}V_{21}(V_{12}V_{22}^{-1}V_{21})^{-1}V_{12}V_{22}^{-1}]\mu_2=2V_{22}^{-1}e_m. \qquad \text{(A29)}$$

For the derivative of δ_m^2 with respect to $\text{vec}(V)$, we write $\delta_m^2=e_m'V_{22}^{-1}e_m$ and use the product rule to obtain

$$\frac{\partial\delta_m^2}{\partial\text{vec}(V)'}=\frac{\partial e_m'V_{22}^{-1}e_m}{\partial\text{vec}(V)'}=2e_m'V_{22}^{-1}\frac{\partial e_m}{\partial\text{vec}(V)'}+(e_m'\otimes e_m')\frac{\partial\text{vec}(V_{22}^{-1})}{\partial\text{vec}(V)'}. \qquad \text{(A30)}$$

For the first term, we use the product rule and the fact that $V_{12}V_{22}^{-1}e_m=0_K$ to obtain

$$\begin{aligned}2e_m'V_{22}^{-1}\frac{\partial e_m}{\partial\text{vec}(V)'} &= -2e_m'V_{22}^{-1}\frac{\partial V_{21}\lambda_m}{\partial\text{vec}(V)'}\\ &= -2e_m'V_{22}^{-1}\left[(\lambda_m'\otimes I_N)\frac{\partial\text{vec}(V_{21})}{\partial\text{vec}(V)'}+V_{21}\frac{\partial\lambda_m}{\partial\text{vec}(V)'}\right]\\ &= -2(\lambda_m'\otimes e_m'V_{22}^{-1})\frac{\partial\text{vec}(V_{21})}{\partial\text{vec}(V)'}. \qquad \text{(A31)}\end{aligned}$$

Writing $V_{21}=[0_{N\times K},\ I_N]V[I_K,\ 0_{K\times N}]'$, we can simplify the first term to

$$\begin{aligned}2e_m'V_{22}^{-1}\frac{\partial e_m}{\partial\text{vec}(V)'} &= -2(\lambda_m'\otimes e_m'V_{22}^{-1})\frac{\partial([I_K,\ 0_{K\times N}]\otimes[0_{N\times K},\ I_N])\text{vec}(V)}{\partial\text{vec}(V)'}\\ &= -2([\lambda_m',\ 0_N']\otimes[0_K',\ e_m'V_{22}^{-1}]). \qquad \text{(A32)}\end{aligned}$$

For the second term, we use the fact that for a nonsingular matrix A, we have $\partial\text{vec}(A^{-1})/\partial\text{vec}(A)'=-(A^{-1}\otimes A^{-1\prime})$. Using this identity and the chain rule, we have

$$\begin{aligned}(e_m'\otimes e_m')\frac{\partial\text{vec}(V_{22}^{-1})}{\partial\text{vec}(V)'} &= (e_m'\otimes e_m')\frac{\partial\text{vec}(V_{22}^{-1})}{\partial\text{vec}(V_{22})'}\frac{\partial\text{vec}(V_{22})}{\partial\text{vec}(V)'}\\ &= -(e_m'\otimes e_m')(V_{22}^{-1}\otimes V_{22}^{-1})([0_{N\times K},\ I_N]\otimes[0_{N\times K},\ I_N])\\ &= -[0_K',\ e_m'V_{22}^{-1}]\otimes[0_K',\ e_m'V_{22}^{-1}]. \qquad \text{(A33)}\end{aligned}$$

Combining these two terms, we have

$$\frac{\partial \delta_m^2}{\partial \mathrm{vec}(V)} = \begin{bmatrix} -2\lambda_m \\ -V_{22}^{-1}e_m \end{bmatrix} \otimes \begin{bmatrix} 0_K \\ V_{22}^{-1}e_m \end{bmatrix}. \tag{A34}$$

This completes the proof.

