

## A Test for the Number of Factors in an Approximate Factor Model

GREGORY CONNOR and ROBERT A. KORAJCZYK\*

### ABSTRACT

An important issue in applications of multifactor models of asset returns is the appropriate number of factors. Most extant tests for the number of factors are valid only for strict factor models, in which diversifiable returns are uncorrelated across assets. In this paper we develop a test statistic to determine the number of factors in an approximate factor model of asset returns, which does not require that diversifiable components of returns be uncorrelated across assets. We find evidence for one to six pervasive factors in the cross-section of New York Stock Exchange and American Stock Exchange stock returns.

THE ARBITRAGE PRICING THEORY (APT) of Ross (1976) has generated an increased interest in the application of linear factor models in the study of capital asset pricing. The APT has the attractive feature that it makes a minimal number of assumptions about the nature of the economy (a factor structure for the returns generating process, a large number of assets, and frictionless trading). The costs of these minimalist assumptions include certain ambiguities such as an approximate pricing relation and an unknown number of pervasive factors.

In order to estimate and test the APT, one must specify the number of pervasive factors in asset returns. The issue of the appropriate number of factors has been the subject of some controversy (see, for example, Roll and Ross (1980, 1984); Dhrymes, Friend, and Gultekin (1984); Luedecke (1984); Trzcinka (1986); Conway and Reinganum (1988); and Brown (1989)). In this paper we propose a new approach to estimating the number of pervasive economic factors generating asset returns. An important feature of our approach is that it is valid when asset returns follow an approximate, rather than a strict, factor model.

A strict factor structure is one in which the idiosyncratic, or diversifiable, components of asset returns have zero correlation across assets. Ross (1976) assumes a strict factor structure in his original development of the APT. However, he notes that this assumption can be weakened. The key requirement for the APT is that nonfactor risk can be diversified away in many-asset

\* Connor is from the London School of Economics and Korajczyk is from Northwestern University. We would like to thank Peter Bossaerts, Gary Chamberlain, Richard Green, David Modest, Jay Shanken, René Stulz, Mark Watson, an anonymous referee, and seminar participants at London Business School, London School of Economics, and Nuffield College, Oxford for helpful comments. We also thank Theodore Sternberg and Robert Uhlauer for research assistance.

portfolios. This diversification criterion does not require strictly zero correlation across idiosyncratic returns. Rather, it requires that the correlations be sufficiently weak so that the law of large numbers applies.

Chamberlain and Rothschild (1983) and Ingersoll (1984) generalize the pricing results of the APT to the case of an approximate factor structure. In an approximate factor structure the idiosyncratic components of returns need not be uncorrelated and, hence, the idiosyncratic covariance matrix need not be diagonal. However, an asymptotic limit is assumed on the amount of covariance between idiosyncratic returns. This is expressed as a bound on the eigenvalues of the idiosyncratic covariance matrix as the number of cross-sections increases. Intuitively, as the number of cross-sections increases, the proportion of total variation explained by any nonpervasive source of risk must approach zero.

In addition to being more general, the use of an approximate factor model is intuitively more appealing. It seems possible that a few firms in the same industry might have industry-specific components to their returns which are not pervasive sources of uncertainty for the whole economy. For example, awarding a defense contract to one aerospace firm might affect the stock prices of several firms in the industry. Assuming a strict factor structure would force us to treat this industry-specific uncertainty as a pervasive factor.

A strict factor structure is assumed in the standard factor analysis literature (see Anderson (1984, chapter 14)). There are well-known likelihood ratio tests for the appropriate number of factors in this case. Intuitively, the test that  $k$  is a sufficient number of factors is a test of whether the idiosyncratic covariance matrix is diagonal under a  $k$ -factor structure. In the presence of an approximate factor model, tests which assume a strict factor model (such as those commonly used in maximum likelihood factor analysis) will tend to identify too many factors.

Since the definition of an approximate factor model relies on the limiting properties of the eigenvalues of the return covariance matrix, a natural test would be to investigate the behavior of the eigenvalues of the sample covariance matrix as the number of assets increases. This type of analysis is done in Luedecke (1984) and Trzcinka (1986). Some difficulties with this approach are discussed in Brown (1989) and in Section II.C below.

