Journal of Empirical Finance 16 (2009) 483–506

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Improvement in finite sample properties of the Hansen–Jagannathan distance test☆

Yu Ren [b], Katsumi Shimotsu [a,*]

[a] Department of Economics, Queen's University, Kingston, Ontario, Canada K7L 3N6
[b] Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, Fujian, 361005 China

ARTICLE INFO

Article history:
Received 22 June 2007
Received in revised form 29 December 2008
Accepted 30 December 2008
Available online 13 January 2009

JEL classification:
C13
C52
G12

Keywords:
Covariance matrix estimation
Factor models
Finite sample properties
Hansen–Jagannathan distance
Shrinkage method

ABSTRACT

Jagannathan and Wang [Jagannthan, R., and Wang, Z., "The conditional CAPM and the cross-section of expected returns." Journal of Finance, 51 (1996), 3–53] derive the asymptotic distribution of the Hansen–Jagannathan distance (HJ-distance) proposed by Hansen and Jagannathan [Hansen, L.P., and Jagannathan, R., Assessing specific errors in stochastic discount factor models." Journal of Finance, 52 (1997), 557–590], 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 [Ahn, S.C., and Gadarowski, C., "Small sample properties of the GMM specification test based on the Hansen–Jagannathan distance." Journal of Empirical Finance, 11 (2004), 109–132] 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 finite sample properties of the HJ-distance test by applying the shrinkage method [Ledoit, O., and Wolf, M., "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." Journal of Empirical Finance, 10 (2003), 603–621] to compute its weighting matrix. The proposed method improves the finite sample performance of the HJ-distance test significantly.

1. Introduction

Asset pricing models are the cornerstone of finance. They reveal how portfolio returns are determined and which factors affect returns. Stochastic discount factors (SDF) describe portfolio returns from another point of view. SDFs display which prices are reasonable given the returns in the current period. Asset prices can be represented as inner products of payoffs and SDFs. If asset pricing models were the true data generating process (DGP) of returns, SDFs could price the returns perfectly.

In reality, asset pricing models are at best approximations. This implies no stochastic discount factors can price portfolios perfectly in general. Therefore, it is important to construct a measure of pricing errors produced by SDFs so that we are able to compare and evaluate SDFs. For this purpose, Hansen and Jagannathan (1997) develop the Hansen–Jagannathan distance (HJ-distance). This measure is in the quadratic form of the pricing errors weighted by the inverse of the second moment matrix of returns. Intuitively, the HJ-distance 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 has already 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 ones. Hereby, the new model can be supported if it offers small pricing

☆ The authors are grateful to three anonymous referees for helpful comments. The authors thank Christopher Gadarowski for providing the dataset for calibrating the Premium-Labor model and Seung Ahn, Masayuki Hirukawa, and Victoria Zinde-Walsh for the helpful comments. Shimotsu thanks the SSHRC and Royal Bank of Canada Faculty Fellowship for financial support.

* Corresponding author. Tel.: +1 613 533 6546; fax: +1 613 533 6668.
E-mail addresses: reny@xmu.edu.cn (Y. Ren), shimotsu@econ.queensu.ca (K. Shimotsu).

0927-5398/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2008.12.003

errors. This type of comparison has been adopted in many recent papers. For instance, by using the HJ-distance, Jagannathan and Wang (1998) discuss cross sectional regression models; Kan and Zhang (1999) study asset pricing models when one of the proposed factors is in fact useless; Campbell and Cochrane (2000) explain why the CAPM and its extensions are better at approximating asset pricing models than the standard consumption-based asset pricing theory; 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 the 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; Jacobs 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.

The HJ-distance has several desirable properties in comparison to the *J*-statistic of Hansen (1982): first of all, it does not reward variability of SDFs. The weighting matrix used in the HJ-distance is the second moment of portfolio returns and independent of pricing errors, while 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. Unlike the *J*-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 statistic.

However, Ahn and Gadarowski (2004) find that the specification test based on the HJ-distance severely overrejects correct models in commonly used sample size, compared with the Hansen test which mildly overrejects correct models. Ahn and Gadarowski (2004) 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 (2004) 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 finite sample properties of the HJ-distance test.

In this paper, we propose to improve the finite sample properties of the HJ-distance test via more accurate estimation of the *weighting matrix*, which is the inverse of the second moment matrix of portfolio returns. We justify our method by showing that poor estimation of the weighting matrix contributes significantly to the poor small sample performance of the HJ-distance test. When the exact second moment matrix is used, the 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 (2003) to obtain a more accurate estimate of the covariance matrix. The basic idea behind shrinkage estimation 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 estimation error due to the small number of parameters to be estimated. The sample covariance matrix provides another estimate which has a small bias, but a large estimation error. 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 serving as the shrinkage target. In many cases, the HJ-distance test is used to examine the pricing errors of the SDF of an asset pricing model. Then, a natural choice of a structural model is the asset pricing model whose SDF is examined by the HJ-distance test. The optimally weighted average is constructed by minimizing the distance between the weighted covariance matrix and the true covariance matrix, and the optimal weight can be estimated consistently from the data.

We allow both possibilities where the target model is correctly specified and misspecified. In the former case, the shrinkage target is asymptotically unbiased. In the latter case, the shrinkage target is biased, but the estimated weight on the shrinkage target converges to zero in probability as the sample size tends to infinity. Therefore, the proposed covariance matrix estimate is consistent in both cases.

Using this covariance matrix estimate greatly improves the finite sample performance of the HJ-distance test. We use similar data sets with Ahn and Gadarowski (2004). 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 in which the sample covariance matrix is used.

We also examined the performance of the proposed method with various models, including both the correctly specified and misspecified models, a nonlinear factor model, and a model with GARCH errors. The proposed method performs well overall, and its performance is mainly determined by the number of portfolios and the degree of misspecification by the target model. In all the cases considered, the proposed method works at least as good as the standard HJ-distance test.

[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 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 the HJ-distance test; Section 4 describes the proposed solution to this problem; and Section 5 reports the simulation results; Section 6 concludes.

2. Hansen–Jagannathan distance

Hansen and Jagannathan (1997) develop a measure of degree of misspecification of an asset pricing model. This measure, called the HJ-distance, is defined as the least squares distance between the stochastic discount factor associated with an asset pricing model and the family of stochastic discount factors that price all the assets correctly. Hansen and Jagannathan (1997) show that the HJ-distance is also equal to the maximum pricing errors generated by a model on the portfolios whose second moments of returns are equal to one.

Consider a portfolio of N primitive assets, and let R_t denote the t-th period gross returns of these assets. R_t is a $1\times N$ vector. A valid stochastic discount factor (SDF), m_t, satisfies $E(m_tR_t')=1_N$, where 1_N is a N-vector of ones. If an asset pricing model implies a stochastic discount factor $m_t(\delta)$, where δ is a $K\times 1$ unknown parameter, then the HJ-distance corresponding to this asset pricing model is given by

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

where $w_t(\delta)=R_t'm_t(\delta)-1_N$ denotes the pricing errors and $G=E(R_t'R_t)$.

We follow Ahn and Gadarowski (2004) and focus on linear factor pricing models. Linear factor pricing models imply the SDF of the linear form $m_t(\delta)=\tilde{X}_t\delta$, where $\tilde{X}_t=[1\ X_t]$ is a $1\times K$ vector of factors including 1; see Hansen and Jagannathan (1997). Note that linear factor pricing models can accommodate nonlinear function of factors because $\tilde{X}_t$ may contain polynomials of factors, and the linearity assumption here is not very restrictive. For example, Bansal et al. (1993), Chapman (1997), and Dittmar (2002) consider nonlinear factor models of this type. In addition, many successful asset pricing models are in linear forms.[2]

The HJ-distance can be estimated by its sample analogue

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

where $w_T(\delta)=T^{-1}\sum_{t=1}^{T} w_t(\delta)=D_T\delta-1_N$, $D_T=T^{-1}\sum_{t=1}^{T} R_t'\tilde{X}_t$ and $G_T=T^{-1}\sum_{t=1}^{T} R_t'R_t$. Following Jagannathan and Wang (1996), the parameter δ is estimated by minimizing the sample HJ-distance $\mathrm{HJ}_T(\delta)$, giving the estimate δ_T as

$$\delta_T = \left(D' T G_T^{-1} D_T\right)^{-1} D' T G_T^{-1} 1_N.$$

The estimator δ_T is equivalent to a GMM estimator with the moment condition $E[w_t(\delta)]=0$ and the weighting matrix G_T^{-1}.

Jagannathan and Wang (1996) prove that, under the hypothesis that the SDF prices the returns correctly, the sample HJ-distance follows

$$T[\mathrm{HJ}_T(\delta_T)]^2 \to_d \sum_{j=1}^{N-K} \lambda_j \upsilon_j,$$

where $\upsilon_1,\ldots\upsilon_{N-K}$ are independent $\chi^2(1)$ random variables, and $\lambda_1,\ldots\ \lambda_{N-K}$ are nonzero eigenvalues of the following matrix:

$$\Lambda = \Omega^{1/2} G^{-1/2}\left[I_N - \left(G^{-1/2}\right)' D\left(D' G^{-1} D\right)^{-1} D' G^{-1/2}\right]\left(G^{-1/2}\right)'\left(\Omega^{1/2}\right)'.$$

Here $\Omega=E[w_t(\delta)w_t(\delta)']$ denotes the variance of pricing errors, and $D=E(R_t'\tilde{X}t)$. It can be proved that A is positive semidefinite with rank N–K. G_T and D_T can be used to estimate G and D consistently. Under the hypothesis that the SDF prices the returns correctly, Ω can be estimated consistently by $\Omega_T=T^{-1}\sum_{t=1}^{T} w_t(\delta_T)w_t(\delta_T)'$.

δ_T is not as efficient as the optimal GMM estimator that uses Ω_T^{-1} (optimal weighting matrix) as the weighting matrix, defined as

$$\delta_{OPT,T} = \left(D' T \Omega_T^{-1} D_T\right)^{-1} D' T \Omega_T^{-1} 1_N.$$

Associated with $\delta_{OPT,T}$ and Ω_T^{-1} is the J-statistic of Hansen (1982)

$$J_T\left(\delta_{OPT,T}\right) = T w_T\left(\delta_{OPT,T}\right)' \Omega_T^{-1} w_T\left(\delta_{OPT,T}\right),$$

which is widely used for specification testing. Under the null hypothesis that the SDF prices the returns correctly, Hansen's J-statistic is asymptotically χ^2-distributed with N–K degrees of freedom.

[2] For example, the Sharpe (1964)-Lintner (1965)-Black (1972) CAPM, Ross (1976), the Breeden (1979) consumption CAPM, the Adler and Dumas (1983) international CAPM, the Chen et al (1986) five macro factor model, and the Fama-French (1992, 1996) three factor model.

The HJ-distance has several desirable properties over the *J*-statistic. First, it does not reward the variability of SDFs. 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 *J*-statistic uses the inverse of the second moment of the pricing errors as the weighting matrix and hence 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 for a given dataset.

3. Finite sample properties of the HJ-distance test

In this section, we investigate the finite sample performances of the specification test based on the HJ-distance (henceforth the HJ-distance test) following the settings of Ahn and Gadarowski (2004).

3.1. Simulation design

We simulate three sets of data comparable to those in Ahn and Gadarowski (2004). The first set is a simple three-factor model with independent factor loadings, where the scale of expected returns and variability of the factors are roughly matched to those of the actual market-wide returns. The statistical properties of the factors and idiosyncratic errors are set to be identical to those in Ahn and Gadarowski (2004). We refer to this model as the Simple model henceforth. The second set of data is calibrated to resemble the statistical properties of the three-factor model in Fama and French (1992). The third set of data is calibrated based on the Premium-Labor model in Jagannathan and Wang (1996). The details of the data generation are provided in the Appendix.