Proof of Lemma 4. It is straightforward to show that

$$\begin{aligned} \frac{\partial \lambda_m}{\partial \mu_1'} &= 0_{K\times K}, && \text{(A35)} \\ \frac{\partial \lambda_m}{\partial \mu_2'} &= HV_{12}V_{22}^{-1}. && \text{(A36)} \end{aligned}$$

For the derivative of λ_m with respect to $\mathrm{vec}(V)$, we use the product rule to obtain

$$\frac{\partial \lambda_m}{\partial \mathrm{vec}(V)'} = (\mu_2' V_{22}^{-1}V_{21} \otimes I_K)\frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(V)'} + (\mu_2' V_{22}^{-1} \otimes H)\frac{\partial \mathrm{vec}(V_{12})}{\partial \mathrm{vec}(V)'} + (\mu_2' \otimes HV_{12})\frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'}. \tag{A37}$$

The last two terms were already derived in the proof of Lemma 3 and they are given by

$$\begin{aligned} (\mu_2' V_{22}^{-1} \otimes H)\frac{\partial \mathrm{vec}(V_{12})}{\partial \mathrm{vec}(V)'} &= [0_K',\ \mu_2' V_{22}^{-1}] \otimes [H,\ 0_{K\times N}], && \text{(A38)} \\ (\mu_2' \otimes HV_{12})\frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'} &= -[0_K',\ \mu_2' V_{22}^{-1}] \otimes [0_{K\times K},\ HV_{12}V_{22}^{-1}]. && \text{(A39)} \end{aligned}$$

For the first term, we use the chain rule to obtain

$$\begin{aligned} & (\mu_2' V_{22}^{-1}V_{21} \otimes I_K)\frac{\partial \mathrm{vec}(H)}{\partial \mathrm{vec}(V)'} \\ = \ & (\mu_2' V_{22}^{-1}V_{21} \otimes I_K)\frac{\partial \mathrm{vec}((V_{12}V_{22}^{-1}V_{21})^{-1})}{\partial \mathrm{vec}(V_{12}V_{22}^{-1}V_{21})'}\frac{\partial \mathrm{vec}(V_{12}V_{22}^{-1}V_{21})}{\partial \mathrm{vec}(V)'} \\ = \ & -(\mu_2' V_{22}^{-1}V_{21} \otimes I_K)(H\otimes H)\left[(V_{12}V_{22}^{-1} \otimes I_K)\frac{\partial \mathrm{vec}(V_{12})}{\partial \mathrm{vec}(V)'}\right. \\ & \left. + (V_{12}\otimes V_{12})\frac{\partial \mathrm{vec}(V_{22}^{-1})}{\partial \mathrm{vec}(V)'} + (I_K \otimes V_{12}V_{22}^{-1})\frac{\partial \mathrm{vec}(V_{21})}{\partial \mathrm{vec}(V)'}\right] \\ = \ & -(\lambda_m' \otimes H)\left[[0_{K\times K},\ V_{12}V_{22}^{-1}] \otimes [I_K,\ 0_{K\times N}]\right. \\ & \left. - [0_{K\times K},\ V_{12}V_{22}^{-1}] \otimes [0_{K\times K},\ V_{12}V_{22}^{-1}] + [I_K,\ 0_{K\times N}] \otimes [0_{K\times K},\ V_{12}V_{22}^{-1}]\right] \\ = \ & -[0_K',\ \lambda_m' V_{12}V_{22}^{-1}] \otimes [H,\ 0_{K\times N}] + [0_K',\ \lambda_m' V_{12}V_{22}^{-1}] \otimes [0_{K\times K},\ HV_{12}V_{22}^{-1}] \\ & - [\lambda_m',\ 0_N'] \otimes [0_{K\times K},\ HV_{12}V_{22}^{-1}]. \end{aligned} \tag{A40}$$

Combining the three terms and using the identity $e_m = \mu_2 - V_{21}\lambda_m$, we have

$$\frac{\partial \lambda_m}{\partial \mathrm{vec}(V)'} = [-\lambda_m',\ 0_N'] \otimes [0_{K\times K},\ HV_{12}V_{22}^{-1}] + [0_K',\ e_m' V_{22}^{-1}] \otimes [H,\ -HV_{12}V_{22}^{-1}]. \tag{A41}$$

Finally, (68) is obtained by using the relation $\lambda_{HJ} = \lambda_m/(1+\delta_m^2)$ given in Lemma 1. This completes the proof.

