Abstract

Jagannathan and Wang (1996) derive the asymptotic distribution of the Hansen-Jagannathan distance (HJ-distance) proposed by Hansen and Jagannathan (1997) and develop a specification test of asset pricing models based on the HJ-distance. While the HJ-distance has several desirable properties, Ahn and Gadarowski (2004) find that the specification test based on the HJ-distance overrejects correct models too severely in commonly used sample size to provide a valid test. This paper proposes to improve the small sample properties of the HJ-distance by applying the shrinkage method (Ledoit and Wolf, 2003) to compute its weighting matrix. The proposed method improves the finite sample performance of the HJ-distance significantly. With this method, the rejection frequencies are close to their nominal levels in most practically relevant cases.
1 Introduction

In asset pricing theory, a stochastic discount factor (SDF) is defined to be a random variable which can be used to compute market prices today by discounting the corresponding returns at a future date. Asset pricing models imply stochastic discount factors, and they describe portfolio returns from a different point of view. Asset pricing models try to reveal how portfolio returns are determined and which factors affect returns, while stochastic discount factors are used to price returns, displaying which prices are reasonable given today’s returns.

In reality, since no asset pricing models are strictly true, stochastic discount factors cannot price portfolios perfectly in general. Therefore, it is of importance to construct a measure of pricing errors produced by the SDFs which enables us to compare different SDFs and the asset pricing models that imply them. Hansen and Jagannathan (1997) develop a distance measure, called HJ-distance. This measure equals the maximum pricing error generated by a model for portfolios with unit second moment. It is also the least-squares distance between a stochastic discount factor and the family of SDFs that price portfolios correctly. The HJ-distance can be estimated by GMM using the inverse of the second moment matrix of returns as the weighting matrix.

In contrast to comparisons based on chi-square statistics, such as the Hansen statistic, the HJ-distance has several desirable properties: first of all, it does not reward variability of SDF. The weighting matrix used in the HJ-distance is the second moment of portfolio returns and independent of pricing errors. On the other hand, the Hansen statistic uses the inverse of the second moment of the pricing errors as the weighting matrix and rewards models with high variability of pricing errors. Second, as Jagannathan and Wang (1996) point out, the weighting matrix of the HJ-distance remains the same across various pricing models, which makes it possible to compare the performances among competitive SDFs by the relative values of the HJ-distances. But unlike Hansen statistic, the HJ-distance does not follow a chi-squared distribution asymptotically. Instead, Jagannathan and Wang (1996) show that, for linear factor models, the HJ-distance is asymptotically distributed as a weighted chi-squared distribution. In addition, they suggest a simulation method to develop the empirical p-value of the HJ-distance.

The HJ-distance has been applied widely in financial studies. Typically, when a new model is proposed, the HJ-distance is employed to compare the new model with alternative models.
Hereby, the model can be supported if it offers small pricing errors. This type of comparison has been employed in many recent papers. For instance, by using the HJ-distance, Jagannathan and Wang (1998) discuss cross-sectional regression model; Kan and Zhang (1999) study asset pricing models when one of the proposed factors is in fact useless; Campbell and Cochrane (2000) explain why CAPM and its extensions are better approximating asset pricing models than the standard consumption based model; Hodrick and Zhang (2001) evaluate the specification errors of several empirical asset pricing models that have been developed as potential improvements on the CAPM; Lettau and Ludvigson (2001) explain the cross section of average stock returns; Jagannathan and Wang (2002) compare the SDF method with Beta method in estimating risk premium; Vassalou (2003) studies models that include a factor that captures news related to future Gross Domestic Product (GDP) growth; Jacob and Wang (2004) investigate the importance of idiosyncratic consumption risk for the cross sectional variation in asset returns; Vassalou and Xing (2004) compute default measures for individual firms; Huang and Wu (2004) analyze the specifications of option pricing models based on time-changed Levy process; and Parker and Julliard (2005) evaluate the consumption capital asset pricing model in which an asset’s expected return is determined by its equilibrium risk to consumption. Some other works test econometric specifications using the HJ-distance, including Bansal and Zhou (2002) and Shapiro (2002). Dittmar (2002) uses the HJ-distance to estimate the nonlinear pricing kernels in which the risk factor is endogenously determined and preferences restrict the definition of the pricing kernel.

However, as Ahn and Gadarowski (2002) find out, the specification test based on the HJ-distance severely overrejects correct models in commonly used sample size, while the Hansen test mildly overrejects correct models. Ahn and Gadarowski attribute this overrejection to poor estimation of the pricing error variance matrix, which occurs because the number of assets is relatively large for the number of time-series observations. Ahn and Gadarowski report that the rejection probability reaches as large as 75% for a nominal 5% level test, demonstrating a serious need for an improvement of the small sample properties of the HJ-distance. Because the true pricing error variance matrix is unknown, how to improve it has remained as an open question.