We simulate each set of the data with 1000 replications. For each replication, we calculate the HJ-distance and test the null 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, the actual rejection frequency is supposed to be close to the nominal level. The critical values of the HJ-distance test are calculated following the algorithm by Jagannathan and Wang (1996). First, draw $M\times(N-K)$ independent random variables from $\chi^2(1)$ distribution. Next, calculate $u_j=\sum_{i=1}^{N-K}\lambda_i v_{ij}$ $(j=1,\ldots,M)$. Then the empirical *p*-value of the HJ-distance is

$$p = M^{-1}\sum_{j=1}^{M} I\left(u_j \geq T[\mathrm{HJ}_T(\delta_T)]^2\right),$$

where $I(\cdot)$ is an indicator function which equals one if the expression in the brackets is true and zero otherwise. In our simulation, we set $M=5{,}000$.

3.2. Simulation results

Table 1 summarizes the results from this simulation with 25 and 100 portfolios and $T=160,330,700$. Panel A of this table corresponds to Table 1 of Ahn and Gadarowski (2004), while Panels B and C correspond to Table 3 of Ahn and Gadarowski (2004). The first column in each panel is the significance level of the tests. The other columns report the actual rejection frequencies for different numbers of observations. The results are comparable to those in Ahn and Gadarowski (2004).

The HJ-distance test overrejects the correct null under all combinations of the DGPs, the number of portfolios, and sample sizes. In the Simple model and Fama–French model, the size distortion is noticeable except for the combination of $T=700$ and 25 portfolios. The size distortion is particularly large with 100 portfolios but improves as *T* increases. In the Premium-Labor model, the HJ-distance test is severely oversized both with 25 and 100 portfolios and for all sample sizes. As suggested by Ahn and Gadarowski (2004), this excessive rejection frequencies for the HJ-distance may be due to a feature of the data based on the Premium-Labor model not present in the other data, possibly the temporal dependence of the factors.

Ahn and Gadarowski (2004) investigate the source of this overrejection and find that one of its sources is the poor estimation of the variance matrix of the pricing errors, Ω. They find that repeating their simulations using the exact pricing error variance matrix Ω removes most of the upward bias in the size of the HJ-distance test. However, the exact pricing error matrix in unknown, and hence it is impossible to use this method in practice and the problem of overrejection has remained unsolved.

We examine other possible sources of overrejection. 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 (e.g., Burnside and Eichenbaum, 1996). We conjecture another possible source of the overrejection is the poorly estimated weighting matrix. Jagannathan and Wang (1996) use $T^{-1}\sum_{t=1}^{T}R_t'R_t=\widehat{Var}(R_t)+\hat{E}(R_t)'\hat{E}(R_t)$ as an estimate of $G=E(R_t'R_t)=Var(R_t)+E(R_t)'E(R_t)$. As pointed out by Jobson and Korkie (1980), the sample covariance matrix can be a very inaccurate estimate of $\mathrm{Var}(R_t)$ when the number of observation is not large enough relative to the number of portfolios. In our case, with 25 portfolios, *G* has $(26\times25)/2=325$ elements. Consequently, the poor estimation of *G* may be another main reason for the poor small sample performance of the HJ-distance test.

We confirm this conjecture by repeating the simulations in Table 1 but replacing G_T with the exact second moment matrix *G*. Table 2 shows the resulting rejection frequencies of the HJ-distance test. We approximate *G* by the sample second moment matrix from 10,000 time-series observations. In all cases, the rejection rates of the HJ-distance test improve dramatically. The HJ-distance test now has good small sample properties in the Simple model and the Fama–French model. In particular, with 25 portfolios, the actual size is close to the nominal size for all *T*. Comparing it with Table 1 suggests that the improvement of the size of the original

Table 1
Rejection frequencies of the HJ-distance test.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	4.5	2.4	1.9
5%	15.1	8.7	7.7
10%	23.8	16.4	13.4
100 Portfolios			
1%	99.6	51.3	11.7
5%	99.9	71.8	27.4
10%	99.9	81.3	39.3
(B) Fama–French model			
25 Portfolios			
1%	5.8	3.3	1.1
5%	15.1	10.6	7.1
10%	23.9	18.9	12.8
100 Portfolios			
1%	99.8	53.7	13.8
5%	100.0	76.0	30.8
10%	100.0	84.3	44.6
(C) Premium-Labor model			
25 Portfolios			
1%	14.9	11.3	9.2
5%	31.9	26.0	19.5
10%	42.7	34.4	29.0
100 Portfolios			
1%	99.7	79.1	36.8
5%	99.9	90.1	59.1
10%	99.9	94.3	69.6

This table shows the rejection rates over 1000 trials using the p-values for the HJ-distance. For Panel (A), factors and returns were simulated to make the mean and variance of gross returns roughly consistent with historical data in the US stock market. For Panels (B) and (C), factors and returns were simulated using either the Fama and French (1996) model or the Premium-Labor model per Jagannathan and Wang (1996).

Table 2
Rejection frequencies of the HJ-distance test with the exact weighting matrix G.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	0.9	1.2	1.3
5%	5.4	5.8	6.8
10%	12.4	11.8	12.1
100 Portfolios			
1%	0.8	1.2	1.2
5%	4.4	5.7	5.7
10%	11.6	10.9	10.9
(B) Fama–French model			
25 Portfolios			
1%	1.0	1.2	0.7
5%	4.9	5.7	5.0
10%	9.9	11.8	11.9
100 Portfolios			
1%	1.1	1.6	1.3
5%	7.4	7.4	5.8
10%	16.8	14.7	13.6
(C) Premium-Labor model			
25 Portfolios			
1%	4.5	6.7	6.9
5%	15.2	20.3	16.3
10%	24.3	29.7	25.9
100 Portfolios			
1%	3.1	7.3	8.2
5%	14.2	22.3	23.0
10%	26.5	35.1	36.3

This table shows the rejection rates over 1000 trials using the P-value of the HJ-distance, but approximating the weighting matrix, G, by the sample second moment matrix from 10,000 time-series observations.

HJ-distance test with large T occurs mainly through a more accurate estimation of G. In the Premium-Labor model, there still remains size distortion, but its magnitude is much smaller than those in Panel C of Table 1.

4. Improved estimation of covariance matrix by shrinkage

The simulation evidence in the previous section reveals that the finite sample performance of the HJ-distance test improves significantly when one employs a better estimate of the second moment matrix of portfolio returns, or equivalently, a better estimate of the covariance matrix of portfolio returns. In this section, we explore the possibility of improved estimation of the portfolio covariance matrix by the shrinkage method following the approach of Ledoit and Wolf (2003).

4.1. Shrinkage method and the HJ-distance

The shrinkage method dates back to the seminal paper by Stein (1956). The basic idea behind the shrinkage method is to balance the trade-off between bias and variance by taking a weighted average of two estimators. If one estimator is unbiased but has a large variance while the other estimator is biased but has a small variance, then taking a properly weighted average of the two estimators can outperform both estimators in terms of accuracy (mean squared error). The biased estimator is called the shrinkage target to which the unbiased estimator with a large variance is shrunk.

In our context, the sample covariance matrix is an unbiased estimator of the true covariance matrix but has a large variance. Note that the purpose of the HJ-distance test is to test if a SDF can price the returns correctly. Therefore, a natural choice of the shrinkage target is the covariance matrix implied by the factor model which implies the SDF of interest. Factor pricing models explain asset returns in terms of a few factors and uncorrelated residuals, thereby imposing a low-dimensional factor structure to the returns. Since the parameters of a factor model can be estimated with a small variance, the estimate of the asset covariance matrix implied by the factor model has a small variance, although it is a biased estimate when the factor model is misspecified.

The HJ-distance test is often used to compare the fit of different SDFs. The shrinkage method is attractive in such circumstances, because one can use the same shrinkage estimate of G in computing the test statistic for different models. In other words, the shrinkage target model does not need to be the same model as the model being tested. Comparing different SDFs by the HJ-distance requires that the same weighting matrix be used across all candidate SDFs, but one does not know which SDF is the correct SDF *a priori*. Using the shrinkage method allows one to use the same weighting matrix across different SDFs without assuming a particular SDF is correct. For this reason, it is undesirable to use the asset covariance matrix implied by the factor model alone, without combining it with the sample covariance.

4.2. Optimal shrinkage intensity

Shrinkage method assigns α weight to a covariance matrix implied by a factor model and the other $1-\alpha$ weight to the sample covariance matrix. Using the shrinkage method requires the determination of α, which is called the shrinkage intensity. Ledoit and Wolf (2003) derive the analytical formula for the optimal α and discuss its estimation when the shrinkage target is a single-factor model and is a misspecified model of asset returns. We extend their method to multiple-factor shrinkage targets as well as to the case where the shrinkage target is the correct model of asset returns.

As in Section 2, let R_t denote a $1\times N$ vector of the t-th period gross returns of N assets, and let X_t denote a $1\times K^*$ vector of factors not including a constant, where $K^*=K-1$. Let R_{ti} denote the t-th period gross return of the i-th asset, so that $R_t=(R_{t1},\ldots,R_{tN})$.

Suppose the following K-factor linear asset pricing model is used to construct the shrinkage target. It is not necessary that the model generates the actual stock returns.

$$R_{ti} = \mu_i + X_t\beta_i + \varepsilon_{ti}, \quad t = 1,\ldots,T, \tag{1}$$

where β_i is a $K^*\times 1$ vector of slopes for the ith asset, and ε_{ti} is the mean-zero idiosyncratic error for asset i in period t. ε_{ti} has a constant variance δ_{ii} across time, and is uncorrelated to ε_{tj} with $j\neq i$ and to the factors. The model (1) may be the asset pricing model corresponding to the SDF we test, but any other linear factor model can be used. Let $\beta=(\beta_1,\ldots,\beta_N)$ denote the $K^*\times N$ matrix of the slopes, $\mu=(\mu_1,\ldots,\mu_N)$ be the $1\times N$ vector of the intercepts, and $\varepsilon_t=(\varepsilon_{t1},\ldots,\varepsilon_{tN})$. Then the factor model (1) is written as

$$R_t = \mu + X_t\beta + \varepsilon_t, \quad Var(\varepsilon_t) = \Delta = \mathrm{diag}(\delta_{ii}), \quad t = 1,\ldots,T. \tag{2}$$

We impose the following assumptions on the stock returns and factors.

Assumption 1. Stock returns R_t and factors X_t are independently and identically distributed over time.

Assumption 2. R_t and X_t have finite fourth moment, and $Var\,(R_t)=\sum$.

We use the iid assumption following Ledoit and Wolf (2003). Section 4 discusses the consequence of allowing R_t and X_t to be heteroscedastic and/or serially correlated.

The asset pricing model (2) implies the following covariance matrix of R_t:

$$\Phi = \beta'\mathrm{Var}(X_t)\beta + \Delta.$$

We can estimate Φ by estimating its components. Regressing the i-th portfolio returns on an intercept and the factors, we obtain the least squares estimate of β_i and the residual variance estimate. Let b_i and d_{ii} denote these estimates of β_i and δ_{ii}, respectively. Let $b=(b_1,\ldots,b_N)$ and $D=\mathrm{diag}(d_{ii})$, then the estimate of Φ is

$$F = b'\widehat{Var}(X_t)b + D, \tag{3}$$

where $\widehat{Var}(X_t)$ is the sample covariance matrix of the factors.

