



ELSEVIER

Journal of Financial Economics 36 (1994) 29–55

JOURNAL OF
Financial
ECONOMICS

Finite sample properties of the Generalized Method of Moments in tests of conditional asset pricing models*

Wayne E. Ferson ^{*,a}, Stephen R. Foerster^b

^a School of Business Administration, University of Washington, Seattle, WA 98195, USA

^b Western Business School, University of Western Ontario, London, Ont. N6A 3K7, Canada

(Received June 1991; final version received November, 1993)

Abstract

We develop evidence on the finite sample properties of the Generalized Method of Moments (GMM) in an asset pricing context. The models imply nonlinear, cross-equation restrictions on predictive regressions for security returns. We find that a two-stage GMM approach produces goodness-of-fit statistics that reject the restrictions too often. An iterated GMM approach has superior finite sample properties. The coefficient estimates are approximately unbiased in simpler models, but their asymptotic standard errors are understated. Simple adjustments for the standard errors are partially successful in correcting the bias. In more complex models the coefficients and their standard errors can be highly unreliable. The power of the tests to reject a single-premium model is higher against a two-premium, fixed-beta alternative than against a conditional Capital Asset Pricing Model with time-varying betas.

Key words: Asset pricing; Finite sample properties; Generalized method of moments; Latent variables

JEL classification: C15; C31; G12

* Corresponding author.

This paper was presented at the Sixth World Congress of the Econometric Society, the American Finance Association, the Northern Finance Association, and the Conference on Financial Economics and Accounting. We thank Dennis Aigner, Ed Allen, Jim Bodurtha, Tim Bollerslev, Philip Braun, John Campbell, John Cochrane, Bjorn Flesaker, Eugene Fama, Ron Gallant, Eric Ghysels, Campbell Harvey, John Huizinga, Ravi Jagannathan, Joe Kairys, Aditya Kaul, Narayana Kocherlakota, Robert Korajczyk, Yum Keung Kwan, Bruce Lehmann, Anthony Lynch, Jiangping Mei, Angelo Melino, Choon-Geol Moon, Peter Rossi, G. William Schwert (the editor), Andy Siegel, Stephen D. Smith, George Tauchen, Harry Turtle, Simon Wheatley, an anonymous referee, and participants in workshops at University of Alberta, the University of Chicago, Emory University, the London Business School, the University of Toronto, and Washington University for helpful discussions and comments. Ferson acknowledges financial assistance from the Center for Research in Security Prices at the University of Chicago and from the Pigott-Paccar professorship at the University of Washington. Foerster acknowledges financial assistance from the Plan for Excellence and the Social Sciences and Humanities Research Council.

1. Introduction

Because of its simplicity, flexibility, and generality Hansen's (1982) Generalized Method of Moments (GMM) has become an important technique for estimating and testing financial asset pricing models. While the asymptotic properties of the GMM are well understood, evidence on its finite sample properties is sparse. Existing studies include Tauchen (1986), Kocherlakota (1990), and Mao (1991), who examine nonlinear consumption-based asset pricing models; Flesaker (1993), who examines nonlinear term structure models; MacKinlay and Richardson (1991), who examine zero restrictions on 'market model' regression coefficients, as implied by the Capital Asset Pricing Model (CAPM); and Nelson and Startz (1990), who examine a simple linear regression model.

This paper develops evidence on the finite sample properties for models that imply nonlinear, cross-equation restrictions on regressions that predict security returns. The cross-equation restrictions can be motivated by *latent variables* asset pricing models (Hansen and Hodrick, 1983; Gibbons and Ferson, 1985). They also arise in versions of the consumption-based Capital Asset Pricing Model (Hansen and Singleton, 1983; Ferson, 1983), and are used in a number of other economic applications (see Aigner et al., 1984; Chamberlain and Goldberg, 1990).

Gibbons and Ferson (1985) do not reject a single-premium, latent-variable model for stock returns. However, subsequent studies, based on large sample theory, reject a single latent variable. Since these rejections are based on the asymptotic distribution of the test statistics, it is important to verify that rejections of the models are not the result of finite sample bias.

We focus on the size and power of the GMM test statistics, the sampling properties of the coefficient estimators, their standard errors, and *t*-ratios. We examine both two-stage and iterated GMM estimators. The two procedures have the same asymptotic properties, and studies typically employ only one of the two. We find that in larger models, the two-stage GMM tests reject the null hypothesis too often, while an iterated GMM test statistic conforms more closely to the asymptotic distribution. We find that the GMM coefficient estimators are approximately unbiased in the simpler models. However, the standard errors for the coefficients are understated, using the asymptotic formula from Hansen (1982). The understatement is more severe in systems with large numbers of assets and small sample sizes. In more complex models, the coefficient estimates and the standard errors can be biased by large amounts. We investigate simple adjustments to reduce the finite sample bias.

We examine the power of the tests for a single latent variable against two alternative models. Our first alternative is a two-latent-variables model (i.e., two time-varying risk premiums with fixed betas). Our second alternative is a conditional Capital Asset Pricing Model (CAPM) with time-varying market betas. We find that the power of the tests against the CAPM with time-varying betas is low and the power against the two-premium alternative is higher.

The paper is organized as follows: Section 2 reviews latent variables models and the cross-equation restrictions. Section 3 describes our methodology. Section 4 describes the data, and Section 5 presents the simulation results. Section 6 concludes.

2. Latent variables models of expected returns

We regress asset returns R_{it} over time on a vector of predetermined variables, \underline{Z}_{t-1} . We use the projections $\delta_i \underline{Z}_{t-1}$ to model the conditional expected returns. We study tests of cross-equation restrictions on the coefficients δ_i to detect reduced dimensionality across assets in the time-varying expected returns. The restrictions can be motivated by a class of beta pricing models, including conditional versions of the CAPM, the Arbitrage Pricing Theory (APT), and the intertemporal asset pricing models of Merton (1973), Long (1974), Breeden (1979), and Cox, Ingersoll, and Ross (1985). Consider the model in (1):

$$E(R_{it} | \underline{Z}_{t-1}) = \lambda_0(\underline{Z}_{t-1}) + \sum_{j=1}^K b_{ij} \lambda_j(\underline{Z}_{t-1}), \quad i = 0, \dots, N, \quad t = 1, \dots, T, \quad (1)$$

where $\lambda_j(\underline{Z}_{t-1})$ is a market-wide expected risk premium and b_{i1}, \dots, b_{iK} are the conditional betas of asset i relative to the K underlying, unobserved risk factors. \underline{Z}_{t-1} is the vector of instruments for the information available when prices are set at time $t - 1$, and $E(R_{it} | \underline{Z}_{t-1})$ is the expected return conditional on this information. In general, the b_{ij} can depend on \underline{Z}_{t-1} . The latent variables models specialize Eq. (1) by assuming that the betas are fixed parameters over time and that the $\lambda_j(\underline{Z}_{t-1})$ are the latent variables.¹

Define the $T \times N$ excess return matrix r , with typical element $r_{it} = R_{it} - R_{0t}$, $i = 1, \dots, N$ and $t = 1, \dots, T$, where R_{0t} is the return of an arbitrarily chosen zeroth asset. Define the $T \times K$ matrix of the $\lambda_j(\underline{Z}_{t-1})$ as $\lambda(\underline{Z})$, where K is the number of latent variables. Define the $T \times L$ matrix of \underline{Z}_{t-1} 's as \underline{Z} . Assuming fixed betas, Eq. (1) implies the following expression for the expected excess

¹ The first studies to interpret such asset pricing models were Hansen and Hodrick (1983) and Gibbons and Ferson (1985). Foerster (1987), Ferson, Foerster, and Keim (1993), and Campbell and Hamao (1992) examine stock returns using similar models. Stambaugh (1988) and Chang and Huang (1990) examine bond returns. Campbell (1987) and Ferson (1989, 1990) study both stock and bond returns. Forward currency premiums are examined by Hansen and Hodrick (1983), Hodrick and Srivastava (1984), Cumby (1987, 1988), Campbell and Clarida (1987), and Jorion and Giovannini (1987). Bessembinder and Chan (1992) study futures. International equity returns are examined by Campbell and Hamao (1992), Harvey (1991), and Chang, Prinegar, and Ravichandran (1992). See Wheatley (1989) for a critique of latent variable models of asset pricing and Ferson (1993) for a recent review.

returns:

$$E(r|Z) = \lambda(Z)\beta, \quad (2)$$

where β is the $K \times N$ matrix of conditional betas for the excess returns ($\beta_{ij} = b_{ij} - b_{0j}$) and $E(r|Z)$ is the $T \times N$ matrix of the $E(r_{it}|\underline{Z}_{t-1})$, for $t = 1, \dots, T$ and $i = 1, \dots, N$.

Partition the excess returns as $r = (r_1 \ r_2)$, where r_1 is a $T \times K$ matrix of reference assets and r_2 is a $T \times (N - K)$ matrix of test assets. Partition the matrix of betas conformably as $\beta = (\beta_1 \ \beta_2)$. The reference assets are chosen so that the $K \times K$ matrix β_1 is nonsingular. From the partitioned Eq. (2), solve for the risk premiums and substitute back $\lambda(Z) = E(r_1|Z)\beta_1^{-1}$, to obtain the following restrictions:

$$E(r_2|Z) = E(r_1|Z)C, \quad (3)$$

where C is a $K \times (N - K)$ matrix equal to $\beta_1^{-1}\beta_2$. Following Gibbons and Ferson (1985), assume that the conditional expected excess returns of the reference assets are linear regression functions of the instruments. Eq. (3) restricts the linear regressions as follows:

$$r_1 = Z\delta_1 + u_1, \quad r_2 = Z\delta_1 C + u_2, \quad (4)$$

where Z includes a constant term, δ_1 is an $L \times K$ matrix of regression coefficients, and $E(u_1|Z) = E(u_2|Z) = 0$. The latent variable model in Eq. (4) implies that if there are K common factors that describe expected excess returns over time, then linear combinations of the regression functions that predict the excess returns of K reference assets are sufficient to capture the predictable variation in all returns.