In this paper, we propose to improve the finite sample properties of the HJ-distance via more accurate estimation of the \textit{weighting matrix}, which is the inverse of the second moment matrix of portfolio returns. First, we show that poor estimation of the weighting matrix contributes
significantly to the poor small sample performance of the HJ-distance. When the exact second moment matrix is used, the small sample properties of the HJ-distance dramatically, and its rejection frequency becomes comparable to its nominal size.\(^1\)

Of course, the true covariance matrix is unknown. We employ the idea of the shrinkage method following Ledoit and Wolf (2002) to obtain a more accurate estimate of the covariance matrix. The basic idea behind shrinkage estimates is to take an optimally weighted average of the sample covariance matrix and the covariance matrix implied by a possibly misspecified structural model. The structural model provides a covariance matrix estimate that is biased but has a small variance due to the small number of parameters to be estimated. The sample covariance matrix provides another estimate which has no bias but a large variance. The shrinkage estimation balances the trade-off between the estimation error and bias by taking a weighted average of these two estimates.

In this shrinkage method, one needs to choose a structural model that serves as the shrinkage target. Here, because testing an asset pricing model is the purpose of using the HJ-distance, a natural choice of the structural model is the asset pricing model whose SDF is tested by the HJ-distance. The optimally weighted average is constructed by minimizing the mean squared error, and the optimal weight can be estimated consistently from the data.

We allow both possibility where the target model is correctly specified and misspecified. In the former case, the shrinkage target has no bias, and the estimated weight converges to the optimal weight. In the latter case, the shrinkage target is biased, and the estimated weight on the shrinkage target converges to zero as the sample size tends to infinity. Therefore, the proposed covariance matrix estimate is consistent under both cases.

Using this covariance matrix estimate greatly improves the small sample performance of the H-J distance. We use a similar data sets to Ahn and Gadarowski (2002). With 25 portfolios, the rejection frequencies are close to the nominal size even for the sample sizes of 160. With 100 portfolios, the rejection frequency is sometimes far from the nominal size, but it is much closer than the case where the sample covariance matrix is used.

The rest of this paper is organized as follows. Section 2 briefly reviews the HJ-distance and the specification test based on it. Section 3 presents the problem of the small sample properties of

\(^1\)Jobson and Korkie (1980) also report poor performance of the sample covariance matrix as an estimate of the population covariance when the sample size is not large enough compared with the dimension of the portfolio.
the HJ-test. Section 4 describes the proposed solution of this problem and report the simulation results. Section 5 concludes.

2 Hansen-Jagannathan distance

We follow Ahn and Gadarowski (2002) and focus on linear discount factor models. Although the HJ-distance is applicable to nonlinear model, successful asset pricing models are most in linear forms, such as the Sharpe (1964)-Lintner (1965)-Black (1972) CAPM, the Breeden (1979) consumption CAPM, the Adler and Dumas (1983) international CAPM, the Chen, Roll, and Ross (1986) five macro factor model, and the Fama-French (1992, 1996) three factor model. A linear asset pricing model implies a stochastic discount factor which is a linear combination of the factors in the model. If the asset-pricing model is correctly specified, the stochastic discount factor prices the returns perfectly.

Suppose we have \( N \) portfolios with \( T \) observations on each of them. Denote the \( t \)-th period returns by \( R_t \). \( R_t \) is a \( N \times 1 \) vector. Assume \( Y_t \) is a \( K \times 1 \) factor vector including a column of 1’s. Linear factor pricing models imply the stochastic discount factor \( m_t \) of the linear form \( m_t = Y_t' \delta \); see Hansen and Jagannathan (1997). Here \( \delta \) is an unknown parameter needing to be estimated. According to Euler equation, if the SDF prices returns correctly, we have \( E(R_t m_t) = 1 \).

We estimate \( \delta \) by minimizing the pricing errors produced by SDF. Since pricing errors form a vector, we need a measure based on vectors to define this misspecification. Hansen (1982) provides a quadratic form of pricing errors weighted by its covariance matrix and a \( \chi^2 \) statistic (Hansen statistic) for detecting such errors. But this statistic suffers from two problems. First, the small statistic may be caused by volatile pricing errors. Second, the weighting matrix varies for different SDFs. In view of these problems, Hansen and Jagannathan (1997) propose to use the HJ-distance to measure the pricing error. For the linear factor pricing model, the HJ-distance takes the form

\[
HJ(\delta) = \sqrt{E[w_t(\delta)']G^{-1}E[w_t(\delta)]},
\]

where \( w_t(\delta) = R_t Y_t' \delta - 1_N \) and \( G = E[R_t R_t'] \).

This distance is equal to the least squares distances between the stochastic discount factors and families of stochastic discount factors that price correctly. \( w_t(\delta) \) is the pricing errors, and \( G \), the second moment of portfolio returns, is used as the weighting matrix. Since the weighting
matrix is independent of the variance of the pricing errors, the HJ-distance would not reward the noisy SDFs. It also implies that we can compare the distances associated with different SDFs for the same data set.