We estimate $\sum$ by a weighted average of F and the sample covariance of R_t, S, with the weight (shrinkage intensity) α assigned to the shrinkage target F. We choose the shrinkage intensity α so that it minimizes a risk function. Let $||Z||$ be the Frobenius norm of an $N\times N$ matrix Z, so

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

Following Ledoit and Wolf (2003), we use the following risk function

$$Q(\alpha) = E[L(\alpha)], \tag{4}$$

where $L(\alpha)$ is a quadratic measure of the distance between the true and estimated covariance matrices

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

Let s_{ij}, f_{ij}, σ_{ij}, and ϕ_{ij} denote the (i,j)-th element of S, F, $\sum$, and Φ, respectively. It follows that

$$\begin{aligned} Q(\alpha) &= \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N} E\left(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}\right)^2 \\ &= \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\mathrm{Var}\left(\alpha f_{ij} + (1-\alpha)s_{ij}\right) + \left[E\left(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}\right)\right]^2\right\} \\ &= \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\alpha^2\mathrm{Var}\left(f_{ij}\right) + (1-\alpha)^2\mathrm{Var}\left(s_{ij}\right) + 2\alpha(1-\alpha)\mathrm{Cov}\left(f_{ij}, s_{ij}\right) + \alpha^2\left(\phi_{ij} - \sigma_{ij}\right)^2\right\}. \end{aligned}$$

The optimal α can be derived by differentiating $Q(\alpha)$ with respect to α. The second order condition is satisfied since $Q(\alpha)$ is concave. Solving the first order condition for α gives the optimal α as

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

which is the same as Eq. (3) in Ledoit and Wolf (2003). Multiplying both the numerator and the denominator by T, we obtain

$$\alpha^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Var}\left(\sqrt{T}s_{ij}\right) - \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\left(\sqrt{T}f_{ij}, \sqrt{T}s_{ij}\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Var}\left(\sqrt{T}f_{ij} - \sqrt{T}s_{ij}\right) + T\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\phi_{ij} - \sigma_{ij}\right)^2}. \tag{5}$$

As in Ledoit and Wolf (2003), define $\pi = \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{AsyVar}\left[\sqrt{T}s_{ij}\right], \rho = \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{AsyCov}\left[\sqrt{T}f_{ij}, \sqrt{T}s_{ij},\right]$ and let $\gamma = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\phi_{ij} - \sigma_{ij}\right)^2$ denote the measure of the misspecification of the factor model (2). Define $\eta = \sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{AsyVar}\left[\sqrt{T}\left(f_{ij} - s_{ij}\right)\right]$. We consider the limit of α^* as $T\to\infty$ in two cases separately, depending on whether $\Phi=\sum$. First, consider the case where $\Phi\neq\sum$. Since S is consistent while F is not, the optimal shrinkage intensity α^* converges to 0 as $T\to\infty$. Ledoit and Wolf (2003) prove in their Theorem 1 that

$$T\alpha^* \to \frac{\pi-\rho}{\gamma}, \quad \text{as } T\to\infty. \tag{6}$$

When $\Phi=\sum$, both S and F are consistent for $\sum$, but they have different variances. In this case, the optimal shrinkage intensity α^* converges to a degenerating limit

$$\alpha^* \to \frac{\pi-\rho}{\eta}, \quad \text{as } T\to\infty. \tag{7}$$

This case is not considered in Ledoit and Wolf (2003), but the proof of Eq. (7) follows from the proof of Theorem 1 of Ledoit and Wolf (2003, pp. 610–611). From Eqs. (6) and (7), the shrinkage estimate $aF+(1-a)S$ is consistent for $\sum$ under both $\Phi\neq\sum$ and $\Phi=\sum$. Note that $(\pi-\rho)/\eta$ does not necessarily equal 1. $(\pi-\rho)/=1$ if F is an asymptotically efficient estimator of $\sum$.

4.3. Estimation of the optimal shrinkage intensity

Since π, ρ, μ and γ in the formula for α^* are unobservable, we must find estimators for them. Define $\pi_{ij} = AsyVar\left[\sqrt{T}s_{ij}\right], \rho_{ij} = AsyCov\left[\sqrt{T}f_{ij}, \sqrt{T}s_{ij}\right]$, $\gamma_{ij} = \left(\phi_{ij} - \sigma_{ij}\right)^2$, and $\eta_{ij} = AsyVar\left[\sqrt{T}(f_{ij} - s_{ij})\right]$, so that $\pi = \sum_{i=1}^{N}\sum_{j=1}^{N}\pi_{ij}$, $\rho = \sum_{i=1}^{N}\sum_{j=1}^{N}\rho_{ij}$, $\gamma = \sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_{ij}$, and $\eta = \sum_{i=1}^{N}\sum_{j=1}^{N}\eta_{ij}$. In the following, we present consistent estimates of these quantities and show the asymptotic behavior of our estimate of a^*.

4.3.1. π_{ij} and γ_{ij}

From Lemma 1 of Ledoit and Wolf (2003), a consistent estimator for π_{ij} is given by

$$p_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left[\left(R_{ti} - m_i\right)\left(R_{tj} - m_j\right) - s_{ij}\right]^2, \tag{8}$$

where $m_i = T^{-1}\sum_{t=1}^{T}R_{ti}$ is the sample average of the return of the i-th asset. Define $c_{ij} = (f_{ij} - s_{ij})^2$, then $c_{ij} \to_p \gamma_{ij}$ follows from Lemma 3 of Ledoit and Wolf (2003).

4.3.2. ρ_{ij}

When $i = j$, note that $f_{ii} = s_{ii}$. Thus we can use p_{ii} to estimate ρ_{ii}. When $i \neq j$, first define $M = I - T^{-1}11'$, where I is a $T\times T$ identity matrix and 1 is a $T\times 1$ vector of ones. Collect the factors into a $T\times K^*$ matrix X: $X = (X_1', \ldots, X_T')'$. We use R_i to denote a $T\times 1$ vector of the i-th asset return. Recall that (see Eq. (3))

$$F = b'\widehat{Var}(X_t)b + D,$$

where $b = (b_1, \ldots, b_N)$, and b_i is given by $b_i = (X'MX)^{-1}X'MR_{\cdot i}$. D is a diagonal matrix of residual variance estimates.

Define $S_{xi} = T^{-1}R_{\cdot i}'MX$, which is the $1\times K^*$ sample covariance vector between X_t and R_{tj}, and define $S_{xx} = T^{-1}X'MX$, which is the $K^*\times K^*$ sample covariance matrix of X_t. Then we can express f_{ij} for $i \neq j$ as

$$f_{ij} = R_{\cdot i}'MX(X'MX)^{-1}T^{-1}(X'MX)(X'MX)^{-1}X'MR_{\cdot j} = S_{xi}(S_{xx})^{-1}\left(S_{xj}\right)'. \tag{9}$$

Let $\bar{X} = T^{-1}\sum_{t=1}^{T}X_t = 1\times K^*$ vector of the sample average of the factors; $\sigma_{xj} = 1\times K^*$ covariance vector between X_t and R_{tj}; $\sigma_{xx} = K^*\times K^*$ covariance matrix of X_t.

The following lemma provides a consistent estimator of ρ_{ij}. Recall that s_{ij} denotes the (i, j)-th element of S and is equal to the sample covariance between R_{ti} and R_{tj}.

Lemma 1. *A consistent estimator of ρ_{ij} is given by r_{ij}, defined as follows: for $i = j$, set $r_{ii} = p_{ii}$, and for $i \neq j$, set r_{ij} as*

$$r_{ij} = Z_i S_{xx}^{-1}\left(S_{xj}\right)' + S_{xi}S_{xx}^{-1}\left(Z_j\right)' - S_{xi}S_{xx}^{-1}Z_x S_{xx}^{-1}\left(S_{xj}\right)',$$

where $Z_{xi,ij}$ and $Z_{xx,ij}$ are consistent estimates of $AsyCov\left[\sqrt{T}S_{xi}, \sqrt{T}s_{ij}\right]$ and $AsyCov\left[\sqrt{T}S_{xx}, \sqrt{T}s_{ij}\right]$, respectively, and they take the form

$$Z_{xi,ij} = T^{-1}\sum_{t=1}^{T}\left[(X_t - \bar{X})(R_{ti} - m_i) - S_{xi}\right]\left[(R_{ti} - m_i)\left(R_{tj} - m_j\right) - s_{ij}\right],$$
$$Z_{xx,ij} = T^{-1}\sum_{t=1}^{T}\left[\left(X_t - \bar{X}\right)'\left(X_t - \bar{X}\right) - S_{xx}\right]\left[(R_{ti} - m_i)\left(R_{tj} - m_j\right) - s_{ij}\right].$$

Proof. For $i = j$, the stated result follows from $f_{ii} = s_{ii}$. For $i \neq j$, from Eq. (9), f_{ij} converges to $\sigma_{xi}\sigma_{xx}^{-1}(\sigma_{xj})'$ in probability. Expanding $\sqrt{T}f_{ij}$ around $\sqrt{T}\sigma_{xi}\sigma_{xx}^{-1}(\sigma_{xj})'$ gives

$$\begin{aligned}\sqrt{T}f_{ij} = {} & \sqrt{T}\sigma_{xi}\sigma_{xx}^{-1}\left(\sigma_{xj}\right)' + \sqrt{T}(S_{xi} - \sigma_{xi})\sigma_{xx}^{-1}\left(\sigma_{xj}\right)' + \sqrt{T}\sigma_{xi}\sigma_{xx}^{-1}\left(S_{xj} - \sigma_{xj}\right)' \\ & - \sigma_{xi}\sigma_{xx}^{-1}\sqrt{T}(S_{xx} - \sigma_{xx})\sigma_{xx}^{-1}\left(\sigma_{xj}\right)' + o_p(1),\end{aligned}$$

where the third term follows from $\partial(X(\theta)^{-1})/\partial\theta = -X(\theta)^{-1}(\partial X(\theta)/\partial\theta)X(\theta)^{-1}$. It follows that

$$\begin{aligned}AsyCov\left[\sqrt{T}f_{ij}, \sqrt{T}s_{ij}\right] = {} & AsyCov\left[\sqrt{T}S_{xi}, \sqrt{T}s_{ij}\right]\sigma_{xx}^{-1}\left(\sigma_{xj}\right)' + \sigma_{xi}\sigma_{xx}^{-1}AsyCov\left[\sqrt{T}\left(S_{xj}\right)', \sqrt{T}s_{ij}\right] \\ & -\sigma_{xi}\sigma_{xx}^{-1}\mathrm{AsyCov}\left[\sqrt{T}S_{xx}, \sqrt{T}s_{ij}\right]\sigma_{xx}^{-1}\left(\sigma_{xj}\right)'.\end{aligned}$$

Since (X_t, R_t) is iid, the three asymptotic covariances on the right-hand side are estimated consistently by Z_i, $(Zj)'$, and Z_x, respectively. The required result follows because S_{xj} and S_{xx} are consistent estimates of σ_{xj} and σ_{xx}. □

4.3.3. η_{ij}

A similar analysis gives the following lemma. Its proof follows from the proof of lemma 1 and hence omitted. Let $\{A\}_{kl}$ denote the (k,l)-th element of matrix A, and let $\{a\}_k$ denote the k-th element of vector a.

