Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Empirical Finance 11 (2004) 109–132
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

Small sample properties of the GMM specification test based on the Hansen–Jagannathan distance

Seung C. Ahn[a], Christopher Gadarowski[b,*]

[a] *Arizona State University, USA*
[b] *Rowan University, 256D Bunce Hall, Glassboro, NJ 08028, USA*

Accepted 18 September 2002

Abstract

Following Hansen and Jagannathan (J. Finance 52 (1997) 557), Jagannathan and Wang (J. Finance 51 (1996) 3) propose a distance measure that estimates the maximum pricing error generated by a linear asset model. Jagannathan and Wang propose a test of this HJ-distance using an empirical *p*-value as an alternative generalized method of moments (GMM) measure to Hansen's (Econometrica 50 (1982) 1029) GMM specification test. Using Monte Carlo analysis, we examine the finite sample properties of these specification tests. While the Hansen test mildly overrejects correct models in commonly used sample size, the empirical *p*-value of the HJ-distance rejects correct models too severely in such samples to provide a valid test of such models.

JEL classification: C15; C52; G12
Keywords: Generalized method of moments; Empirical *p*-value; Pricing error; Hansen-Jagannathan distance; Finite sample properties

1. Introduction

Empirical relevance of an asset pricing model can be examined by testing for statistical significance of the pricing errors made by the model for individual assets. The generalized method of moments (GMM) developed by Hansen (1982) provides a convenient χ^2 statistic for detecting such errors, which we hereafter call the Hansen statistic. Asset pricing models imply stochastic discount factors (SDFs) that can be used to determine

* Corresponding author.
E-mail addresses: gadarowski@rowan.edu (C. Gadarowski), miniahn@asu.edu (S.C. Ahn).

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2002.09.001

current market prices of portfolios by discounting future payoffs state by state (Duffie, 1996). The stochastic discount factor (SDF) implied by a model is a function of data and some unknown parameters. Using GMM, a researcher can estimate the SDF parameters by minimizing a quadratic function of pricing errors weighted by an arbitrary positive definite matrix. With the variance matrix of pricing errors used as a weighting matrix, the minimized value of the quadratic function becomes the Hansen statistic, which is asymptotically χ^2-distributed under the null hypothesis that the model estimated is correctly specified. The inverse of the variance matrix of pricing errors is an optimal GMM weighting matrix in the sense that use of it minimizes the asymptotic variances of all GMM parameter estimates. Although the variance matrix is not observed, a consistent estimate of it is often used in practice.

Use of the Hansen statistic however has been criticized for at least three reasons. First, the Hansen statistic favors models with highly variable pricing errors (Hansen and Jagannathan, 1997; Jagannathan and Wang, 1996). This is so because the statistic is inversely related with the variances of pricing errors. Second, as Jagannathan and Wang (1996) point out, the Hansen statistic cannot be used to compare the relative performances of different models because the statistic uses different weighting matrices for different models. Although the Hansen statistic can be used to test whether the pricing errors generated by a model are zero or not, it cannot be used to determine which of competing models generates smaller pricing errors. Third, as Ferson and Foerster (1994) and Burnside and Eichenbaum (1996) have shown, sizes of the Hansen statistics are often too large with finite samples.

In response to these problems, Hansen and Jagannathan (1997) propose a measure which equals the least squares distance between a stochastic discount factor (SDF) implied by a model and the family of SDFs that correctly price assets. Jagannathan and Wang (1996) refer to this measure as the HJ-distance. Hansen and Jagannathan show that the HJ-distance equals the maximum pricing error generated by a model for portfolios whose second moments of returns are equal to one. They also show that the HJ-distance can be conveniently estimated by GMM using the inverse of the sample second moment matrix of returns as the weighting matrix (instead of an estimated variance matrix of pricing errors). Jagannathan and Wang (1996) develop an alternative GMM specification test based on this distance measure.

The HJ-distance possesses several desirable properties. First, since the same weighting matrix is used for different asset pricing models, the relative performance of competing models can be evaluated by comparing the relative sizes of the HJ-distances for each model. Second, unlike the Hansen statistic, the HJ-distance does not reward models with noisy pricing errors. This is so because the HJ-distance does not depend on the variances of pricing errors. Finally, similarly to the Hansen statistic, the HJ-distance estimated by GMM can be used to test whether a given asset pricing model is correctly specified. In general, though, because the HJ-distance does not use an optimal weighing matrix in GMM estimation, it is not χ^2-distributed under the null hypothesis of correct model specification. Jagannathan and Wang, however, show that for linear factor models, the HJ-distance is asymptotically equivalent to a random variate from a weighted χ^2 distribution. Based on this finding, they develop a simulation method to empirically estimate the p-value for the HJ-distance, hereonafter referred to the “empirical p-value”.

Their test method can be viewed as an alternative GMM specification test. Both Jagannathan and Wang (1996) and Jagannathan et al. (1998) evaluate several linear asset pricing models using this method. Recently, Hodrick and Zhang (2001) apply the HJ-distance to compare a wide range of linear asset pricing models, while Dittmar (2002) applies it to nonlinear models. Kan and Zhou (2002) conduct a more theoretical analysis where they provide an economic interpretation of the HJ-distance in terms of the mean-variance frontier.

Despite the theoretical merits of the HJ-distance, finite-sample properties of the GMM test based on the HJ-distance and its empirical *p*-value are unknown. The motivation of this paper is to fill this gap in the literature. Our main concern in this paper is the empirical performance of the HJ-distance *p*-value. Recent studies of GMM have shown that the optimal GMM weighting matrix is poorly estimated in small samples (e.g. Andersen and Sørensen, 1996; Burnside and Eichenbaum, 1996). Use of this estimated weighting matrix leads to poor finite-sample performance of the Hansen statistic. An advantage of using the HJ-distance is that it does not depend on the optimal weighting matrix. Zhou (1994) also provides some evidence that GMM with nonoptimal weighting matrices could have better finite-sample properties. However, like the Hansen statistic, the exact distribution of the HJ-distance depends on the variance matrix of pricing errors, and its asymptotic *p*-value computation in practice requires use of an estimate of this matrix. Thus, there is no guarantee that the specification test based on the HJ-distance and its empirical *p*-value has better small-sample properties than the Hansen statistic.

In this paper, we conduct some limited Monte Carlo experiments to compare the finite-sample performances of the HJ-distance and the Hansen statistic. The results from our Monte Carlo experiments can be summarized as follows. First, the empirical *p*-value of the HJ-distance rejects correct models too frequently. The rejection rate by the *p*-value becomes reasonable only if sample size is unrealistically large. This over-rejection problem persists even if the optimal weighting matrix is replaced by the true parameter values instead of their estimates. For most of the cases we investigate, we find that the Hansen statistic, which also tends to over-reject correct models, outperforms the HJ-distance *p*-value. Second, we apply the degrees-of-freedom adjustment proposed by Ferson and Foerster (1994) to improve the finite-sample performance of the optimal GMM to the HJ-distance *p*-value. Our results indicate that this adjustment mitigates some of the bias in the HJ-distance *p*-value, but it still remains severe for most cases when a larger number of assets are used. Third, we examine how finite sample estimation of the GMM weighting matrix contributes to the over-rejection by examining the impact of using estimated versus true or asympotic parameters in the calculations. Here, the optimal GMM weighting matrix is the inverse of the pricing error variance matrix. There are two possible reasons for poor estimation of this variance matrix using a finite sample: (a) the pricing errors could be estimated poorly because of errors in estimating the stochastic discount factor (SDF) parameters, and (b) this variance matrix could be estimated poorly even if the true pricing errors are based on the true SDF parameters because the number of assets is large relative to the number of time-series observations. Our analysis shows that the over-rejection of this test is due to second of these issues. These results are consistent with previous studies that show that GMM estimation of such variance matrices is sensitive to sample size. Our analysis also isolates the problem as being due to the second stage of analysis: finite estimation of the variance matrix even given true pricing errors.

This paper is organized as follows. Section 2 briefly discusses the estimation and model specification test methods based on the HJ-distance. Section 3 discusses the simulation methodology we adopt in this paper. Section 4 reports our Monte Carlo results. Some concluding remarks follow in Section 5.

2. The Hansen–Jagannathan distance

Hansen and Jagannathan (1997) derive a measure, called the HJ-distance, of how badly a general asset pricing model is misspecified using an estimate of its stochastic discount factor.[1] In this section, we raise some econometric issues related with the use of this measure as a model specification test. For simplicity, we only consider linear discount factor models in what follows. We do so, because, to the extent that many nonlinear models can be well approximated by linear ones, the conclusion of this paper should also shed some light on the use of the HJ-distance for nonlinear models.