In Jagannathan and Wang (1996), this distance can be estimated by

$$HJ_T(\delta) = \sqrt{w_T(\delta)'G_T^{-1}w_T(\delta)},$$

where $D_T = T^{-1}\sum_{t=1}^{T} R_t Y_t'$, $w_T(\delta) = T^{-1}\sum_{t=1}^{T} w_t(\delta) = D_T\delta - 1_N$ and $G_T = T^{-1}\sum_{t=1}^{T} R_t R_t'$. Since $\delta$ is unknown, it is estimated by $\hat{\delta}_T$ that minimizes the estimated the HJ distance

$$w_T(\hat{\delta}_T)'G_T^{-1}w_T(\hat{\delta}_T).$$

The first order condition gives

$$\hat{\delta}_T = (D_T'G_T^{-1}D_T)^{-1}D_T'G_T^{-1}1_N.$$

Since the HJ-distance uses the second moment of asset returns, $G_T$, as the weighting matrix, it does not follow a chi-squared distribution. However, Jagannathan and Wang (1996) prove that the asymptotic distribution of the HJ-distance is

$$T[HJ_T(\hat{\delta}_T)]^2 \overset{d}{\rightarrow} \sum_{j=1}^{N-K} \lambda_j v_j,$$

where $v_1, \ldots, v_{N-K}$ are independent $\chi^2(1)$ random variables, and $\lambda_1, \ldots, \lambda_{N-K}$ are nonzero eigenvalues of the following matrix:

$$\psi = S^{1/2}G^{-1/2}[I_N - (G^{-1/2})'D(D'G^{-1}D)^{-1}D'G^{-1/2}](G^{-1/2})'(S^{1/2})'.$$

Here $S = Ew_t(\delta)w_t(\delta)'$ denotes the variance of pricing errors, and $D = E(R_t'Y_t)$. It can be proved that $\psi$ is a positive semi-definite matrix with the rank of $N - K$. Hereby, it has $N - K$ positive eigenvalues. $G_T$ and $D_T$ are used to estimate $G$ and $T$, respectively. Under the hypothesis that the SDF prices the returns correctly, $S$ can be estimated consistently by $S_T = T^{-1}\sum_{t=1}^{T} w_t(\delta_T)w_t(\delta_T)'$.

Jagannathan and Wang (1996) provide the following algorithm to compute the $p$-value of the HJ-distance test. Draw $N - K$ random variables from $\chi^2(1)$ distribution $M$ times. For each draw, calculate $u_j = \sum_{i=1}^{N-K} \lambda_i v_{ij} (j = 1, 2, \ldots, M)$. Then the empirical $p$-value of the HJ-distance is

$$p = M^{-1}\sum_{j=1}^{M} I(u_j \geq T[HJ_T(\hat{\delta}_T)]^2),$$

5
where \( I(\cdot) \) is an indicator function which equals one if the expression in the bracket is true and zero otherwise. In practice, \( M \) can be any large number. We set \( M = 5,000 \).

3 Small sample properties of the HJ-distance

Following Ahn and Gadarowski (2002), we examine the small sample performances of the HJ-distance using Monte Carlo simulations.

We simulate two sets of data comparable to those in Ahn and Gadarowski (2002). The first set is a simple three-factor model with independent factor loadings. The statistical properties of the factors and idiosyncratic errors are set to be identical to those in Ahn and Gadarowski (2002). We refer to this model as a simple model henceforth. The second set of data is calibrated to resemble the statistical properties of the three-factor model in Fama-French (1992). The details of the data generation are provided in the Appendix.

We simulate each set of the data for 1000 trials. For each trial, we calculate the HJ-distance and test the hypothesis that the stochastic discount factor implied by the DGP prices portfolio returns correctly. Since the stochastic discount factors are derived from the true DGPs, they are supposed to price portfolios perfectly. Following Jagannathan and Wang (1996), we compute the \( p \)-value of the HJ-distance and record the rejection frequencies of the correct models.

Table 1 summarizes the results. This table corresponds to Table 1 of Ahn and Gadarowski (2002). The first columns in both panels are the significant levels of the tests. The other columns are the empirical rejection rates for different numbers of observations. We can see that for both 25 portfolio returns and 100 portfolio returns, the rejection rates of correct models are very high. The test based on the HJ-distance \( p \)-values overrejects severely in small sample. This makes it difficult to use the HJ-distance to test whether a stochastic discount factor prices the observed returns correctly.

Ahn and Gadarowski (2002) discuss the empirical problem of the HJ-distance and argue that its overrejection is due to the poor estimation of the variance matrix of the pricing errors, \( S \). They replace \( S_T \) with the exact variance matrix of the pricing errors, which is approximated by the sample variance matrix computed from 10,000 time-series observations. Since the variance matrix computed from 10,000 observations is not materially different from the true variance matrix, this would eliminate the effect of poor estimation of the variance matrix.
We repeat this experiment of Ahn and Gadarowski (2002). Table 2 shows the results and corresponds to their Table 6. It turns out this reduces the rejection rates of the HJ-distance p-value to reasonable levels. Overall, the rejection rates are close to their nominal levels. For example, the rejection rate at the 1% level changes from 4.5% to 1.3% for 25 portfolios with 160 observations. However, this method is not feasible in practice.