Lemma 2. *A consistent estimator of η_{ij} is given by $h_{ij}=w_{ij}+p_{ij}-2r_{ij}$, where w_{ij} is a consistent estimator of* $\mathrm{AsyVar}\sqrt{T}f_{ij}$. *For $i=j$, we set $w_{ii}=p_{ii}$. For $i\neq j$, w_{ij} is given by*

$$\begin{aligned} w_{ij} &= S_{xj}S_{xx}^{-1}Z_{ii}^{a}S_{xx}^{-1}\left(S_{xj}\right)' + S_{xi}S_{xx}^{-1}Z_{jj}^{a}S_{xx}^{-1}(S_{xi})' + 2S_{xi}S_{xx}^{-1}Z_{ji}^{a}S_{xx}\left(S_{xj}\right)' \\ &\quad + \sum_{k=1}^{K}\sum_{l=1}^{K}\left[\left\{S_{xi}S_{xx}^{-1}\right\}_k\left\{S_{xj}S_{xx}^{-1}\right\}_l S_{xi}S_{xx}^{-1}Z_{kl}^{b}S_{xx}^{-1}\left(S_{xj}\right)'\right] \\ &\quad - 2\sum_{k=1}^{K}\sum_{l=1}^{K}\left[\left\{S_{xi}S_{xx}^{-1}\right\}_k\left\{S_{xj}S_{xx}^{-1}\right\}_l\left(Z_{i,kl}^{c}S_{xx}^{-1}\left(S_{xj}\right)' + Z_{j,kl}^{c}S_{xx}^{-1}(S_{xi})'\right)\right] \end{aligned}$$

where Z_{ij}^{a}, Z_{kl}^{b}, $Z_{i,kl}^{c}$ are consistent estimates of $\mathrm{AsyCov}\left[\sqrt{T}S_{xi},\sqrt{T}S_{xj}\right]$, $\mathrm{AsyCov}\left[\sqrt{T}S_{xx},\sqrt{T}\{S_{xx}\}_{kl}\right]$, *and* $\mathrm{AsyCov}\left[\sqrt{T}S_{xi},\sqrt{T}\{S_{xx}\}_{kl}\right]$, *respectively, and they take the form*

$$\begin{aligned} Z_{ij}^{a} &= T^{-1}\sum_{t=1}^{T}\left[(R_{ti}-m_i)\left(X_t-\bar{X}\right)' - (S_{xi})'\right]\left[\left(R_{tj}-m_j\right)\left(X_t-\bar{X}\right) - S_{xj}\right], \\ Z_{kl}^{b} &= T^{-1}\sum_{t=1}^{T}\left[\left(X_t-\bar{X}\right)'\left(X_t-\bar{X}\right) - S_{xx}\right]\left[\left\{\left(X_t-\bar{X}\right)'\left(X_t-\bar{X}\right) - S_{xx}\right\}_{kl}\right], \\ Z_{i,kl}^{c} &= T^{-1}\sum_{t=1}^{T}\left[(R_{ti}-m_i)\left(X_t-\bar{X}\right) - S_{xi}\right]\left[\left\{\left(X_t-\bar{X}\right)'\left(X_t-\bar{X}\right) - S_{xx}\right\}_{kl}\right]. \end{aligned}$$

If we assume R_t and X_t are normally distributed, we can simplify the estimation of α^* because p_{ij} in Eq. (8) and Z matrices in Lemmas 1 and 3 can be replaced with a function of sample first and second moments.

4.3.4. Estimate of α^* and its asymptotic behavior

We construct an estimate of the optimal shrinkage intensity by replacing the unknowns in α^* in Eq. (5) with their estimates:

$$\hat{\alpha} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}p_{ij} - \sum_{i=1}^{N}\sum_{j=1}^{N}r_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_{ij} + T\sum_{i=1}^{N}\sum_{j=1}^{N}c_{ij}}. \tag{10}$$

We analyze the asymptotic behavior of $\hat{a}$ for the following two cases:

Case 1. $\Phi\neq\sum$.

Case 2. The stock returns are generated by the factor model (2), and ε_t is independently and identically distributed over time with finite fourth moment.

These two cases cover most situations of practical interest. They leave out only a small case in which $\Phi=\sum$ but the stock returns are not generated by the factor model (2). Note that Case 1 does not assume that returns are generated by the factor model (2). It only assumes Assumptions 1 and 2, thus the idiosyncratic errors can be cross-sectionally correlated.

The following lemma shows that, in Case 1, $T\hat{\alpha}$ converges in probability to the limit of $T\alpha^*$, while in Case 2, $\hat{a}$ converges to a random variable which is smaller than α^*. Since $0<\alpha^0<\alpha^*$ and the risk function $Q(\alpha)$ given by Eq. (4) is concave, the shrinkage estimator has a smaller risk than the sample covariance matrix. The simulations in the following section show that using the shrinkage estimator leads to a substantial improvement of the finite sample performance of the HJ-distance test.

Lemma 3. As $T\to\infty$, we have

$$\begin{aligned} T\hat{\alpha} &\to_p \frac{\pi-\rho}{\gamma} = \lim_{T\to\infty}T\alpha^* \quad \text{in Case 1}, \\ \hat{\alpha} &\to_d \alpha^0 = \frac{\pi-\rho}{\eta+\xi} < \alpha^*, \qquad \text{in Case 2}, \end{aligned}$$

where $\xi=\sum_{i=1,i\neq j}^{N}\sum_{j=1}^{N}(\xi_{ij})^2$ and $\{\xi_{ij}\}_{i,j=1,\ldots,N,I\neq j}$ are jointly normally distributed with mean zero.
Indeed, an estimate of α^* that is consistent in both Case 1 and Case 2 is given by

$$\tilde{\alpha} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}p_{ij} - \sum_{i=1}^{N}\sum_{j=1}^{N}r_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_{ij} + T^{a}\sum_{i=1}^{N}\sum_{j=1}^{N}c_{ij}}, \quad a\in(0,1). \tag{11}$$

By downweighting $\sum_{i=1}^{N}\sum_{j=1}^{N}c_{ij}$, this estimate favors the possibility that $\Phi=\sum$. From the proof of Lemma 3, it follows straightforwardly that $\tilde{\alpha}\to_p 0$ in Case 1 and $\tilde{\alpha}\to_p(\pi-\rho)/\eta$ in Case 2. However, $\tilde{\alpha}$ converges to 0 at a slower rate than α^* in Case 1. This reflects a trade-off between the consistency in both cases and the higher-order consistency in Case 1. Our preference of $\hat{a}$ over $\tilde{a}$ and our choice of a conservative position regarding this trade-off seems appropriate, because we expect that simple factor models are used as a shrinkage target in practice and those models are neither likely to nor meant to provide a complete description of the observed data.

Proof. In Case 1, rewrite $T\hat{\alpha}$ as

$$T\hat{\alpha}=\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}p_{ij}-\sum_{i=1}^{N}\sum_{j=1}^{N}r_{ij}}{(1/T)\sum_{i=1}^{N}\sum_{j=1}^{N}h_{ij}+\sum_{i=1}^{N}\sum_{j=1}^{N}c_{ij}}.$$

Then the stated result follows from $p_{ij}\to_p\pi_{ij}$ and $c_{ij}\to_p\gamma_{ij}$ (Ledoit and Wolf (2003), Lemmas 1 and 3), and Lemmas 1 and 3. In Case 2, it follows from $p_{ij}\to_p\pi_{ij}$ and Lemmas 1 and 3 that

$$\hat{\alpha}=\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}p_{ij}-\sum_{i=1}^{N}\sum_{j=1}^{N}r_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_{ij}+T\sum_{i=1}^{N}\sum_{j=1}^{N}c_{ij}}=\frac{\pi-\rho+o_p(1)}{\eta+o_p(1)+\sum_{i=1}^{N}\sum_{j=1}^{N}Tc_{ij}}.$$

We proceed to derive the asymptotic distribution of $Tc_{ij}=T(f_{ij}-s_{ij})^2$. Recall $c_{ij}=0$ for $i=j$. For $i\neq j$, we have, from the definition of s_{ij} and Eq. (9),

$$s_{ij}=T^{-1}R'_{\cdot i}MR_{\cdot j},\quad f_{ij}=T^{-1}R'_{\cdot i}MX(X'MX)^{-1}X'MR_{\cdot j}. \tag{12}$$

Define $\varepsilon_{\cdot i}=(\varepsilon_{1i},\ldots,\varepsilon_{Ti})'$, and rewrite the model (2) as $R_{\cdot i}=\mu_i 1+X\beta_i+\varepsilon_{\cdot i}$ for $i=1,\ldots,N$. Substituting this into Eq. (12), we can express the difference between f_{ij} and s_{ij} as

$$\begin{aligned}f_{ij}-s_{ij}&=T^{-1}(X\beta_i+\varepsilon_{\cdot i})'MX(X'MX)^{-1}X'M\left(X\beta_j+\varepsilon_{\cdot j}\right)-T^{-1}(X\beta_i+\varepsilon_{\cdot i})'M\left(X\beta_j+\varepsilon_{\cdot j}\right)\\&=T^{-1}\varepsilon'_{\cdot i}MX(X'MX)^{-1}X'M\varepsilon_{\cdot j}-T^{-1}\varepsilon'_{\cdot i}M\varepsilon_{\cdot j}\\&=T^{-1}\left(T^{-1/2}\varepsilon'_{\cdot i}MX\right)\left(T^{-1}X'MX\right)^{-1}\left(T^{-1/2}X'M\varepsilon_{\cdot j}\right)-T^{-1}\varepsilon'_{\cdot i}M\varepsilon_{\cdot j}.\end{aligned}$$

Since $T^{-1/2}\varepsilon'_{\cdot i}MX=T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti}X_t-(T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti})(T^{-1}\sum_{t=1}^{T}X_t)=O_p(1)$, $T^{-1}X'MX\to_p\sigma_{xx}$, and $T^{-1/2}\varepsilon'_{\cdot i}M\varepsilon_{\cdot j}=T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti}\varepsilon_{tj}-(T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti})(T^{-1}\sum_{t=1}^{T}\varepsilon_{tj})=T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti}\varepsilon_{tj}+O_p(1)$, it follows that

$$\sqrt{T}\left(f_{ij}-s_{ij}\right)=T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti}\varepsilon_{tj}+o_p(1).$$

Since $\varepsilon_{ti}\varepsilon_{tj}$ is iid with mean 0 and finite variance, an $(N^2-N)\times 1$ vector $\{T^{-1/2}\sum_{t=1}^{T}\varepsilon_{ti}\varepsilon_{tj}\}_{i,j=1,\ldots,N,i\neq j}$ converges to a normally distributed random vector in distribution. □

The following theorem is a simple consequence of Lemma 3:

Theorem 1. *Define the shrinkage covariance matrix estimate* $\hat{\sum}$ *as*

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

Then $\hat{\sum}\to_p\sum$ *as* $T\to\infty$, *because if* $\Phi\neq\sum$ *then* $\hat{\alpha}\to 0$ *and if* $\Phi=\sum$ *then both F and S are consistent for* $\sum$.

4.4. Shrinkage under heteroscedasticity and/or serial correlation

In practice, the returns and factors are likely to be heteroscedastic and/or serially correlated. In this subsection, we relax Assumption 2 (the iid-assumption) and analyze how the results in the above are affected.