To begin with, consider a portfolio consisting of N primitive assets. Let R_t be the tth period vector of gross returns for these assets. Asset pricing models typically imply the Euler condition $E(R_t m_t(\delta)) = 1_t$, where $m_t(\delta)$ denotes a stochastic discount factor (SDF), δ denotes a unknown $K\times 1$ parameter vector to be estimated, and 1_N is an $N\times 1$ vector of ones. Different asset pricing models imply different SDFs. For example, linear factor pricing models, which are our particular interest in this paper, imply SDFs of the form, $m_t(\delta) = Y_t'\,\delta$, where Y_t is the $K\times 1$ vector of common factors and one. Other SDFs can be found in Hansen and Jagannathan (1997).

Jagannathan and Wang (1996) consider the HJ-distance for linear factor pricing models. Using sample data on returns and factors, this HJ-distance can be estimated as follows:

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

with $D_T = T^{-1}\sum_{t=1}^{T} R_t Y_t' = T^{-1}R'Y$, $w_T(\delta) = T^{-1}\sum_{t=1}^{T} w_t(\delta) = D_T\delta - 1_N$, $G_T = T^{-1}\sum_{t=1}^{T} R_t R_t' = T^{-1}R'R$, where $R = [R_1,\ldots, R_T]'$, $Y = [Y_1,\ldots, Y_T]'$, and $w_t(\delta) = R_t Y_t'\delta - 1_N$ is the $N\times 1$ vector of pricing errors at time t. Jagannathan and Wang show the minimizer of this sample HJ-distance equals

$$\delta_{\mathrm{HJ},T} \equiv (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} 1_N, \tag{2}$$

which is essentially a GMM estimator based on the moment conditions $E[w_t(\delta)] = 0_N$ with the weighting matrix G_T^{-1}. With this estimate, the HJ-distance can be consistently estimated by $\mathrm{HJ}_T(\delta_{\mathrm{HJ},T})$.

The asymptotic results from Hansen (1982) imply that $\delta_{\mathrm{HJ},T}$ is not an efficient (minimum variance) estimator among those utilizing the moment conditions $E[w_t(\delta)] = 0$. This is so

[1] See Appendix A.1 for further details.

because the matrix G_t^{-1} is not an optimal weighting matrix in the sense that use of it does not produce the minimum-variance estimator among other GMM estimators based on the moment conditions $E[w_t(\delta)]=0$. Under the assumption that data are serially uncorrelated, Hansen (1982) implies that the optimal weighting matrix is S_T^{-1}, where $S_T^{-1} = T^{-1} \sum_{t=1}^{T} w_t(\delta_{\mathrm{HJ},t}) w_t(\delta_{\mathrm{HJ},T})'$ is a consistent estimate of the variance matrix of model pricing errors, $\mathrm{Var}[\sqrt{T} w_T(\delta)]$.[2] With this weighting matrix, the optimal GMM estimator of δ is given by

$$\delta_{\mathrm{OGMM},T} = (D_T' S_T^{-1} D_T)^{-1} D_T' S_T^{-1} 1_N. \tag{3}$$

An advantage of using this optimal GMM estimator $\delta_{\mathrm{OGMM},T}$ is that it provides a convenient model specification test statistic, which we call the Hansen statistic,[3] is:

$$J_T = T w_T(\delta_{\mathrm{OGMM},T})' \, S_T^{-1} w_T(\delta_{\mathrm{OGMM},T}). \tag{4}$$

Under the null hypothesis that the model used to estimate δ is correctly specified, this statistic is asymptotically χ^2-distributed with degrees of freedom equal to $(N-K)$.

Despite simplicity and convenience, the Hansen test has several undesirable properties compared to the HJ-distance. First, observe that the Hansen statistic is inversely related with the matrix S_T, which is an estimate of the variance matrix of pricing errors. This implies that the Hansen test rewards models creating highly variable pricing errors across different assets (Hansen and Jagannathan, 1997; Jagannathan and Wang, 1996). This problem does not apply to the HJ-distance because it uses G_T^{-1} instead of S_T^{-1} as a weighting matrix. Second, although the Hansen statistic can be used to determine whether a particular model is correctly specified or not, it is an inappropriate tool for comparing relative performances of competing models. This is so because the matrix S_T varies depending on the model estimated. The Hansen statistics computed for different models are not comparable. In contrast, the HJ-distance uses the same weighting matrix G_T^{-1} for different models. Thus, it can be used to compare the relative sizes of pricing errors generated by different models.

Recent studies of GMM raise another concern about the Hansen test. Ferson and Foerster (1994) and Burnside and Eichenbaum (1996) show that the size of the Hansen test is often too large in small samples, and thus, rejects correct models more frequently than the asymptotic theory of GMM predicts. Zhou (1994) and Altonji and Segal (1996) also find that nonoptimal GMM procedures often outperform optimal GMM procedures. An important reason for this problem is that the optimal GMM weighting matrix required for the Hansen test is generally poorly estimated in small samples, especially when too many moment conditions are used (Andersen and Sørensen, 1996). In our case, this is a concern when the number of assets, N, is large. Whether the Hansen test for linear factor pricing models, defined in Eq. (4), has desirable finite-sample properties crucially depends on how

[2] If data are serially correlated, S_T is no longer a consistent estimator of S. An autocorrelation-robust estimate can be obtained following Newey and West (1987) or Andrews (1991). In this paper, we use S_T because only serially uncorrelated data are used for our experiments. See Appendix A for further details.

[3] This statistic is also called the J-statistic in the literature.

well the estimated variance matrix of pricing errors, S_T, can approximate the true variance matrix of pricing errors.

These findings naturally lead to a conjecture that the (nonoptimal) GMM procedure based on the HJ-distance may have better finite-sample properties than the optimal GMM procedure. To take advantage of the favorable properties of the HJ-distance relative to the Hansen statistic, Jagannathan and Wang develop an alternative GMM model specification test based on the distance measure. However, as shown below, this alternative test is not entirely free of the problems that the Hansen test suffers from.

To understand the Jagannathan-Wang method, and the potential problems with the finite sample properties of their test, define

$$\mathrm{SHJ}_T = T \times \mathrm{HJ}_T(\delta_{\mathrm{HJ},T})^2. \tag{5}$$

Although J_T and SHJ_T are similar, the latter is not χ^2-distributed because it uses the nonoptimal weighting matrix G_T^{-1}. Jagannathan and Wang show that SHJ_T is asymptotically equivalent to a weighted average of $(N-K)$ independent χ^2-distributed random variables. The weights for this average are the positive eigenvalues of a matrix defined as

$$\Phi = S^{1/2}G^{-1/2}[I_N - (G^{-1/2})'D(D'G^{-1}D)^{-1}D'G^{-1/2}](G^{-1/2})'(S^{1/2})', \tag{6}$$

where I_N is the $N\times N$ identity matrix, $D=E(R_tY_t')$, $G=E(R_tR_t')$, and $S^{1/2}$ and $G^{-1/2}$ are the upper-triangle matrices from the Cholesky decompositions of S and G^{-1}, respectively. Furthermore, the matrix Φ can be consistently estimated by replacing D, S and G with D_T, S_T and G_T.

Jagannathan and Wang show how the eigenvalues of the matrix Φ can be used in a simulation method to estimate an empirical p-value of the HJ-distance test for the hypothesis that $E[w_t(\delta)]=0_N$.[4] Herein lies the potential finite sample problems of this method: While the HJ-distance does not depend on the variance matrix of model pricing errors (S), its asymptotic distribution does. Thus, the asymptotic p-value computation (empirical p-value) in practice requires the estimation of this matrix (S_T). More specifically, the empirical p-value depends on the eigenvalues from an estimate of Φ defined in Eq. (6). But estimation of the matrix Φ in turn requires use of the matrix S_T, which is suspected to contribute to poor small-sample performances of the Hansen test. Accordingly, there is no guarantee that the specification test based on the HJ-distance and its empirical p-value outperforms the Hansen test in small samples.

3. Monte Carlo setup

As discussed in the previous section, inspection of the empirical p-value of the HJ-distance described by Jagannathan and Wang suggests that this test may suffer from size distortions at least as much as the Hansen test may do. To test this conjecture, we perform several Monte Carlo analyses of the size of the several specification tests, including the

[4] See Appendix A.2 for details.