We take a somewhat different approach. We develop a simple statistic which is based on the result that, if  $k$  is the correct number of pervasive factors, then there should be no significant decrease (adjusting for degrees of freedom) in the cross-sectional mean square of idiosyncratic returns in moving from  $k$  to  $k + 1$  factors. We develop statistics for testing for a significant decrease. The tests allow for cross-sectional dependence and heteroskedasticity in the idiosyncratic returns.

In Section I we relate the approximate factor model assumptions to the statistical concept of strong ( $\alpha$ ) mixing. We show that the nonpervasiveness of idiosyncratic risk can be characterized as strong mixing of the cross-section of idiosyncratic variates. In Section II we use the assumption of

strong mixing to derive our statistic and compare our approach to previous research. In Section III we present the empirical results. We find evidence for between one and six pervasive factors. Section IV concludes.

### I. Nonpervasiveness as a Mixing Condition on Idiosyncratic Returns

In this section we show how the theory of strong mixing can be applied to the analysis of approximate factor structures. We begin by defining strict and approximate factor structures. Let  $r^n$  denote an  $n$ -vector of asset returns in period  $t$ . We assume that returns are independently and identically distributed through time,<sup>1</sup> and have the usual linear factor structure:

$$\tilde{r}^n = c^n + B^n \tilde{f} + \tilde{\varepsilon}^n \tag{1}$$

$$E[\tilde{\varepsilon}^n | \tilde{f}] = 0 \tag{2}$$

Here  $c^n$  denotes an  $n$ -vector of constants,  $\tilde{f}$  the  $k$ -vector of pervasive factors,  $B^n$  the  $n \times k$ -matrix of factor betas, and  $\tilde{\varepsilon}^n$  the  $n$ -vector of idiosyncratic returns. A strict factor structure imposes the condition that the covariance matrix of idiosyncratic returns,  $V^n = E[\tilde{\varepsilon}^n \tilde{\varepsilon}^{n'}]$ , is a diagonal matrix. Under a strict factor model, the APT also requires that the diagonal elements of  $V^n$  are bounded for all  $n$ .

Chamberlain and Rothschild (1983) and Ingersoll (1984) show that the diagonality of  $V^n$  is not necessary for the proof of the APT. Rather than assuming diagonality of  $V^n$  they assume that:

$$\lim_{n \rightarrow \infty} \frac{1}{n} B^{n'} B^n = \Xi; \Xi \text{ nonsingular} \tag{3}$$

$$\lim_{n \rightarrow \infty} \|V^n\| \leq \gamma < \infty \tag{4}$$

where  $\| \cdot \|$  is the matrix 2-norm.<sup>2</sup> Using the terminology of Chamberlain and Rothschild (1983), we call a model obeying (1) to (4) an approximate  $k$ -factor model. Note that, from (4),  $V^n$  is not required to be diagonal but is required to have bounded eigenvalues as  $n$  increases. This limits the amount of cross-sectional correlation in the idiosyncratic returns.

The strict factor model, assumed in the standard factor analysis literature, allows for cross-sectional heterogeneity in idiosyncratic variability (i.e., different diagonal elements in  $V$ ) but does not allow for cross-sectional dependence (i.e., all off-diagonal elements in  $V$  are assumed to equal zero). An approximate factor structure allows both heterogeneity and limited amounts of dependence.

<sup>1</sup> It is possible to allow some heterogeneity across time. We have chosen not to do so because it complicates the analysis without changing the ultimate form of the tests.

<sup>2</sup> For symmetric, positive semidefinite matrices the matrix 2-norm is the largest eigenvalue of the matrix. See Stewart (1973, pp. 179, 180) for definitions of various matrix norms.

We incorporate heterogeneity and dependence in idiosyncratic returns into our model by assuming that the cross-sectional sequence  $\{\tilde{\varepsilon}_i\}_{i=1}^\infty$  is a mixing process. Our brief description of mixing processes relies on the more detailed discussion provided in White (1984). Let  $\{\tilde{\varepsilon}_i\}_{i=1}^\infty$  be a sequence of random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$  and let  $\mathcal{F}_a^b$  be the information set ( $\sigma$ -algebra) generated by  $(\tilde{\varepsilon}_a, \dots, \tilde{\varepsilon}_b)$ . A measure of the dependence between  $\sigma$ -fields  $\mathcal{G}$  and  $\mathcal{H}$  is given by  $\alpha(\mathcal{G}, \mathcal{H}) = \sup_{\{G \in \mathcal{G}, H \in \mathcal{H}\}} |P(G \cap H) - P(G)P(H)|$  which is the maximal difference between the joint probability of events in  $\mathcal{G}$  and  $\mathcal{H}$  and the product of the probabilities of each event. The mixing coefficients of the process  $\{\tilde{\varepsilon}_i\}$  are defined as  $\alpha(m) = \sup_n \alpha(\mathcal{F}_{-\infty}^n, \mathcal{F}_{n+m}^\infty)$ . If  $\alpha(m) \rightarrow 0$  as  $m \rightarrow \infty$  the sequence  $\{\tilde{\varepsilon}_i\}$  is called strong mixing or  $\alpha$ -mixing. We assume that the cross-sectional heterogeneity and dependence is of the following form:

**MIXING ASSUMPTION:** *There exists an ordering of the sequence  $\{\tilde{\varepsilon}_i\}_{i=1}^\infty$  such that  $\{\tilde{\varepsilon}_i\}$  is a strong mixing process with mixing coefficients,  $\alpha(m)$  that are  $O(m^{-\lambda})$  for  $\lambda > 2$ .*

The notation  $O(m^{-\lambda})$  denotes that the sequence  $\alpha(m)$  is "at most of order  $(m^{-\lambda})$ ." That is, there exists a real number  $M$  ( $M < \infty$ ) such that  $m^\lambda |\alpha(m)| \leq M$  for all  $m$ .

An intuitive description of alpha mixing might be useful for the reader unfamiliar with the concept. Note that the independence of two events  $G$  and  $H$  implies  $P(G \cap H) = P(G)P(H)$ . Roughly speaking, the mixing condition says that if  $m$  is very large, then for any two random variables  $\varepsilon_i$  and  $\varepsilon_{i+m}$ ,  $P(G \cap H) \approx P(G)P(H)$  where  $G$  is the realization of  $\varepsilon_i$  and  $H$  is the realization of  $\varepsilon_{i+m}$ . Intuitively, if two random variables are "far enough apart" then they are "almost independent." While the mixing conditions are most naturally interpreted in a time series context we are applying them in a cross-sectional problem. Our statistics are structured such that we do not need to know the appropriate ordering, we only need to assume that one exists.<sup>3</sup>

Consider some specific examples of idiosyncratic covariance matrices,  $V$ , which would satisfy the mixing assumptions. If the  $(i, j)$  element of  $V$  is equal to  $\sigma^2 \cdot \rho^{|i-j|}$  and  $-1 < \rho < 1$  then the idiosyncratic returns have dependence which "dies off" sufficiently fast for  $\{\tilde{\varepsilon}_i\}$  to be  $\alpha$ -mixing. An alternative covariance structure might be a block diagonal structure where the blocks are defined by industry. Let  $I < \infty$  be the maximum number of firms in a given industry. Then the assumption of block diagonality implies  $\alpha(m) = 0$  for all  $m > I$ , hence the process is  $\alpha$ -mixing.

The mixing processes explicitly allow a tradeoff between the dependence in the process and moment restrictions required for laws of large numbers and central limit theorems. As  $\lambda$  decreases, the series is allowed to exhibit more

<sup>3</sup> Mei and Newey (1990) consider a test which relies on knowing the appropriate ordering of assets.

dependence while the existence of more moments is required. For our purposes, assuming the sequence  $\{\tilde{\varepsilon}_i\}$  is mixing allows us to have cross-sectionally dependent and heteroskedastic idiosyncratic returns but still maintain the conditions necessary for an approximate factor structure. This is shown in the following theorem.

**THEOREM 1:** *Assume the above Mixing Assumption for the sequence of idiosyncratic returns and that  $E[|\tilde{\varepsilon}_i|^{2(1+\delta)}]$  is bounded for all  $i$  and for some  $\delta > 2/(\lambda - 2)$ . Then the eigenvalues of  $V^n$  are bounded as  $n$  approaches infinity.*