To allow for heteroscedasticity and/or serial correlation, we redefine G and $\sum$ as $G=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E(R_t'R_t)$ and $\sum=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}Var(R_t)$. The sample covariance matrix, S, is a consistent estimate of $\sum$ under suitable regularity conditions (for example, mixing). Defining the factor model as in Eq. (2), and defining $V_x=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}Var(X_t)$ and $\Delta=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}Var(\varepsilon_t)$, it follows that $\sum$ implied by the factor model can be written as $\Phi=\beta'V_x\beta+\Delta$. Consequently, we can define the optimal shrinkage intensity, α^*, by replacing the variances and covariances in formula (5) with their asymptotic counterpart. However, $\hat{\alpha}$ as defined in Eq. (10) does not converge to the limit stated in Lemma 3 because (p,r,h) are not consistent

Table 3
Rejection frequencies of the HJ-distance test with shrinkage estimation of G.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	1.6	1.3	0.8
5%	6.6	6.8	5.4
10%	13.4	12.8	10.4
100 Portfolios			
1%	3.9	1.3	0.9
5%	15.1	7.2	5.3
10%	28.4	13.7	11.1
(B) Fama–French model			
25 Portfolios			
1%	1.3	1.2	0.7
5%	5.8	5.1	4.0
10%	9.9	9.8	9.6
100 Portfolios			
1%	23.3	7.7	2.8
5%	49.9	22.9	11.2
10%	64.3	33.6	20.5
(C) Premium-Labor model			
25 Portfolios			
1%	6.6	9.7	5.4
5%	18.8	20.8	13.5
10%	28.7	32.4	23.4
100 Portfolios			
1%	31.5	19.5	11.2
5%	59.0	42.4	28.8
10%	73.0	58.4	40.0

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance, but approximating the weighting matrix, G, by a shrinkage method, which averages the sample covariance and the structure covariance with an optimal weight.

for (π,ρ,η). In Case 1, we still obtain $\hat{\alpha}+O_p(T^{-1})$ because $(p,r,h)=O_p(1)$. In Case 2, $\hat{\alpha}$ converges to a random variable that may or may not be smaller than α^*. In both cases, $\hat{\sum}=\alpha F+(1-\hat{\alpha})S$ is consistent for $\sum$, and Theorem 1 holds.

5. Simulation results with the shrinkage method

With the shrinkage covariance matrix estimate $\hat{\sum}$constructed in the previous section, we define the shrinkage estimate of the second moment of the asset returns as

$$\hat{G}=\hat{\Sigma}+\left(\frac{1}{T}\sum_{t=1}^{T}R_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T}R_t\right).$$

In this section, we examine the finite sample performance of the HJ-distance test when the inverse of $\hat{G}$ is used as the weighting matrix. The other settings of the Monte Carlo experiments are the same as in Section 2. In order to avoid overshrinkage or negative shrinkage, we set 0 and 1 as the lower and upper bound for $\hat{\alpha}$.

5.1. Shrinkage with a correctly specified target model

Table 3 reports the rejection frequencies of the HJ-distance test with $\hat{G}$. Compared with Table 1, we find that the rejection frequencies improve in all cases. For example, for the Simple model and Fama–French model with 25 portfolios, the rejection frequency of the HJ-distance test in Table 1 is more than twice the nominal level for $T=160$ and 330 whereas the rejection frequencies in Table 3 are close to the nominal level for all T. With 100 portfolios, the HJ-distance test with $\hat{G}$ still tends to overreject the correct null, but the degree of overrejection is much smaller than in Table 1.

With 25 portfolios, the rejection frequencies in Table 3 are almost as good as those in Table 2. One possible explanation for this is that an accurate estimate of G also contributes to an accurate estimate of Ω, because Ω_T depends on δ_T, which in turn depends on G_T. Table 4 reports the bias and MSE (measured by the Frobenius norm) of the estimate of Ω associated with the shrunk and non-shrunk estimate of G. With the Simple model and Fama–French model, applying shrinkage to G improves the accuracy of the estimate of Ω in many cases, whereas shrinkage on G has little effect on the estimate of Ω with the Premium-Labor model. Therefore, improved estimation of Ω seems to give a partial explanation.

On the other hand, according to Table 6 of Ahn and Gadarowski (2004), using the exact Ω and sample second moment matrix for G gives a close to nominal size for the Simple model with both 25 and 100 portfolios. One possible reason may be that the

Table 4
Bias and MSE of the estimates of Ω.

Number of observations		$T=160$	$T=330$	$T=700$
(A) Simple Model				
25 Portfolios				
Sample covariance	Bias	4.9631	2.2638	0.9872
	MSE	149.14	27.436	4.0172
Shrinkage on G	Bias	4.4209	2.1646	1.0631
	MSE	73.646	17.502	3.5766
100 Portfolios				
Sample covariance	Bias	67.755	29.960	13.545
	$\text{MSE}\times 10^{-3}$	46.667	4.5576	0.9139
Shrinkage on G	Bias	86.477	36.594	12.882
	$\text{MSE}\times 10^{-3}$	36.915	5.2441	0.4710
(B) Fama–French Model				
25 Portfolios				
Sample covariance	Bias	1.9197	0.7675	0.2654
	MSE	22.016	2.4207	0.3133
Shrinkage on G	Bias	1.5981	0.6342	0.2122
	MSE	11.957	1.5074	0.2433
100 Portfolios				
Sample covariance	Bias	129.09	68.458	28.420
	$\text{MSE}\times 10^{-3}$	28.350	9.4855	2.4478
Shrinkage on G	Bias	113.76	61.986	24.529
	$\text{MSE}\times 10^{-3}$	25.028	8.1222	1.4210
(C) Premium-Labor Model				
25 Portfolios				
Sample covariance	Bias	640.86	463.86	259.74
	$\text{MSE}\times 10^{-5}$	4.4285	2.6716	0.9872
Shrinkage on G	Bias	632.33	458.26	254.24
	$\text{MSE}\times 10^{-5}$	4.4137	2.5621	0.9171
100 Portfolios				
Sample covariance	$\text{Bias}\times 10^{-3}$	3.9618	3.1428	1.6773
	$\text{MSE}\times 10^{-7}$	1.6823	1.0647	0.3456
Shrinkage on G	$\text{Bias}\times 10^{-3}$	4.2411	3.1674	1.7347
	$\text{MSE}\times 10^{-7}$	1.8524	1.0675	0.3428

This table shows the bias and MSE (measured by the Frobenius norm) of the estimate of Ω under two estimation methods of G, a shrinkage method and the sample covariance.

Table 5
Summary statistics of $\hat{\alpha}$.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
Mean	0.8290	0.8757	0.8981
Standard deviation	0.1287	0.1040	0.0895
100 Portfolios			
Mean	0.8180	0.8722	0.8951
Standard deviation	0.0953	0.0679	0.0532
(B) Fama–French model			
25 Portfolios			
Mean	0.9280	0.9462	0.9605
Standard deviation	0.0780	0.0631	0.0533
100 Portfolios			
Mean	0.6324	0.6443	0.6514
Standard deviation	0.0909	0.0875	0.0844
(C) Premium-Labor model			
25 Portfolios			
Mean	0.8120	0.8152	0.8199
Standard deviation	0.0832	0.0780	0.0753
100 Portfolios			
Mean	0.7304	0.7326	0.7340
Standard deviation	0.0297	0.0238	0.0206

This table shows the mean and the standard deviation of the estimated optimal shrinkage intensity $\hat{\alpha}$ for each model.

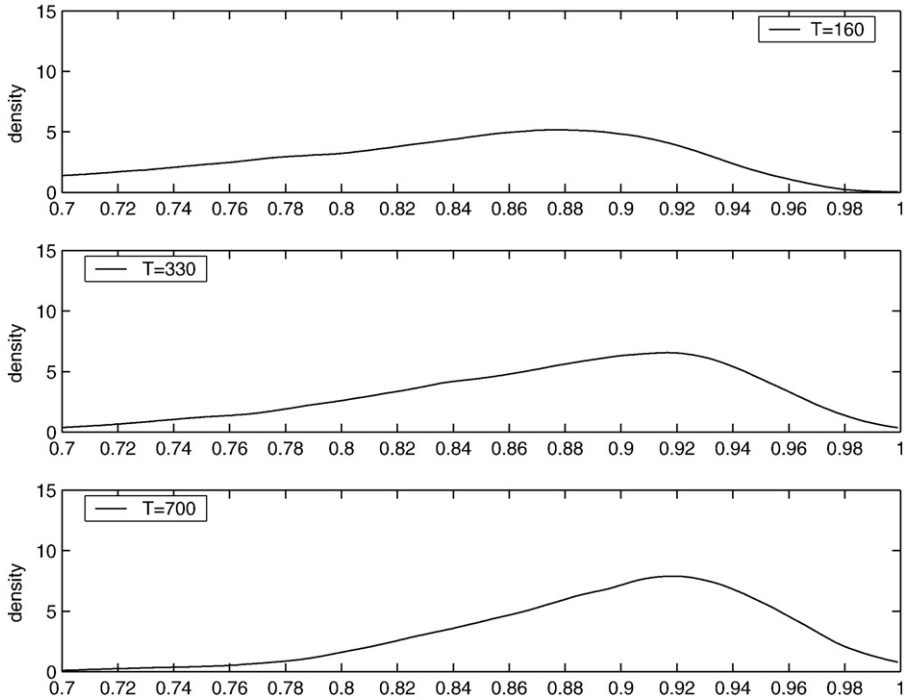

Fig. 1. The density functions for $\hat{\alpha}$ in the Simple model of 100 Portfolios.

sampling errors in estimated G and Ω create some undesirable correlations in finite samples. We examine this possibility by conducting the HJ-distance test with setting $G=I$, an identity matrix. The resulting statistic is no longer a measure of the maximum pricing error, but the test is valid, and its p-value can be computed by the same way as the p-value of the HJ-test is

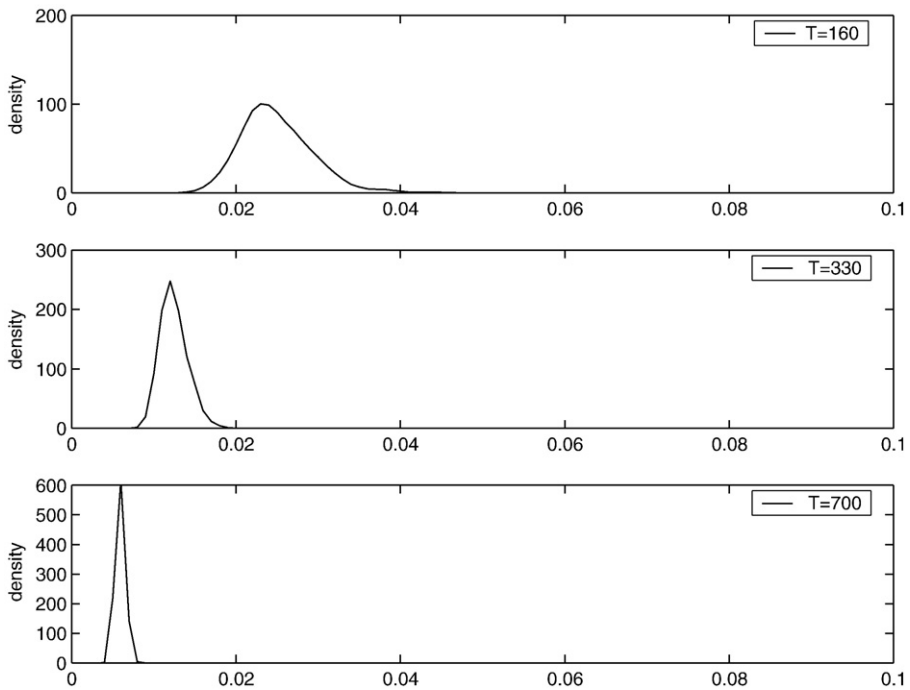

Fig. 2. The density functions for $\hat{\alpha}$ in a misspecified Simple model of 100 Portfolios.