HJ-distance, using different types of simulated data. We also conduct some additional simulation exercises to identify the sources of finite-sample size distortions in the specification tests we consider.

The first type of data in our analysis is the data simulated under quite restrictive assumptions such as factors are independently and identically distributed over time. However, to be nominally consistent with the factor models examined by Jagannathan and Wang, we use a three-factor model in this simple simulation, with factor loadings randomly distributed among the assets. The data is generated so that the scale of expected returns and variability of the factors are roughly equivalent to those of market-wide returns, and the estimation of our hypothetical three-factor model results in cross-sectional explanatory power roughly equal that found in the three-factor models examined by Jagannathan and Wang.[5] This data is hereafter referred to as the simple data. Appendix A.3 provides a detailed discussion about how this simple data is generated.

The second type of data uses factors and returns calibrated to statistically resemble the factors and asset returns used by Jagannathan and Wang, hereafter the calibrated data. With the calibrated data, returns are simulated using estimated parameters from either the Fama and French (1993) model or the Premium-Labor model introduced by Jagannathan and Wang, and factors simulated to mimic the actual factors in the models. The Fama-French model includes the following factors: VWRET, the return on the Center for Research on Security Prices (CSRP) value-weighted market index; SMB, the return on a portfolio of small firms less the return on a portfolio of large firms; and HML, the return on a portfolio of high book-to-market firms less the return on a portfolio of low book-to-market firms. The Premium-Labor model includes VWRET, PREM, which is the market yield-to-maturity on BBB corporate bonds less the market yield-to-market on the AAA corporate bonds, and LABOR, which is the two-month average change in personal income less dividends per the National Income and Product Accounts.

With both the simple and calibrated data, we limit our analysis to non-autocorrelated data. We do so for two reasons. First, as later results show, we find the problems with the HJ-distance based specification test even under this simplifying assumption. It is quite unlikely that the HJ-distance test would perform better in autocorrelated data. Second, by restricting our analysis to non-autocorrelated data, we avoid potential complications associated with finite-sample estimation of autocorrelation-consistent variance matrix of pricing errors. The main reason why the HJ-distance test over-rejects too much is due to the problems in estimating the variance matrix of average pricing errors (the optimal GMM weighting matrix). With a general distributional structure, it is even more difficult to estimate this matrix. Thus, the over-rejection problem with the HJ-distance test is unlikely to be less severe in autocorrelated data. While our simulations are limited to no-autocorrelation cases, we consider cross-sectional correlations among both factors and asset returns when calibrating the simulated data to that of the data provided by Jagannathan and Wang. Further details of these data sets are provided in Appendix A.4.

[5] Specifically, the cross-sectional R-squared of the mean returns on factor loadings is roughly equal to 0.50 using 100 assets. This is roughly equal to the R-squared reported by Jagannathan and Wang for Fama-French and Premium-Labor models for 100 assets.

With both the simple and calibrated data, we examine the size of several GMM specification tests based on the SDF of a linear factor model for a number of combinations of the number of assets (N) and time-series observations (T). These specification tests include the HJ distance per Jagannathan and Wang (HJ), the Hansen statistic using two-step GMM (J), and the Hansen statistic using iterative GMM (JI), using the same simulation trial to estimate all test parameters.[6] With the simple data we examine cases with 25 and 100 assets, with the time-series observations ranging from 160 to 3000 periods. These ranges allow us to see how the relative number of assets versus time-series observations affects the size of the empirical p-value over ranges commonly used in studies (Jagannathan and Wang: $N=100$, $T=330$; and Jagannathan et al., 1998: $N=25$, $T=160$) versus samples with more time-series observation. With the calibrated data, we examine cases identical to that used by Jagannathan and Wang: 25 and 100 assets with 330 time-series observations.[7] We generate 1000 different simple data sets (1000 trials) and 1000 different calibrated data sets (1000 trials). Then, we count the proportion of times that the empirical p-value is less than commonly used asymptotic size levels (1%, 5%, and 10%) across 1000 trials. For each trial, following Jagannathan and Wang (1996), we use 5000 random draws (M) from a weighted $\chi^2(1)$ distribution to estimate the empirical p-value of the HJ-distance. The detailed estimation procedure for the empirical p-value is given in Appendix A.2.

In addition to using the asymptotic formulations of the Hansen and HJ-distance tests, we also examine the size of these specification tests using a degrees-of-freedom adjustment for finite sample size suggested by Ferson and Foerster (1994). While this adjustment is ad hoc, Ferson and Foerster find that it improves the empirical size of the Hansen statistic in GMM estimation for finite samples. Our intent here is to examine whether this adjustment has a similar affect for the HJ distance p-value and Hansen statistic p-values in our case. To the extent that it mitigates finite sample problems with the Hansen statistic, it may do likewise with the HJ-distance.

The foundation of Ferson and Foerster's (1994) experiments is a latent-variables asset pricing model developed by Hansen and Hodrick (1983) and Gibbons and Ferson (1985), among others, in which the risk free rate and risk factor prices are linear functions of L instrumental variables. For such models, Ferson and Foerster find that the Hansen statistic rejects correct models too often in small samples, and that the asymptotic standard errors for optimal GMM estimates are understated. In response to this problem, they propose a degrees-of-freedom adjustment which accounts for the number of time series observations (T) corresponding to each of assets and the L instrumental variables, and for the number of model parameters (K) plus the number of elements in the weighting matrix used for the test ($[NL(NL+1)]/2$). Specifically, their

[6] For example, in each trial we recorded the HJ-distance results, the two-stage GMM result of the Hansen statistic (J), then continued the iterations to compute the iterative GMM Hansen statistic (JI).

[7] Strictly, Jagannathan and Wang provide results for only 100 assets but report that similar results are found using 25 assets for the same number of observations.

method is equivalent to inflating the inverse of the optimal GMM weighting matrix (in our case, S_T as defined in Eq. (3)) by multiplying it by an adjustment factor:

$$F = \frac{(N+L)T}{(N+L)T-Q}, \tag{7}$$

where $Q=K+[NL(NL+1)]/2$. Note that $F>1$. Their simulation results indicate that this adjustment leads to better performances of the optimal GMM procedures, provided that the number of instruments (L) and assets (N) is not too large, and that the denominator of F is not too close to zero.

We apply this degrees-of-freedom adjustment to our simulations. Since the linear model we simulate has no instrumental variables, we replace $(N+L)T$ in F by NT. Also, because there are only N moment conditions, the number of terms in the weighting matrix that are estimated using the time-series of returns is equal to $N(N+1)/2$. Further, the number of parameters we estimate (δ_0, δ_1, δ_2 and δ_3) equals 4 (i.e., $K=4$). Accordingly, we calculate Q in F by $4+N(N+1)/2$. We use this modified adjustment to estimate the variance matrix of pricing errors (S_T defined above Eq. (3)).

Lastly, we perform a diagnostic analysis of the size of the HJ-distance and Hansen statistics by separately substituting the exact pricing-error variance matrix (S) for the estimated variance matrix (S_T) or the variance matrix estimated with exact SDF parameters for S_T. The intent of this analysis is to determine whether finite sample bias in SDF parameter estimates is responsible for excess test sizes. We use the simple data for this analysis because we have complete knowledge of both the exact values of the SDF parameters. Details on these derivations are provided in the Appendix A.5.

4. Simulation results

4.1. Analysis of test size using simple data

Jagannathan and Wang (1996) conjecture that the GMM specification test based on the HJ-distance may have better finite-sample properties than the Hansen statistic because estimating the HJ-distance does not require estimating the variance matrix of pricing errors (S_T defined in Eq. (3)). However, as we discuss in Section 2, the exact distribution of the HJ-distance depends on the variance matrix, and computation of the empirical p-value of the HJ-distance requires estimating the variance matrix. Thus, the empirical p-value is likely to be biased in finite samples for the same reason the Hansen statistic is biased. Results of analysis using simple data are consistent with this conjecture.

Table 1 shows the rejection rates for the Hansen statistic and HJ distance using simple data using various combinations of time-series observations and number of assets. Similar to Ferson and Foerster (1994), we estimate the Hansen statistic using both two-step and iterative GMM.[8] The Hansen statistics, based on either two-step or iterative GMM, are

[8] Iterative GMM does not necessarily produce lower rejection rates than two-step GMM does The iterative and two-step GMM estimators use different weighting matrices. Thus, the Hansen statistic computed by iterative GMM is not necessarily smaller than that computed by two-step GMM.