Proof of Proposition 3. From Muirhead (1982, p. 42, p. 49), we know that when Y_t follows a multivariate elliptical distribution with finite fourth moments, we have

$$\mathrm{Acov}[\hat{V}_{ij}, \hat{V}_{kl}] = \kappa V_{ij}V_{kl} + (1+\kappa)(V_{ik}V_{jl} + V_{il}V_{jk}). \tag{A42}$$

Using this and the symmetric property of multivariate elliptical distribution, we can write $S_0 = \mathrm{Avar}[\hat{\phi}]$ compactly as

$$S_0 = \begin{bmatrix} V & 0_{p\times p^2} \\ 0_{p^2\times p} & (1+\kappa)(I_{p^2}+K_p)(V\otimes V) + \kappa\mathrm{vec}(V)\mathrm{vec}(V)' \end{bmatrix}, \tag{A43}$$

where $p = N+K$ and K_p is a $p^2\times p^2$ commutation matrix such that $K_p\mathrm{vec}(A) = \mathrm{vec}(A')$ for a $p\times p$ matrix A. Denoting $A_1 = [-\lambda_m',\ 0_N']\otimes[0_{K\times K},\ HV_{12}V_{22}^{-1}]$ and $A_2 = [0_K',\ e_m'V_{22}^{-1}]\otimes[H,\ -HV_{12}V_{22}^{-1}]$ and using the identity $V_{12}V_{22}^{-1}e_m = 0_K$, it is easy to verify the following identities

$$\begin{aligned} (A_1+A_2)(V\otimes V)(A_1+A_2)' &= (\lambda_m'V_{11}\lambda_m)H + \delta_m^2 H(V_{11} - V_{12}V_{22}^{-1}V_{21})H, && \text{(A44)} \\ (A_1+A_2)K_p(V\otimes V)(A_1+A_2)' &= \lambda_m\lambda_m', && \text{(A45)} \\ (A_1+A_2)\mathrm{vec}(V) &= -\lambda_m. && \text{(A46)} \end{aligned}$$

It follows that

$$\begin{aligned} & \frac{\partial\lambda_m}{\partial\phi'}S_0\left[\frac{\partial\lambda_m}{\partial\phi'}\right]' \\ = \ & [0_{K\times K},\ HV_{12}V_{22}^{-1}]V[0_{K\times K},\ HV_{12}V_{22}^{-1}]' \\ & + (A_1+A_2)[(1+\kappa)(I_{p^2}+K_p)(V\otimes V) + \kappa\mathrm{vec}(V)\mathrm{vec}(V)'](A_1+A_2)' \\ = \ & H + (1+\kappa)[(A_1+A_2)(V\otimes V)(A_1+A_2)' + (A_1+A_2)K_p(V\otimes V)(A_1+A_2)'] \\ & + \kappa(A_1+A_2)\mathrm{vec}(V)\mathrm{vec}(V)'(A_1+A_2)' \\ = \ & H + (1+\kappa)[(\lambda_m'V_{11}\lambda_m)H + \delta_m^2H(V_{11} - V_{12}V_{22}^{-1}V_{21})H + \lambda_m\lambda_m'] + \kappa\lambda_m\lambda_m' \\ = \ & [1+(1+\kappa)\lambda_m'V_{11}\lambda_m]H + (1+2\kappa)\lambda_m\lambda_m' + (1+\kappa)\delta_m^2H(V_{11} - V_{12}V_{22}^{-1}V_{21})H. \qquad \text{(A47)} \end{aligned}$$

Using (68), $\mathrm{Avar}[\hat{\lambda}_{HJ}]$ is given by

$$(1-\delta_{HJ}^2)^2\left(\mathrm{Avar}[\hat{\lambda}_m] + \mathrm{Avar}[\hat{\delta}_m^2]\lambda_{HJ}\lambda_{HJ}' - \frac{\partial\lambda_m}{\partial\phi'}S_0\frac{\partial\delta_m^2}{\partial\phi}\lambda_{HJ}' - \lambda_{HJ}\frac{\partial\delta_m^2}{\partial\phi'}S_0\left[\frac{\partial\lambda_m}{\partial\phi'}\right]'\right). \tag{A48}$$

