Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance

Seung C. Ahn
Arizona State University

Christopher Gadarowski
Arizona State University

This Version: August 1999

Abstract
This paper examines finite-sample properties of a model test method developed by Hansen and Jagannathan (1997) and Jagannathan and Wang (1996). As a model comparison tool, Hansen and Jagannathan (1997) propose a measure which estimates the maximum pricing error generated by an asset pricing model for portfolios with unit second moments of gross returns. Jagannathan and Wang (1996) show that this measure has an asymptotic p-value that can be empirically estimated using a weighted average of draws from a $\chi^2$ distribution. Our simulation results indicate that the empirical p-value rejects correct models too often for commonly used sample sizes.

Acknowledgment
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. We gratefully acknowledge seminar participants at Arizona State University, especially Hank Bessembinder, John Griffin and Mike Lemmon. All remaining errors are of course our own. Corresponding author: Christopher Gadarowski, Department of Finance, Tempe, AZ 85287; email: Chris.Gadarowski@asu.edu.
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 $\chi^2$ statistic for detecting such errors, which we hereafter call the Hansen statistic. Asset pricing models imply stochastic discount factors that can be used to determine 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 $\chi^2$-distributed under the null hypothesis that the model estimated is correctly specified. The variance matrix of pricing errors is an optimal GMM weighting matrix in the sense that use of it minimizes the asymptotic variances of GMM parameter estimates. Although the variance matrix is not observed, a consistent estimate of it can used in practice.

Use of the Hansen test however has been criticized for at least three reasons. Firstly, the Hansen test favors models with highly variable pricing errors (Hansen and Jagannathan, 1997; and Jagannathan and Wang, 1996). This is so because the statistic is inversely related with the variances of pricing errors. Secondly, 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. Thirdly, as Ferson and Foerster (1994) and Burnside and Eichenbaum (1996) have shown, sizes of the Hansen tests are often too large with finite samples.

In response to these problems, Hansen and Jagannathan (1997) and Jagannathan and Wang (1996) 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 in data.
Jagannathan and Wang (1996) call this measure the Hansen-Jagannathan distance or simply, HJ-distance. Hansen and Jagannathan (1997) 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).

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 test, 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 test, the HJ-distance estimated by GMM can be used to test whether a given asset pricing model is correctly specified. While the HJ-distance is not \( \chi^2 \)-distributed under the null hypothesis of correct model specification, Jagannathan and Wang (1996) show that it is asymptotically equivalent to a random variate from a weighted \( \chi^2 \) distribution. Based on this finding, they develop a simulation method to empirically estimate the p-value for the HJ-distance. Jagannathan and Wang (1996) and Jagannathan, Kutoba and Takehara (1998) evaluate several linear asset pricing models using this method.

Despite these theoretical merits, finite-sample properties of the HJ-distance and its empirical p-value are unknown. The motivation of this paper is to fill this gap in the literature. In this paper, we conduct some limited Monte Carlo experiments to compare the finite-sample performances of the HJ-distance and the Hansen test. The foundation of our experiments is a linear three-factor pricing model. Empirical studies of linear models often use the \( R^2 \) from the regression of mean returns on estimated betas, which we call the two-pass \( R^2 \), to compare specifications. We also briefly compare the HJ-distance with the two-pass \( R^2 \).

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; and Burnside and Eichenbaum, 1996). Use of this estimated weighting matrix leads to poor finite-sample performance of the Hansen test. 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 empirical p-value of the HJ-distance requires estimation of the optimal weighting matrix. Thus, there is no guarantee that the specification test based on the HJ-distance p-value has better small-sample properties than the Hansen test.

The results from our Monte Carlo experiments can be summarized as follows. First, unlike the two-pass $R^2$, the HJ-distance does not reward models using factors highly correlated with returns. A model using factors more highly correlated between returns does not necessarily generate smaller pricing errors. Accordingly, the HJ-distance can be used as a supplementary tool for model comparison. Second, 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. For most of the cases we investigate, we find that the Hansen test, which also tends to over-reject correct models, outperforms the HJ-distance p-value. Third, we apply the degrees-of-freedom adjustment proposed by Ferson and Foerster (1994), which is originally designed 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 remain severe for most cases when a larger number of assets (50 and 100) are used.

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

This section introduces the measure of model fit developed by Hansen and Jagannathan (1997) and Jagannathan and Wang (1996), and explains how the measure and its empirical p-value can be estimated. We also consider the link between this measure and the generalized method of moments (GMM). Most of our discussions closely follow the above two studies.
To begin with, consider a portfolio consisting of $N$ primitive assets. Let $R_t$ be the $t$'th period vector of gross returns for these assets. Asset pricing models imply certain stochastic discount factors which can be used to price assets. Let $m_t(\delta)$ denote a stochastic discount factor (SDF) implied by a particular model, where $\delta$ denotes a unknown $K \times 1$ parameter vector to be estimated. Then, the vector of asset prices implied by the model is given by $E[R_t m_t(\delta)]$.