Table 1
Model rejection rates using asymptotic specification tests of the Hansen statistic and HJ-distance using simple simulated data

Time-series observations	25 Assets			100 Assets		
	J (%)	JI (%)	HJ (%)	J (%)	JI (%)	HJ (%)
	Rejection at the 1% level					
160	2.0	1.7	5.5	3.6	1.5	99.3
330	2.0	2.0	2.3	5.9	5.3	51.0
700	1.5	1.6	1.9	3.3	3.5	11.9
1500	1.5	1.5	1.9	2.5	2.7	4.2
3000	0.8	0.8	0.8	1.8	1.6	2.4
	Rejection at the 5% level					
160	8.5	8.1	15.1	31.9	20.6	100.0
330	6.9	7.1	8.8	23.1	22.7	73.3
700	5.4	5.2	6.2	14.0	13.5	28.6
1500	5.4	5.5	6.1	9.3	9.6	13.5
3000	4.7	4.6	4.6	7.0	7.0	8.8
	Rejection at the 10% level					
160	15.9	14.3	24.7	55.4	42.1	100.0
330	13.2	12.6	15.3	37.7	36.2	81.6
700	10.4	10.4	12.1	23.9	23.5	43.0
1500	10.5	10.4	11.1	17.6	17.7	23.6
3000	9.7	9.6	10.1	13.2	13.8	16.5

This table shows the rejection rates over 1000 trials using the p-values for the Hansen statistic using two-step GMM (J), Hansen statistic using iterative GMM (JI), and Hansen-Jagannathan distance (HJ). Factors and returns, which are simulated returns for this analysis, are generated using the following model:

$$R_{it} = \alpha + \beta_i' F_t + e_{it}.$$

R_{it} the return for asset i for period t; α the risk free return for period t; β_i is vector of factor betas for asset i randomly distributed between 0 and 2 across assets; F_t is vector of three homogeneous and independently simulated risk factors for period t; and e_{it} is the idiosyncratic error for asset i for period t, with the ratio of the variance of each risk factor to the variance of the error term $\mathrm{Var}(F_{kt})/\mathrm{Var}(e_{it})$ equal to 1.0.

generally over-sized, especially when the number of time-series observations is small or the number of assets is large. For example, with 25 assets and 160 observations, the Hansen statistic is rejected at the 1% level 2.0% of the time using two-step GMM (J) and 1.7% of the time using iterative GMM (JI), while the rejection rates are close to the asymptotic levels with 3000 observations (both 0.8%). Similar results for the Hansen statistic methods are seen with rejection rates at the 5% and 10% levels. Noticeably, the HJ-distance p-values have higher rejection rates than those of the Hansen statistics.[9] These results are consistent with the results reported by Jagannathan et al. (1998). Using 25

[9] In unreported experiments, we used 5000 trials for some selected cases with 25 assets The model rejection rates were similar to those reported in this paper but approached asymptotic levels in a more monotonic way as the number of time-series observations increases.

Table 2
Model rejection rates of degrees-of-freedom adjusted specification rests of the Hansen statistic and HJ-distance using simple simulated data

Time-series observations	25 Assets			100 Assets		
	JF (%)	JIF (%)	HJF (%)	JF (%)	JIF (%)	HJF (%)
	Rejection at the 1% level					
160	0.3	0.4	3.3	0.0	0.0	81.6
330	1.2	1.3	1.7	0.2	0.0	15.2
700	1.2	1.2	1.7	0.6	0.7	4.8
1500	1.5	1.4	1.6	0.9	1.2	2.4
3000	0.8	0.8	0.8	1.0	0.9	1.6
	Rejection at the 5% level					
160	4.7	4.3	9.2	0.0	0.0	92.4
330	5.7	5.7	6.9	1.5	1.3	34.2
700	4.8	4.8	5.1	4.7	4.2	14.7
1500	5.3	5.4	5.7	5.7	5.6	9.2
3000	4.6	4.5	4.4	5.2	5.2	7.1
	Rejection at the 10% level					
160	8.9	8.3	17.0	0.0	0.0	95.1
330	10.6	10.1	12.8	3.1	2.6	48.9
700	9.0	8.7	11.1	9.6	9.3	23.9
1500	9.6	9.7	10.0	11.5	11.4	16.6
3000	9.5	9.5	9.9	11.3	11.4	13.1

This table shows the rejection rates over 1000 trials using the p-values for these model specification tests using an degrees-of-freedom adjustment for finite sampling similar to Ferson and Foerster (1994): Hansen statistic using two-step GMM (JF), Hansen statistic using iterative GMM (JIF), and the Hansen-Jagannathan distance (HJF). Factors and returns are simulated identically to those used in Table 1 using the following model:

$$R_{it} = \alpha + \beta_i' F_t + e_{it}.$$

R_{it} the return for asset i for period t; α the risk free return for period t; β_i is vector of factor betas for asset i randomly distributed between 0 and 2 across assets; F_t is vector of three homogeneous and independently simulated risk factors for period t; and e_{it} is the idiosyncratic error for asset i for period t, with the ratio of the variance of each risk factor to the variance of the error term Var(F_{kt})/Var(e_{it}) equal to 1.0.

assets with 148 monthly observations for Japanese stock market data, they report consistently lower p-values for the HJ-distance than for the Hansen statistic.[10] Of more critical concern, however, are the results using 100 assets. Note that the rejection rates of the HJ-distance are extremely high with as many as 330 time-series observations (51.0% at the 1% level, 73.3% at the 5% level, and 81.6% at the 10% level). Jagannathan and Wang report that the Fama-French model is rejected using the HJ-distance with an identical sample size. However, our simulation results indicate that their finding could be due to the excessive size of the HJ-distance test.

[10] For eight models using the same time-series/cross-section of Japanese stock market returns, Jagannathan et al. (1998) report the following p-values using the HJ-distance (the Hansen statistic): 1.46% (9.85%), 66.65% (69.21%), 3.61% (14.39%), 49.86% (64.96%), 59.65% (68.35%), 56.31% (62.97%), 74.49% (83.48%) and 69.85% (77.20%).

Ferson and Foerster (1994) suggest that a degrees-of-freedom adjustment to the estimated variance matrix of pricing errors (S_T) can improve the finite test size of the Hansen statistic. Table 2 shows results of applying this adjustment to tests of the Hansen statistic using two-step GMM (JF) and iterative GMM (JIF), and the HJ-distance (HJF).

Table 3
Model rejection rates using specification tests of the Hansen statistic and HJ-distance using calibrated simulation data

Test size (%)	Asymptotic tests			Degrees-of-freedom adjusted tests		
	J (%)	JI (%)	HJ (%)	JF (%)	JIF (%)	HJF (%)
(A) Fama-French (1993) model						
25 Assets						
1	1.3	1.3	2.4	0.5	0.5	1.8
5	5.0	5.0	8.7	3.4	3.3	6.1
10	10.0	10.2	16.5	7.3	7.2	12.9
100 Assets						
1	0.7	0.7	54.3	0.0	0.0	19.5
5	4.4	4.5	75.0	0.3	0.3	37.3
10	11.5	10.8	83.0	0.5	0.4	49.0
(B) Premium-Labor model (Jagannathan and Wang, 1996)						
25 Assets						
1	8.2	7.4	11.1	6.1	6.2	8.8
5	22.3	21.4	24.1	17.4	17.7	20.1
10	34.3	34.3	35.7	28.1	29.0	31.4
100 Assets						
1	25.6	23.6	77.0	1.3	1.4	43.4
5	55.6	53.4	89.9	9.0	7.8	65.5
10	71.0	69.5	94.2	19.7	17.8	75.6

This table shows rejection rates over 1000 trials using the p-values of the Hansen statistic and the HJ-distance with calibrated simulation data. Factors and returns are simulated using either the Fama-French (1993) model (Panel A) or the Premium-Labor model per Jagannathan and Wang (1996) (Panel B). First, using actual asset and factors provided by or based on data per Jagannathan and Wang (1996), asset betas and factor risk price components are estimated using two-pass estimation following Shanken (1992):

$$E(R_{it}) = \sum_{k=0}^{K}(E(F_{kt}) + \lambda_k)\beta_{ki},$$

where $F_{0t}=1$, and F_{kt} where $k>0$ is the realization of factor k in period t, and β_{ki} is the multivariate time-series factor beta associated with F_{kt}. Next, all factors are simultaneously simulated for 330 periods using time-series means, standard deviations and correlations among factors estimated using actual data. Next, expected asset returns are simulated using the estimated betas and factor risk price components and simulated Fama-French factors. Lastly, for each factor model, we estimate the error variance matrix using the regression residuals from the estimation of the equation above, simulate idiosyncratic error terms for all assets for each period using this estimated variance matrix, and add them to the simulated expected returns. For each trial, we use the actual 100 assets provided by Jagannathan and Wang and 25 value-weighted portfolios constructed from the original 100 assets in Jagannathan and Wang's data set using adjacent portfolios sorted by average portfolio size and beta. The test statistics are as follows: The Hansen-Jagannathan distance (HJ), and adjusted for Ferson-Foerster degrees of freedom (HJF); Hansen statistic using two-step GMM (J), and adjusted for Ferson-Foerster degrees of freedom (JF); and Hansen statistic using iterative GMM (JI), and adjusted for Ferson-Foerster degrees of freedom (JIF).