The only two terms that we need to obtain are $\mathrm{Avar}[\hat{\delta}_m^2]$ and $(\partial\lambda_m/\partial\phi')S_0(\partial\delta_m^2/\partial\phi)$. For the first term, since Y_t is multivariate elliptically distributed, u_t and y_t are bivariate elliptically distributed because both of them are linear combinations of the elements of Y_t. Using the properties of the multivariate elliptical distribution (see Muirhead (1982, p. 41)), we have $E[u_t] = 0$, $E[u_t^2] = e_m' V_{22}^{-1} e_m = \delta_m^2$, $E[u_t^3] = 0$, $E[u_t^4] = 3(1+\kappa)E[u_t^2]^2 = 3(1+\kappa)\delta_m^4$, $E[y_t^2] = 1 + \lambda_m' V_{11}\lambda_m$, where κ is the kurtosis parameter of the elliptical distribution. In addition, using the identity $V_{12}V_{22}^{-1}e_m = 0_K$, we have

$$E[u_t y_t] = E[u_t] - E[e_m' V_{22}^{-1}(r_t - \mu_2)(f_t - \mu_1)'\lambda_m] = 0 - e_m' V_{22}^{-1} V_{21}\lambda_m = 0, \tag{A49}$$

implying that u_t and y_t are uncorrelated. It follows that $E[u_t^2 y_t^2] = (1+\kappa)\delta_m^2 \lambda_m' V_{11}\lambda_m + \delta_m^2$ and $E[u_t^3 y_t] = 0$. Using these moments of u_t and y_t, we have

$$\begin{aligned} \mathrm{Avar}[\hat{\delta}_m^2] = E[q_t^2(\phi)] &= 4E[u_t^2 y_t^2] + E[u_t^4] + \delta_m^4 - 4E[u_t^3 y_t] - 4E[u_t y_t]\delta_m^2 - 2E[u_t^2]\delta_m^2 \\ &= 4[1 + (1+\kappa)\lambda_m' V_{11}\lambda_m]\delta_m^2 + (2+3\kappa)\delta_m^4. \end{aligned} \tag{A50}$$

For the second term, let $A_3 = [-2\lambda_m',\ -e_m' V_{22}^{-1}] \otimes [0_K',\ e_m' V_{22}^{-1}]$. It is easy to verify that

$$(A_1 + A_2)(V \otimes V)A_3' = 0_K, \tag{A51}$$

$$(A_1 + A_2)K_p(V \otimes V)A_3' = 2\delta_m^2\lambda_m - 2\delta_m^2 H V_{11}\lambda_m, \tag{A52}$$

$$A_3\mathrm{vec}(V) = -\delta_m^2. \tag{A53}$$

It follows that

$$\begin{aligned} \frac{\partial\lambda_m}{\partial\phi'} S_0 \frac{\partial\delta_m^2}{\partial\phi} &= [0_{K\times K},\ H V_{12} V_{22}^{-1}]V[0_K',\ 2e_m' V_{22}^{-1}]' \\ &\quad + (A_1 + A_2)[(1+\kappa)(I_{p^2} + K_p)(V\otimes V) + \kappa\mathrm{vec}(V)\mathrm{vec}(V)']A_3' \\ &= 0_{K\times K} + (1+\kappa)[(A_1 + A_2)(V\otimes V)A_3' + (A_1+A_2)K_p(V\otimes V)A_3'] \\ &\quad + \kappa(A_1 + A_2)\mathrm{vec}(V)\mathrm{vec}(V)'A_3' \\ &= (2+3\kappa)\delta_m^2\lambda_m - 2(1+\kappa)\delta_m^2 H V_{11}\lambda_m \\ &= \frac{(2+3\kappa)\delta_{HJ}^2\lambda_{HJ} - 2(1+\kappa)\delta_{HJ}^2 H V_{11}\lambda_{HJ}}{(1-\delta_{HJ}^2)^2}. \end{aligned} \tag{A54}$$

Substituting (A50) and (A54) into (A48), we obtain our expression of $\mathrm{Avar}[\hat{\lambda}_{HJ}]$. This completes the proof.