Different asset pricing models imply different SDFs. For example, linear factor pricing models, which are our particular interest, imply SDFs of the form, $m_t(\delta) = R_t Y_t \delta$, where $Y_t$ is the $K \times 1$ vector of common factors and one. Other SDFs can be found in Hansen and Jagannathan (1997).

A way to check whether or not a given model is correctly specified is to test whether or not the SDF implied by the model generates statistically significant pricing errors for individual assets. If we normalize all of the $N$ asset prices to ones, the $N \times 1$ vector of pricing errors can be denoted by $E[w(\delta)]$, where $w(\delta) = R_t m_t(\delta) - 1_N$ and $1_N$ is a $N \times 1$ vector of ones (see Hansen and Jagannathan, 1997; or Duffie, 1996). Obviously, a correctly specified model should not generate pricing errors. Thus, the null hypothesis of correct model specification implies $N$ moment conditions, $E[w(\delta)] = 0_N$, where $0_N$ indicates a $N \times 1$ vector of zeros.

The measure of model fit proposed by Hansen and Jagannathan (1997), the HJ-distance, is defined as a quadratic function of pricing errors weighted by the inverse of the second moments of gross returns:

$$HJ(\delta) = \sqrt{E[w(\delta)]^\top G^{-1} E[w(\delta)]}, \quad (1)$$

where $G = E(R_t R_t')$ is a positive definite matrix.¹ Hansen and Jagannathan show that for a given $\delta$, this measure equals the maximum pricing error generated by a given asset pricing model. To see why, let $x$ be the $N \times 1$ vector of weights given to individual assets by a portfolio. Then, the pricing error corresponding to this portfolio equals $x' E[w(\delta)]$. Consider the problem of finding a pricing-error maximizing portfolio among those having the unit second moment. This problem is equivalent to the problem of finding a weight vector $x$ which maximizes $x' E[w(\delta)]$ subject to the constraint $E[(x' R_t)^2] = x' E(R_t R_t') x = 1$. The solution of this problem is $x^* = \ldots$

¹Jagannathan, Kubota and Takehara (1998) refer to this measure as the ‘least square distance’.
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. However, this alternative measure is much more complicated to estimate than the HJ-distance.

Computation of the HJ-distance requires estimation of the unknown parameter vector $\delta$. GMM provides a convenient way to estimate $\delta$. Jagannathan and Wang (1996) explain in detail how we can estimate the HJ-distance for a linear factor pricing model. We briefly summarize their method here. Define the following sample analogs:

$$D_T = T^{-1}\sum_{t=1}^{T}R_t^\prime Y_t^\prime = T^\prime R^\prime Y;$$

$$G_T = T^{-1}\sum_{t=1}^{T}R_tR_t^\prime = T^\prime R^\prime R;$$

$$w_T(\delta) = T^{-1}\sum_{t=1}^{T}w_t(\delta) = D_T\delta - 1_N,$$

where $R$ is the $T\times N$ matrix of returns, $Y$ is the $T\times K$ matrix of factors and ones, and $w_t(\delta) = R_tY_t^\prime\delta - 1_N$ is the $N\times 1$ vector of pricing errors at time $t$. This leads to a sample analog of the HJ-distance:

$$HJ_T(\delta) = \sqrt{w_T(\delta)^\prime G_T^{-1}w_T(\delta)}. \quad (2)$$

Jagannathan and Wang (1996) show the minimizer of this sample HJ-distance equals

$$\delta_{HJ,T} = (D_T^\prime G_T^{-1}D_T)^{-1}D_T^\prime G_T^{-1}1_N, \quad (3)$$

which is essentially a GMM estimator based on the moment conditions $E[w(\delta)] = 0_N$. With this estimate, the HJ-distance can be consistently estimated by $HJ_T(\delta_{HJ,T})$. Whenever there is no notational confusion, we hereafter use the term "HJ-distance" to refer to this estimated HJ-distance.

The asymptotic variance matrix of the GMM estimator $\delta_{HJ,T}$ can be easily estimated

---

\footnote{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) derives an alternative measure which restricts correct SDFs to be positive. However, this alternative measure is much more complicated to estimate than the HJ-distance.}
following Hansen (1982). Specifically, a consistent estimate of the variance matrix can be obtained by the formula:

\[ \text{Var}(\delta_{\text{HJ},T}) = T^{-1}(D_T'G_T^{-1}D_T)^{-1}D_T'G_T^{-1}S_T^{-1}G_T^{-1}D_T(D_T'G_T^{-1}D_T)^{-1}, \] (4)

where \( S_T = T^{-1}\sum_{t=1}^{T} W_t(\delta_{\text{HJ},t})W_t(\delta_{\text{HJ},t})' \) is an estimate of the variance matrix of pricing errors \( S = \text{Var} [\sqrt{T}W_T(\delta)] \). Here, we implicitly assume that data are serially uncorrelated. 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.