3. Methodology

3.1. GMM estimation

Define the $T \times N$ matrix of error terms from eq. (4) as $u = (u_1 \ u_2)$. The model implies $E(u|Z) = 0$, therefore $E(u'Z) = 0$. Define an $N \times L$ matrix of sample orthogonality conditions: $G_T = (u'Z/T)$. Partition G_T into rows of length L and stack these into a column vector, g_T , with a length equal to the number of orthogonality conditions, NL . Obtain the GMM estimators by searching for the parameter vector θ , consisting of the elements of δ and C , that minimizes a quadratic form $g_T'Wg_T$. The $NL \times NL$ weighting matrix W is the inverse of a consistent estimate of the covariance matrix of the orthogonality conditions. We use the sample weighting matrix described by Hansen (1982), $W = [(1/T)\sum_t (u_t u_t') \otimes (Z_{t-1}' Z_{t-1})]^{-1}$, where \otimes denotes the Kronecker product.

Hansen showed that at the minimizing parameter vector, $Tg'_T Wg_T$ is asymptotically chi-square distributed. Its degrees of freedom are equal to the difference between the number of orthogonality conditions and the number of parameters: $NL - [KL + (N - K)K] = (N - K)(L - K)$. A test of the model requires L information variables, where $L > K$, and N assets, where $N > K$. The GMM estimator's asymptotic variance matrix is $[T(\partial g/\partial \theta)' W(\partial g/\partial \theta)]^{-1}$.

Hansen and Singleton (1982) describe a two-step approach for implementing the GMM. In the first step, they substitute the identity matrix for W to obtain initial estimates of the parameters, and then use these parameters to form an estimate of W . They use the estimate of W in the quadratic form, $g'_T Wg_T$, to obtain second-stage estimates of the parameters, which they use to form a second-stage estimate of the weighting matrix and the quadratic form. We call this a *Two-Stage GMM* approach. In practice, it may be desirable to iterate, repeatedly updating the weighting matrix until the procedure converges. We call this approach *Iterated GMM*.

We find that the asymptotic standard errors are understated in finite samples, and we therefore examine adjustment factors for the standard errors. These are analogous to the usual bias adjustment for the maximum likelihood estimator of a covariance matrix (e.g., Hinkley, 1977). The *traditional adjustment* is to multiply the asymptotic variance by $[T/(T - P)]$, where T is the number of time series observations and $P = KL + (N - K)K$ is the number of model parameters. We also evaluate an *alternative adjustment*, which is to multiply the asymptotic variance by $[(N + L)T/((N + L)T - Q)]$, where $Q = P + [(NL)^2 + NL]/2$. The alternative adjustment accounts for the number of time series observations provided by the instruments plus the assets, and for the number of model parameters plus the number of elements in the weighting matrix.

3.2. The simulations

This paper's approach complements that of Tauchen (1986), Cecchetti, Lam, and Mark (1990), Kandel and Stambaugh (1990), Gallant and Tauchen (1989), and others who simulate model economies. Those studies use discrete state processes to approximate the forcing equations, and they calibrate the models by matching selected moments of the data. As the GMM is likely to be sensitive to moments in the data not matched by the artificial economies, we use a more direct approach, resampling the data in a manner similar to the bootstrap methods of Efron (1982). We make the artificial samples satisfy a given model by restricting the particular moments that are the focus of the asset pricing hypothesis. The procedure does not require us to completely specify a model economy for each hypothesis, and it attempts to retain many of the statistical properties of the original data. Our approach, which is described in more detail in the appendix, can also be used to examine finite sample issues in other contexts.

3.3. A single-premium economy

The following regression system describes our first economy, a conditional CAPM with fixed betas:

$$r_{it} = \beta_{im} E(r_{mi} | \underline{Z}_{t-1}) + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (5)$$

$$E(e_{it} | \underline{Z}_{t-1}) = 0,$$

$$E(r_{mi} | \underline{Z}_{t-1}) = \delta'_m \underline{Z}_{t-1}, \quad t = 1, \dots, T,$$

where r_{it} is the excess return of an asset, r_{mt} is the excess return of the market portfolio, and β_{im} is the market beta coefficient of asset i . The instrument set is \underline{Z}_{t-1} and δ_m is a vector of coefficients. The restricted regression system, Eq. (4), specializes when $K = 1$ as follows:

$$r_{1t} = \delta'_1 \underline{Z}_{t-1} + u_{1t}, \quad (6)$$

$$r_{jt} = C_{1j} (\delta'_1 \underline{Z}_{t-1}) + u_{jt}, \quad t = 1, \dots, T, \quad j = 2, \dots, N,$$

where r_1 is the reference asset excess return. The test asset excess returns are r_2, \dots, r_N , and C_{12}, \dots, C_{1N} are the test assets' 'relative betas'. The CAPM implies that $C_{1j} = \beta_{jm} / \beta_{1m}$, $j = 2, \dots, N$. The return of the Center for Research in Security Prices (CRSP) value-weighted common stock index, in excess of a one-month Treasury bill return, is our proxy for r_{mt} . While it is used to generate the data for our simulations, we assume that the econometrician does not observe the market index. Therefore, it is not included in the regression system (6). We provide a more detailed description in the appendix.

In some of our simulations we reuse the \underline{Z}_{t-1} , $t = 1, \dots, T$, from the actual data. This allows the artificial data to retain both the autocorrelation and the cross-correlation properties of the instruments. For most of the experiments, we resample at random with replacement to generate the vector of the error terms $\{u_{jt}\}$, $j = 1, \dots, N$. This preserves the covariance structure across the assets, but breaks the link between the error terms and \underline{Z}_{t-1} so that the artificial data will satisfy the condition $E(u_{jt} | \underline{Z}_{t-1}) = 0$. As a consequence, the artificial data will not display conditional heteroskedasticity of the residual.

Of course, the GMM estimators do not assume homoskedasticity. One reason for using the GMM in financial models is its generality, and in particular its ability to handle conditional heteroskedasticity. We therefore examine several refinements of the resampling scheme. In one we generate artificial data that display conditional heteroskedasticity, which we describe in the appendix. We also examine the sensitivity to conditioning on a particular sample path of the instruments by modelling $\{r_{mt} \underline{Z}_{t-1}\}$ as a first-order vector autoregression (VAR). We resample from the vector of residuals, using the unconditional means as starting values, to generate a different series of the $\{\underline{Z}_{t-1}\}$ for each sample of the artificial data.

4. The data

We design our experiments to be representative of the many latent-variable studies that have used the GMM. We conduct simulations where the number of observations T ranges from 60 to 720, the number of assets N ranges from 3 to 14, and the number of instruments L is either 3 or 8. Since the single-premium model has received the most attention in the literature, we concentrate on this case, although we also examine models with $K = 2$ latent variables.

The assets are the monthly returns on portfolios of common stocks and bonds, measured in excess of the one-month U.S. Treasury bill rate. The data are provided by CRSP. We use ten value-weighted New York Stock Exchange common stock 'size' portfolios [grouped on market equity capitalization at the beginning of each year] and twelve value-weighted industry portfolios [grouped by two-digit SIC codes, following Breeden, Gibbons, and Litzenberger (1989)]. We also include one long-term corporate and one government bond portfolio. These are Ibbotson Associates data provided by CRSP.

The conditioning information available at time $t - 1$, denoted by \underline{Z}_{t-1} , includes a constant, the Treasury bill return for month t , and the lagged CRSP value-weighted market return. This *small instrument set* contains the minimal number of instruments to obtain overidentification in a two-latent-variable model. We are interested in the sensitivity of the finite sample properties to a strategy of using a larger set of instruments. The *large instrument set* includes the small instrument set plus these additional variables: the yield spread of a three-month over a one-month Treasury bill, the spread between the yields-to-maturity of AAA-rated corporate bonds and the three-month bill, the spread between the yields-to-maturity of BAA corporate bonds and the Composite of corporate bond yields, the annual dividend yield of the CRSP value-weighted common stock index, and a dummy variable for the month of January.

5. Simulation results

5.1. Evaluating the simulation methodology

The simulation technique may produce unreliable finite sample distributions, especially when the number of assets, N , is relatively large and the number of observations, T , is small. Therefore, we conduct an experiment to evaluate the methodology.

Choosing $N = 12$ and $T = 60$, we generate 5,000 samples of artificial data that satisfy the single-latent-variable model. We use the small instrument set. We use a Monte Carlo approach similar to that described in the appendix, except that a normal random number generator is used for the error terms. These 5,000 samples determine an idealized 'true' sampling distribution for the test statistic.

Table 1

Experiments to assess the accuracy of the bootstrap procedure

Results of bootstrapping when the finite sample distribution of the test statistic is known. The idealized 'true' sampling distribution is based on 5,000 simulation trials. The data for each trial satisfy the single-latent-variable model. The sample size is $T = 60$, the number of assets is $N = 12$, and the number of instruments is $L = 3$. The test statistic is the two-stage or iterated GMM goodness-of-fit statistic for a single-latent-variable model. Five of the idealized data samples are chosen at random, and each one is used (in experiments 1 through 5) as if it were the actual sample data, to calibrate a bootstrap experiment with 1,000 trials.