*Proof:* By corollary 6.16 of White (1984) the assumptions above imply that there exists a constant  $M^* < \infty$  such that  $|V_{ij}^n| \leq M^*|i - j|^{-\lambda\delta/(2+2\delta)}$ ,  $i \neq j$ , where  $V_{ij}^n$  is the  $(i, j)$  element of  $V^n$ . Let  $s = \lambda\delta/(2 + 2\delta)$ . The restrictions on  $\lambda$  and  $\delta$  imply that  $s > 1$ . Consider the matrix 1-norm (sometimes referred to as the column-sum norm) of  $V^n$ . This norm is defined as

$$\max_j \left[ \sum_{i=1}^n |V_{ij}^n| : j = 1, 2, \dots, n \right].$$

The bound placed on the absolute value of the elements of  $V^n$  by the mixing conditions implies that the column-sum norm of  $V^n$ , denoted  $\|V^n\|_1$ , is bounded by  $M^*[1 + 2\zeta(s)]$  where  $\zeta(s)$  is the Riemann zeta-function:

$$\zeta(s) = \sum_{\nu=1}^{\infty} \frac{1}{\nu^s}$$

which converges for  $s > 1$ . Thus,  $\|V^n\|_1 < \infty$ . The column-sum norm of a matrix is an upper bound on the eigenvalues of that matrix (see Schwarz, Rutishauser, and Stiefel (1973, pp. 10, 11)). Hence, the eigenvalues of  $V^n$  are bounded as  $n \rightarrow \infty$ . Since the largest eigenvalue is increasing in  $n$ , all eigenvalues are uniformly bounded for all  $n$ .  $\square$

The representation of returns in (1) and the bounded eigenvalues of  $V^n$  are conditions which imply an approximate factor structure (see Chamberlain and Rothschild (1983)). Thus, the assumptions of Theorem 1 are sufficient for asset returns to follow an approximate factor structure. We believe that Theorem 1 is a useful characterization since it ties together the approximate factor model conditions and the mixing conditions.

Some approaches to estimating pervasive factors (e.g., Connor and Korajczyk (1986)) assume that the average squared idiosyncratic returns,  $\tilde{\varepsilon}^{n'}\tilde{\varepsilon}^n/n$ , converge in probability. The mixing assumptions imply this convergence.

**THEOREM 2:** *Under the assumptions of Theorem 1,  $\tilde{\varepsilon}^{n'}\tilde{\varepsilon}^n/n$  converges in probability to  $\sigma_\varepsilon$  as  $n \rightarrow \infty$ , where  $\sigma_\varepsilon = \lim_{n \rightarrow \infty} E[\tilde{\varepsilon}^{n'}\tilde{\varepsilon}^n/n]$ .*

*Proof:* The moment restrictions assumed in Theorem 1 imply that  $V_{ii}^n = E(\tilde{\varepsilon}_i^2)$  is bounded. Therefore,  $\sigma_\varepsilon$  exists and is finite. Since  $\{\tilde{\varepsilon}_i\}$  is mixing,  $\{\tilde{\varepsilon}_i^2\}$

is mixing, (White (1984, theorem 3.49)). Since the sequence  $\{\tilde{\varepsilon}_t^2\}$  is mixing  $\tilde{\varepsilon}^{n'}\tilde{\varepsilon}^n/n$  converges almost surely to  $\sigma_\varepsilon$  (White (1984, corollary 3.48)). Almost sure convergence implies convergence in probability.  $\square$

## II. A Test Statistic for the Number of Factors in Approximate Factor Models

In Section II.A we derive the test statistic under the assumption that we observe the diversifiable returns on assets without error. This serves to provide the main intuition behind the test without the added complications of dealing with estimation error. In Section II.B we generalize the test to the case where the factor returns and idiosyncratic returns are estimated. In Section II.C we relate our approach to other approaches that focus on the behavior of the eigenvalues of return covariance matrices as the number of cross sections increases, such as Luedecke (1984), Trzcinka (1986), and Brown (1989).

### A. Test Statistics with Observed Idiosyncratic Returns

In this section we present the basic statistic. We assume that asset returns are given by an approximate factor structure and that the idiosyncratic returns are strong mixing. Let  $\tilde{f}^*$  denote some random variable which is not one of the pervasive factors and is not perfectly correlated with them. We will call this variable “factor  $k + 1$ ” of the approximate  $k$ -factor model. This pseudofactor can be some combination of the idiosyncratic variates or some industry-wide influence which is not widespread enough to be a pervasive factor. We want a test which can distinguish this pseudofactor from the  $k$  true factors.

Note that we can rewrite the factor model in its original form (equation (1)) and in a form which includes the “ $k + 1$ st” factor:

$$\tilde{r}^n = c^n + B^n\tilde{f} + \beta^n\tilde{f}^* + \tilde{\varepsilon}^{*n}$$

where  $\beta^n$  is the  $n \times 1$  vector of least squares projections of  $\tilde{r}^n$  on  $\tilde{f}^*$  (controlling for  $\tilde{f}$ ). A  $k$ -factor approximate factor structure implies that  $\beta^{n'}\beta^n \leq \omega < \infty$  for all  $n$  (Ingersoll (1984)).