Another possible source of the overrejection is the poorly estimated weighting matrix. Jagannathan and Wang (1996) use $T^{-1} \sum_{t=1}^{T} R_t R_t' = \hat{Cov}(R_t) + \hat{E}(R_t)' \hat{E}(R_t)$ as an estimate of $G = E(R_t R_t') = Cov(R_t) + E(R_t)' E(R_t)$. While $E(R_t)$ can be estimated accurately by the sample mean for the sample size of our interest, the sample covariance matrix can be a very inaccurate estimate of $Cov(R_t)$ when $N/T$ is not negligible, as pointed out by Jobson and Korkie (1980). Consequently, the poor estimation of $G$ could be another main reason for the poor small sample performance of the HJ-distance.

It is well known that the accuracy of the weighting matrix has a significant effect on the finite sample property of the GMM-based Wald tests (Burnside and Eichenbaum, 1996). We expect this is also applicable to the HJ-distance. We replace $G_T$ with the exact covariance matrix $G$, which is approximated by the sample covariance matrix from 10,000 time-series observations, and compute the p-values using $G$.

Table 3 shows the results of this experiment. As we expect, the rejection rates of the HJ-distance p-values improve dramatically, and the rejection rates are comparable to those in Table 2. Finally, we repeat the above simulations with both exact $S$ and $G$. The results are summarized in Table 4. Using both exact $S$ and $G$ improve the rejection rates for the 25 portfolios, but the improvement is not substantial. For 100 portfolios, using both exact $S$ and $G$ leads to underrejection of the correct null.

The above experiments reveal that the performance of the HJ-distance p-value improves greatly when a better estimate of $G$ is used. Furthermore, it is not necessary to substitute both $S$ and $G$ with better estimates. Only one of them suffices. In the subsequent sections, we explore this possibility and focus our attention to improving the estimation of the second moment of portfolio returns.
4 Improving the finite sample properties by shrinkage

We apply shrinkage technique to estimate the covariance matrix of the asset returns. Shrinkage is proposed by Stein (1956). It assigns $\alpha$ weight to a covariance matrix estimated under some structure assumptions and the other $1 - \alpha$ weight to the sample covariance matrix. In the setting of this paper, each asset pricing model is a structure model for portfolio returns since it is assumed that portfolio returns are generated in the way the model specifies. And based on that structure model, we can estimate the covariance of the asset returns. The structure model covariance matrix has a large bias from the stringent structure assumption, but small estimation error because OLS is employed in estimation. On the contrary, the sample covariance imposes little structure, and has little bias or misspecified structure assumption, but it has a large estimation error. The intuition behind shrinkage is to trade off between bias and estimation error by choosing an optimal weight $\alpha$.

Ledoit and Wolf (2003) provide us an optimal $\alpha$ of the shrinkage towards the single-index model. We extend their method to $K$-factor asset pricing models, and apply it in the HJ-test.

Suppose the asset pricing model corresponding to the SDF we test is

$$R = \alpha + X\beta + \epsilon,$$

where $R$ is $T \times N$ matrix, and each column of $R$ is the observations of a portfolio return. $X$ is $T \times K$ matrix with $T$ data points and $K$ factors. $\beta$ is the beta of factors with the dimension of $K$ by $N$, and $\epsilon$ are uncorrelated error terms with dimension of $T$ by $N$. The covariance matrix of $R$ implied by this asset pricing model is

$$\Phi = \beta'\text{Cov}(X_t)\beta + \Delta,$$

Here, $\text{Cov}(X_t)$ is the covariance matrix of the factors and $\Delta$ is the diagonal covariance matrix of $\epsilon$. We use $\Sigma$ to denote the true covariance matrix of portfolio returns. Since one SDF is not uniquely determined by an asset pricing model, under the null hypothesis that the SDF prices the portfolios correctly, the model may or may not be correctly specified.

Regressing the $i$-th portfolio returns ($i$-th column of $R$) on the factors, we obtain the estimates of $\beta_i$, $b_i$, and the estimate of the variance of the residuals $d_{ii}$. So the estimate of the covariance matrix for the structure model, denoted by $F$, is

$$F = b'\text{Cov}(X)b + D,$$
where $\bar{Cov}(X)$ is the sample covariance matrix of the factors and $D$ is the estimate of $\Delta$. $b$ is a $K \times N$ matrix whose $i$-th column is $b_i$ i.e. $b = (b_1, b_2, \ldots, b_n)$. $D$ is a $N \times N$ diagonal matrix whose diagonal elements are $\{d_{ii}\}$. The typical element of $F$ is $f_{ij}$.