	Critical values are set to the following right-tail areas of chi-square distribution						Summary statistics of the goodness-of-fit tests		
	0.500	0.250	0.100	0.050	0.025	0.010	Mean	Median	Std. dev.
	Fraction exceeding critical values								
<i>Two-stage results</i>									
Idealized 'true'	0.836	0.560	0.263	0.128	0.058	0.015	27.169	26.992	5.853
Bootstrap simulation ^a									
Experiment 1	0.809	0.535	0.279	0.138	0.062	0.014	27.023	26.716	6.145
Experiment 2	0.794	0.508	0.235	0.113	0.037	0.007	26.356	26.206	5.873
Experiment 3	0.939	0.818	0.640	0.523	0.375	0.206	33.825	34.362	7.850
Experiment 4	0.939	0.806	0.618	0.480	0.345	0.191	33.215	33.520	7.838
Experiment 5	0.832	0.579	0.295	0.148	0.069	0.017	27.418	27.186	5.989
<i>Iterated results</i>									
Idealized 'true'	0.534	0.191	0.035	0.011	0.005	0.002	22.015	21.761	4.730
Bootstrap simulation ^a									
Experiment 1	0.566	0.215	0.041	0.011	0.007	0.002	22.276	22.060	4.838
Experiment 2	0.550	0.182	0.031	0.006	0.000	0.000	21.952	21.925	4.582
Experiment 3	0.549	0.208	0.067	0.040	0.025	0.016	22.566	21.837	5.742
Experiment 4	0.556	0.220	0.043	0.024	0.017	0.011	22.345	22.075	5.392
Experiment 5	0.557	0.200	0.047	0.013	0.005	0.002	22.308	22.112	4.873

^aAn approximate standard error of the difference between the two fractions exceeding a critical value α , based on n_1 and n_2 binomial trials, is $[\alpha(1-\alpha)/n_1 + \alpha(1-\alpha)/n_2]^{1/2}$. For $n_1 = 5,000$ and $n_2 = 1,000$, and α values of 0.500, 0.250, 0.100, 0.050, 0.025, 0.010, the standard errors are 0.017, 0.015, 0.010, 0.008, 0.005, 0.003 respectively.

We select five of these samples at random from each of which we generate 1,000 replications of the bootstrap simulation. If the bootstrap simulation procedure is reliable, the idealized true distribution should be revealed by the five bootstrap experiments.

Table 1 reports the results of the experiments. The distributions of the statistics are summarized by their means, medians, standard deviations, and the tail areas to the right of various critical values. In most of the experiments, the idealized true sampling distribution and the distributions generated by the bootstrap procedures lead to similar impressions about the test statistics. The

distributions conform better for iterated than for two-stage GMM. However, some of the evidence (especially experiments 3 and 4) shows that the bootstrapped p -values can differ from those of the true sampling distribution by a significant amount. We therefore check the sensitivity of our bootstrap results by using the traditional Monte Carlo simulation approach in a number of experiments.

5.2. Coefficient estimators and goodness-of-fit statistics

Table 2 summarizes the results of using the GMM to estimate and test a single-latent-variable model when the single-premium economy generates the data. The table shows systems with $N = 3$ to $N = 14$ assets and samples of $T = 60, 120,$ and 720 time-series observations. The first (second) row reports results for two-stage (iterated) GMM. The first column of Table 2 shows the average percentage bias in the C coefficient estimates; the second column shows the mean absolute percentage bias. We compute for each asset the average, over the 1,000 bootstrap replications, of the percentage difference between the coefficient estimate and the true coefficient. The mean (mean absolute) percentage bias is the average (average absolute) value across the assets. When $T = 720$, the bias is small. The two-stage and iterated GMM results are similar. Even with $T = 60$, the average bias of the estimators is small. Using the small instrument set with three to twelve assets, the average bias is no larger than 3.4% of the true coefficient. The mean absolute bias is usually close to the mean bias. This shows that when one asset's coefficient is biased in a particular direction, the other assets' coefficients are usually biased in the same direction. With larger numbers of assets, the bias tends to increase in absolute magnitude.

The right-hand columns of Table 2 evaluate the finite sample distributions of the test statistics, and report fractions rejected at various nominal significance levels. The fractions are the portion of the 1,000 replications of an experiment in which the test statistic exceeded a critical value from the chi-square distribution. With only three assets, the number of stages for convergence of iterated GMM is small, the difference between the two statistics is small, and the results for either set of instruments are similar. However, the results are sensitive to the number of assets.

With larger numbers of assets, the number of stages for convergence is larger and the differences between iterated and two-stage GMM are greater. With $T = 60$ observations, the accuracy of the two-stage test statistic decreases markedly as the number of assets increases. At a nominal 10% significance level, the rejection rates for two-stage GMM are 19.2% when $N = 3$, increasing to 36.2% when $N = 14$ (small instrument set). While two-stage GMM rejects too often, the iterated test statistic rejects too infrequently. This tendency is more pronounced as the number of assets increases. In many of the experiments where $T = 60$, the correct rejection frequency is 'bracketed' by the two test statistics.

Table 2

Finite sample properties of the Generalized Method of Moments (GMM) in latent-variables models for expected returns: properties of the coefficient estimators and the goodness-of-fit statistics

Results of estimating models with $K = 1$ latent variable, using artificial data generated so that a single-latent-variable model is the true model. For the *small instrument set*, Z consists of a constant, the one-month nominal Treasury bill rate, and the one-month lagged return on the CRSP value-weighted stock index. For the *large instrument set*, Z comprises the small instrument set and these additional variables: the lagged yield spread between a three-month and a one-month Treasury bill, the spread between the yield-to-maturity of AAA-rated corporate bonds and the lagged three-month Treasury bill return, the spread between the yield-to-maturity of BAA corporate bonds and the composite of corporate bond yields, the dividend yield of the CRSP value-weighted stock index, and a dummy variable for the month of January. The model is

$$r_1 = Z\delta + u_1, \quad r_2 = Z\delta C + u_2,$$

where r_1 represents the returns of the reference asset in excess of the one-month Treasury bill rate and r_2 is the excess returns of $N - 1$ test assets. N is the number of assets. δ is an L vector, and C is an $N - 1$ vector of coefficients. Each C_i element of C is the relative beta of asset i with respect to the reference asset r_1 . The latent-variable model implies that $E(u_1 u_2 | Z) = 0$. The results are based on 1,000 bootstrap replications. For each replication, the test statistics are evaluated for both two-stage and iterated GMM. The first row of numbers corresponds to two-stage GMM, the second row of numbers corresponds to iterated GMM. The bootstrap experiments are calibrated using monthly data. $D1-D10$ are the returns of value-weighted stock portfolios from market-value-ranked deciles. $D1$ contains the smallest firms and $D10$ the largest firms. GB is a long-term government bond return and CB is a long-term corporate bond return. $I1-I12$ are value-weighted industry-grouped portfolios, based on two-digit SIC codes. The returns are measured in excess of a one-month Treasury bill return. Results for various subsets of the assets are presented, indicated as follows:

$N:$	3	5	10	12	14
Assets:	$D1, D5, D10$	$D1, D3, D5, D10, GB$	$D1, \dots, D10$	$D1, \dots, D10, GB, CB$	$I1, \dots, I12, GB, CB$

\hat{C} is the GMM estimate of C . The average values (and average absolute values) across the N assets of the average over the 1,000 replications for each asset are shown under $\text{mean}((\hat{C} - C)/C)$ and $\text{mean}(|(\hat{C} - C)/C|)$.

N	% bias in \hat{C}		Actual % rejected at the following nominal significance levels					
	Mean % bias: $\text{Mean}((\hat{C} - C)/C)$	Mean absolute: $\text{Mean}((\hat{C} - C)/C)$	0.500	0.250	0.100	0.050	0.025	0.010
Using the small instrument set								
$T = 60$ time series observations (calibrated using 1982:1–1987:12)								
3	- 0.01%	1.00%	0.642	0.375	0.192	0.096	0.047	0.022
	- 1.24%	1.32%	0.573	0.272	0.089	0.027	0.015	0.003
5	0.34	0.67	0.622	0.338	0.141	0.088	0.043	0.013
	0.15	1.39	0.518	0.219	0.072	0.027	0.009	0.001
10	- 1.23	1.23	0.823	0.574	0.292	0.148	0.074	0.027
	- 3.33	3.33	0.559	0.214	0.052	0.016	0.005	0.000
12	- 0.71	0.86	0.855	0.612	0.320	0.180	0.084	0.024
	- 2.11	3.04	0.570	0.196	0.035	0.015	0.005	0.002
14	2.33	7.60	0.886	0.641	0.362	0.206	0.104	0.035
	7.99	21.6	0.550	0.193	0.044	0.017	0.004	0.002

T = 120 time series observations (calibrated using 1978:1–1987:12)

3	1.46	1.46	0.533	0.255	0.098	0.049	0.023	0.005
	1.79	1.92	0.527	0.247	0.090	0.032	0.018	0.003
5	2.66	3.71	0.510	0.254	0.091	0.051	0.024	0.008
	5.98	7.73	0.488	0.221	0.069	0.030	0.008	0.001
10	0.08	0.43	0.581	0.301	0.123	0.051	0.027	0.011
	– 0.22	1.00	0.509	0.216	0.056	0.020	0.006	0.003
12	2.13	3.51	0.615	0.309	0.114	0.042	0.017	0.001
	6.14	7.17	0.505	0.207	0.054	0.013	0.004	0.000
14	4.80	5.14	0.647	0.339	0.106	0.050	0.020	0.006
	9.72	15.7	0.533	0.210	0.052	0.012	0.005	0.001

T = 720 time series observations (calibrated using 1928:1–1987:12)