Similar to Ferson and Foerster, we find that this adjustment does a good job of moving the sizes of the Hansen statistics close to asymptotic levels when assets are limited, but is not nearly as effective with the HJ-distance. For example, the rejection rates for the adjusted two-step GMM Hansen statistic with 25 assets and 160 observations are 0.3%, 4.7%, and 8.9% at the 1%, 5%, and 10% levels. For the adjusted HJ-distance test, the rejection rates at these levels are 3.3%, 9.2%, and 17.0%. Furthermore, while the degrees-of-freedom adjustment over-corrects the test size of the Hansen statistic with 100 assets, it has only a limited effect on the test size of the HJ-distance.[11] With 100 assets and 330 observations, the rejection rates of the adjusted HJ-distance test at the 1%, 5%, and 10% levels are substantially larger than the asymptotic levels: 15.2%, 34.2%, and 48.9%. These results suggest that the finite sample limitations of the HJ-distance test cannot be easily corrected with a simple adjustment for degrees of freedom.

4.2. Analysis of test size using calibrated data

The analysis presented in the previous section suggests that the test size of the HJ-distance is too large in finite samples of comparable size to those used in previous studies (i.e., Jagannathan and Wang, 1996; Jagannathan et al., 1998). Because that analysis uses data simulated with a simple factor model and restrictive assumptions, the results may not be applicable to the data used in actual studies. In this section we report an analysis using data calibrated to resemble actual data used by Jagannathan and Wang.

Table 3 shows rejection rates for tests of the Hansen statistic and HJ-distance using the calibrated data. As can be seen in Panel A, the results using data calibrated based on the Fama-French model are similar in a material manner to those using the simple data. For each test size, the HJ-distance (HJ) rejects more often than the Hansen statistics (J and JI) using as few as 25 assets with 330 time-series observations. Furthermore, the rejection rates for the HJ-distance are extremely large relative to the asymptotic test sizes when using 100 assets: 54.3% at the 1% level, to 83.0% at the 10% level. Lastly, testing with the adjustment for degrees of freedom (HJF) only partially reduces the excessive size of the HJ-distance test even though it over-adjusts the test size of the two-step and iterative GMM Hansen statistics (JF and JIF).

The analysis of data calibrated to the Fama-French model supports the conclusions suggested by the analysis using the simple data: The size of the HJ-distance test is excessive in finite samples, and a simple adjustment based on degrees of freedom does not appear to correct this problem. It is interesting to observe that both the Hansen tests based on two-step and iterative GMM are well sized regardless of the number of assets. In fact, the Hansen tests perform better with the data calibrated to the Fama-French models than with our simple data, regardless of the number of assets. The degrees-of-freedom adjustment by Ferson and Foerster rather distorts the sizes of these tests, when asset returns are generated by the Fama-French factors. This result indicates that the degrees-of-freedom adjustment by Ferson and Foerster should be used with caution in practice.

[11] Ferson and Foerster also find that their method is less effective when the number of assets is large.

Analysis using data calibrated to the Premium-Labor model provides even more distracted results for the GMM test based on the HJ-distance. With only 25 assets (and 330 observations), the HJ-distance is rejected 11.1% of the time at the 1% level, 24.1% at the 5% level, and 35.7% at the 10% level. With 100 assets, the rejection rates of the HJ-distance rise to 77.0%, 89.9%, and 94.2%. Adjusting for degrees of freedom reduces these rejection rates only marginally with 25 assets and only partially with 100 assets. Notably, the Hansen tests based on two-pass and iterative GMM are now seriously biased toward rejecting correct model specifications, especially when a large number of assets are analyzed relative to the cases with the simple data and the data calibrated to the Fama-

Table 4
Model rejection rates of the Hansen statistic and HJ-distance p-value with the pricing-error variance matrix estimated using exact SDF parameters

Time-series observations	25 Assets				100 Assets			
	J (%)	JF (%)	HJ (%)	HJF (%)	J (%)	JF (%)	HJ (%)	HJF (%)
	Rejection at the 1% level							
160	3.6	1.8	7.3	4.7	3.1	0.0	99.3	80.9
330	1.9	1.2	2.6	2.0	11.1	0.4	51.3	21.7
700	1.3	1.2	1.6	1.5	4.8	1.4	15.9	6.5
1500	1.2	1.1	1.1	0.9	3.3	1.9	4.7	3.1
3000	0.6	0.5	0.8	0.7	1.5	1.0	2.1	1.7
	Rejection at the 5% level							
160	11.0	6.4	18.8	13.1	28.5	0.0	99.9	92.2
330	8.7	6.4	10.8	8.5	27.8	3.2	73.1	39.7
700	5.5	4.5	6.1	5.6	15.8	6.8	32.1	19.1
1500	5.1	4.6	5.4	5.2	9.2	6.9	14.0	9.8
3000	5.2	4.7	5.4	5.3	6.7	5.3	9.6	7.3
	Rejection at the 10% level							
160	19.9	12.3	28.1	20.1	53.1	0.0	99.9	95.0
330	15.8	12.4	18.5	15.4	41.0	7.4	82.1	48.8
700	11.6	10.8	13.4	12.0	27.3	12.5	43.3	27.7
1500	11.8	11.3	12.1	11.3	17.3	11.0	23.4	16.8
3000	10.6	10.3	11.3	11.0	15.0	11.9	16.7	13.7

This table shows the rejection rates over 1000 trials using the p-values of the non-iterative Hansen and HJ-distance specification tests examined in Tables 1 and 2 (J, JF, HJ, and HJF), but substituting the exact parameters of the stochastic discount factor (SDF) for the estimated ones when estimating the pricing-error variance matrix. The SDF parameters for the three-factor model used in the simple data are δ_0, δ_1, δ_2 and δ_3, in the following model:

$$E[R_{it}(\delta_0 + \delta_1 F_{1t} + \delta_2 F_{2t} + \delta_3 F_{3t})] = 1.$$

Based on the known parameters used to simulate the simple data (see Appendix A.3), the exact SDF parameters are:

δ_0	δ_1	δ_2	δ_3
1.21	−31.89	−31.89	−31.89

Other than substituting the exact parameters for the estimated ones, the analysis shown here for the non-iterative Hansen and HJ-distance tests is identical to those used for results shown in Tables 1 and 2.

French model.[12] Adjusting the two-step and iterative GMM Hansen statistics for degrees of freedom (JF and JIF) reduces their over-rejection rates, but only marginally. As such, the excessive rejection rates for the HJ-distance may be due to a feature of data based on the Premium-Labor model not present in the other data. This result suggests that the finite sample size of the HJ-distance based test is sensitive to the choice of factor models in a manner similar to, but more exaggerated than, the other GMM-based specification tests.

4.3. Diagnosis of size overstatement in finite samples

As discussed in Section 2, previous studies of GMM indicate that Hansen statistics are biased in small samples because they use poorly estimated optimal weighting matrices. In this subsection, we demonstrate that this problem also applies to the empirical p-value of the HJ-distance. In our simulations, the optimal GMM weighting matrix is the inverse of the variance matrix of pricing errors. We estimate this variance matrix by the matrix S_T defined in Eq. (3). There are two possible reasons why S_T may be a poor estimate of the true variance matrix. First, the computation of the matrix S_T requires use of estimated SDF parameters. If these estimates are not precise enough, the S_T matrix could be poorly estimated. Second, S_T may be a poor estimate even if the true SDF parameters are used. The variance matrix of pricing errors has $N(N+1)/2$ distinct elements that should be estimated using only T time-series observations. It is well known that models with too many parameters compared to sample size are poorly estimated. For the same reason, we can expect that the S_T matrix, even if it is computed with the true SDF parameters, will have poor finite-sample properties when the number of assets (N) is too large compared to the number of time-series observations (T).