The asymptotic results from Hansen (1982) imply that \( \delta_{\text{HJ},T} \) is not an efficient (minimum variance) estimator among those utilizing the moment conditions \( \mathbb{E}[w(\delta)] = 0 \). This is so because the matrix \( G_T^{-1} \) is not an optimal weighting matrix. Hansen (1982) implies that the optimal weighting matrix which minimizes the asymptotic variance matrix (4) is \( S_T^{-1} \). With this weighting matrix, the optimal GMM estimator of \( \delta \) is given by

\[ \delta_{\text{OGMM},T} = (D_T'S_T^{-1}D_T)^{-1}D_T'S_T^{-1}1_N, \] (5)

where \( \text{Var}(\delta_{\text{OGMM},T}) = T^{-1}(D_T'S_T^{-1}D_T)^{-1} \). An advantage of using this optimal GMM estimator \( \delta_{\text{OGMM},T} \) is that it provides a convenient model specification test statistic which we call the Hansen (1982) statistic:

\[ J_T = TW_T(D_{\text{OGMM},T})'(S_T^{-1}W_T(\delta_{\text{OGMM},T}). \] (6)

Under the null hypothesis that the model used to estimate \( \delta \) is correctly specified, this statistic is asymptotically \( \chi^2 \)-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. Firstly, 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

---

3This statistic is also called the J-statistic in the literature.
Hansen test rewards models creating highly variable pricing errors across different assets (Hansen and Jagannathan, 1997; and 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. Secondly, 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 compatible. 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. 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. Zhou (1994) and Atonji and Segal (1996) also find that nonoptimal GMM procedures often outperform optimal GMM procedures. 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. Whether the Hansen test for linear factor pricing models, defined in (6), has desirable finite-sample properties crucially depends on how well the estimated variance matrix of pricing errors, $S_T$, can approximate the true variance matrix $S = \text{Var}\{\sqrt{T}w_T(\delta)\}$.

As an alternative to the Hansen test, Jagannathan and Wang (1996) develop a model specification test using the HJ-distance. To understand their method, define

$$SHJ_T = T \times HJ_T(\delta_{HJ,T})^2. \quad (7)$$

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

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

where $I_N$ is the $N \times N$ identity matrix, $D = \text{E}(R \tilde{y})$, $G = \text{E}(R \tilde{y} \tilde{y}')$, and $S^{1/2}$ and $G^{-1/2}$ are the upper-triangle matrices from the Cholesky decompositions of $S$ and $G^{-1}$, respectively. The matrix $\Phi$ can be consistently estimated by replacing $D$, $S$, and $G$ with $D_T$, $S_T$, and $G_T$. Let $\lambda_j$ $(j = 1, \ldots, N-K)$ denote the $(N-K)$ positive eigenvalues. With this notation, Jagannathan and Wang (1996) show that under the null hypothesis of correct model specification,

$$
SHJ_T \rightarrow u = \sum_{j=1}^{N-K} \lambda_j v_j,
$$

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 (1996) propose use of 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

$$M^{-1} \sum_{i=1}^{M} I(u_i \leq SHJ_T),$$

where $M$ is the number of draws from the $\chi^2$ distribution, and $I(\bullet)$ 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, we need to choose an appropriate size of $M$. Following Jagannathan and Wang (1996), we set $M$ to 5,000 in our simulations.5

---

4The matrix $\Phi$ 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.

5In unreported simulations, we have tried values of $M$ ranging from 1000 to 10,000. We found the standard deviation of the sample $p$-value around its asymptotic value was about 1.0% at
Although the model specification test based on the empirical p-value appears theoretically sound, the small-sample properties of the test has not been explored in the literature. In fact, there is a good reason to question the finite-sample performance of the HJ-distance p-value. Notice that the p-value depends on the eigenvalues from an estimate of the matrix $\Phi$ defined in (8). Estimation of the matrix $\Phi$ requires use of the matrix $S_T$, which is suspected to contribute to poor small-sample performances of the Hansen test. The HJ-distance itself is independent of the matrix $S_T$, but its empirical p-value is not. Accordingly, there is no guarantee that the specification test based on the HJ-distance outperforms the Hansen test in small samples.

3. Monte Carlo Setup

This section explains the design of our Monte Carlo experiments. To investigate the finite-sample properties of the HJ-distance estimate and its empirical p-value, we construct hypothetical returns using the following data generating process:

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

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 $t$, $\alpha$ is an overall intercept term, $F_{ji}$ ($j = 1, 2, 3$) are the common factors for period $t$, $\beta_{ji}$ are the corresponding factor betas for asset $i$, and $e_{it}$ is the idiosyncratic error with mean zero for asset $i$ for period $t$. Most empirical studies based on linear factor models use monthly time-series data. Accordingly, we can view the subscript "t" as indexing months.