References

Balduzzi, P., Kallal, H., 1997. Risk premia and variance bounds. Journal of Finance 52, 1913–1949.

Balduzzi, P., Robotti, C., 2008. Asset pricing models and economic risk premia: a decomposition. Working Paper. Boston College and Federal Reserve Bank of Atlanta.

Balduzzi, P., Yao, T., 2007. Testing heterogeneous-agent models: an alternative aggregation approach. Journal of Monetary Economics 54, 369–412.

Berndt, E. R., Savin, N. E., 1977. Conflict among criteria for testing hypotheses in the multivariate linear regression model. Econometrica 45, 1263–1278.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–454.

Burnside, C., 2007. Empirical asset pricing and statistical power in the presence of weak risk factors. Working Paper. Duke University.

Campbell, J. Y., 1996. Understanding risk and return. Journal of Political Economy 104, 298–345.

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

Cochrane, J. H., 2005. Asset Pricing. Princeton University Press, New Jersey.

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

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

Farnsworth, H., Ferson, W. E., Jackson, D., Todd, S., 2002, Performance evaluation with stochastic discount factors. Journal of Business 75, 473–503.

Ferson, W. E., Foerster, S. R., Keim, D., 1993. General tests of latent variable models and mean-variance spanning. Journal of Finance 48, 131–156.

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Hall, A. R., Inoue, A., 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. Journal of Econometrics 114, 361–394.

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

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

Hansen, L. P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

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

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

Hou, K., Kimmel, R., 2006. On the estimation of risk premia in linear factor models. Working Paper. Ohio State University.

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

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

Jagannathan, R., Wang, Z., 2002. Empirical evaluation of asset pricing models: a comparison of the SDF and beta methods. Journal of Finance 57, 2337–2367.

Jagannathan, R., Wang, Y., 2007. Lazy investors, discretionary consumption, and the cross-section of stock returns. Journal of Finance 62, 1623–1661.

Kan, R., Robotti, C., 2008. Model comparison using the Hansen-Jagannathan distance. Working Paper. University of Toronto and Federal Reserve Bank of Atlanta.

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

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

Kan, R., Zhou, G., 1999. A critique of the stochastic discount factor methodology. Journal of Finance 54, 1221–1248.

Kan, R., Zhou, G., 2001. Empirical asset pricing: the beta method versus the stochastic discount factor method. Working Paper. University of Toronto and Washington University in St. Louis.

Kan, R., Zhou, G., 2004. Hansen-Jagannathan distance: geometry and exact distribution. Working Paper. University of Toronto and Washington University in St. Louis.

Kirby, C., 1998. The restrictions on predictability implied by rational asset pricing models. Review of Financial Studies 11, 343–382.

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

Muirhead, R. J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

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

Shanken, J., 1985. Multivariate tests of the zero-beta CAPM. Journal of Financial Economics 14, 327–348.

Shanken, J., Zhou, G., 2007. Estimating and testing beta pricing models: alternative methods and their performance in simulations. Journal of Financial Economics 84, 40–86.

Table 1
A comparison of the performance of the Jagannathan and Wang (1996) and Fama and French (1993) models on 100 size-beta sorted portfolios using a linear specification of the stochastic discount factor

Panel A: Original factors

	JW model				FF3 model			
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\delta}_{HJ}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{HJ}$
Estimate	0.09	48.18	59.92	0.1442	3.31	1.02	8.96	0.5494
t-ratio	0.25	13.11	9.25		2.31	0.51	3.19	
p-value	0.805	0.000	0.000	0.965	0.021	0.614	0.001	0.264
Estimate of $E[y]$	0.1228				0.9478			

Panel B: Transformed factors

	JW model				FF3 model			
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\delta}_{HJ}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{HJ}$
Estimate	0.03	−0.24	2.70	0.5832	0.20	0.14	0.65	0.0136
t-ratio	0.02	−0.01	0.09		7.05	3.37	18.09	
p-value	0.984	0.991	0.931	0.140	0.000	0.008	0.000	0.011
Estimate of $E[y]$	1.0005				0.0050			