Table 6
The mean squared error of the HJ-distance from two estimation methods of G.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
Sample second moment	0.0084	0.0017	0.00046
Shrinkage estimation	0.0033	0.0009	0.00032
100 Portfolios			
Sample second moment	0.5630	0.0394	0.0040
Shrinkage estimation	0.0424	0.0067	0.0009
(B) Fama–French model			
25 Portfolios			
Sample second moment	0.00440	4.9618×10^{-4}	6.4341×10^{-5}
Shrinkage estimation	0.00098	1.1522×10^{-4}	1.4838×10^{-5}
100 Portfolios			
Sample second moment	0.5606	0.0385	0.0036
Shrinkage estimation	0.0511	0.0077	0.0009
(C) Premium-Labor model			
25 Portfolios			
Sample second moment	0.0049	7.4934×10^{-4}	1.1823×10^{-4}
Shrinkage estimation	0.0009	1.6685×10^{-4}	2.9242×10^{-5}
100 Portfolios			
Sample second moment	0.5139	0.0366	0.0038
Shrinkage estimation	0.0331	0.0054	0.0007

This table compares the HJ-distances when the sample second moment or the shrinkage estimate was used as the weighting matrix. Here, we report the mean squared error.

computed. The rejection frequencies of the resulting specification test are similar to those in Table 2.[3] This and Table 2 give an indirect evidence of correlations between the sampling errors in estimated G and Ω.

Kan and Zhou (2004) derive the exact distribution of the HJ-distance under the normality assumption. Their Tables 1 and 3 report the rejection frequency of the asymptotic HJ-distance test and that of the feasible version of their exact test, respectively. We compare their Table 3 with the results for the Fama–French model in our Table 3.[4] With 25 portfolios, the HJ-distance test with shrinkage performs as well as the exact test, and the actual sizes of both tests are close to the nominal size. With 100 portfolios, the exact test performs substantially better than the shrinkage version. This is probably due to a poor chi-squared approximation. However, very few applications use as many as 100 portfolios; of the applications of the HJ-distance tests surveyed in the Introduction, all of them but Jagannathan and Wang (1996) use fewer than 25 portfolios. Therefore, we may conclude that the HJ-distance test with shrinkage performs as well as the exact test for most portfolio sizes of practical interest.

Table 5 reports the summary statistics of the estimated optimal shrinkage intensity $\hat{\alpha}$. Fig. 1 shows the kernel density estimate of $\hat{\alpha}$ for the Simple model. This corresponds to the case where $\Phi=\sum$ in Lemma 3. The results with the other factor models are similar and thus not reported here. From Table 5 and Fig. 1, we can see that $\hat{\alpha}$ is centered around 0.8 ~ 1 and the estimated covariance matrices are much closer to F than the sample covariance matrix when $\Phi=\sum$. Fig. 2 shows the kernel density estimate of $\hat{\alpha}$ when the data are simulated from the Simple model with 100 portfolios, but only two of the three factors are used in constructing F. This corresponds to the case where $\Phi\neq\sum$ in Lemma 3. Fig. 2 shows that $\hat{\alpha}$ is converging to zero, corroborating Lemma 3.

One important feature of the shrinkage method is that it provides a better estimate of the HJ-distance itself. Table 6 reports the MSE of the HJ-distance with two estimates of G relative to the HJ-distance computed with the true value of G. The MSE of the HJ-distance with shrinkage is less than half of and substantially smaller than the MSE of the HJ-distance with sample covariance. Therefore, the shrinkage method provides a more accurate comparison of the HJ-distance across different models. Note that this feature is not present with the exact distribution approach.

5.2. Additional shrinkage

We might expect to improve the size of the HJ-distance test further by applying the shrinkage method to $E(R_t')$ and/or Ω in addition to G. For example, since $E(R_t')$ is a $N\times1$ vector, applying shrinkage to $E(R_t')$ would give a more accurate estimate when N is large. In this subsection, we discuss the simulation results with additional shrinkage.

First, we consider applying shrinkage to G and $E(R_t')$. We use the same shrinkage on G as in Table 3, and a target model for $E(R_t')$ that has the common mean and slope for all portfolios: $R_{ti}=\mu^{\diamond}+X_t\beta^{\diamond}+\varepsilon_{ti}$, where $\mu^{\diamond}$ is a scalar and $\beta^{\diamond}$ is a $K\times1$ vector. The

[3] The results are available from the authors upon request.
[4] Although Kan and Zhou (2004) describe their factors as the Premium-Labor factors when $K=3$, the rejection frequencies of the asymptotic test in their Table I for $K=3$ are too small compared with the results with the Premium-Labor model in our Tables 1 and 3 of Ahn and Gadarowski (2004). In fact, their results for $K=3$ in Table I are more compatible with our results with the Fama-French model.

Table 7
Rejection frequencies of the HJ-distance test with shrinkage estimation of G and $E(R_t)$.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	0.7	1.1	0.8
5%	5.6	4.8	4.3
10%	11.9	11.0	11.0
100 Portfolios			
1%	3.8	1.4	1.4
5%	16.2	6.2	7.8
10%	27.8	12.2	12.2
(A) Fama–French model			
25 Portfolios			
1%	1.0	0.7	1.0
5%	5.5	4.5	4.1
10%	11.0	10.1	9.6
100 Portfolios			
1%	20.0	7.8	3.2
5%	44.6	21.8	10.8
10%	59.8	34.0	19.8
(C) Premium-Labor Model			
25 Portfolios			
1%	6.2	6.4	6.2
5%	19.4	18.7	15.8
10%	29.7	28.3	24.5
100 Portfolios			
1%	31.2	18.2	14.4
5%	57.4	40.4	30.8
10%	71.0	55.0	44.2

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance when the shrinkage method was applied to G and $E(R_t)$.

Table 8
Rejection frequencies of the HJ-distance test with shrinkage estimation of G and Ω.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	0.7	0.4	0.6
5%	5.0	3.6	5.1
10%	10.2	7.9	9.8
100 Portfolios			
1%	1.8	1.0	1.2
5%	12.4	7.2	4.2
10%	22.6	13.0	11.2
(B) Fama–French model			
25 Portfolios			
1%	0.7	0.5	1.0
5%	5.1	3.7	3.8
10%	8.7	9.7	9.2
100 Portfolios			
1%	19.6	6.2	2.8
5%	45.2	18.2	11.0
10%	63.0	29.6	16.2
(C) Premium-Labor model			
25 Portfolios			
1%	5.0	5.8	6.3
5%	15.6	16.7	16.1
10%	25.3	26.5	23.7
100 Portfolios			
1%	31.8	19.0	10.4
5%	58.2	46.2	27.6
10%	72.6	58.2	37.8

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance when the shrinkage method was applied to G and Ω.

Table 9
Rejection frequencies of the HJ-distance test with shrinkage estimation of G using five factors.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	1.3	1.3	0.8
5%	6.6	6.4	5.3
10%	12.8	12.3	10.4
100 Portfolios			
1%	3.5	1.3	0.9
5%	14.9	7.1	5.2
10%	28.0	13.2	11.0
(B) Fama–French model			
25 Portfolios			
1%	1.2	1.2	0.7
5%	5.4	4.9	3.9
10%	9.6	9.6	9.4
100 Portfolios			
1%	23.4	7.8	2.5
5%	49.7	22.3	10.8
10%	63.0	33.7	20.4
(C) Premium-Labor model			
25 Portfolios			
1%	6.3	9.6	5.4
5%	17.9	20.5	13.4
10%	28.8	31.4	23.2
100 Portfolios			
1%	28.5	21.0	11.6
5%	55.5	44.9	30.0
10%	70.4	57.0	42.0

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. We simulated the factors and returns in the same manner as in Table 1, but we used two additional factors when we estimated the factor model and computed the shrinkage target. They were simulated with the same statistic properties of the first two factors in the original three factors.

Table 10
Rejection frequencies of the HJ-distance test with shrinkage estimation of G using two factors.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	4.2	1.8	1.2
5%	13.1	9.8	7.1
10%	23.0	17.2	11.8
100 Portfolios			
1%	97.2	45.0	11.0
5%	99.6	69.8	28.4
10%	99.9	81.3	41.4
(B) Fama–French model			
25 Portfolios			
1%	3.3	2.1	0.9
5%	10.6	7.8	5.9
10%	18.5	13.2	12.0
100 Portfolios			
1%	48.5	23.5	6.0
5%	66.3	38.9	19.6
10%	79.3	50.5	29.6
(C) Premium-Labor model			
25 Portfolios			
1%	7.5	10.4	5.4
5%	20.9	22.6	14.7
10%	29.3	31.7	24.7
100 Portfolios			
1%	53.7	46.8	18.7
5%	78.2	65.0	38.6
10%	85.8	74.7	51.4

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. We simulated the factors and returns in the same manner as in Table 1, but we used only two randomly selected factors when we estimated the factor model and computed the shrinkage target.

Table 11
Rejection frequencies of the HJ-distance test with the estimation of G using a three-factor structure model.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	1.1	1.3	0.8
5%	6.3	6.7	5.3
10%	12.1	12.4	10.5
100 Portfolios			
1%	0.9	1.0	0.9
5%	9.8	6.4	5.4
10%	20.5	13.1	10.9
(B) Fama–French model			
25 Portfolios			
1%	1.3	1.3	0.7
5%	5.5	5.1	4.1
10%	9.6	9.8	9.5
100 Portfolios			
1%	7.1	3.6	2.3
5%	29.0	15.1	8.4
10%	46.5	25.6	16.0
(C) Premium-Labor model			
25 Portfolios			
1%	6.2	9.7	5.4
5%	17.8	20.7	13.5
10%	28.6	32.3	23.5
100 Portfolios			
1%	21.5	18.5	9.4
5%	47.3	40.4	28.0
10%	62.5	53.3	39.5

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. We used three (correct) factors to estimate the structure model, and used the structure covariance as the weighting matrix.

estimate of $E(R_t')$ from this model is $1_N N^{-1} T^{-1} \sum_{i=1}^{T} \sum_{t=1}^{T} R_{ti}$. To determine the shrinkage intensity, we use a similar risk function to the one used for G. Then, the optimal shrinkage intensity estimate is derived analogously.

Table 7 reports the rejection frequencies of the resultant HJ-distance test. As compared with Table 3, we find a small improvement in the rejection frequencies over the shrinkage on only G, particularly when T is small. However, the improvement is rather small in general.

Second, we apply shrinkage to G and $\Omega = \mathrm{E}[w_t(\delta) w_t(\delta)']$. From the definition, $w_t(\delta) = R_t' m_t(\delta) - 1_N$ with $m_t(\delta) = \tilde{X}_t \delta$, which is a scalar. We then use the following two-factor model as the shrinkage target model:

$$w' t = \mu_w + m_t(\delta_T)\beta_w + \varepsilon_t,$$

where $m_t(\delta_T)$ is the factor, and μ_w and β_w are $1 \times N$ vector. We then use the same formula as that used for G to obtain the shrinkage intensity for Ω, in addition to using the same shrinkage on G as in Table 3.

Table 8 reports the rejection frequencies when we apply shrinkage to both G and Ω. We find some improvement over Table 3. The size distortion is noticeably smaller in some cases, for example in the Simple model with $T=160$. In many cases, however, the size improvement over Table 3 is not substantial. When we apply shrinkage only to Ω (but not to G), the results are close to those in Table 1.[5] In this case, using shrinkage does not improve the size of the HJ-distance test, but it does not deteriorate the size either.