Jagannathan and Wang (1996) use the HJ-distance method to analyze their own model (so-called Premium-Labor model) and the model of Fama and French (1992), both of which utilize three factors. Accordingly, our simulation exercises use three factors hoping to mitigate potential biases in comparing our simulation results and Jagannathan and Wang’s empirical results with actual data. In our simulations, we consider 15 different combinations of $N$ and $T$, with $N = 25, 50$ and $100$, and $T = 160, 330, 700, 1,500$ and $3,000$.

A three-factor linear model implies the following restriction:

---

sampling with 5,000 draws.
Note that the risk factors we use to generate the time-series of returns are not mean-adjusted as is commonly specified in an APT-style model. Instead, 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 factor prices corresponding to $F_{1t}$, $F_{2t}$ and $F_{3t}$, respectively. Clearly, our data generating process (11) 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, the value of the overall intercept term $\alpha$ is fixed at 1.0033. This number is chosen because the historical average of the risk free component for annual U.S. stock (net) returns roughly equals 4% (= 0.33% × 12 months). The three betas are random draws from a uniform distribution with the range between zero and two, so that the cross-sectional mean of each beta equals one. The three factors are random draws from a normal distribution with mean equal to 0.0022 and variance equal to $6.944 \times 10^{-5}$. 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% × 3 factors × 12 months). This number is consistent with the historical average of risk premiums in the U.S. stock market. The value $6.944 \times 10^{-5}$ 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 U.S.

We draw the idiosyncratic error terms $e_{it}$ from normal distributions with common mean zero but with different variances. The size of $\text{Var}(e_{it})$ determines the signal-to-noise ratio (SNR) which equals to the ratio between the two components of the return variation: one caused by the factor variations; and the other by the error variation. Different values for $\text{Var}(e_{it})$ are used to

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

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 (11) 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.

6 Note that the risk factors we use to generate the time-series of returns are not mean-adjusted as is commonly specified in an APT-style model. Instead, 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.

7If we follow the traditional definition of signal-to-noise ratio, the ratio for our data generating process should be defined as $\text{Var}(F_{1t} \beta_{1i} + F_{2t} \beta_{2i} + F_{3t} \beta_{3i})/\text{Var}(e_{it})$. However, since factor variances are the same in our simulations, we hereafter use the term "signal-to-noise ratio" to refer to...
generate return data with different SNRs.

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

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

where

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

for j = 1, 2 and 3. The term ($\delta_0 + \delta_1 F_{1t} + \delta_2 F_{2t} + \delta_3 F_{3t}$) in (13) is the SDF implied by our three-factor model, and is equivalent to our notation $Y_t$ in the previous section. Our Monte Carlo experiments exploit the moment conditions (13) to compute Hansen statistics and HJ-distances.

The risk-free rate and factor risk prices ($c_0$, $c_1$, $c_2$ and $c_3$, defined in equation (12)) can be estimated by the two-pass cross-sectional regression method developed by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973). This method requires two sequential regressions: First, each asset’s betas are estimated by time-series linear regression of the asset’s return on a set of common factors. Then, the risk free rate and factor risk prices are estimated by a cross-sectional regression of mean returns on the estimated betas. The $R^2$ from this second-stage regression can be used to evaluate the explanatory power of the model. This $R^2$ measure, which we call the two-pass $R^2$, is positively related with SNR: The more highly the common factors are correlated with returns, the higher is the two-pass $R^2$.

The HJ-distance can be used as a complementary tool to compare relative performances of competing linear factor models. A pitfall of the two-pass $R^2$ is that it overly favors models using factors highly correlated with returns. Misspecified linear factor models that can explain returns

$$\text{Var}(F_{jt})/\text{Var}(e_{jt}).$$
well may generate large pricing errors. Accordingly, use of both the $R^2$ and the HJ-distance would lead to more convincing model comparisons. This is so because the HJ-distance is more directly related with pricing errors. As a model comparison tool supplementary to the two-pass $R^2$, the HJ-distance would be even more desirable if it is robust to the degree of the correlation between returns and factors. Accordingly, we investigate the sensitivity of the HJ-distance to changes in SNR. To do so, we generate data with 10 different values for SNR (0.1, 0.25, 0.5, 1, 1.5, 5, 10, 25, 50 and 100). Thus, our simulations consider in total 150 combinations of $N$, $T$ and SNR. For each of these 150 cases, we compute the mean value of the HJ-distances from 1,000 simulations and count the percentage of times across 1000 simulations the HJ-distance p-values are less than commonly used asymptotic size levels (10%, 5% and 1%).