Let $S$ denote the sample covariance. We form a weighted average of $F$ and $S$ that minimizes the distance between the true covariance and the estimate. Following Ledoit and Wolf (2003), we use Frobenius norm to measure this distance, which is defined by, for any $N \times N$ matrix $Z$,

$$||Z||^2 = Trace(Z^2) = \sum_{i=1}^{N} \sum_{j=1}^{N} z_{ij}^2.$$  \hfill (4)

The loss function is

$$L(\alpha) = ||\alpha F + (1 - \alpha)S - \Sigma||^2,$$  \hfill (5)

We minimize the expected loss function to get the optimal $\alpha$,

$$Q(\alpha) = E[L(\alpha)]$$

This gives to

$$Q(\alpha) = \sum_{i=1}^{N} \sum_{j=1}^{N} E(\alpha f_{ij} + (1 - \alpha)s_{ij} - \sigma_{ij})^2$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \{Var(\alpha f_{ij} + (1 - \alpha)s_{ij} + [E(\alpha f_{ij} + (1 - \alpha)s_{ij} - \sigma_{ij})]^2\}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \{\alpha^2 Var(f_{ij}) + (1 - \alpha)^2 Var(s_{ij}) + 2\alpha(1 - \alpha)Cov(f_{ij}, s_{ij}) + \alpha^2(\phi_{ij} - \sigma_{ij})^2\}$$

The optimal $\alpha$ can be derived by differentiating $Q(\alpha)$ with respect to $\alpha$. The second order condition is satisfied since $Q(\alpha)$ is concave with respect to $\alpha$.

Solving the first order condition for $\alpha$ gives the optimal $\alpha$ as

$$\alpha^* = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} Var(s_{ij}) - \sum_{i=1}^{N} \sum_{j=1}^{N} Cov(f_{ij}, s_{ij})}{\sum_{i=1}^{N} \sum_{j=1}^{N} Var(f_{ij} - s_{ij}) + \sum_{i=1}^{N} \sum_{j=1}^{N} (\phi_{ij} - \sigma_{ij})^2}.$$  \hfill (6)

Multiplying both the numerator and denominator by $\sqrt{T}$, we get

$$\alpha^* = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} Var(\sqrt{T}s_{ij}) - \sum_{i=1}^{N} \sum_{j=1}^{N} Cov(\sqrt{T}f_{ij}, \sqrt{T}s_{ij})}{\sum_{i=1}^{N} \sum_{j=1}^{N} Var(\sqrt{T}f_{ij} - \sqrt{T}s_{ij}) + T \sum_{i=1}^{N} \sum_{j=1}^{N} (\phi_{ij} - \sigma_{ij})^2}.$$  \hfill (7)

We define $\pi = \sum_{i=1}^{N} \sum_{j=1}^{N} AsyVar[\sqrt{T}s_{ij}]$, $\rho = \sum_{i=1}^{N} \sum_{j=1}^{N} AsyCov[\sqrt{T}f_{ij}, \sqrt{T}s_{ij}]$, $\mu = \sum_{i=1}^{N} \sum_{j=1}^{N} AsyVar[\sqrt{T}f_{ij} - \sqrt{T}s_{ij}]$ and $\gamma = \sum_{i=1}^{N} \sum_{j=1}^{N} (\phi_{ij} - \sigma_{ij})^2$, and assume I.I.D. data and finite fourth moments.
When the structure model we are using is not the true DGP, $\Phi \neq \Sigma$. Theorem 1 in Ledoit and Wolf (2003) tells us the optimal $\alpha^*$ converges to $^2$

$$T \alpha^* \rightarrow \frac{\pi - \rho}{\gamma}.$$  

(8)

Notice that unlike Ledoit and Wolf (2003), we do not make an assumption $\Phi \neq \Sigma$ because under the hypothesis there is a chance that the asset pricing is the true DGP. When this is the case, $\Phi = \Sigma$ and $\mu$ dominates $T\gamma$ in convergence. And the convergence of $\alpha^*$ changes to

$$\alpha^* \rightarrow \frac{\pi - \rho}{\mu}.$$  

(9)

We drop the dominated part of the denominator so that we can increase the accuracy of estimation on the denominator.

As $\pi$, $\rho$, $\mu$ and $\gamma$ in the formula for $\alpha^*$ are unobservable, we must find consistent estimators for them. As Lemma 1 in Ledoit and Wolf (2003), we have a consistent estimator for $\pi_{ij}$

$$p_{ij} = \frac{1}{T} \sum_{t=1}^{T} [(R_{it} - m_i)(R_{jt} - m_j) - s_{ij}]^2,$$  

(10)

where $m_i$ is the mean value for the $i$-th column of $R$.

Because on the diagonal $f_{ii} = s_{ii}$, the consistent estimator for the diagonal elements of $\rho$ is $p_{ii}$. For the off-diagonal elements of $\rho$, we have to make a little effort. The OLS regression estimate of $\beta$ is

$$b = (X'X)^{-1}X'R.$$  

Plug this into the formula for the structure covariance,

$$F = R'X(X'X)^{-1}Cov(X)(X'X)^{-1}X'R + D$$

$$= R'X(X'X)^{-1} \frac{1}{T} X'(I - \frac{1}{T} 11')X(X'X)^{-1}X'R + D.$$  

If we define $\tilde{R} = X(X'X)^{-1}X'R$, we have

$$F = \frac{1}{T} \tilde{R}'(I - \frac{1}{T} 11')\tilde{R} + D.$$  

(11)

We can see the covariance matrix of the structure model is equivalent to the sample covariance of the artificial portfolio returns $\tilde{R}$ plus $D$. So we have the following three lemmas:

---

$^2$The proof can be found in Ledoit and Wolf (2003).
**Lemma 1**  
$p_{ii}$ is the consistent estimator of the diagonal of $\rho$. For $i \neq j$, a consistent estimator of $\rho_{ij}$ is given by:

\[
r_{ij} = \frac{1}{T} \sum_{t=1}^{T} [(R_{it} - m_i)(R_{jt} - m_j) - s_{ij}][(\tilde{R}_{it} - \tilde{m}_i)(\tilde{R}_{jt} - \tilde{m}_j) - f_{ij}]
\]

where $\tilde{R} = X(X'X)^{-1}X'R$, $\tilde{m}_i$ is the mean value for the column $i$ of $\tilde{R}$.