5.3. Shrinkage with a misspecified target model

We are also interested in the sensitivity of the shrinkage method to the overspecification and/or underspecification of the factor model used in constructing F. Tables 9 and 10 report the results of the following simulation experiment. For each model, we conduct the HJ-distance test as in Table 3 but we use an overspecified or underspecified factor model to estimate the factor model (1) and construct the shrinkage target F. For the overspecified case, we generate two additional factors with the same statistical properties as the original factors, and use the five-factor model to estimate F. In the underspecified case, we drop one factor randomly from the original three factors, and use the two-factor model to estimate F. In the overspecified case, F is still consistent for $\sum$ but suffers from extra sampling error, while F is inconsistent for $\sum$ and the shrinkage estimate should converge to the sample covariance in the underspecified case.

[5] The results are available from the authors upon request.

Table 12
Rejection frequencies of the HJ-distance test with the estimation of G using a two-factor structure model.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model			
25 Portfolios			
1%	0.0	0.0	0.0
5%	0.0	0.0	0.0
10%	0.1	0.2	0.0
100 Portfolios			
1%	0.0	0.0	0.0
5%	0.0	0.0	0.0
10%	0.0	0.0	0.0
(B) Fama–French model			
25 Portfolios			
1%	0.0	0.0	0.0
5%	0.0	0.1	0.0
10%	0.0	0.3	0.0
100 Portfolios			
1%	0.0	0.0	0.0
5%	0.6	0.6	0.4
10%	3.3	2.5	1.6
(C) Premium-Labor model			
25 Portfolios			
1%	3.5	5.9	3.7
5%	12.1	13.4	9.3
10%	17.7	20.7	15.6
100 Portfolios			
1%	12.6	10.6	6.2
5%	29.8	24.0	15.6
10%	42.8	38.8	25.6

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. In forming the weighting matrix, we randomly selected two factors from the three factors, estimated a two-factor model, and used the implied second moment matrix as the weighting matrix.

Table 9 reports the results with the overspecified target factor model. The rejection frequencies reported in Table 9 are close to those in Table 3, and using an overspecified shrinkage target causes little deterioration in the performance of the HJ-distance test. On the other hand, the results with the underspecified target factor model reported in Table 10 are substantially worse than those in Table 3. However, they still improve upon those in Table 1, in particular with the Fama–French and Premium-Labor models. A researcher can benefit significantly from using the shrinkage method with a possibly overspecified shrinkage target to estimate G.

Tables 9 and 10 suggest that the size distortion from an overspecified shrinkage target is smaller than that from an underspecified shrinkage target. In view of these results, when the models being compared are nested, we recommend using the largest model as the shrinkage target in constructing the (common) estimate of G. When some models are non-nested, we recommend using a shrinkage target that includes all the factors used in the models being compared. Heuristically, using an overspecified target model would still be an improvement over the sample second moment matrix unless the target model has very many factors. This is because the sample second moment matrix corresponds to the shrinkage estimate from the most overspecified factor model, in which there are as many factors as the number of observations.

We also examine the rejection frequencies when we use the factor models to estimate G without shrinkage. Table 11 reports the results with a three-factor model, i.e., the correct model, to estimate G. Not surprisingly, the results are better than the results with shrinkage (Table 3). However, the difference between Tables 3 and 11 is small for models with 25 portfolios. This is also consistent with Fig. 1, which shows that the estimated shrinkage intensity $\hat{\alpha}$ is close to one when the factor model is correctly specified. Table 12 reports the results with an underspecified model in which one factor is randomly dropped. Using an underspecified model to estimate G severely distorts the size of the HJ-distance test. Therefore, it is important to guard against misspecification by using the shrinkage.

The results from Tables 9–12 need to be interpreted with a caution, because only a limited variety of models were tested. It is very interesting to examine the extent to which these results can be generalized. However, such an examination would require extensive simulations with many different combinations of the true and target factor models, and has been left for future investigation.

5.4. GARCH errors

Conducting simulations using two models with GARCH(1,1) errors allows us to examine how our shrinkage procedure performs under heteroscedasticity. The first model is a Simple model with GARCH(1,1) errors, in which ε_t was generated by $\sigma_t^2 = 8.6352\times10^{-7} + 0.9118\sigma_{t-1}^2 + 0.0797\varepsilon_{t-1}^2$. The parameter values were estimated from 3028 daily observations of the NYSE Composite Index from January 2, 1990 to December 31, 2001. The second model used is the Fama–French model with GARCH(1,1)

Table 13
Rejection frequencies of the HJ-distance test when the error follows GARCH (1,1).

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple model without shrinkage			
25 Portfolios			
1%	5.0	2.0	2.0
5%	15.3	9.2	6.8
10%	24.8	17.6	12.2
100 Portfolios			
1%	99.3	51.2	13.2
5%	99.9	73.1	29.5
10%	99.9	82.3	42.8
(B) Simple model with shrinkage			
25 Portfolios			
1%	1.1	0.8	1.3
5%	4.9	5.5	5.1
10%	10.3	11.0	10.1
100 Portfolios			
1%	3.2	0.6	0.8
5%	18.6	9.8	4.2
10%	32.8	19.4	10.2
(C) Fama–French model without shrinkage			
25 Portfolios			
1%	6.2	1.6	2.0
5%	17.6	6.6	7.6
10%	26.6	14.6	11.0
100 Portfolios			
1%	99.4	56.0	15.2
5%	99.8	77.0	35.0
10%	99.8	84.8	47.8
(D) Fama–French model with shrinkage			
25 Portfolios			
1%	0.8	2.0	0.8
5%	5.6	6.6	3.6
10%	10.8	11.4	9.6
100 Portfolios			
1%	4.6	2.8	1.8
5%	19.6	11.4	10.0
10%	29.4	22.6	17.6

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. Factors were simulated in the same manner as in Table 1. ε_t was generated by GARCH (1,1).

errors. For each portfolio, ε_{ti} was generated by a GARCH(1,1) model whose parameter values were calibrated with the Fama–French data.

The first and second panels of Table 13 report the rejection frequencies of the HJ-distance test for the Simple model with and without shrinkage. In both cases, the rejection frequencies are similar to those in Tables 1 and 3. The shrinkage method improves the size of the HJ-distance test substantially.

The third and fourth panels of Table 13 report the rejection frequency of the HJ-distance test for the Fama–French model with and without shrinkage. Again, the shrinkage method improves the size. With 25 portfolios, the rejection frequencies are similar to those in Tables 1 and 3. With 100 portfolios, the rejection frequencies with shrinkage are closer to the nominal size than in the iid case. This is because the serial correlation in variance affects the terms in Eq. (10), and, consequently, affects the shrinkage intensity estimate $\hat{\alpha}$. In this case, $\hat{\alpha}$ under GARCH(1,1) errors turns out to be higher than it is under iid errors; the average of $\hat{\alpha}$ is around 0.9 under GARCH(1,1), while 0.6 it is under iid.

5.5. Nonlinear stochastic discount factor model

We also examine the performance of the shrinkage method when the returns are generated by a nonlinear stochastic discount factor model. We consider a nonlinear version of the Premium-Labor model as an example since Dittmar (2002) points out the potential nonlinearity in the Premium-Labor model. We simulate the following quadratic factor model without cross-products of the factors:

$$R_{ti} = \mu_i + X_{t1}\beta_{1i} + X_{t2}\beta_{2i} + X_{t3}\beta_{3i} + X_{t1}^2\beta_{4i} + X_{t2}^2\beta_{5i} + X_{t3}^2\beta_{6i} + e_{ti}.$$

Table 14
Rejection frequencies of the HJ-distance test using the nonlinear Premium-Labor model.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Sample second moment			
25 Portfolios			
1%	4.8	4.0	4.7
5%	14.1	10.3	12.1
10%	22.2	16.9	19.0
100 Portfolios			
1%	100.0	63.0	19.2
5%	100.0	84.8	44.0
10%	100.0	90.6	56.6
(B) Exact weighting matrix			
25 Portfolios			
1%	1.3	2.1	4.1
5%	6.5	9.4	12.3
10%	12.9	15.7	19.9
100 Portfolios			
1%	1.7	3.6	4.5
5%	10.3	14.5	15.2
10%	21.7	26.9	29.0
(C) Shrinkage with correctly specified target			
25 Portfolios			
1%	1.5	2.0	3.8
5%	6.6	8.4	11.1
10%	12.8	13.6	17.7
100 Portfolios			
1%	8.0	6.0	4.5
5%	30.1	20.2	15.2
10%	45.0	33.6	27.2
(D) Shrinkage with misspecified target			
25 Portfolios			
1%	1.7	2.0	3.4
5%	6.1	8.1	10.4
10%	13.0	13.1	17.2
100 Portfolios			
1%	12.0	7.4	4.8
5%	40.4	23.6	16.8
10%	57.4	36.0	28.8

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. The factors and returns were simulated using a quadratic version of the Premium-Labor model. The rejection frequencies were calculated using different estimates of the weighting matrix.

Table 15
Power of the HJ-distance test from two estimation methods of G.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Sample second moment			
25 Portfolios			
1%	29.0	52.3	91.9
5%	51.6	74.8	98.0
10%	61.7	82.8	98.9
100 Portfolios			
1%	99.6	83.0	87.4
5%	100	94.6	96.4
10%	100	97.6	98.2
(B) Shrinkage estimation			
25 Portfolios			
1%	29.9	53.8	92.3
5%	52.4	75.6	98.0
10%	61.8	83.5	99.0
100 Portfolios			
1%	99.0	81.6	87.8
5%	99.8	93.8	96.4
10%	99.8	97.0	98.2

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. The true DGP is a four-factor model. The SDF we tested, as well as our shrinkage target, were based on a three-factor model.

Table 14 summarizes the results from this experiment. The first and second panels show the rejection frequencies of the HJ-distance test when the sample second moment matrix and exact second moment matrix are used to estimate G. The third panel shows the rejection frequencies with the shrinkage estimate of G.

The results in the first three panels of Table 14 are roughly comparable to the corresponding results from the linear Premium-Labor model reported in Tables 1–3. The HJ-distance test with the sample second moment matrix is severely oversized, with 100 portfolios. Using the exact second moment matrix results in a dramatic improvement in size. The shrinkage estimate also improves the size substantially. With 25 portfolios, its performance is very similar to that of the exact weighting matrix. Overall, the size distortion is smaller than in the linear Premium-Labor model, which most likely reflects the ability of the additional regressors to fit the data better.

The fourth panel of Table 14 reports the rejection frequencies when the linear Premium-Labor model is used as the shrinkage target model. Namely, the model we test is the true DGP, but the target factor model is misspecified. With 100 portfolios, the size distortion with the misspecified target model is larger than that with the correctly specified target model, but it is still a substantial improvement over the sample second moment matrix. With 25 portfolios, the correctly specified and misspecified target models give very close results.

5.6. Effects of shrinkage on power

In this subsection, we examine the effect of the shrinkage on the power of the HJ-distance test. Heuristically speaking, using a more precise estimate of G should give a less noisy test statistic and improve the power.

Table 15 reports the power of the HJ-distance test when the true DGP is a four-factor version of the Simple model and the null of a three-factor SDF is tested. Here, both the factor model being tested and the shrinkage target are misspecified.

The power of the two tests is very similar. Although the shrinkage method does not significantly improve power, the HJ-distance test with the shrinkage method is at least as powerful as the one with the sample covariance matrix. This suggests that the shrinkage estimate is at least as precise as the sample covariance matrix when the shrinkage target is misspecified.