3	– 0.30	0.30	0.503	0.230	0.090	0.049	0.022	0.009
	– 0.34	0.34	0.502	0.229	0.090	0.048	0.021	0.009
5	1.06	1.72	0.504	0.254	0.095	0.045	0.022	0.011
	1.10	1.77	0.504	0.251	0.095	0.044	0.022	0.011
10	– 0.12	0.32	0.546	0.260	0.116	0.048	0.024	0.009
	– 0.17	0.35	0.541	0.253	0.106	0.043	0.021	0.008
12	– 0.59	1.55	0.542	0.266	0.104	0.058	0.027	0.008
	– 0.59	1.71	0.533	0.253	0.097	0.053	0.022	0.005
14	– 1.85	2.21	0.526	0.263	0.111	0.046	0.025	0.017
	– 2.12	2.55	0.513	0.248	0.098	0.038	0.019	0.011

Using the large instrument set

T = 60 time series observations (calibrated using 1982:1–1987:12)

3	0.12%	2.08%	0.721	0.405	0.164	0.074	0.024	0.005
	1.44%	3.60%	0.513	0.163	0.038	0.012	0.002	0.001
5	– 14.9%	16.9	0.941	0.738	0.368	0.156	0.056	0.004
	– 9.5%	18.4	0.808	0.415	0.124	0.041	0.014	0.001

T = 120 time series observations (calibrated using 1978:1–1987:12)

3	0.15	0.41	0.523	0.263	0.081	0.037	0.011	0.003
	0.34	0.34	0.478	0.209	0.047	0.011	0.002	0.001
5	– 2.78	2.78	0.616	0.308	0.109	0.046	0.015	0.003
	– 4.09	4.09	0.459	0.167	0.028	0.005	0.002	0.001
10	0.16	0.42	0.913	0.669	0.295	0.122	0.044	0.006
	– 0.54	0.66	0.780	0.376	0.091	0.030	0.012	0.001

T = 720 time series observations (calibrated using 1927:2–1987:12)

3	0.57	0.57	0.525	0.248	0.093	0.044	0.020	0.007
	0.57	0.57	0.522	0.246	0.088	0.043	0.020	0.005
5	0.01	1.77	0.509	0.254	0.096	0.047	0.020	0.005
	– 0.26	1.86	0.492	0.239	0.090	0.040	0.014	0.004
10	0.81	0.81	0.574	0.295	0.114	0.057	0.023	0.006
	0.73	0.73	0.538	0.249	0.087	0.040	0.014	0.004
12	– 0.01	1.90	0.591	0.308	0.122	0.060	0.033	0.017
	– 0.24	2.20	0.538	0.238	0.090	0.041	0.023	0.007
14	– 2.23	2.23	0.559	0.281	0.096	0.047	0.026	0.009
	– 3.07	3.07	0.492	0.197	0.062	0.026	0.010	0.005

Increasing the number of time series observations to $T = 120$, Table 2 shows that the overrejection by the two-stage GMM has essentially vanished. With $T = 720$ observations, the rejection frequencies are generally accurate.

Our experiments uncover another practical reason to prefer an iterated GMM approach. Repeatedly updating the weighting matrix and searching to find new parameter estimates reduces the chances that the algorithm will settle on a local minimum. We infer this because some of our experiments for the largest ($N = 14$) system were sensitive to the choice of the reference asset. If the global minimum is attained, the sample value of the test statistic is invariant to the reference asset (see Ferson, 1993).

Hansen (1982) derived the asymptotic properties of the GMM estimators and test statistics by assuming ergodicity and strict stationarity of the data vector $\{r_t, Z_{t-1}\}$. Our instruments include variables that are highly autocorrelated and may be nearly nonstationary (e.g., the one-month Treasury bill). Other authors have used similar instruments with the GMM in a number of empirical studies. It is therefore comforting to find that the GMM coefficient estimators and test statistics conform well to some of the theoretical asymptotic properties, even when the instruments may be nearly nonstationary.

5.3. *Standard errors and t-statistics*

Using 1,000 simulation trials, we compare the empirical standard deviations of the estimation errors in the individual-asset C coefficients with the mean of the 1,000 values of Hansen's (1982) asymptotic standard errors. If the asymptotic standard errors are reliable, the two should be similar. The second column of Table 3 reports ratios of the mean GMM to the empirical standard errors, averaged across the assets. Using the small instrument set and $T = 720$, we find that the mean reported standard errors differ from the empirical standard deviations only in the third decimal place, and that results for two-stage and iterated GMM are virtually identical. The asymptotic standard errors are understated by an amount that grows from approximately 3% to about 17% when T is reduced from 720 to 60 ($N = 3$). The bias increases as N is increased. For $T = 60$ and $N = 14$, the reported standard errors average about two-thirds of the correct magnitudes.

Table 3 shows that the bias in Hansen's (1982) standard errors is greater when using the large instrument set. Even when $T = 720$, the ratios of the asymptotic to the empirical standard errors are between 0.70 and 0.86. Thus, in models with large numbers of equations or instruments, there is a serious risk of overstating the significance of parameter estimates by relying on the asymptotic standard errors. This sensitivity should not be surprising since the number of orthogonality conditions, and therefore the complexity of the covariance matrix of the parameters, is determined by the number of equations and the number of instruments per equation.

The understated standard errors motivate an investigation of the multiplicative adjustment factors for the covariance matrix of the parameters. We examine the traditional adjustment factor $[T/(T - P)]$ and the alternative adjustment factor $[T(N + L)/\{T(N + L) - Q\}]$.

For each of 177 cases (where a case is defined by a given asset for a given N , T , and L) we compute the absolute difference between the competing standard errors and the empirical standard deviations. Using the traditional adjustment, the standard errors are closer to the empirical standard deviations than the unadjusted ones in 98% of the cases. The alternative adjustment is superior to the traditional adjustment in 90% of the cases. We divide the cases into subsamples based on the choice of the instruments. Using the small (large) instrument set, the alternative adjustment is superior to no adjustment in 86% (100%) of the cases and superior to the traditional adjustment in 85% (100%) of the cases.

Table 3 illustrates the effects of the adjustment factors. Each adjusted standard error is expressed as a fraction of the empirically determined standard error. Ratios less than 1.0 indicate downward bias in the standard errors. The table shows that the traditional adjustment factor is too small. The alternative adjustment is better in most cases, but it overadjusts the standard error in some cases. Those are extreme cases where $T(N + L)$ is close to Q , and the adjustment is too large because the effective degrees of freedom in the denominator are small (e.g., $T = 60$, $N = 14$, $L = 3$).

A standard error is typically used in conjunction with the point estimate of a coefficient. Table 3 therefore provides an analysis of t -ratios. For each simulation trial, we take the difference between the point estimate and the true value of each individual C coefficient and divide this by a standard error. For each adjustment we form a frequency distribution of the ratios, pooling them across the assets and simulation trials. Asymptotically, the t -ratios should be normally distributed with mean zero and unit variance.

Using $T = 720$ and the small instrument set, the unadjusted t -ratios have means near zero and variances close to 1.0. The degrees-of-freedom adjustments have little effect. The fractiles of the empirical distribution of the t -ratios (not shown) are close to the values from the normal distribution. Iterated and two-stage GMM results are similar. Decreasing the sample to $T = 60$, the standard deviations of the unadjusted t -ratios are about 1.2 for $N = 3$ and increase to 2.2 for $N = 14$. When the large instrument set is used, the empirical distribution of the t -ratios becomes more fat-tailed.

The traditional adjustment helps to reduce these biases, but the alternative adjustment is better. Even the alternative adjustment is inaccurate in extreme cases. The adjustment factor is too close to 1.0 for the large instrument set (the adjusted t -ratios have standard deviations between 1.4 and 1.9 when $T = 60$), and it is too extreme when $T = 60$ and $N = 14$, where the standard error of the adjusted t -ratio is only 0.7.

Table 3

Finite sample properties of the Generalized Method of Moments (GMM) standard errors in a single-latent-variable model for expected returns

Results for models with $K = 1$ latent variable, using artificial data generated so that a single-latent-variable model is the true model. The empirical standard errors are the sample standard deviations of the errors in the coefficient estimates. The *unadjusted* standard errors are the average across the 1,000 simulation trials of the asymptotic standard errors. The *traditional adjusted* standard errors use the asymptotic variances multiplied by the adjustment factor $[T/(T - P)]$, where T is the number of time series observations and P is the number of model parameters. The *alternative adjusted* standard errors use the asymptotic variances multiplied by the adjustment factor $[T(N + L)/\{T(N + L) - Q\}]$, where N is the number of asset equations, L is the number of instruments, and Q is the number of model parameters plus the number of unique elements in the GMM weighting matrix. The left-hand column for each case shows the average across the assets of the mean ratio for 1,000 simulation trials of a particular standard error (*Se:avg*) to the empirical (*True*) standard error. The *Mean t-ratio* is the average of 1,000 ratios, each formed as the difference between the coefficient estimate and the true coefficient, each divided by the relevant standard error of the coefficient estimate. *Std. of t-ratio* is the sample standard deviation of the 1,000 *t*-ratios. For the *small instrument set*, Z consists of a constant, the one-month nominal Treasury bill rate, and the one-month lagged return on the CRSP value-weighted stock index. For the *large instrument set*, Z comprises the small instrument set and these additional variables: the lagged yield spread between a three-month and a one-month Treasury bill, the spread between the yield-to-maturity on AAA-rated corporate bonds and the lagged three-month Treasury bill return, the spread between the yield-to-maturity of BAA corporate bonds and the composite of corporate bond yields, the dividend yield of the CRSP value-weighted stock index, and a dummy variable for the month of January. The model is

$$r_1 = Z\delta + u_1, \quad r_2 = Z\delta C + u_2,$$

where r_1 represents the returns of K reference assets in excess of the one-month Treasury bill rate and r_2 is the excess returns of $N - K$ test assets. δ is an L vector, and C is an $N - 1$ vector of coefficients. Each element C_i of C is the relative beta of asset i with respect to the reference asset r_1 . The latent-variable model implies that $E(u_1 u_2 | Z) = 0$. The results are based on 1,000 bootstrap replications for iterated GMM. The bootstrap experiments are calibrated using the actual monthly data. $D1-D10$ are the returns of value-weighted stock portfolios from market-value-ranked deciles. $D1$ contains the smallest firms and $D10$ the largest firms. GB is a long-term government bond return and CB is a long-term corporate bond return. $I1-I12$ are value-weighted industry-grouped portfolios, based on two-digit SIC codes. The returns are in excess of a one-month Treasury bill return. Results for various subsets of the assets are presented, indicated as follows:

<i>N</i> :	3	5	10	12	14
Assets:	<i>D1, D5, D10</i>	<i>D1, D3, D5, D10, GB</i>	<i>D1, ..., D10</i>	<i>D1, ..., D10, GB, CB</i>	<i>I1, ..., I12, GB, CB</i>

<i>N</i>	Unadjusted			Traditional adjusted			Alternative adjusted		
	Se:avg /True	Mean t-ratio	Std. of t-ratio	Se:avg /True	Mean t-ratio	Std. of t-ratio	Se:avg /True	Mean t-ratio	Std. of t-ratio

Using the small instrument set

$T = 60$ time series observations (calibrated using 1982:1–1987:12)

3	0.829	0.242	1.213	0.866	0.232	1.161	0.893	0.225	1.125
5	0.758	0.257	1.274	0.806	0.242	1.197	0.884	0.220	1.092

10	0.647	0.313	1.589	0.723	0.280	1.421	1.037	0.195	0.990
12	0.506	0.374	1.891	0.578	0.327	1.656	1.024	0.185	0.935
14	0.678	0.101	2.208	0.792	0.086	1.891	2.154	0.032	0.695

T = 120 time series observations (calibrated using 1978:1–1987:12)

3	0.700	– 0.147	1.108	0.715	– 0.144	1.085	0.726	– 0.142	1.069
5	0.773	– 0.086	1.096	0.797	– 0.084	1.063	0.830	– 0.080	1.021
10	0.765	– 0.127	1.183	0.807	– 0.121	1.122	0.919	– 0.106	0.986
12	0.619	– 0.127	1.183	0.651	– 0.125	1.153	0.785	– 0.105	0.968
14	0.665	– 0.121	1.247	0.714	– 0.126	1.161	0.897	– 0.101	0.925

T = 720 time series observations (calibrated using 1928:1–1987:12)

3	0.970	0.018	1.019	0.973	0.018	1.016	0.975	0.018	1.013
5	0.960	– 0.020	1.039	0.965	– 0.020	1.034	0.971	– 0.020	1.027
10	0.950	0.039	1.049	0.958	0.039	1.040	0.975	0.038	1.022
12	0.951	0.034	1.035	0.960	0.033	1.025	0.983	0.033	1.002
14	0.954	– 0.044	1.035	0.965	– 0.044	1.026	0.992	– 0.042	0.999

Using the large instrument set

T = 60 time series observations (calibrated using 1982:1–1987:12)

3	0.216	– 0.059	2.611	0.237	– 0.054	2.383	0.297	– 0.043	1.901
5	0.247	– 0.179	5.308	0.276	– 0.160	4.748	0.955	– 0.046	1.371

T = 120 time series observations (calibrated using 1978:1–1987:12)

3	0.546	– 0.213	1.595	0.570	– 0.204	1.527	0.624	– 0.186	1.395
5	0.406	– 0.516	2.002	0.428	– 0.490	1.900	0.595	– 0.353	1.368
10	0.196	– 1.032	7.149	0.212	– 0.956	6.623	0.275	– 0.735	5.095

T = 720 time series observations (calibrated using 1928:1–1987:12)

3	0.707	0.206	1.166	0.717	0.203	1.150	0.735	0.198	1.121
5	0.853	0.174	1.179	0.860	0.173	1.169	0.893	0.166	1.125
10	0.796	0.209	1.259	0.806	0.207	1.244	0.920	0.181	1.090
12	0.753	0.215	1.351	0.711	0.212	1.333	0.916	0.176	1.110
14	0.701	– 0.281	1.366	0.763	– 0.277	1.346	0.906	– 0.218	1.057

In summary, the empirical distribution of the *t*-ratios has more dispersion than the normal in small samples. The alternative adjustment factor is useful, provided that the number of instruments and assets is not too large, and that $T(N + L)$ is not too close to Q . However, the results should still be interpreted with caution. This motivates future research to develop alternative variance estimators with better small-sample properties.

5.4. Robustness of the results

We conduct a number of experiments to check the sensitivity of the results to variations in the experimental design. In the first exercise, the artificial data display conditional heteroskedasticity. Similar to the results in Table 2, we find

that the two-stage GMM tests reject the model too often using the small instrument set. The overrejection by the two-stage test statistics is worse than before in the larger systems. The iterated GMM test statistics are more reliable, but have a slight tendency to reject the model too infrequently. Thus, the results reinforce the importance of using iterated GMM for goodness-of-fit tests in the larger models.

When we introduce heteroskedasticity, the sampling distributions of the t -ratios are more fat-tailed in the smaller samples than they are without the heteroskedasticity (the t -ratios are well specified when $T = 720$). The downward bias of the asymptotic standard errors and the average absolute bias in the coefficient estimates are both worse than before. The alternative degrees-of-freedom adjustment is useful for small N , but it does not completely eliminate the fat tails of the t -ratios.

We replicate some of the experiments in Tables 2 and 3 using a traditional Monte Carlo approach instead of resampling the residuals. The rejection rates and the general patterns of the results are similar to those in Table 2, while the standard errors and t -ratios are slightly better behaved than in Table 3. We also replicate some of the experiments using a vector autoregression to generate random samples of the lagged instruments. We do this as an alternative to reusing the sample values in each replication. The results are similar to those in tables 2 and 3.

5.5. *The power of the tests*

The empirical literature typically fails to reject latent-variables models with small numbers of premiums. This suggests that either a small number of common factors determine conditional expected returns or the tests are low in power. To evaluate power we generate artificial data from two alternative economies. In the first, expected returns are determined by a two-beta model ($K = 2$) with fixed betas and two time-varying premiums. In the second, a conditional version of the CAPM holds in which the conditional market betas are time-varying. The appendix describes these alternative economies and summarizes the finite sample properties of the GMM when the null hypothesis is the two-latent-variable model.

Figs. 1 through 4 provide some feel for the properties of the alternative economies. Fig. 1 shows a time series plot of the expected market premium, using the small instrument set and the 720 monthly observations of the actual instruments from 1928 through 1987. In the single-latent-variable model, each asset's expected excess return is a constant beta multiplied by a similar expected market premium, based on the simulated instruments. Fig. 2 shows the expected interest rate premium in the two-latent-variables model. In the two-premium model, each asset's expected excess return is a constant linear combination of two expected premiums, similar to Figs. 1 and 2. Fig. 3 plots examples of the

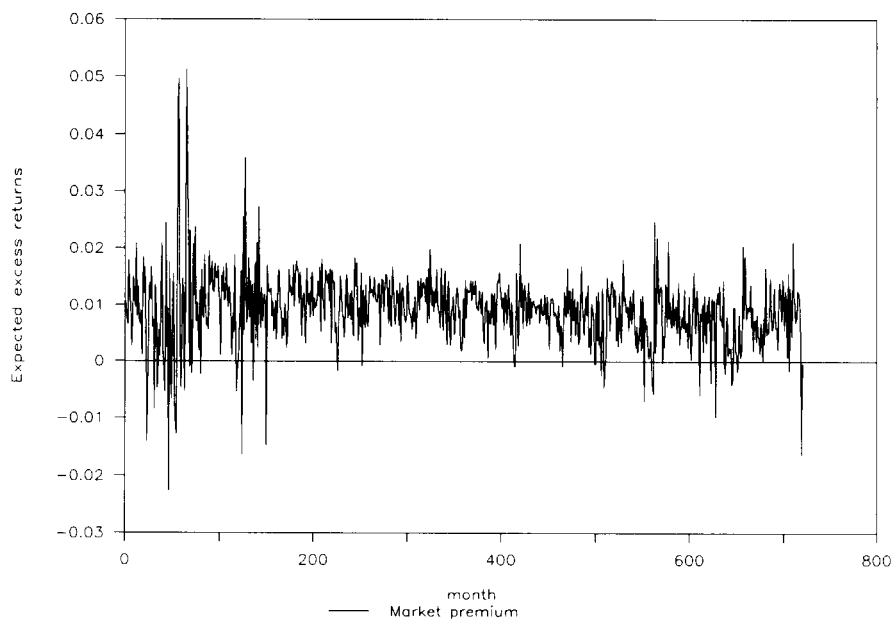


Fig. 1. Time series of the expected market premium. The monthly fitted expected excess return is obtained by regressing the return of the CRSP value-weighted index in excess of a one-month Treasury bill on the small instrument set. The small instrument set consists of a constant, the lagged level of the one-month bill, and the lagged excess return of the CRSP index. The 720 fitted values are from 1928 to 1987. These fitted values are used as the true expected returns in the simulation experiments.

betas for the time-varying beta CAPM economy. In this economy, an asset's expected excess return is the product of a changing beta and an expected market premium, similar to Fig. 1. Fig. 4 illustrates differences in the expected returns under two alternative models. We present two time series plots of fitted expected return differences. Deviations from the horizontal line at zero indicate differences between the expected returns under the single-factor model and an alternative. These are based on the actual sample of instruments. The series denoted by the squares represents the two-factor, constant-beta alternative. The mean difference is -0.73% and the standard deviation of the difference is 0.33% . The series denoted by the '+' signs represents the time-varying beta model. The mean difference is 0.27% and the standard deviation is 0.40% . The changing-beta expected returns differ from the null hypothesis by a smaller average amount, and the variance of the departure is larger.