**Proof:** $r_{ij}$ is the sample estimator for the asymptotic covariance between $s_{ij}$ and $f_{ij}$. It converges in probability to $\rho_{ij}$. □

**Lemma 2**  
Define

\[
w_{ij} = \frac{1}{T} \sum_{t=1}^{T} [(\tilde{R}_{it} - \tilde{m}_i)(\tilde{R}_{jt} - \tilde{m}_j) - f_{ij}]^2
\]

where $\tilde{R} = X(X'X)^{-1}X'R$, $\tilde{m}_i$ is the mean value for the column $i$ of $\tilde{R}$. The consistent estimator of $\gamma_{ij}$ is given by

\[g_{ij} = w_{ij} + p_{ij} - 2r_{ij}.
\]

**Proof:** $w_{ij}$ is the sample estimator for the asymptotic variance of $f_{ij}$. And

\[\text{Var}(f_{ij} - s_{ij}) = \text{Var}(f_{ij}) + \text{Var}(s_{ij}) - 2\text{Cov}(f_{ij}, s_{ij})
\]

So,

\[\text{AsyVar}(\sqrt{T}f_{ij} - \sqrt{T}s_{ij}) = \text{AsyVar}(\sqrt{T}f_{ij}) + \text{AsyVar}(\sqrt{T}s_{ij}) - 2\text{AsyCov}(\sqrt{T}f_{ij}, \sqrt{T}s_{ij})
\]

From equation (13) and Lemma 1, we know

\[g_{ij} = w_{ij} + p_{ij} - 2r_{ij}.
\]

□

**Lemma 3**  
A consistent estimator for $\gamma_{ij} = (\phi_{ij} - \sigma_{ij})^2$ is its sample counterpart $c_{ij} = (f_{ij} - s_{ij})^2$.

After deriving all the consistent estimators for $\pi, \rho$ and $\gamma$, we have the consistent estimator for the general form of the optimal shrinkage weight.
Theorem 1  Define an estimate of $\alpha$ as

$$\hat{\alpha} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij} - \sum_{i=1}^{N} \sum_{j=1}^{N} r_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} (g_{ij} + Tc_{ij})}$$

Then if $\Phi \neq \Sigma$,

$$T \hat{\alpha} \rightarrow \frac{\pi - \rho}{\gamma};$$

if $\Phi = \Sigma$,

$$\hat{\alpha} \rightarrow \frac{\pi - \rho}{\mu}.$$

Proof: Combine Lemma 1, 2 and 3 with equation (9). □

The shrinkage covariance matrix $V_s$ is:

$$V_s = \hat{\alpha}F + (1 - \hat{\alpha})S.$$

Shrinkage estimation balances the trade-off between bias and estimation error. In the following, we show several empirical results when the shrinkage estimation is used to estimate the covariance matrix of portfolio returns.

When shrinkage estimation is employed in the HJ-test,

$$G_T = V_s + \left(\frac{1}{T}R'1\right)\left(\frac{1}{T}R'1\right)'$$

We repeat the Monte Carlo experiments in Section 2, but substitute the sample covariance with the shrinkage estimation. And we check the rejection rates of correct model for the HJ-test. Table 5 reports the rejection frequencies. We also report the statistics of $\hat{\alpha}$ in Table 6. In addition, we estimate the density function of $\hat{\alpha}$ by Epanechnikov kernel, as shown in Figure 1 to Figure 4.

Let us compare Table 1 with Table 5 for 25 portfolios first. In Table 1, for the Fama-French model, the rejection rate of the HJ-test can be as high as 23.9% when $T = 160$ and the significant level is 10%. While, in Table 5, we can see the largest difference between the rejection rates and the significant levels is 2.5%, and all the other rejection rates are very close to the significant levels. This improvement is more obvious when we look at 100 portfolios. In Table 1, when $T = 160$, the HJ-test rejects the hypothesis for almost every trial. While, in Table 5, those

3In practice, in order to make $V_s$ positive semi-definite, we set 0 and 1 as the lower bound and upper bound for $\hat{\alpha}$. 12
numbers decrease greatly. We notice that the HJ-test under-rejects for the simple simulation case when $T = 160$. This may be due to estimation errors in $\hat{\alpha}$.