6. Conclusion

The HJ-distance test rejects correct SDFs too often in the finite sample, which limits its practical use. We find that one reason for this phenomenon is a poorly estimated covariance matrix of the asset returns. We propose to use the shrinkage method to construct an improved estimate of this matrix.

The sample covariance matrix is often used to estimate the covariance matrix of asset returns. When the number of portfolios is large, however, this estimate suffers from a large estimation error. The shrinkage method uses another estimate that imposes some structure onto this high dimensional estimation problem, and combines it optimally with the sample covariance matrix. Our simulation results show that the shrinkage method significantly mitigates the overrejection problem of the HJ-distance test.

A few questions remain to be addressed in future research. First, the shrinkage method mitigates but does not completely solve the overrejection problem of the HJ-distance test, in particular when the portfolio size is as large as 100. A further improvement would be desirable. Second, it would be interesting to investigate how to choose the shrinkage target optimally and how to obtain a better estimate of the optimal shrinkage intensity. Third, the estimation of the covariance matrix plays an important role in many tests in empirical finance. It would be worthwhile to examine whether the method proposed in this paper can improve the finite properties of those tests.

Appendix A

A.1. Simple model

The first Simple model is generated by following the procedure of Ahn and Gadarowski (2004). Specifically, the data are generated by the following data generating process:

$$R_{ti} = \mu + X_{t1}\beta_{1i} + X_{t2}\beta_{2i} + X_{t3}\beta_{3i} + e_{ti},$$

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

A.2. Fama–French and Premium-Labor models

We follow the procedure of Ahn and Gadarowski (2004) to generate data sets calibrated to resemble the statistical properties of the Fama–French and Premium-Labor models. First, we collect 330 time-series observations of monthly returns of the Fama–French portfolios and the Fama–French factors between July 1963 and December 1990[6]. For the Premium-Labor model, we follow the steps in Jagannathan and Wang (1996) to obtain the portfolio returns and the factors.

[6] URL is http://mba.tuck.dartmouth.edu/pages/faulty/ken.french/data_library.htm.

Second, we apply the two-pass estimation following Shanken (1992). Specifically, we regress the portfolio returns on the corresponding factors by OLS, obtain the estimates of β_{ki}, and collect the residuals. We then compute the diagonal sample covariance matrix of the residuals. Subsequently, we run the following cross sectional regression:

$$E[R_{ti}] = \mu + (E(X_{t1}) + \eta_1)\beta_{1i} + (E(X_{t2}) + \eta_2)\beta_{2i} + (E(X_{t3}) + \eta_3)\beta_{3i}.$$

This gives the estimates of the risk-free rate, μ, and the factor-mean adjusted risk prices, η_k.

Finally, 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 the corresponding factors. The error terms, e_{ti}, 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:

$$R_{ti} = \mu + (X_{t1} + \eta_1)\beta_{1i} + (X_{t2} + \eta_2)\beta_{2i} + (X_{t3} + \eta_3)\beta_{3i} + e_{ti}$$

The risk-adjusted prices are incorporated in order to simulate the portfolio return close to the true data.

The nonlinear version in Section 5 is formulated as

$$R_{ti} = \mu + (X_{t1} + \eta_1)\beta_{1i} + (X_{t2} + \eta_2)\beta_{2i} + (X_{t3} + \eta_3)\beta_{3i} + \left(X_{t1}^2 + \eta_4\right)\beta_{4i} + \left(X_{t2}^2 + \eta_5\right)\beta_{5i} + \left(X_{t3}^2 + \eta_6\right)\beta_{6i} + e_{ti}.$$

The parameters are estimated in the same manner as in the linear model.

Table A.1
Rejection frequencies of the HJ-distance test with setting G = *I*.

Number of observations	T = 160	T = 330	T = 700
(A) Simple model			
25 Portfolios			
1%	0.5	0.7	1.2
5%	4.4	4.6	4.5
10%	9.0	9.2	9.8
100 Portfolios			
1%	0.0	0.6	1.0
5%	1.6	3.4	3.8
10%	5.4	8.8	9.2
(B) Fama–French Model			
25 Portfolios			
1%	0.8	0.6	1.1
5%	3.6	4.4	4.4
10%	8.4	10.2	9.0
100 Portfolios			
1%	1.0	0.8	1.2
5%	5.6	5.2	4.2
10%	11.6	11.0	8.8
(C) Premium-Labor Model			
25 Portfolios			
1%	2.4	3.8	4.7
5%	8.2	13.3	12.8
10%	15.7	21.2	20.5
100 Portfolios			
1%	2.4	6.6	6.8
5%	13.6	19.4	20.4
10%	25.0	30.6	30.0

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance. We set G = *I*.

Table A.2
Bias and MSE of the estimates of Ω.

Number of observations		*T* = 160	*T* = 330	*T* = 700
(A) Simple Model				
25 Portfolios				
Sample covariance	Bias	4.9631	2.2638	0.9872

Table A.2 (*continued*)

Number of observations		$T=160$	$T=330$	$T=700$
	MSE	149.14	27.436	4.0172
Shrinkage on G	Bias	4.4209	2.1646	1.0631
	MSE	73.646	17.502	3.5766
100 Portfolios				
Sample covariance	Bias	67.755	29.960	13.545
	$\text{MSE}\times 10^{-3}$	46.667	4.5576	0.9139
Shrinkage on G	Bias	86.477	36.594	12.882
	$\text{MSE}\times 10^{-3}$	36.915	5.2441	0.4710
(B) Fama–French Model				
25 Portfolios				
Sample covariance	Bias	1.9197	0.7675	0.2654
	MSE	22.016	2.4207	0.3133
Shrinkage on G	Bias	1.5981	0.6342	0.2122
	MSE	11.957	1.5074	0.2433
100 Portfolios				
Sample covariance	Bias	129.09	68.458	28.420
	$\text{MSE}\times 10^{-3}$	28.350	9.4855	2.4478
Shrinkage on G	Bias	113.76	61.986	24.529
	$\text{MSE}\times 10^{-3}$	25.028	8.1222	1.4210
(C) Premium-Labor Model				
25 Portfolios				
Sample covariance	Bias	640.86	463.86	259.74
	$\text{MSE}\times 10^{-5}$	4.4285	2.6716	0.9872
Shrinkage on G	Bias	632.33	458.26	254.24
	$\text{MSE}\times 10^{-5}$	4.4137	2.5621	0.9171
100 Portfolios				
Sample covariance	$\text{Bias}\times 10^{-3}$	3.9618	3.1428	1.6773
	$\text{MSE}\times 10^{-7}$	1.6823	1.0647	0.3456
Shrinkage on G	$\text{Bias}\times 10^{-3}$	4.2411	3.1674	1.7347
	$\text{MSE}\times 10^{-7}$	1.8524	1.0675	0.3428

This table shows the bias and MSE (measured by the Frobenius norm) of the estimate of under two estimation methods of G, a shrinkage method and the sample covariance.

Table A.3
Rejection frequencies of the HJ-distance test with shrinkage on Ω.

Number of observations	$T=160$	$T=330$	$T=700$
(A) Simple Model			
25 Portfolios			
1%	4.4	2.1	1.7
5%	12.5	8.9	5.9
10%	21.1	15.1	12.0
100 Portfolios			
1%	99.2	49.6	9.2
5%	99.8	68.6	27.0
10%	99.9	81.0	37.4
(B) Fama–French Model			
25 Portfolios			
1%	3.5	1.6	1.9
5%	12.3	7.9	8.6
10%	19.9	14.2	14.5
100 Portfolios			
1%	99.2	48.8	13.4
5%	100.0	72.8	32.6
10%	100.0	82.4	42.4
(C) Premium-Labor Model			
25 Portfolios			
1%	12.2	13.0	7.4
5%	26.4	26.4	18.3
10%	36.4	36.6	26.4
100 Portfolios			
1%	99.8	77.8	30.6
5%	99.8	89.8	54.4
10%	99.8	93.6	66.4

This table shows the rejection rates over 1000 trials using the p-value of the HJ-distance.

References

Adler, M., Dumas, B., 1983. International portfolio choice and corporate finance: a synthesis. Journal of Finance 38, 925–984.
Ahn, S.C., Gadarowski, C., 2004. Small sample properties of the GMM specification test based on the Hansen–Jagannathan distance. Journal of Empirical Finance 11, 109–132.
Bansal, R., Hsieh;, D.A., Viswanathan, S., 1993. A new approach to international arbitrage pricing. Journal of Finance 48, 1719–1747.
Bansal, R., Zhou, H., 2002. Term structure of interest rate with regime shifts. Journal of Finance 57, 1997–2043.
Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45, 444–454.
Breeden, D.T., 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. Journal of Financial Economics 7, 265–296.
Burnside, C., Eichenbaum, M., 1996. Small-sample properties of GMM-based Wald tests. Journal of Business & Economic Statistics 7, 265–296.
Campbell, J.Y., Cochrane, J.H., 2000. Explaining the poor performance of consumption-based asset pricing models. Journal of Finance 55, 2863–2878.
Chapman, D.A., 1997. Approximating the asset pricing kernel. Journal of Finance 52, 1383–1410.
Chen, N.F., Roll;, R., Ross, S.A., 1986. Economic forces and the stock market. Journal of Business 59, 383–403.
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., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–466.
Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.
Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.
Hansen, L.P., Jagannathan, R., 1997. Assessing specific errors in stochastic discount factor models. Journal of Finance 52, 557–590.
Hodrick, R.J., Zhang, X.Y., 2001. Evaluating the specification errors of asset pricing models. Journal of Financial Economics 62, 327–376.
Huang, J.Z., Wu, L., 2004. Specification analysis of option pricing models based on time-changed levy process. Journal of Finance 59, 1405–1439.
Jacobs, K., Wang, K.Q., 2004. Idiosyncratic consumption risk and the cross section of asset returns. Journal of Finance 59, 2211–2252.
Jagannthan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.
Jagannthan, R., Wang, Z., 1998. An asymptotic theory for estimating beta-pricing models using cross-sectional regression. Journal of Finance 53, 1285–1309.
Jagannthan, R., Wang, Z., 2002. Empirical evaluation of asset-pricing models: a comparison of the SDF and beta methods. Journal of Finance 57, 2337–2367.
Jobson, J.D., Korkie, B., 1980. Estimation for Markowitz efficient portfolios. Journal of the American Statistical Association 75, 544–554.
Kan, R., Zhang, C., 1999. GMM tests of stochastic discount factor models with useless factors. Journal of Financial Economics 54, 103–127.
Kan, R., Zhou, G., 2004. Hansen–Jagannathan distance: geometry and exact distribution. Working paper. University of Toronto.
Ledoit, O., Wolf, M., 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. Journal of Empirical Finance 10, 603–621.
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.
Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.
Parker, J., Julliard, C., 2005. Consumption risk and the cross section of expected returns. Journal of Political Economy 113, 185–222.
Ross, S., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.
Shanken, J., 1992. On the estimation of beta-pricing models. Review of Finance Studies 5, 1–34.
Shapiro, A., 2002. The investor recognition hypothesis in a dynamic equilibrium: theory and evidence. The Review of Financial Studies 15, 97–141.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.
Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Berkeley, CA (Ed.), Proceedings of the Third Berkeley Symposium on Mathematical and Statistical Probability. University of California, Berkeley.
Vassalou, M., 2003. News related to future GDP growth as a risk factor in equity returns. Journal of Financial Economics 68, 47–73.
Vassalou, M., Xing, Y.H., 2004. Default risk in equity returns. Journal of Finance 59, 831–868.