Table 4 presents the empirical power of the $K = 1$ test. Under the null hypothesis $K = 1$, the sampling distribution of the two-stage GMM test statistic is more disperse than the iterated GMM (see Table 1). If we adjust the two-stage

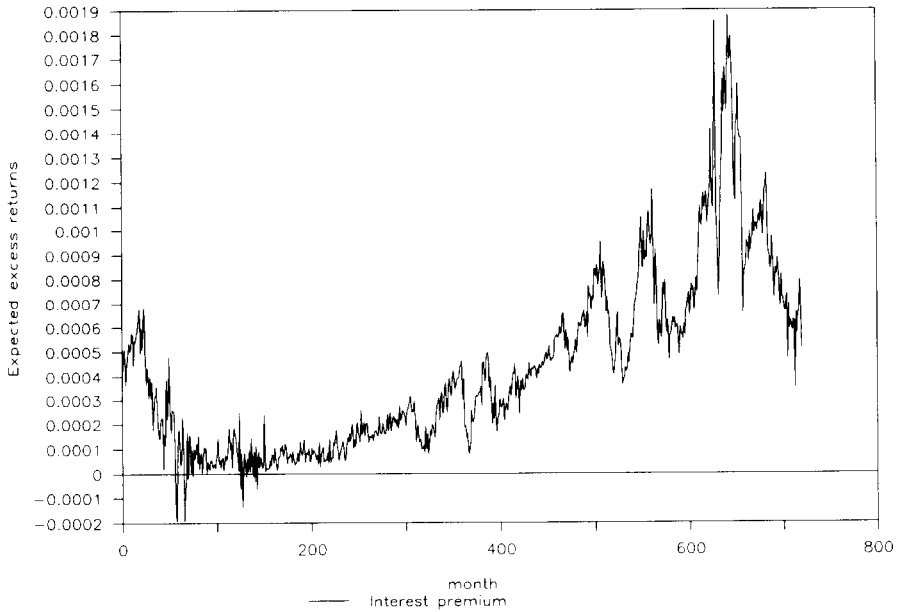


Fig. 2. Fitted expected interest rate risk premium. The monthly fitted expected excess return is obtained by regressing the return of a two-month Treasury bill in excess of a one-month bill on the small instrument set. The small instrument set consists of a constant, the lagged level of the one-month bill, and the lagged excess return of the CRSP index. The 720 fitted values are from 1928 to 1987. These fitted values are used as the expected interest rate risk premiums in the simulation experiments.

statistic for the correct size, we expect it to have inferior power. We therefore concentrate on the power of iterated GMM. The power is computed for tests with sizes of $\alpha = 0.10, 0.05,$ and 0.01 . We find an adjusted critical value for the test statistic for each case in the table. The adjusted critical value is the value which is exceeded by α fraction of the statistics simulated under the null hypothesis. The null hypothesis is the single-premium, constant-beta economy, similar to that shown in Table 2. The empirical power is the fraction of 500 trials in which a test using the adjusted critical value rejects the $K = 1$ hypothesis when the data are generated from the alternative economies. To hold the alternative hypothesis fixed while the sample size T is varied, we bootstrap from the $T = 720$ sample for each alternative. We also use the vector autoregression approach, which generates a different time series of the small instrument set in each replication. We experiment with other methods of generating the alternatives, and the overall impressions are similar.

Table 4 shows that the tests have more power to detect the two-premium alternative than to detect the CAPM with time-varying betas. The power of a 10% test to detect changing betas is less than 25% in all but two of the

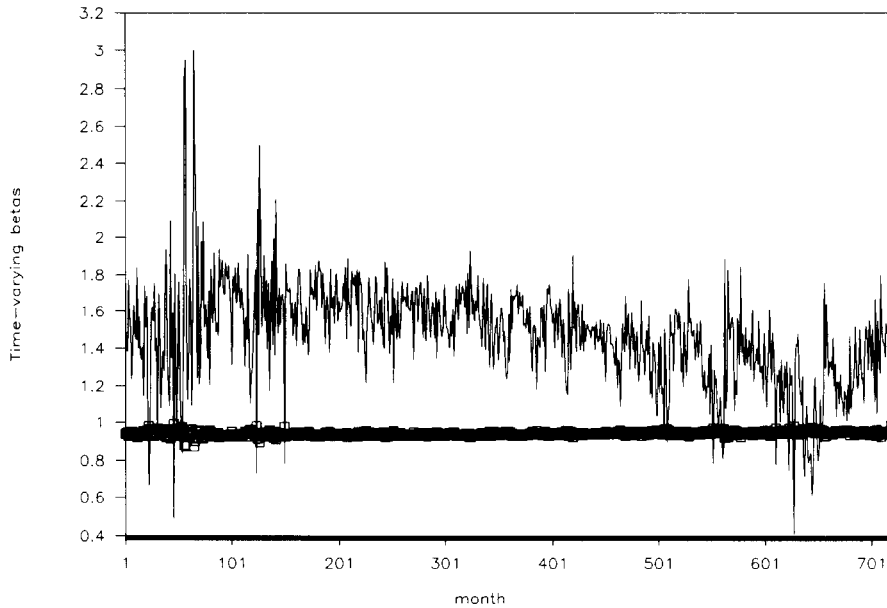


Fig. 3. Time-varying betas. Time series plots of time-varying betas on the CRSP value-weighted market index are shown for the 720 months from 1928 to 1987. The plot in the center that uses squares represents the market capitalization (size) portfolio of the largest firms. The betas for the smallest decile are shown as a solid line. The betas are assumed to be a linear function of the small instrument set, and are estimated by a regression of the returns on the CRSP value-weighted index and the products of the instruments and the index. The small instrument set consists of a constant, the lagged level of the one-month Treasury bill, and the lagged excess return of the CRSP value-weighted stock index.

15 cases. In eight cases the power is less than 10%. The power of a 10% test for the two-premium alternative exceeds 25% in eight of the 15 cases. In five cases it exceeds 60%.

It is remarkable that the power can be high, since the tests focus on regressions that explain a small fraction of the variance of the data – less than 10% in most cases and often less than 1%. The explained variance is small because the regressors are predetermined. In tests of unconditional beta pricing restrictions, the regressors are contemporaneous and they explain much larger fractions of the variance. The power of the GMM tests here to detect the two-factor alternative is comparable to the power in a number of studies of unconditional beta pricing restrictions.²

² See Gibbons (1982), Jobson and Korkie (1982), MacKinlay (1987), Gibbons and Shanken (1987), and Affleck-Graves and McDonald (1990).

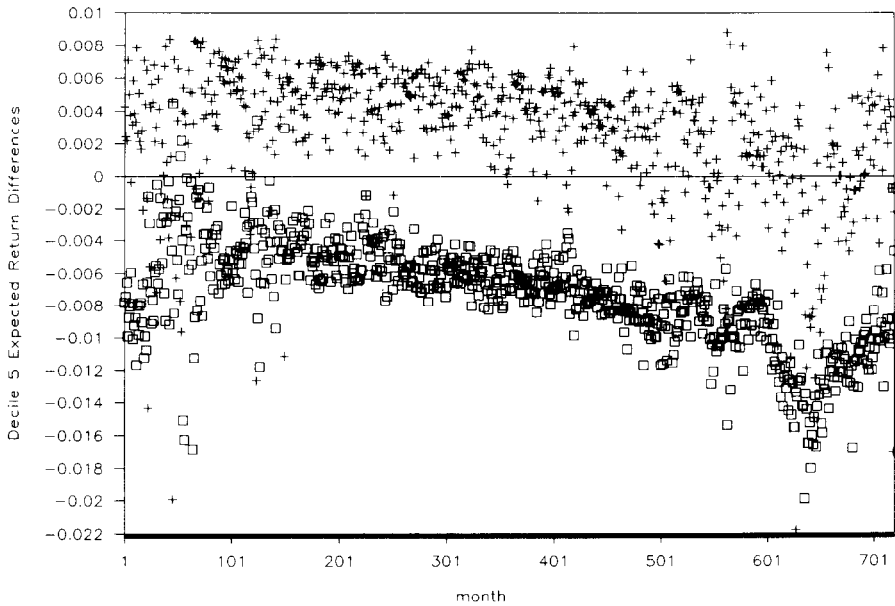


Fig. 4. Differences in expected returns under two alternative models. The squares represent the differences between the expected returns for the common stocks of market capitalization (size) portfolio five, under the single-latent-variable model versus the two-premium alternative. The '+' signs represent the expected return differences for the same portfolio, between the single-latent-variable model and the changing-beta CAPM alternative. We use the small instrument set, which consists of a constant, the lagged level of the one-month Treasury bill, and the lagged excess return of the CRSP value-weighted index. The 720 monthly observations are for 1928 to 1987.

5.6. Implications of the simulation evidence

Our results have important implications for studies that use the GMM. Recent studies reject single-latent-variable models for the expected returns of stock and bond portfolios. With two-stage GMM, the tests are likely to be biased against the model in small samples if the size of the system is large. For example, using the fourteen-asset system (small instrument set) and the data in this study, a two-stage GMM test of the single-latent-variable model produces an asymptotic p -value of 0.01. The bootstrapped p -value is 0.24. Our results show that using iterated GMM is important to avoid such a bias. Fortunately, much of the literature testing latent-variable models with large systems has used iterated GMM.³ Our results suggest that the rejections of single-latent-variable

³ Campbell (1987) used a two-stage GMM approach. He was kind enough to send us his data, so we conducted his tests again using iterated GMM. The results were essentially identical. This is not surprising, given our results and the small size of the systems in his paper.