Shrinkage estimation is better than sample covariance, but it is definitely not the best. This can be shown when we add Table 3 into the comparison. If the shrinkage estimate was very close to the true covariance matrix, the first panel of Table 5 should be similar with Table 3, where we use the sample covariance in large sample as the estimate of the covariance matrix in small sample. Although the differences between these two sets of tables are not very large, they are not negligible either. Those differences tell us there is still a distance between shrinkage estimation and true covariance matrix. This leaves much room for future research work.

5 Conclusion

The HJ-test rejects correct SDFs too often in small sample, which limits its practical use. This paper studies this phenomena, and finds out the reason is due to the poor estimates of the weighting matrix and the covariance matrix of the pricing error. We propose to use shrinkage method to improve the estimation of the weighting matrix of the HJ-test.

In small sample, sample covariance matrix can have a large estimation error and little bias. While, structure covariance matrix has the opposite small sample property. Shrinkage estimation balances the trade off between bias and estimation errors, and chooses an optimal weight to combine sample covariance with structure covariance. So shrinkage estimate would have better small sample property than sample covariance or structure covariance. When this kind of method is imposed on the estimation of the second moment of portfolio returns, the rejection rate of the HJ-test in small sample decreases significantly.

Of course, shrinkage estimation is not the only method which can improve the HJ-test. It could be future work to compare the performances of the HJ-test when other competitive covariance estimations are employed.
A Appendix

A.1 Simple model

The simple model is generated by the following data generating process (DGP):

\[ R_{it} = \alpha + X_{1t}\beta_{1i} + X_{2t}\beta_{2i} + X_{3t}\beta_{3i} + e_{it} \]  

(12)

where \( i \) is the index of individual portfolio return, and \( t \) is the index of time. \( R_{it} \) is the gross return for portfolio \( i \) at time \( t \). \( X_{jt} \) \((j=1, 2 \text{ and } 3)\) is the common factor for time \( t \), drawn from a normal distribution with mean equal to 0.0022 and variance equal to 6.944 \times 10^{-5}. \( \beta_{ki} \) \((k=1, 2 \text{ and } 3)\) is the corresponding beta of factor \( X_k \) for portfolio \( i \), and they are drawn from \( U[0, 2] \). \( e_{it} \) is the idiosyncratic error that is normally distributed with mean zero and variance 6.944 \times 10^{-5}. \( \alpha, \beta \) and \( X \) are chosen at values which make the mean and variance of gross returns roughly consistent with historical data in the US stock market.

A.2 Fama-French model

The data are generated from a three-factor model

\[ R_{it} = \beta_{0i} + X_{1t}\beta_{1i} + X_{2t}\beta_{2i} + X_{3t}\beta_{3i} + e_{it} \]

First, we regress the portfolio returns in Fama-French (1992) on the Fama-French factors by two-step OLS, obtain the estimates of \( \beta_{ki} \), and collect the residuals. Then we compute the sample covariance of the residuals. Second, we run the following cross-sectional regression:

\[ E[R_{it}] = \alpha + (E(X_{1t}) + \eta_1)\beta_{1i} + (E(X_{2t}) + \eta_2)\beta_{2i} + (E(X_{3t}) + \eta_3)\beta_{3i} \]

This gives estimates of the risk-free rate, \( \alpha \), and the factor-mean adjusted risk prices, \( \eta_k \). Third, we simulate the factors from normal distribution with the mean and the covariance equal to the sample mean and the sample covariance matrix derived from the actual data of Fama-French factors. The error terms, \( e_{it} \), are drawn from normal distribution with the mean equal to zero and the variance equal to the sample covariance of the residuals. The calibrated portfolio returns are generated by the following equation:

\[ R_{it} = \alpha + (X_{1t} + \eta_1)\beta_{1i} + (X_{2t} + \eta_2)\beta_{2i} + (X_{3t} + \eta_3)\beta_{3i} + e_{it} \]  

(13)

We incorporate the risk adjusted price in order to simulate the portfolio return close to the true data.
B Tables and Figures

Table 1: Rejection frequencies of the specification test using the HJ-distance

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>25 Portfolios</th>
<th>100 Portfolios</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T=160</td>
<td>T=330</td>
</tr>
<tr>
<td><strong>(A) Simple Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>4.5</td>
<td>2.4</td>
</tr>
<tr>
<td>5%</td>
<td>15.1</td>
<td>8.7</td>
</tr>
<tr>
<td>10%</td>
<td>23.8</td>
<td>16.4</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>99.6</td>
<td>51.3</td>
</tr>
<tr>
<td>5%</td>
<td>99.9</td>
<td>71.8</td>
</tr>
<tr>
<td>10%</td>
<td>99.9</td>
<td>81.3</td>
</tr>
<tr>
<td><strong>(B) Fama-French Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>5.8</td>
<td>3.3</td>
</tr>
<tr>
<td>5%</td>
<td>15.1</td>
<td>10.6</td>
</tr>
<tr>
<td>10%</td>
<td>23.9</td>
<td>18.9</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>99.8</td>
<td>53.7</td>
</tr>
<tr>
<td>5%</td>
<td>100</td>
<td>76</td>
</tr>
<tr>
<td>10%</td>
<td>100</td>
<td>84.3</td>
</tr>
</tbody>
</table>
Table 2: Rejection frequencies of the specification test using the HJ-distance with the exact pricing-error variance matrix $S$