The table presents the estimation results of two asset pricing models. The first model (JW) is from Jagannathan and Wang (1996), which assumes that the stochastic discount factor is

$$y_t = 1 - \lambda_{vw} R_t^{vw} - \lambda_{prem} R_{t-1}^{prem} - \lambda_{lab} R_t^{lab},$$

where R_t^{vw} is the return on the CRSP value-weighted index, R_{t-1}^{prem} is the yield spread between low and high-grade corporate bonds, and R_t^{lab} is the growth rate in per capita income. The second model (FF3 model) is from Fama and French (1993), which assumes that the stochastic discount factor is

$$y_t = 1 - \lambda_{vw} r_t^{vw} - \lambda_{smb} r_t^{smb} - \lambda_{hml} r_t^{hml},$$

where r_t^{vw} is the excess return (in excess of 1-month T-bill rate) on the CRSP value-weighted index, r_t^{smb} is the return difference between portfolios of small and large stocks, and r_t^{hml} is the return difference between portfolios of high and low book-to-market ratios. The models are estimated using monthly excess returns on 100 size and beta sorted portfolios of the NYSE and AMEX over the period 1963/7–1990/12. Panel A reports the estimates of the λ and the HJ-distances (δ_{HJ}) for the two models. In addition, it reports the estimated means of the two stochastic discount factors. Panel B reports the estimation results of the two models after performing an affine transformation of the original factors. For the JW model, we subtract 0.0171, 0.0121, and 0.0059 from R_t^{vw}, R_{t-1}^{prem} and R_t^{lab}, respectively. For the FF3 model, we add one to all three factors. The standard errors of $\hat{\lambda}$ are computed assuming no serial correlation. The p-values for $\hat{\lambda}$ are two-tailed p-values.

Table 2
A comparison of the performance of the Jagannathan and Wang (1996) and Fama and French (1993) models on 100 size-beta sorted portfolios using a linear de-meaned specification of the stochastic discount factor

Panel A: Traditional HJ-distance

	JW model				**FF3 model**			
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\delta}_{HJ}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{HJ}$
Estimate	1.07	50.77	51.74	0.5624	2.00	1.01	4.75	0.5726
t-ratio	0.82	2.46	1.63		1.33	0.48	1.57	
p-value	0.410	0.014	0.103	0.441	0.184	0.630	0.117	0.225

Panel B: Modified HJ-distance

	JW model				**FF3 model**			
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\delta}_{m}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{m}$
Estimate	1.56	74.26	75.68	0.6802	2.98	1.51	7.06	0.6984
t-ratio	1.16	3.53	2.29		1.93	0.71	2.30	
p-value	0.246	0.000	0.022	0.016	0.054	0.475	0.022	0.000

The table presents the estimation results of two asset pricing models. The first model (JW) is from Jagannathan and Wang (1996), which assumes that the stochastic discount factor is

$$y_t = 1 - \lambda_{vw}(R_t^{vw} - E[R_t^{vw}]) - \lambda_{prem}(R_{t-1}^{prem} - E[R_{t-1}^{prem}]) - \lambda_{lab}(R_t^{lab} - E[R_t^{lab}]),$$

where R_t^{vw} is the return on the CRSP value-weighted index, R_{t-1}^{prem} is the yield spread between low and high-grade corporate bonds, and R_t^{lab} is the growth rate in per capita income. The second model (FF3 model) is from Fama and French (1993), which that assumes the stochastic discount factor is

$$y_t = 1 - \lambda_{vw}(r_t^{vw} - E[r_t^{vw}]) - \lambda_{smb}(r_t^{smb} - E[r_t^{smb}]) - \lambda_{hml}(r_t^{hml} - E[r_t^{hml}]),$$

where r_t^{vw} is the excess return (in excess of 1-month T-bill rate) on the CRSP value-weighted index, r_t^{smb} is the return difference between portfolios of small and large stocks, and r_t^{hml} is the return difference between portfolios of high and low book-to-market ratios. The models are estimated using monthly excess returns on 100 size and beta sorted portfolios of the NYSE and AMEX over the period 1963/7–1990/12. Panel A reports the estimates of the λ and the traditional HJ-distance (δ_{HJ}) for the two models using the inverse of the second moment matrix of excess returns as the weighting matrix. Panel B reports the estimation results of the λ and the modified HJ-distance (δ_m) for the two models using the inverse of the covariance matrix of excess returns as the weighting matrix. The standard errors of $\hat{\lambda}$ are computed assuming no serial correlation. The p-values for $\hat{\lambda}$ are two-tailed p-values.