Table 4

Power of tests for a single latent variable

Fraction of 500 trials in which the test for $K = 1$ latent variable rejects the null hypothesis when alternative models generate the artificial data. The instruments are a constant, the one-month nominal Treasury bill rate, and the one-month lagged return on the CRSP value-weighted stock index.

N	T	Alternative hypothesis: $K = 2$, fixed betas			Alternative hypothesis: $K = 1$, changing betas		
		Actual % rejected at the following nominal significance levels			Actual % rejected at the following nominal significance levels		
		0.10	0.05	0.01	0.10	0.05	0.01
3	60	0.186	0.114	0.026	0.102	0.062	0.008
	120	0.194	0.130	0.044	0.084	0.050	0.010
	720	0.618	0.474	0.332	0.178	0.086	0.018
5	60	0.186	0.086	0.048	0.058	0.012	0.006
	120	0.216	0.128	0.036	0.042	0.012	0.002
	720	0.656	0.596	0.406	0.242	0.178	0.050
10	60	0.206	0.094	0.034	0.040	0.020	0.002
	120	0.268	0.134	0.028	0.050	0.014	0.004
	720	0.736	0.640	0.376	0.205	0.159	0.023
12	60	0.234	0.126	0.018	0.072	0.028	0.000
	120	0.314	0.210	0.086	0.046	0.020	0.006
	720	0.764	0.688	0.340	0.290	0.186	0.026
14	60	0.226	0.102	0.038	0.104	0.038	0.006
	120	0.300	0.222	0.066	0.082	0.040	0.004
	720	0.830	0.730	0.540	0.396	0.262	0.104

models using iterated GMM have a mild conservative bias because iterated GMM rejects a latent-variable-model too infrequently.

Hansen's (1982) asymptotic standard errors should be adjusted in small samples to remove a downward bias. Studies that do not use such an adjustment have probably overstated the significance of their estimates. The adjustments can be applied *ex post facto* to evaluate the magnitude of the problem. Ferson (1990, Table IV) reports coefficients for a single-latent-variable model with $T = 151$ observations, $N = 7$ assets, and $L = 8$ instruments. Our alternative adjustment factor is 3.46, so Ferson's reported standard errors are probably far too small. Dividing his t -ratios by the adjustment factor, we find that two of the four t -statistics that are reported larger than 2.0 are no longer significant. Harvey (1991, Table VI) reports coefficients for a single-latent-variable model with $T = 232$, $N = 15$, and $L = 6$. All but one of the fifteen coefficients are reported as more than 1.96 standard errors from zero. In Harvey's case the

adjustment factor is 6.44. Dividing his t -ratios by this factor, we find that all of the fifteen coefficients have t -ratios below 2.

6. Concluding remarks

This paper develops evidence on the finite sample properties of the Generalized Method of Moments (GMM). The GMM is used to estimate predictive regressions for security returns and to test nonlinear cross-equation restrictions. Subject to the caveat that it is hazardous to extrapolate the simulations beyond this context, our experiments lead to several conclusions.

- (1) In simple models with a small number of equations and instruments, the GMM is reliable with as few as 60 time series observations. The coefficient estimates are approximately unbiased, and the goodness-of-fit statistics conform well to the asymptotic distribution.
- (2) In more complex models it is important to employ an iterated version of the GMM, since a two-stage procedure rejects the models too often in small samples. Iterated GMM test statistics conform more closely to the asymptotic distribution, but they have a mild tendency to reject too infrequently.
- (3) Hansen's (1982) standard errors are biased toward zero in small samples. A simple adjustment factor reduces the bias in simpler models. The adjustment multiplies the asymptotic variance by a ratio greater than unity, which accounts for the number of parameters plus the elements in the GMM weighting matrix. A traditional adjustment factor that does not account for the size of the weighting matrix is too small.
- (4) When the numbers of instruments and assets are small, t -ratios that are formed by using the adjustment factor can be evaluated with the unit normal distribution. However, the total number of observations must not be too close to the number of parameters plus the number of elements in the GMM weighting matrix. When the number of assets and instruments is large even the adjusted t -ratios are unreliable.
- (5) In more complex models, while the iterated GMM goodness-of-fit statistics are reasonably well specified, the coefficient estimates should be viewed with suspicion. They can be highly unreliable in small samples.
- (6) Even though the regression systems explain a small fraction of the variance in returns, the power to detect the more complex cross-equation structure of the coefficients implied by a two-beta model can be high. Given a changing-beta CAPM that induces nonlinearities and heteroskedasticity, we found that this alternative is closer to the single-factor null hypothesis and that the GMM tests have low power. Low power against some alternatives is one cost of the generality of the GMM approach.

Appendix: The three economies

We generate data from the economy to satisfy the null hypothesis of a single latent variable, using the following procedure:

- (a) r_{mt} is regressed on monthly data for \underline{Z}_{t-1} . The OLS estimates of the coefficients, δ_m , become the ‘true’ parameter values in Eq. (5) for the simulations.
- (b) Excess asset returns, r_{it} , $i = 1, \dots, N$, are regressed on an intercept, and the fitted expected excess market returns, $\delta'_m \underline{Z}_{t-1}$ from (a). The slope coefficients, β_{im} , become the ‘true’ values of the assets’ betas on the market portfolio.
- (c) We form a $T \times N$ sample of errors: $e_{it} = r_{it} - \beta_{im} (\delta'_m \underline{Z}_{t-1})$, $t = 1, \dots, T$, $i = 1, \dots, N$. The vectors of the mean-centered error terms determine the population distribution of the errors in the bootstrap experiments.
- (d) To generate an observation of artificial returns for each date t ($t = 1, \dots, T$) we draw a random integer t^* , $1 \leq t^* \leq T$, and select e_{it}^* as the error vector for time t . The artificial returns for asset i for time t are generated as $r_{it} = \beta_{im} (\delta'_m \underline{Z}_{t-1}) + e_{it}^*$.
- (e) The restricted regression system of Eq. (6) is estimated by GMM on the artificial returns. A two-stage and an iterated GMM test statistic are calculated for each replication. We set the maximum number of stages to 30. The criterion for convergence within a stage is a change in the objective function less than 0.0001. The criterion for convergence across the stages is 0.001. If we encounter a nonsingularity in the GMM weighting matrix, we start over at step (d) with a new replication. Such instances are rare.
- (f) Steps (d) and (e) are repeated for a total of 1,000 replications.

To generate artificial data that display conditional heteroskedasticity and satisfy a single-latent-variable model, we assume that the conditional correlations of the portfolio returns are fixed, and we specify a factor structure for the conditional heteroskedasticity. We estimate a model for the conditional variance of the market portfolio return, $\{\sigma_{mt}^2\}$, as a function of the instruments, \underline{Z}_{t-1} , using the methods of Davidian and Carroll (1987). We then obtain the implied conditional standard deviation series, for each portfolio return i , from the identity: $\sigma_{it} = \sigma_{mt}(\beta_{im}/\rho_{im})$, where β_{im} is the conditional beta used in the simulations and ρ_{im} is the correlation of the OLS residuals from the regressions of r_{it} and r_{mt} on \underline{Z}_{t-1} in the sample data. The conditional covariance matrix for date t , Ω_t , is formed from the products of the $\{\sigma_{it}\}$ and the $\{\rho_{ij}\}$. The vectors of the unanticipated returns, e , that we randomly draw for date t in our original experiments have a covariance matrix denoted by Σ . To generate conditionally heteroskedastic unanticipated returns with covariance matrix Ω_t , we use the transformation $e' \Sigma^{-1/2} \Omega_t^{1/2}$, where $\Sigma^{-1/2}$ and $\Omega_t^{1/2}$ are the Cholesky factorizations of Σ^{-1} and Ω_t , respectively.

In our changing-beta economy the conditional market betas are constructed by assuming the betas are linear functions of the instruments. The coefficients of the linear functions are estimated by regressing the asset return on the market index return and the product of the market index return with the lagged instruments. We generate the artificial data using a strategy similar to the previous experiments, except that we use the time-varying betas instead of fixed betas.

In our two-factor economy, the first factor is the (CRSP value-weighted) excess market return and the second is the excess return of a three-month Treasury bill. The conditioning information for the two expected premiums is the same as before, and both are estimated by linear regressions on the sample data. We generate artificial data using the two fitted premiums in step (b) and two fixed betas in steps (c) and (d).

Mean-variance efficiency implies that the two-beta alternative is one in which $E(r_{it}|Z_{t-1}) = \beta_{im}E(r_{mt}|Z_{t-1}) + \beta_{ir}E(r_{rt}|Z_{t-1})$ can be expressed as another single-factor, variable-beta model:

$$E(r_{it}|Z_{t-1}) = \beta_{ipt} E(r_{pt}|Z_{t-1}), \quad i = 1, \dots, N,$$

where

$$r_{pt} = \{w_{1t}r_{mt} + w_{2t}r_{rt}\}/(w_{1t} + w_{2t}),$$

and the w_{jt} are time-varying expected excess return-to-variance ratios for the two-factor portfolios r_{mt} and r_{rt} . The portfolio r_{pt} is a minimum variance combination of the two-factor portfolios, and the betas with respect to r_{pt} will vary over time. Therefore, the results on power may be interpreted as providing evidence about power against two different models of time-varying betas. The power of the test should depend upon how much the betas vary over time, but the distance of an alternative model from the null hypothesis cannot be measured solely by variation in the betas. In a previous version of this paper, we provided an analysis of the determinants of the power. These results are available by request.