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>$T=160$</th>
<th>$T=330$</th>
<th>$T=700$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>1.3</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>5%</td>
<td>6.0</td>
<td>6.5</td>
<td>6.7</td>
</tr>
<tr>
<td>10%</td>
<td>10.6</td>
<td>11.0</td>
<td>11.3</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>1.2</td>
<td>0.8</td>
<td>1.4</td>
</tr>
<tr>
<td>5%</td>
<td>4.9</td>
<td>5.7</td>
<td>5.7</td>
</tr>
<tr>
<td>10%</td>
<td>9.5</td>
<td>11.1</td>
<td>11.7</td>
</tr>
</tbody>
</table>
Table 3: Rejection frequencies of the specification test using the HJ-distance with the exact weighting matrix $G$

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>$T=160$</th>
<th>$T=330$</th>
<th>$T=700$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.9</td>
<td>1.2</td>
<td>1.3</td>
</tr>
<tr>
<td>5%</td>
<td>5.4</td>
<td>5.8</td>
<td>6.8</td>
</tr>
<tr>
<td>10%</td>
<td>12.4</td>
<td>11.8</td>
<td>12.1</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.8</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>5%</td>
<td>4.4</td>
<td>5.7</td>
<td>5.7</td>
</tr>
<tr>
<td>10%</td>
<td>11.6</td>
<td>10.9</td>
<td>10.9</td>
</tr>
</tbody>
</table>
Table 4: Rejection frequencies of the specification test using the HJ-distance with the exact $S$ and $G$

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>$T=160$</th>
<th>$T=330$</th>
<th>$T=700$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.8</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>5%</td>
<td>4.4</td>
<td>5.5</td>
<td>6.7</td>
</tr>
<tr>
<td>10%</td>
<td>8.8</td>
<td>10.3</td>
<td>11.3</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.2</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>5%</td>
<td>1.4</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>10%</td>
<td>2.4</td>
<td>5.5</td>
<td>6.7</td>
</tr>
</tbody>
</table>
Table 5: Rejection frequencies of the specification test using the HJ-distance with shrinkage estimation of $G$

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>$T=160$</th>
<th>$T=330$</th>
<th>$T=700$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(A) Simple Model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.6</td>
<td>1.4</td>
<td>1.0</td>
</tr>
<tr>
<td>5%</td>
<td>5.6</td>
<td>7.4</td>
<td>5.8</td>
</tr>
<tr>
<td>10%</td>
<td>10.8</td>
<td>12.5</td>
<td>11</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.4</td>
<td>1.0</td>
<td>1.4</td>
</tr>
<tr>
<td>5%</td>
<td>6.6</td>
<td>6.1</td>
<td>6.1</td>
</tr>
<tr>
<td>10%</td>
<td>16.5</td>
<td>14.0</td>
<td>10.6</td>
</tr>
<tr>
<td><strong>(B) Fama-French Model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0.9</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>5%</td>
<td>5.3</td>
<td>6.0</td>
<td>4.8</td>
</tr>
<tr>
<td>10%</td>
<td>10.7</td>
<td>11.6</td>
<td>10.7</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>4.9</td>
<td>2.2</td>
<td>1.7</td>
</tr>
<tr>
<td>5%</td>
<td>23.9</td>
<td>11.5</td>
<td>7.3</td>
</tr>
<tr>
<td>10%</td>
<td>39.9</td>
<td>22.1</td>
<td>15.0</td>
</tr>
</tbody>
</table>
Table 6: Summary statistics of $\hat{\alpha}$

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>$T=160$</th>
<th>$T=330$</th>
<th>$T=700$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Simple Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.8965</td>
<td>0.8986</td>
<td>0.9014</td>
</tr>
<tr>
<td>variance</td>
<td>0.0024</td>
<td>0.0012</td>
<td>0.0006</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.8968</td>
<td>0.8988</td>
<td>0.9005</td>
</tr>
<tr>
<td>variance</td>
<td>$6.369 \times 10^{-4}$</td>
<td>$3.629 \times 10^{-4}$</td>
<td>$1.649 \times 10^{-4}$</td>
</tr>
<tr>
<td>(B) Fama-French Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.9368</td>
<td>0.9478</td>
<td>0.9592</td>
</tr>
<tr>
<td>variance</td>
<td>0.0063</td>
<td>0.0035</td>
<td>0.002</td>
</tr>
<tr>
<td>100 Portfolios</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.8074</td>
<td>0.8098</td>
<td>0.8134</td>
</tr>
<tr>
<td>variance</td>
<td>0.0021</td>
<td>0.0016</td>
<td>0.0016</td>
</tr>
</tbody>
</table>
Figure 1: The density functions for $\hat{\alpha}$ in Simple Model of 25 Portfolios

Figure 2: The density functions for $\hat{\alpha}$ in Simple Model of 100 Portfolios
Figure 3: The density functions for $\hat{\alpha}$ in Fama-French 25 Portfolios Model

Figure 4: The density functions for $\hat{\alpha}$ in Fama-French 100 Portfolios Model
References