Table 3
A comparison of the performance of the Jagannathan and Wang (1996) and Fama and French (1993) models on 100 size-beta sorted portfolios using a linear de-meaned specification of the stochastic discount factor and misspecification robust standard errors

Panel A: Traditional HJ-distance

	JW model				**FF3 model**			
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\delta}_{HJ}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{HJ}$
Estimate	1.07	50.77	51.74	0.5624	2.00	1.01	4.75	0.5726
t-ratio	1.12	2.22	1.50	14.92	1.84	0.69	2.04	16.14
p-value	0.261	0.026	0.134		0.066	0.490		
95% conf. interval for δ_{HJ}		(0.4311, 0.7603)					(0.4439, 0.7646)	

Panel B: Modified HJ-distance

	JW model				**FF3 model**			
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\delta}_{m}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{m}$
Estimate	1.56	74.26	75.68	0.6802	2.98	1.51	7.06	0.6984
t-ratio	1.12	2.32	1.53	10.20	1.85	0.69	2.07	10.85
p-value	0.263	0.021	0.127		0.064	0.492	0.039	
95% conf. interval for δ_m		(0.5655, 0.8311)					(0.5870, 0.8433)	

The table presents the estimation results of two asset pricing models. The first model (JW) is from Jagannathan and Wang (1996), which assumes that the stochastic discount factor is

$$y_t = 1 - \lambda_{vw}(R_t^{vw} - E[R_t^{vw}]) - \lambda_{prem}(R_{t-1}^{prem} - E[R_{t-1}^{prem}]) - \lambda_{lab}(R_t^{lab} - E[R_t^{lab}]),$$

where R_t^{vw} is the return on the CRSP value-weighted index, R_{t-1}^{prem} is the yield spread between low and high-grade corporate bonds, and R_t^{lab} is the growth rate in per capita income. The second model (FF3 model) is from Fama and French (1993), which assumes the stochastic discount factor is

$$y_t = 1 - \lambda_{vw}(r_t^{vw} - E[r_t^{vw}]) - \lambda_{smb}(r_t^{smb} - E[r_t^{smb}]) - \lambda_{hml}(r_t^{hml} - E[r_t^{hml}]),$$

where r_t^{vw} is the excess return (in excess of 1-month T-bill rate) on the CRSP value-weighted index, r_t^{smb} is the return difference between portfolios of small and large stocks, and r_t^{hml} is the return difference between portfolios of high and low book-to-market ratios. The models are estimated using monthly excess returns on 100 size and beta sorted portfolios of the NYSE and AMEX over the period 1963/7–1990/12. Panel A reports the estimates of the λ and the traditional HJ-distance (δ_{HJ}) for the two models using the inverse of the second moment matrix of excess returns as the weighting matrix. Panel B reports the estimation results of the λ and the modified HJ-distance (δ_m) for the two models using the inverse of the covariance matrix of excess returns as the weighting matrix. The reported t-ratios and p-values are robust to model misspecification. The standard errors of $\hat{\lambda}$ are computed assuming no serial correlation. The p-values for $\hat{\lambda}$ are two-tailed p-values. In addition, Panels A and B report the 95% confidence intervals for δ_{HJ} and δ_m, respectively.