Finally, our simulations examine a degrees-of-freedom adjustment method suggested by Ferson and Foerster (1994). They investigate small-sample properties of the optimal GMM estimation and model testing procedures. The foundation of their simulation experiments is a latent-variables asset pricing model developed by Hansen and Hodrick (1983) and Gibbon and Ferson (1985), among others, which is of the form (12) except that the risk free rate $c_0$ and risk factor prices $c_j$ are linear functions of $L$ instrumental variables. Ferson and Foerster find that the Hansen test 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 method is equivalent to inflating the inverse of the optimal GMM weighting matrix (in our case, $S_T$ as defined in (4)) by multiplying it by an adjustment factor:

$$F = \frac{(N+L)T}{(N+L)T - Q}, \quad (15)$$

---

8Ferson and Foerster use NL moment conditions to estimate the latent variables asset pricing model.
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 (\( \delta_0, \delta_1, \delta_2 \) and \( \delta_3 \) in (13)) equals four (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 in (4)).

4. Simulation Results

4.1. Preliminaries

We begin by considering the sensitivity of the HJ-distance (\( HJ_T \) defined in (2)) to sample size. For a given combination of \( N \) and \( T \), 1,000 data sets are generated for each of the 10 different SNRs. Then, 10 different averages of HJ-distances (two-pass \( R^2 \)) are computed using each of these 1,000 data sets. Table 1 reports the average of these 10 different average HJ-distances (two-pass \( R^2 \)). It appears that the HJ distance is sensitive to the number of time-series observations. Since we use the data generated by a correctly specified model, we can expect that the HJ-distance will be closer to zero as the number of time-series observations (\( T \)) increases. Table 1 supports this conjecture. The table also shows that the HJ-distance converges somewhat slowly as the number of time-series observations increases. The HJ-distance falls by less than 50\% even if the number of time-series observations is doubled. For example, as the number of time-series observations increases by 101\% from 160 to 330, the average HJ-distance fell only by 34\% from 0.393 to 0.260 when 25 assets are used. This result indicates that the model specification test based on the HJ-distance p-value may show poor performances in samples with a small number of time-series observations.
When more assets are cross-sectionally available, it is more likely that any particular portfolio of assets will have a larger (finite-sample) pricing error for a given number of time-series observations. These finite sample features of the HJ-distance are evident in Table 1. The level of the estimated HJ-distance is highly dependent on the number of assets. For example, as the number of assets increases from 25 to 100, the average HJ-distance increases from 0.393 to 1.272. This result suggests that the HJ-distance tends to be larger when more assets are used for the same number of time-series observations.

As discussed in Section 3, it is desirable that the HJ-distance be insensitive to changes in SNR. Varying SNR for the factors relative to the idiosyncratic errors should allow us to examine this feature. Not surprisingly, variation in SNR leads to variation in the two-pass $R^2$. As summarized in Table 1, the standard deviations of the average values of the $R^2$ across the SNR range for all the sample size couplets is substantial. For example, the standard deviation of these average $R^2$ range from a low of 0.254 (for 25 assets and 3,000 time-series observations) to a high of 0.329 (for 100 assets and 160 time-series observations). Figure 1 shows a graph of the average $R^2$ across different SNRs for some selected sample size couplets comparable to those used in recent studies. At the lowest SNR, 0.10, the $R^2$ is 0.118 for 100 assets and 330 time-series observations and 0.191 for 25 assets and 160 time-series observations. At the highest SNR, 100, the $R^2$ is 0.979 for 100 assets and 330 time-series observations and 0.980 for 25 assets and 160 time-series observations. While there is some variation in $R^2$ across sample size couplets for each SNR, the average $R^2$ monotonously increases with SNR.

Contrary to the variation in the $R^2$ across SNRs, the standard deviations of the average HJ-distances across the range of SNRs are negligible. As can be seen in Table 1, the standard deviation of the average HJ-distances across SNRs ranging from 0.10 to 100 for 160 time-series observations and 25 assets is less than 1% of the average across these ratios. The other sample size couplets have similar or smaller relative variation in the HJ-distance across this range of SNRs. As an example, the average HJ-distances for each SNR are for 100 assets/330 time-series observations and 25 assets/160 time-series observations are also shown in Figure 1 for direct comparison to the two-pass $R^2$. Clearly, the HJ-distance appears insensitive to the degree of the correlation between returns and factors.

We now turn to finite-sample properties of the model specification test based on the HJ-
distance p-value. Table 2 reports the average model rejection rates from data with 10 different SNRs and their standard errors. The rejection rates are computed for three commonly used test sizes: 1%, 5% and 10%. Our simulation results indicate that the empirical p-value of the HJ-distance rejects the null hypothesis of correct specification many times more than the specified size across the usual significance levels, especially when 100 assets are used. For example, with 330 months of returns and 100 assets, the null hypothesis is rejected at the 1% significance level 58% of the time, at the 5% level 76%, at the 10% level 84% of the time. Decreasing the number of assets from 100 to 25 substantially weakens this size distortion. But even with 25 assets and 330 time-series observations, the null hypothesis is rejected about twice as often as indicated by the significance levels (2.5% of the time for the 1% level, 9.4% for the 5% level and 17.0% for the 10% level).