We use the two-premium economy to investigate the properties of a test for $K = 2$ latent variables under the null hypothesis. (Tables are available by request.) We find that, as the number of assets increases, a two-stage GMM test of a given nominal size rejects the model more and more often. The overrejection becomes dramatic in the larger systems, even for the larger sample sizes. The iterated GMM statistic is more accurate, tending to reject somewhat too infrequently. Changing the reference assets has little effect on the rejection rates. We also find that the coefficient estimates are biased. Even when $N = 3$ and $T = 720$, the average absolute bias in one case exceeds 14%. We observed biases in excess of 600% with smaller sample sizes and larger numbers of equations. The empirical standard errors of the coefficient estimates are not always understated in the $K = 2$ model, and the degrees-of-freedom adjustments do not rectify the poor performance of the standard errors.

The relatively poor performance of the estimators in the $K = 2$ economy could be related to high correlation between the two expected premiums. We therefore conduct an experiment in which we orthogonalize the two risk premiums. The results are similar. In another experiment we use a maximum correlation portfolio for the monthly growth rates of the total U.S. industrial production index (from Citibase) as an alternative second risk factor. We construct the portfolio by regressing the growth rates of the industrial production index on the 25 monthly returns over the 1947.2–1987.12 period. The regression slope coefficients are normalized to sum to unity, and the normalized coefficients define a portfolio weight vector for the 25 assets. We use this return and repeat a subset of the simulations. The results are again similar. If the importance of a second premium is small in the actual data, the $K = 2$ model may be overparameterized. It is therefore interesting to find that the ($K = 1$) tests can have high enough power to detect a second premium in these data.

References

- Aigner, Dennis J., Cheng Hsiao, Arie Kapteyn and Tom Wansbeek, 1984, Latent variables in econometrics, Ch. 6 in: Kenneth J. Arrow and Michael D. Intriligator, eds., *Handbook of econometrics II* (North-Holland, Amsterdam).
- Amsler, Christine and Peter Schmidt, 1985, A Monte Carlo investigation of the accuracy of multivariate CAPM tests, *Journal of Financial Economics* 14, 359–375.
- Affleck-Graves, J. and William MacDonald, 1990, Multivariate tests of asset pricing: The comparative power of alternative statistics, *Journal of Financial and Quantitative Analysis* 25, 163–185.
- Bessembinder, Hendrik and Kalok Chan, 1992, Forecastability of futures returns, *Journal of Financial Economics* 32, 169–193.
- Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7, 265–296.
- Breeden, Douglas T., Michael R. Gibbons, and Robert H. Litzenberger, 1989, Empirical tests of the consumption-oriented CAPM, *Journal of Finance* 44, 231–262.
- Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.
- Campbell, John Y. and Richard H. Clarida, 1987, The term structure of Euromarket interest rates: An empirical investigation, *Journal of Monetary Economics* 19, 25–44.
- Campbell, John Y. and Yasushi Hamao, 1992, Predictable bond and stock returns in the United States and Japan: A study of long-term capital market integration, *Journal of Finance* 47, 43–70.
- Cecchetti, Stephen G., Pok-Sang Lam, and Nelson Mark, 1990, Mean reversion in equilibrium asset prices, *American Economic Review* 80, 398–418.
- Chamberlain, Gary and Arthur S. Goldberger, 1990, Latent variables in econometrics, *Journal of Economic Perspectives* 4, 125–152.
- Chan, K.C., Nai-fu Chen, and David Hsieh, 1985, An exploratory investigation of the firm size effect, *Journal of Financial Economics* 14, 451–472.
- Chang, Eric C. and Roger Huang, 1990, Time-varying return and risk in the corporate bond market, *Journal of Financial and Quantitative Analysis* 25, 323–340.
- Chang, Eric C., J. Michael Pinegar, and Ravi Ravichandran, 1991, Latent variable tests of the integration of European equity markets, Working paper (University of Maryland, College Park, MD).

- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–408.
- Cumby, Robert E., 1987, Consumption risk and international asset returns: Some empirical evidence, National Bureau of Economic Research working paper no. 2383.
- Cumby, Robert E., 1988, Is it risk? Explaining deviations from uncovered interest parity, *Journal of Monetary Economics* 22, 279–299.
- Davidian, Marie and R.J. Carroll, 1987, Variance function estimation, *Journal of the American Statistical Association* 82, 1079–1091.
- Efron, Bradford, 1982, The jackknife, the bootstrap, and other resampling plans (Society for Industrial and Applied Mathematics, Philadelphia, PA).
- Ferson, Wayne E., 1983, Expectations of real interest rates and aggregate consumption: Empirical tests, *Journal of Financial and Quantitative Analysis* 18, 477–497.
- Ferson, Wayne E., 1989, Changes in expected security returns, risk, and the level of interest rates, *Journal of Finance* 44, 1191–1217.
- Ferson, Wayne E., 1990, Are the latent variables in time-varying expected returns compensation for consumption risk?, *Journal of Finance* 45, 397–430.
- Ferson, Wayne E., 1993, Theory and empirical testing of asset pricing models, Ch. A6 in: Robert Jarrow, Voljisy Maksimovich, and William T. Ziemba, eds., *Finance* (North-Holland, Amsterdam) forthcoming.
- Ferson, Wayne E., Stephen R. Foerster, and Donald B. Keim, 1993, General tests of latent variable models and mean variance spanning, *Journal of Finance* 48, 131–156.
- Ferson, Wayne E. and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385–415.
- Ferson, Wayne E., Shmuel A. Kandel, and Robert F. Stambaugh, 1987, Tests of asset pricing with time-varying expected risk premiums and market betas, *Journal of Finance* 62, 201–220.
- Flesaker, Bjorn, 1993, Testing the Heath–Jarrow–Morton/Ho–Lee model of interest rate contingent claims pricing, *Journal of Financial and Quantitative Analysis* 28, 483–496.
- Foerster, Stephen R., 1987, Asset pricing models with changing expectations: An empirical study. Unpublished Ph.D. dissertation (The Wharton School, University of Pennsylvania, Philadelphia, PA).
- Gallant, Ronald A. and George Tauchen, 1989, Semiparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications, *Econometrica* 57, 1091–1120.
- Gibbons, Michael R., 1982, Multivariate tests of financial models, *Journal of Financial Economics* 10, 3–27.
- Gibbons, Michael R. and Wayne E. Ferson, 1985, Testing asset pricing models with changing expectations and an unobservable market portfolio, *Journal of Financial Economics* 14, 217–236.
- Gibbons, Michael R. and Jay Shanken, 1987, Subperiod aggregation and the power of multivariate tests of portfolio efficiency, *Journal of Financial Economics* 19, 389–394.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1987, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.
- Hansen, Lars P., 1982, Large sample properties of the generalized method of moments estimators, *Econometrica* 50, 1029–1054.
- Hansen, Lars P. and Robert J. Hodrick, 1983, Risk averse speculation in the forward foreign exchange market: An econometric analysis of linear models, in: J. Frenkel, ed., *Exchange rates and international macroeconomics* (University of Chicago Press, Chicago, IL).
- Hansen, Lars P. and Kenneth Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50, 1269–1285.
- Hansen, Lars P. and Kenneth J. Singleton, 1983, Stochastic consumption, risk aversion and the temporal behavior of asset returns, *Journal of Political Economy* 91, 249–266.
- Harvey, Campbell R., 1991, The world price of covariance risk, *Journal of Finance* 46, 111–158.

- Hinkley, D.V., 1977, Jackknifing in unbalanced situations, *Technometrics* 19, 285–292.
- Hodrick, R. and S. Srivastava, 1984, An investigation of risk and return in forward foreign exchange, *Journal of International Money and Finance* 3, 5–29.
- Jobson, J.D. and Robert Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Economics* 10, 433–466.
- Jorion, Phillippe and Alberto Giovannini, 1987, Interest rates and risk premia in the stock market and in the foreign exchange market, *Journal of International Money and Finance* 6, 107–123.
- Kandel, Shmuel and Robert F. Stambaugh, 1990, Expectations and volatility of consumption and asset returns, *Review of Financial Studies* 3, 207–232.
- Kocherlakota, Narayana, 1990, On tests of representative consumer asset pricing models, *Journal of Monetary Economics* 26, 285–304.
- Long, John B., 1974, Stock prices, inflation and the term structure of interest rates, *Journal of Financial Economics* 1, 131–170.
- MacKinlay, A. Craig, 1987, On multivariate tests of the CAPM, *Journal of Financial Economics* 18, 341–371.
- MacKinlay, A. Craig, and Matthew P. Richardson, 1991, Using the generalized method of moments to test mean–variance efficiency, *Journal of Finance* 46, 511–528.
- Mao, Ching-Sheng, 1991, Hypothesis testing and finite sample properties of generalized method of moments estimators: A Monte Carlo study, Working paper (Federal Reserve Bank of Richmond, VA).
- Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Nelson, Charles R. and Richard Startz, 1990, The distribution of the instrumental variables estimator and its *t*-statistic when the instrument is a poor one, *Journal of Business* 63, 125–140.
- Stambaugh, Robert F., 1988, The information in forward rates: Implications for models of the term structure, *Journal of Financial Economics* 21, 41–70.
- Tauchen, George, 1986, Statistical properties of generalized method-of-moment estimators of structural parameters obtained from financial market data, *Journal of Business and Economic Statistics* 4, 397–425.
- Wheatley, Simon, 1989, A critique of latent variable tests of asset pricing models, *Journal of Financial Economics* 23, 325–338.