To explore the possibility that finite sample bias in estimated SDF parameters distorts the finite sample size properties of the GMM specification tests, we substitute the exact values of the SDF parameters for estimated SDF parameters when computing the S_T matrix. Because we use the exact values of the parameters, iterative GMM is not appropriate in computing the Hansen statistic. As such, the only results we report for the Hansen statistic are those using two-step GMM. Table 4 reports our simulation results with this adjustment. Comparing the results from Table 4 with those of Tables 1 and 2, we can clearly see that using the true SDF parameter values does not improve the finite-sample performances of the HJ-distance, with or without an adjustment for degrees of freedom. In fact, for some cases, rejection rates for the HJ-distance p-value with exact SDF parameters are slightly farther from asymptotic levels than rejection rates with estimated SDF parameters. For example, with 25 assets and 160 observations, the HJ-distance p-value is rejected at the 1% level 5.5% of the time with estimated SDF parameters but 7.3% of the time when using the exact SDF parameters (Table 4, 25 assets, HJ). Similar but small increases in rejection rates are often observed across the other test methods and different finite sample pairings examined. These results indicate

[12] This over-rejection problem seems to be related to the fact that the PREM factor used in the Premium-Labor model is highly autocorrelated. In Ahn and Gadarowski (2001), we find that the Dickey-Fuller test is unable to reject the presence of unit root in the PREM factor. It is well known that estimators and test statistics based on GMM do not perform well in highly persistent data.

that use of estimated SDF parameters for the variance matrix of pricing errors is not a major reason for finite-sample size overstatements of the HJ-distance p-value and the Hansen statistic.

As further support for the notion that estimation of the SDF parameters is not affecting the finite-sample size of the HJ-distance test, we perform analysis of the bias and root mean square error of the estimated SDF parameters when computing both the two-step Hansen statistic and HJ-distance tests. This analysis directly complements that reported in Table 4 using the unadjusted versions of these tests. Table 5 compares finite-sample biases in the SDF parameters estimated the optimal two-step GMM ($\delta_{OGMM,T}$ defined in Eq. (3)) and by the HJ-distance based ($\delta_{HJ,T}$ defined in Eq. (2)). We report the biases and the root mean square errors (RMSE) of the estimates relative to the true (absolute) parameter values. The boxes labeled as "δ_0" report results for the estimates of the constant term of SDF (δ_0 defined in Eq. (15)), while the boxes labeled as δ_x report the average relative

Table 5
Monte Carlo analysis of the bias and root mean square error of stochastic discount factor parameter estimates by the HJ-distance-based and optimal GMM

Time-series observations	Bias/exact value				RMSE/Abs(exact value)			
	25 Assets				25 Assets			
	δ_0		δ_x		δ_0		δ_x	
	J (%)	HJ (%)	J (%)	HJ (%)	J (%)	HJ (%)	J (%)	HJ (%)
160	1.6	1.6	0.8	1.1	7.2	7.0	38.0	36.5
330	1.1	1.1	0.9	1.0	4.8	4.8	25.8	25.4
700	0.3	0.4	0.1	0.2	3.2	3.2	17.7	17.6
1500	0.2	0.2	0.3	0.3	2.2	2.2	11.9	11.8
3000	0.1	0.1	0.1	0.1	1.4	1.4	8.3	8.3
Time-series observations	100 assets				100 Assets			
	δ_0		δ_x		δ_0		δ_x	
	J (%)	HJ (%)	J (%)	HJ (%)	J (%)	HJ (%)	J (%)	HJ (%)
160	2.3	2.2	3.0	2.6	7.8	7.3	38.0	35.5
330	1.1	1.0	1.2	1.0	5.0	4.6	25.8	24.1
700	0.4	0.4	0.3	0.3	3.2	3.1	17.1	16.3
1500	0.1	0.1	− 0.2	− 0.2	2.2	2.2	11.6	11.2
3000	0.2	0.2	0.3	0.3	1.4	1.4	8.0	7.9

This table shows the bias and root mean square error of stochastic discount factor parameter estimates using either the Two-step GMM method of Hansen (J), or the Hansen–Jagannathan distance method (HJ) over 1000 trials. For either method, the stochastic discount factor (SDF) parameters, δ_0, δ_1, δ_2 and δ_3, are estimated by GMM using the moment condition. Here, R_{it} is the asset return, and F_{1t}, F_{2t} and F_{3t} are factors. For 1000 trials of randomly generated time-series observations of risk factor realizations and factor-generated asset returns using a signal-to-noise ratio of 1.0, the SDF parameters were estimated and compared to their exact values:

δ_0	δ_1	δ_2	δ_3
1.21	−31.89	−31.89	−31.89

Because exact values of δ_1, δ_2 and δ_3 in this analysis are the same, the table below shows average values across δ_1, δ_2 and δ_3 under the variable name δ_x. We here report the sizes of biases and root mean square errors (RMSE) of parameter estimates relative to exact (absolute) values of the true parameters.

biases and RMSE of the estimates for the risk factor parameters in SDF (δ_1, δ_2 and δ_3 defined in Eq. (15)).

As seen in Table 5, the sizes of the biases in SDF parameter estimates appear to be similar for the HJ-distance based (HJ) and two-step GMM (J). Furthermore, the biases from the two different estimation methods are quite small even if the number of time-series observations is small. The biases appear larger with 100 assets than with 25 assets, but not materially so. Both the HJ-distance-based and two-step GMM with 160 time-series observations bias the SDF parameter estimates on average by 3.0% or less for cases with either 25 and 100 assets. The biases become less than 1.1% when more time-series observations are used. Furthermore, the relative RMSE behaves similarly to the relative bias across the different estimation methods and sample pairings. Across both methods, the RMSE monotonically decreases with the number of time-series observations. A notable observation from Table 5 is that unlike with the relative bias, the relative RMSE is substantially larger for the risk factor parameters estimates than for the constant term estimates across all the cases we consider. Kan and Zhou (1999) find similar results. In

Table 6
Model rejection rates of the Hansen statistic and HJ-distance p-values using the exact pricing-error variance matrix

Time-series observations	25 Assets				100 Assets			
	J (%)	JF (%)	HJ (%)	HJF (%)	J (%)	JF (%)	HJ (%)	HJF (%)
	Rejection at the 1% level							
160	0.2	0.0	0.9	0.1	0.0	0.0	0.3	0.0
330	1.0	0.7	1.6	1.0	0.0	0.0	0.7	0.0
700	1.0	0.9	1.1	1.0	0.1	0.1	1.5	0.2
1500	0.6	0.5	0.9	0.8	0.2	0.0	1.0	0.6
3000	0.6	0.6	0.6	0.6	1.0	0.7	1.5	1.0
	Rejection at the 5% level							
160	1.5	0.5	4.2	2.3	0.0	0.0	4.1	0.0
330	3.3	2.5	4.5	3.3	0.3	0.0	3.8	0.3
700	3.9	2.9	4.7	4.2	0.9	0.1	5.1	2.1
1500	3.6	3.2	4.2	3.7	3.0	1.5	6.0	3.4
3000	3.7	3.6	4.1	4.0	4.6	3.6	5.9	4.8
	Rejection at the 10% level							
160	4.5	1.6	7.9	5.1	0.1	0.0	9.9	0.1
330	6.1	4.4	9.4	6.6	0.9	0.0	8.8	0.8
700	7.9	7.2	9.0	7.9	2.3	0.7	10.3	3.6
1500	8.5	8.2	9.2	8.7	6.6	3.9	11.0	7.3
3000	8.8	8.5	9.3	8.3	9.0	6.9	11.3	9.1

This table shows the rejection rates over 1000 trials using the p-values of the non-iterative Hansen and HJ-distance specification tests examined in Tables 1 and 2 (J, JF, HJ, and HJF), but substituting the exact pricing-error variance matrix for the estimated variance matrix. The exact variance matrix is computed using estimated variance matrix with the indicated number of assets but with 10,000 time-series observations. Other than substituting the exact variance matrix for the estimated one for the indicated number of observations, the analysis shown here for the non-iterative Hansen and HJ-distance tests is identical to those used for results shown in Tables 1 and 2.

summary, we find little evidence that the HJ-distance-based GMM produces significantly less reliable SDF parameter estimates than two-step GMM.[13]

In contrast to the possibility that estimated SDF parameters are materially biasing the test size of the HJ-distance, we find direct evidence that the finite-sample overstatement of its test size is caused by the poorly estimated variance matrix S_T. By substituting the exact variance matrix (S) of pricing errors for the estimated matrix (S_T), we eliminate the bias due to the bias in S_T. Because it is computationally complicated to obtain the exact value for each element in the exact variance matrix, we first estimate the variance matrix of pricing errors using 10,000 time-series observations and true SDF parameter values, then substitute this large sample estimate in place of the smaller sample estimate in the calculations where appropriate. As shown in Tables 1 and 2, the test size of the HJ-distance is close to asymptotic levels with about 3000 observations. By using 10,000 observations, it is highly likely that the estimated variance matrix we use is not materially different from the exact variance matrix.