Table 4

A comparison of the performance of the Jagannathan and Wang (2007) and Fama and French (1993) models on 25 size and book-to-market sorted portfolios using the raw and the de-meaned linear specifications of the stochastic discount factor

Panel A: Traditional HJ-distance with linear specification

	CCAPM		**FF3 model**			
	$\hat{\lambda}_{cg}$	$\hat{\delta}_{HJ}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_{HJ}$
Estimate	33.01	0.2888	1.90	0.56	2.61	0.6316
t-ratio	13.36		3.26	0.68	3.31	
p-value	0.000	0.702	0.001	0.497	0.001	0.108
Estimate of $E[y]$	0.1948		0.6912			

Panel B: Modified HJ-distance with de-meaned linear specification

	CCAPM		**FF3 model**			
	$\hat{\lambda}_{cg}$	$\hat{\delta}_m$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\delta}_m$
Estimate	69.98	1.4239	2.87	0.81	3.60	1.4048
t-ratio	2.48		2.74	0.67	2.48	
p-value	0.013	0.000	0.006	0.502	0.013	0.000

The table presents the estimation results of two asset pricing models: the CCAPM and the FF3 model. The models are estimated using value-weighted annual real excess returns on the Fama-French 25 portfolios. Per capita consumption growth is denoted by cg, while the three Fama-French factors are vw (excess market returns), smb (small minus big) and hml (high minus low). All factors are in real terms. The sample period is 1954–2003. Panel A reports the estimates of the λ and the traditional HJ-distance (δ_{HJ}) for the two models using the raw SDF specification and the inverse of the second moment matrix of excess returns as the weighting matrix. In addition, Panel A reports the estimated means of the two SDFs. Panel B reports the estimation results of the λ and the modified HJ-distance (δ_m) for the two models using the inverse of the covariance matrix of excess returns as the weighting matrix. The reported t-ratios and p-values in Panel B are robust to model misspecification, while the ones in Panel A are not. The standard errors of $\hat{\lambda}$ are computed assuming no serial correlation. The p-values for $\hat{\lambda}$ are two-tailed p-values.

Table 5
Parameter estimates and t-ratios for different de-meaned linear stochastic discount factors under correctly specified and potentially misspecified models

	CAPM	CCAPM
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{cg}$
Estimate	3.39	73.18
t-ratio$_{cs}$	3.33	2.96
t-ratio$_{m}$	3.32	1.91

	JW model			CAMP model				
	$\hat{\lambda}_{jvw}$	$\hat{\lambda}_{prem}$	$\hat{\lambda}_{lab}$	$\hat{\lambda}_{rvw}$	$\hat{\lambda}_{clab}$	$\hat{\lambda}_{div}$	$\hat{\lambda}_{rtb}$	$\hat{\lambda}_{trm}$
Estimate	1.79	72.83	−162.54	−0.15	1.04	−43.13	−23.27	−19.69
t-ratio$_{cs}$	1.46	1.35	−2.89	−0.89	2.28	−0.85	−2.40	−1.70
t-ratio$_{m}$	1.27	0.66	−1.39	−0.55	1.43	−0.54	−1.79	−1.20

	FF3 model			FF5 model				
	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\lambda}_{vw}$	$\hat{\lambda}_{smb}$	$\hat{\lambda}_{hml}$	$\hat{\lambda}_{term}$	$\hat{\lambda}_{def}$
Estimate	4.99	2.47	8.91	2.59	2.76	7.39	11.23	20.98
t-ratio$_{cs}$	4.37	1.68	5.60	1.24	1.41	3.61	1.43	0.85
t-ratio$_{m}$	4.37	1.68	5.61	0.64	0.91	2.06	0.73	0.34

The table presents the estimation results of six asset pricing models: the capital asset pricing model (CAPM), the consumption-based capital asset pricing model (CCAPM), the conditional CAPM of Jagannathan and Wang (JW model), the five-factor model of Campbell (CAMP model), the Fama-French three-factor model (FF3 model), and the Fama-French five-factor model (FF5 model). The set of test assets includes the monthly returns on the 25 Fama-French size and book-to-market ranked portfolios in excess of the one-month T-bill rate. Most of the data are from 1952/1 to 2006/12, but the data for the C-CAPM model start in 1959/2 and the data for the CAMP model cover only the period 1952/2–1990/12. We report parameter estimates $\hat{\lambda}$ and their t-ratios under correctly specified models (t-ratio$_{cs}$) as well as under potentially misspecified models (t-ratio$_{m}$). The estimation is based on GMM using the inverse of the covariance matrix of the excess returns as the weighting matrix and the t-ratios are computed assuming no serial correlation.