In Table 2, the rejection rate by the HJ-distance p-value appears insensitive to changes in SNR. Note that the standard deviations of the average rejection rates over different SNRs are small in both absolute and relative terms. The largest standard deviation is 2.0% for the 10% significance level with 25 assets and 160 time-series observations. This is about 12% of the average rejection rate for this sample size couplet and is one of the larger relative variations among all sample size couplets examined. Casual inspection of the rejection rates for this couplet indicate that the rejection rate at the 10% level rose somewhat but not monotonously as SNR increases from 0.10 (21.7%) to 100 (27.6%). This increasing relationship seems to be present at other significance levels for the 25 asset/160 time-series observation trials but does not appear to be systematic across the other sample size couplets.

4.2. Results with Degrees-of-Freedom Adjustment

While the validity of the HJ-distance p-value is based on asymptotic theory, our simulation results reported in Table 2 indicate that it suffers from finite-sample bias by systematically rejecting the null hypothesis too frequently. This finding motivates the use of the degrees-of-freedom adjustment proposed by Ferson and Foerster (1994) discussed in Section 3. Table 3

---

9This is the sample size couplet used by Jagannathan and Wang (1996).
reports the simulation results with this adjustment. It appears that the adjustment corrects some but not all of the finite-sample biases. The rejection rate for the 10% level falls from 24.8% to 17.6%, from 15.7% to 10.5% for the 5% level and from 5.5% to 3.0% for the 1% level when 25 assets are used with 160 time-series observations. With an increase to 330 time-series observations, the bias shrinks considerably. The 10% level is rejected 13.5% of the time, the 5% level 7.4%, and the 1% level 1.8%. Note, though, that increasing the number of assets to 100 substantially increases the model rejection rate even with this adjustment. While the unadjusted p-value rejects the null hypothesis at the 10% level 84% of the time, the adjusted p-value still rejects the null hypothesis at this level 53% of the time. Similarly severe over-rejection rates occur for the 5% (41.6%) and 1% levels (22.8%). Accordingly, it appears that effectiveness of the adjustment method falls as the number of assets increases.

4.3. The HJ-Distance P-Value Versus the Hansen Test

Jagannathan and Wang (1996) conjecture that the specification test based on the HJ-distance may have better finite-sample properties than the Hansen test because estimating the HJ-distance does not require estimating the variance matrix of pricing errors ($S_T$ defined in (4)). However, as we discuss in Section 2, 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 small samples for the same reason the Hansen test is biased. Table 4-A and 4-B explore this possibility.

The optimal GMM procedure discussed in Section 2 is often called two-step GMM, because it requires two stages of estimation: In the first stage, the unknown parameters are estimated using an arbitrary positive definite weighting matrix (usually, an identity matrix). The resulting estimates are used to estimate the optimal GMM weighting matrix (in our case, $S_T^{-1}$ as defined in (4)). Then, in the second stage, this estimated optimal weighting matrix is used to obtain the efficient (minimum-variance) parameter estimates. This procedure can be repeated by updating the weighting matrix until the parameter estimates converge. This alternative procedure is called iterative GMM. Ferson and Foerster (1994) report that iterative GMM outperforms two-step GMM. Accordingly, we consider both of these two procedures.

In Table 4-A and 4-B, the columns of HJ, J and JI report the rejection rates by the HJ-distance p-value and the two Hansen tests based on the two-step and the iterative GMM,
respectively, while the columns of HJ, JF and JIF report the rejection rates by the same tests but with the Ferson-Foerster adjustment. In unreported experiments, we conducted simulations with 10 different SNRs similarly to those reported in Tables 1-3. However, the simulation results were similar across the various SNRs. Thus, we only report the results obtained with SNR of 1.0. As seen in Figure 1, this level of SNR is roughly comparable with a two-pass $R^2$ of 50% for the sample couplets commonly used in studies of asset prices and monthly returns. This range for $R^2$ is also roughly comparable to those reported by Jagannathan and Wang (1996) for their analysis of their Premium-Labor model, Fama and French’s (1993) three-factor model, and Chen, Roll and Ross’ (1986) multi-factor model.