As can be seen in Table 6, the rejection rates of the unadjusted HJ-distance p-value (HJ) are much closer to asymptotic levels with the large sample estimate of the variance matrix for cases with both 25 and 100 assets. For example, with 25 assets and 160 observations, the HJ-distance is rejected 0.9% of the time at the 1% level when the exact pricing-error variance matrix is used, while it rejected 4.2% of the time at the 5% level, and 7.9% at the 10% level. Similar results are seen using 100 assets. Using the Ferson–Foerster adjustment further exacerbates the understatement of test size. As the Ferson–Foerster adjustment is designed for the estimation of both model parameters and the optimal GMM weighting matrix, it is not surprising to see that this adjustment is inappropriate for GMM with the exact optimal weighting matrix. In summary, substituting the exact pricing error variance matrix removes most of the bias in the test size of the HJ-distance. It is of course impossible to use exact variance matrices of pricing errors in practices. Thus, our exercise here only demonstrates the problem in estimation of the HJ-distance p-value using finite samples but offers no solution.

5. Concluding remarks

As an alternative model comparison tool, Hansen and Jagannathan (1997) propose use of a measure, the HJ-distance, which estimates the maximum pricing error generated by an asset pricing model. Extending this study, Jagannathan and Wang (1996) develop a GMM model specification test based on the HJ-distance. Jagannathan and Wang show that for linear models, the p-value for the HJ-distance can be empirically estimated using a weighted average of random draws from a χ^2 distribution. While the HJ-distance measure has many potential advantages over the Hansen (1982) statistic, the empirical p-value of it

[13] Altonji and Segal (1996) show that optimal GMM estimates could be more biased than nonoptimal GMM estimates in small samples. To the extent that two-step GMM is optimal, our findings from Table 5 are not consistent with those from Altonji and Segal (1996) and Ferson and Foerster (1994). These difference in results, however, could be due to different model specifications.

estimated following Jagannathan and Wang (1996) relies upon estimation of variance matrix of pricing errors, as much as the Hansen statistic does. Prior studies of the Hansen statistic (e.g., Ferson and Foerster, 1994; Burnside and Eichenbaum, 1996) show that it rejects correct models too often.

Our Monte Carlo analysis demonstrates that the empirical p-value of the HJ-distance is seriously biased toward rejecting correct models, much more than the Hansen test is. For example, with 100 assets and 330 time-series observations, the sample size used by Jagannathan and Wang (1996) in rejecting the Fama-French model, we find that the HJ-distance overstates the test size at the 1% level about 50% of the time. The major reason for this problem is that the empirical p-value of the HJ-distance is computed with poorly estimated variance matrix of pricing errors. The degrees-of-freedom adjustment by Ferson and Foerster (1994) fails to correct overstatement of test size of the HJ-distance. In short, this study suggests that researchers should be extremely cautious with the use of the GMM specification test based on the HJ-distance proposed by Jagannathan and Wang, especially in samples with too few time-series observations relative to the number of assets.

Nonetheless, the HJ-distance has several attractive properties, as discussed in Section 2. Given these attractive properties, it may be possible to develop an alternative test procedure based on the HJ-distance that has better finite-sample size properties than the method proposed by Jagannathan and Wang. For example, bootstrap-based HJ-distance tests might have better finite sample properties. Further study of this research line would be useful.

Acknowledgements

The first author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. For their comments in the development of this paper, we gratefully acknowledge seminar participants at Arizona State University, especially Hank Bessembinder, John Griffin and Mike Lemmon, as well as the anonymous referee on the paper. All remaining errors are our own.

Appendix A

A.1. The HJ-distance for a general model of returns

Hansen and Jagannathan (1997) propose the following measure to investigate how badly an asset pricing model is misspecified:

$$\mathrm{HJ}(\delta) \equiv \sqrt{E[w_t(\delta)]'\, G^{-1} E[w_t(\delta)]}, \tag{8}$$

where $w_t(\delta) \equiv R_t m_t(\delta) - 1_N$ and 1_N is a $N \times 1$ vector of ones, R_t is the tth period vector of gross returns for N assets, and $G \equiv E(R_t R_t')$.[14] Here, $m_t(\delta)$ denotes a stochastic discount

[14] Jagannathan et al. (1998) refer to this measure as the 'least square distance'.

factor (SDF) implied by a particular model, where δ denotes a unknown parameter vector to be estimated, and the vector of asset prices implied by the model is given by $E[R_t m_t(\delta)]$, which are normalized to one. Hansen and Jagannathan show that for a given δ, this measure equals the maximum pricing error generated by a given asset pricing model. A correctly specified model should not generate pricing errors. Thus, the null hypothesis of correct model specification implies N moment conditions, $E[w_t(\delta)] = 0_N$, where 0_N indicates a $N \times 1$ vector of zeros. It follows that the square root of the quadratic of this measure, when properly weighted, provides the least square distance between a SDF implied by a model and the family of SDFs correctly pricing assets. In order to rule out the possibility of arbitrage, the correct SDFs should be strictly positive with probability one. However, the HJ-distance does not rule out nonpositive SDFs. Hansen and Jagannathan (1997) derive an alternative measure which restricts correct SDFs to be positive.

A.2. Empirical p-value of the HJ-distance

The matrix Φ is not of full column. It is rather a positive semidefinite matrix with rank equal to $(N-K)$. Accordingly, the matrix has only $(N-K)$ positive eigenvalues. Let λ_j $(j=1, \ldots, N-K)$ denote these positive eigenvalues. With this notation, Jagannathan and Wang show that under the null hypothesis of correct model specification,

$$\mathrm{SHJ}_T \rightarrow u \equiv \sum_{j=1}^{N-K} \lambda_j v_j, \tag{9}$$

where " $\rightarrow$ " means "converges in distribution" and the v_j are $(N-K)$ independent $\chi^2(1)$ random variates.

Based on this result, Jagannathan and Wang propose using an empirical p-value which can be obtained using an algorithm that empirically compares SHJ_T to the eigenvalue-weighted sample of random draws from a $\chi^2(1)$ distribution. Specifically, let v_{ij} $(I=1, \ldots,$ M and $j=1, \ldots, N-K)$ denote independent $\chi^2(1)$ random draws; and let $u_i = \sum_{j=1}^{N-K} \lambda_j v_{ij}$. Then, the empirical p-value for the HJ-distance is defined as

$$M^{-1} \sum_{i=1}^{M} \boldsymbol{1}(u_i \geq \mathrm{SHJ}_T), \tag{10}$$

where M is the number of draws from the χ^2 distribution, and $\mathbf{1}(\bullet)$ is an index function which equals one if the expression in the parentheses is true and 0 otherwise. The Law of Large Numbers guarantees that this empirical p-value converges to the true p-value for the SHJ_T statistic as M goes to infinity. In practice, one needs to choose an appropriate size of M to simulate asymptotic results. For example, Jagannathan and Wang (1996) set M to 5000 in their analysis. In unreported simulations, we have tried values of M ranging from 1000 to 10,000, and found the standard deviation of the sample p-value around its asymptotic value was about 1.0% at sampling with 5000 draws.