Tables 4-A and 4-B report our results with 25 and 100 assets. Noticeably, both the adjusted and unadjusted HJ-distance p-values have higher rejection rates than all four of the Hansen tests.\footnote{In unreported experiments, we used 5,000 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.} The unadjusted Hansen tests, based on either two-step or iterative GMM, are generally over-sized, especially when the number of time-series observations is small or the number of assets is large. The rejection rates for the unadjusted Hansen tests, however, are closer to asymptotic size across all observation levels than those for both the unadjusted and adjusted HJ-distance p-values. There is little evidence that the model specification tests based on the HJ-distance p-value is better sized than the Hansen tests. These results are consistent with Jagannathan, Kubota and Takehara (1998). Using 25 assets with 148 monthly observations for Japanese stock market data, they report consistently lower p-values for the HJ-distance than for the Hansen test.\footnote{For eight models using the same time-series/cross-section of Japanese stock market returns, Jagannathan, Kubota and Takehara (1998) report the following p-values using the HJ-distance (the Hansen test): 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) find strong evidence that Hansen tests are better sized when iterative instead of two-step GMM is used. In contrast, our simulation results indicate much weaker evidence for the finite-sample superiority of iterative GMM. Except for the cases with 160 time-series observations, the Hansen tests based on the two different optimal GMM
procedures produce quite compatible model rejection rates. Even with 160 time-series observations, the two Hansen tests have almost identical rejection rates when the Ferson-Foerster adjustment is used. This difference from Ferson and Foerster’s results is likely to be our using a linear model while they use a highly nonlinear model. It appears that the iterative GMM is less effective in estimating linear asset pricing models.

The Ferson-Foerster adjustment does a remarkable job of moving the sizes of the Hansen tests close to asymptotic levels for 25 assets with as few as 160 time-series observations. Unfortunately, it overcorrects with 100 assets at 160 and 330 observations. Ferson and Foerster (1994) also find that their method is less effective when the number of assets is too large. Accordingly, researchers are advised to avoid using this particular adjustment for this range of asset/time-series observation pairings for similar analysis with actual data.

4.4. Finite-Sample Biases in GMM Parameter Estimates

Altonji and Segal (1996) show that optimal GMM estimates could be more biased than nonoptimal GMM estimates in small samples. Their result suggest that the HJ-distance based GMM may generate more reliable SDF parameter estimates than the two-step or iterative GMM. We here examine this possibility.

Table 5 compares finite-sample biases in the SDF parameters estimated by the HJ-distance based GMM ($\hat{\delta}_{HJ,T}$ defined in (3)) and the optimal two-step GMM ($\hat{\delta}_{OGMM,T}$ defined in (5)). The biases in the iterative GMM estimates are also considered. 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 "$\hat{\delta}_o$" report results for the estimates of the constant term of SDF ($\delta_0$ defined in (12)), while the boxes labeled as "$\hat{\delta}_x$" report the average relative biases and RMSE of the estimates for the risk factor parameters in SDF ($\delta_1$, $\delta_2$ and $\delta_3$ defined in (12)). The columns of HJ, J and JI report results for the HJ-distance based, two-step and iterative GMM estimates, respectively.

The sizes of the biases in SDF parameter estimates appear to be similar for the HJ-distance based and two-step GMM. The biases in the estimates obtained by minimizing the HJ-distance are sometimes smaller than the biases in the two-step estimates. But the differences are extremely small. 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.

Table 5 indicates that the iterative GMM parameter estimates are not superior to the two-step estimates. The bias using iterative GMM is markedly higher with 100 assets than with 25 assets. For example, the bias in the constant term estimate is 10.9% with 100 assets, but only 1.9% with 25 assets. Similarly, the bias in the risk factor parameter estimates average 19.2% with 100 assets, but only 1.8% with 25 assets. With 100 assets and 160 time-series observations, the iterative GMM estimates are biased more than 10% relative to true parameter values, more than four times as much as the two-step estimates are. With 100 assets, the biases in the iterative GMM estimates shrink rapidly as the number of time-series observations increases. This result indicates that iterative GMM is less reliable for data with a large number of assets and a short time interval.

The relative RMSE behaves similarly to the relative bias across the different estimation methods and sample pairings. Across all methods, the RMSE monotonously decreases with the number of time-series observations. Similarly to the relative bias, the relative RMSE is markedly higher with 100 assets and 160 observations using iterative GMM than with the other methods and with 25 assets. A notable observation from Table 5 is that unlike with the relative bias, the relative RMSE is substantially larger for the risk factor parameter estimates than for the constant term estimates across all the cases we consider. Kan and Zhou (1999) find similar results.

In summary, there are two major findings from Table 5. First, we find little evidence that the HJ-distance based GMM produces significantly more reliable SDF parameter estimates than two-step GMM. Both GMM methods produce only small biases in parameter estimates. Second, use of iterative GMM could result in misleading statistical inferences when too many assets and too few time-series observations are used. However, we note that these results are likely to be model specific. Our findings from Table 5 are not consistent with those from Altonji and Segal (1996) and Ferson and Foerster (1994) which use different model specifications. Extensive simulations
based on a variety of different asset pricing models would provide more concrete results.

4.5. Diagnosis of Size Overstatement in Finite Samples

As discussed in Section 2, previous studies of GMM indicate that Hansen tests 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 GMM estimation of linear factor models.

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 (4). There are two possible reasons why $S_T$ may be a poor estimate of the 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 the $N(N+1)/2$ 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 ($T$).