A.3. Simple data generation

To generate sets of homogeneous returns and factors, we use the following data generating process:

$$R_{it} = \alpha + F_{1t}\beta_{1i} + F_{2t}\beta_{2i} + F_{3t}\beta_{3i} + e_{it}, \tag{11}$$

where $I=1, \ldots, N$ indexes individual assets, $t=1, \ldots, T$ indexes time, R_{it} is the gross return for asset I for period t, α is an overall intercept term, F_{jt} ($j=1$, 2, and 3) are the common factors for period t, β_{ji} are the corresponding factor betas for asset I, and e_{it} is the idiosyncratic error with mean zero for asset I for period t. We use three factors because both models examined by Jagannathan and Wang with HJ-distance method have three factors (the Fama-French model and their Premium-Labor model). Because of the limited availability of time-series data, studies to date using the HJ-distance that we are aware of have used monthly data (Jagannathan and Wang, 1996; Jagannathan et al., 1998). Accordingly, we view the subscript t as indexing months. In our simulations, we consider 15 different combinations of N and T, with $N=25$, 50 and 100, and $T=160$, 330, 700, 1500 and 3000, although the results for the cases with $N=50$ are not reported here. More detailed simulation results can be found in Ahn and Gadarowski (2000), which is an earlier version of this paper.

A three-factor linear model implies the following restriction:

$$E(R_{it}) = c_0 + c_1\beta_{1i} + c_2\beta_{2i} + c_3\beta_{3i}, \tag{12}$$

where c_0 equals the risk free rate, the coefficients of betas, c_1, c_2 and c_3, are the risk factor prices corresponding to F_{1t}, F_{2t} and F_{3t}, respectively. Clearly, our data generating process (Eq. (12)) satisfies this restriction with $c_0=\alpha$ and $c_j=E(F_{jt})$, for $j=1$, 2, and 3. By construction, factor means in our analysis are equal to risk premiums.

In our simulations of these simple returns, the value of the overall intercept term α is fixed at 1.0033. This number is chosen because the historical average of the risk free component for annual US stock (net) returns roughly equals 4% ($=0.33\% \times 12$ months). The three betas are randomly drawn once from a uniform distribution with the range between zero and two, so that the cross-sectional mean of each betas equals one. The three factors are random draws from a normal distribution with mean equal to 0.0022 and variance equal to 6.944×10^{-5}. Here, we follow the spirit of a CAPM-style model of expected returns, where c_0 is the risk free rate and the other parameters c_j are the risk premiums associated with the risk factor-mimicking portfolios. The factors are mutually independent, and each factor is independently and identically distributed over time. With the means of factors and betas, the risk premium component for annualized returns in our generated data equals 8% ($=0.22\% \times 3$ factors $\times$ 12 months). This number is consistent with the historical average of risk premiums in the US stock market. The value 6.944×10^{-5} for the variances of factors is chosen to make the variance of returns in our generated data roughly consistent with the historical average of stock return variability in the US.

For the realizations of simple returns, we then add homogeneous idiosyncratic error terms, e_{it}, from normal distributions with common mean zero, but examine cases with different variances, Var(e_{it}), for all error terms. The size of Var(e_{it}) determines the signal-to-noise ratio (SNR), which equals the ratio between the two components of the return variation: one caused by the factor variations; and the other by the error variation. Since factor variances are the same in our simulations, we use the term "signal-to-noise ratio" to refer to Var(F_{1t})/Var(e_{it}). Different values for Var(e_{it}) are used to generate return data with different SNRs. We only report the results with SNR = 1.0. The results obtained using different SNRs can be found in Ahn and Gadarowski (2000).

A.4. Calibrated data generation

We also simulate factors and asset returns mimicking the actual factors and returns from 100 assets that are provided by Jagannathan and Wang (1996). To do so, we do the following steps. First, using the actual data for the Fama-French and Premium-Labor model factors and asset returns, we estimate asset betas and factor-mean adjusted risk prices (λ_k) are estimated using two-pass estimation following Shanken (1992):

$$E(R_{it}) = \sum_{k=0}^{3} (E(F_{kt}) + \lambda_k)\beta_{ki}, \tag{13}$$

where $F_{0t} = 1$ for all $t = 1, \ldots T$, and F_{kt} ($k = 1, 2, 3$) is the realization of factor k in period t, and β_{ki} is the multivariate time-series factor beta associated with F_{kt}. Next, all the (five) factors from the two factor models are simultaneously simulated for 330 periods using time-series means, standard deviations and covariances estimated from actual data. Then, using the simulated Fama-French or Premium-Labor model factors with the estimated betas and factor-mean adjusted risk prices for these models, we compute simulated expected returns using Eq. (13). Lastly, for each factor model, we estimate the error variance matrix using the regression residuals from the estimation of Eq. (13). Using this estimated variance matrix, we simulate idiosyncratic error terms for all assets for the each period and add them to the simulated expected returns. Using this procedure repeatedly, we generate 1000 different data sets for each of the Fama-French and PL-models. For a sensitivity analysis, we also calibrate data using only 25 assets. We use 25 value-weighted portfolios constructed from the original 100 assets in Jagannathan and Wang's data set using adjacent portfolios sorted by average portfolio size and beta.

A.5. Exact SDF parameters

As Jagannathan and Wang (1996) point out (following Dybvig and Ingersoll, 1982), when expected returns are linear functions of factor betas as given in Eq. (12), the SDF is a linear function of the factor realizations. Specifically, the restriction (Eq. (12)) generates the following moment conditions:

$$E[R_{it}(\delta_0 + \delta_1 F_{1t} + \delta_2 F_{2t} + \delta_3 F_{3t})] = 1 \tag{14}$$

where

$$\delta_0 = \frac{1}{c_0} + \frac{1}{c_0}\left(\sum_{j=1}^{3} c_j \frac{E(F_{jt})}{\mathrm{Var}(F_{jt})}\right); \quad \delta_j = -\frac{1}{c_0}\frac{c_j}{\mathrm{Var}(F_{jt})}, \tag{15}$$

for $j = 1$, 2 and 3. The term $(\delta_0 + \delta_1 F_{1t} + \delta_2 F_{2t} + \delta_3 F_{3t})$ in Eq. (14) is the SDF implied by our three-factor model, and is equivalent to our notation $Y_t'\delta$ in Section 2. Our Monte Carlo experiments exploit the moment conditions (Eq. (14)) to compute Hansen statistics and HJ-distances. With the simple data, we know the exact values of all the terms in Eqs. (14) and (15), eliminating any chance of error in inferring the exact values from estimation.

References

Ahn, S.C., Gadarowski, C., 2000. Small sample properties of the specification test based on the Hansen-Jagannathan distance. Mimeo. Arizona State University, Tempe, Arizona.

Ahn, S.C., Gadarowski, C., 2001. Two-pass cross-sectional regression of factor pricing models. Mimeo. Arizona State University, Tempe, Arizona.

Altonji, J.G., Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance structures. Journal of Business and Economic Statistics 14, 353–366.

Andersen, T.G., Sørensen, B.E., 1996. GMM estimation of a stochastic volatility model: a Monte Carlo study. Journal of Business and Economic Statistics 14, 328–352.

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Burnside, C., Eichenbaum, M., 1996. Small-sample properties of GMM-based Wald tests. Journal of Business and Economic Statistics 14, 294–308.

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

Duffie, D., 1996. Dynamic Asset Pricing Model, 2nd ed. Princeton University Press, Princeton, NJ.

Dybvig, P.H., Ingersoll Jr., J.E., 1982. Mean-variance theory in complete markets. Journal of Business 55, 233–251.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on bonds and stocks. Journal of Financial Economics 32, 3–56.

Ferson, W.E., Foerster, S.R., 1994. Finite sample properties of the Generalized Method of Moments in tests of conditional asset pricing models. Journal of Financial Economics 36, 29–55.

Gibbons, M.R., Ferson, W.E., 1985. Testing asset pricing models with changing expectations and an unobservable market portfolio. Journal of Financial Economics 14, 217–236.

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

Hansen, L.P., Hodrick, R.J., 1983. Risk averse speculation in the forward foreign exchange market: an econometric analysis of linear models. In: Frenkel, J. (Ed.), Exchange Rates and International Macroeconomics. University of Chicago Press, Chicago, IL, pp. 113–142.

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., 2001. Evaluating the specification errors of asset pricing models. Journal of Financial Economics 62, 327–376.

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

Jagannathan, R., Kubota, K., Takehara, H., 1998. Relationship between labor-income risk and average return: empirical evidence from the Japanese stock market. Journal of Business 71, 319–347.

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

Kan, R., Zhou, G., 2002. Hansen–Jagannathan Distance: Geometry and Exact Distribution, Working Paper, Olin School of Business. Washington University in St. Louis.

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

Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5, 1–34.

Zhou, G., 1994. Analytical GMM tests: asset pricing with time-varying risk premiums. Review of Financial Studies 7, 687–709.