In order to explore the first possibility, we substitute the exact values of the SDF parameters for 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, we report results only for the HJ-distance and the Hansen test based on two-step GMM. Tables 6-A and 6-B reports our simulation results with this adjustment. Comparing the results from Tables 4-A and 4-B with those from Tables 6-A and 6-B, we can clearly see that use of true SDF parameter value does not improve the finite-sample performances of the HJ-distance p-value and the Hansen test. Rather, for many cases, rejection rates for both the HJ-distance p-value and the Hansen test with exact SDF parameters are slightly farther from asymptotic levels than rejection rates with estimated SDF parameters. For example, Table 4-A shows that the null hypothesis of correct specification using the HJ-distance p-value with estimated SDF parameters is rejected 5.5% of the time at the 1% level with 25 assets and 160 time-series observations. In contrast, Table 6-A shows that the HJ-distance p-value with exact SDF parameters rejects the null hypothesis 7.3% of the time.
Similar but small increases in rejection rates are often observed across different test methods and different finite sample pairings examined. These results indicate 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 test.

Turning to Tables 7-A and 7-B, we find direct evidence that the finite-sample overstatement of size is caused by the poorly estimated variance matrix \( S_T \). In Table 7-A and 7-B, we use the "exact" variance matrix of pricing errors to compute the HJ-distance p-value and the Hansen test statistic. It is computationally complicated to obtain the exact value for each element in the variance matrix. Instead, 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. We do so hoping that the variance matrix estimated with 10,000 time-series observations is close enough to the exact variance matrix.

As can be seen in Tables 7-A and 7-B, the rejection rates are much closer to asymptotic levels with the large sample estimate of the variance matrix for cases with both 25 and 100 assets. In fact, using the exact variance matrix tends to drive rejection rates down below asymptotic levels. For example, using the large sample estimate of the variance matrix resulted in rejection rates below the asymptotic 5% and 10% levels for both the HJ-distance and the Hansen statistic with 25 assets and below all three rejection rates using 100 assets. However, in most cases, the understatement of size is not severe unless the Ferson-Foerster adjustment is used. 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.

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) show that the p-value for the HJ-distance can be empirically estimated using a weighted average of random draws from a \( \chi^2 \) distribution. In this paper, we explore the finite-sample properties of the HJ-distance and its
empirical p-value by conducting some limited Monte Carlo experiments based on a linear three-factor model.

Our simulation results indicate that the HJ-distance does not reward high correlations between returns and factors. Given that the $R^2$ from the two-pass regression tends to overly favor models using factors highly correlated with returns regardless of pricing errors, use of the HJ-distance as a supplementary model comparison tool appears promising. Unfortunately, the empirical p-value for the HJ-distance is seriously biased downward, especially when the number of time-series observations is small and the number of assets is large. Our simulation results also provide little evidence that GMM based on the HJ-distance approach produce more reliable SDF parameter estimates than usual optimal GMM does.

The empirical p-value of the HJ-distance rejects correct models too often for sample sizes commonly used in studies of the cross-sectional return studies (100 assets and 330 observations through 25 assets and 160 observations). An adjustment for degrees of freedom suggested by Ferson and Foerster (1994) corrects some of this bias but does not eliminate it. The main reason for the severe size distortion in the HJ-distance p-value is that it requires the estimation of the variance of pricing errors, but this matrix is poorly estimated in finite samples. Use of the exact variance matrix can remove most of the size distortion. The conventional GMM $\chi^2$ test also over-rejects correct models but outperforms the test based on the HJ-distance p-value.

Consistent with Ferson and Foerster (1994), we also find that the GMM $\chi^2$ test adjusted by their method retain desirable size properties in small samples unless too many assets (i.e., 100 assets) are used. However, differently from their prediction, we do not find that iterative GMM is more effective than two-step GMM in estimating and testing linear asset pricing models.

This paper suggests some future research agenda. First, the Monte Carlo design used in this paper is undoubtedly limited. In performing simulations, we use hypothetical returns generated from a known set of common return factors and normally distributed idiosyncratic errors. More general results could be obtained using bootstrapped data from actual economic data, or returns generated from non-normal distributions. It would also be interesting to see if our simulation results could be generalized to more sophisticated asset pricing models. Second, the degrees of freedom adjustment by Ferson and Foerster (1994) does not appear to be sufficient to correct for the finite-sample bias in the HJ-distance p-value. It would be useful to consider alternative
adjustment methods that can lead to estimation of the p-value with less bias. Finally, it would be interesting to investigate the power properties of the specification test based on the HJ-distance p-value. Given the severe size distortion in the empirical p-value for the HJ-distance reported in this paper, we were unable to find a reasonable way to investigate the power of the p-value. Once an appropriate way to fix this size distortion is developed, a proper power comparison between the HJ-distance p-value and the Hansen test would be possible.
References