Note that  $\tilde{\varepsilon}^n = \beta^n\tilde{f}^* + \tilde{\varepsilon}^{*n}$ . If asset returns are described by an approximate factor structure with  $k$  factors then the  $k + 1$ st factor can have nontrivial factor loadings for some assets (since  $V^n$  is not diagonal) but not for a significant proportion of assets. The basic logic behind our tests is that:

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \frac{1}{n} \tilde{\varepsilon}^{n'}\tilde{\varepsilon}^n &= f^{*2} \cdot \lim_{n \rightarrow \infty} \frac{1}{n} \beta^{n'}\beta^n + \text{plim}_{n \rightarrow \infty} \frac{1}{n} \tilde{\varepsilon}^{*n'}\tilde{\varepsilon}^{*n} \\ &= \text{plim}_{n \rightarrow \infty} \frac{1}{n} \tilde{\varepsilon}^{*n'}\tilde{\varepsilon}^{*n} \end{aligned}$$

so that  $\text{plim}_{n \rightarrow \infty} (\tilde{\varepsilon}^n \tilde{\varepsilon}^n - \tilde{\varepsilon}^{*n} \tilde{\varepsilon}^{*n})/n = 0$ , where  $\text{plim}$  denotes the limit in probability. The cross-sectional mean square of  $\tilde{\varepsilon}^n$  approaches the mean square of  $\tilde{\varepsilon}^{*n}$  because  $\beta^n \beta^n/n \leq \omega/n$  which approaches zero as  $n$  increases. That is, the cross-sectional average asset variability explained by the  $k + 1$ st factor is asymptotically zero.<sup>4</sup>

In this subsection we assume that the idiosyncratic returns are observable. In Section II.B we will generalize to the case in which they are estimated via time series ordinary least squares. Previously,  $\varepsilon$  and  $\varepsilon^*$  denoted  $n$ -vectors of idiosyncratic returns. We will now use them to denote the  $n \times T$  matrices of observed idiosyncratic returns for  $n$  assets over  $T$  time periods using  $k$  factors, and using  $k$  factors plus one pseudofactor, respectively. We let  $\varepsilon_t$  denote the  $t$ th column of  $\varepsilon$  and  $\varepsilon_i$  denote the  $i$ th row of  $\varepsilon$ . Consider the following statistics:

$$\mu_t^n = \varepsilon_t' \varepsilon_t/n; \mu_t^{n*} = \varepsilon_t^{*'} \varepsilon_t^*/n, t = 1, \dots, T.$$

A natural statistic to investigate is  $\mu_t^n - \mu_t^{n*} = f_t^{*2} \beta^{n'} \beta^n/n$ . The difficulty with this statistic is that, under the null hypothesis, it converges to a degenerate distribution as  $n$  approaches infinity.<sup>5</sup> Because of this, we devise an alternative statistic which has a mean of zero (under the null) and a nondegenerate asymptotic distribution. We do this by taking the difference between  $\mu$  in one time period and  $\mu^*$  in the next period. We then calculate means and variances using every other observation. Define the  $(T/2)$ -vector<sup>6</sup>  $\hat{\Delta}^n$  by:

$$\hat{\Delta}_s^n = \mu_{2s-1}^n - \mu_{2s}^{n*} \quad s = 1, \dots, T/2.$$

Note that, by the assumed return structure,  $\varepsilon_{it}$  and  $\varepsilon_{j\tau}$  are independent for  $t \neq \tau$ , and identically distributed. This implies that the  $(T/2)$ -vector  $\hat{\Delta}^n$  consists of independent random variables.  $E[\hat{\Delta}_s^n]$  converges to  $f_{2s-1}^{*2} \beta^{n'} \beta^n/n$  so that the expectation of  $\hat{\Delta}_s^n$  goes to zero as  $n$  approaches infinity, under the null hypothesis that there are  $k$  factors. Under the alternative hypothesis that there are  $k + 1$  pervasive factors, we expect the values of the  $\hat{\Delta}_s^n$ 's to be positive because the term  $f_{2s-1}^{*2} \beta^{n'} \beta^n/n$  will not converge to zero. We are now ready to state our main result. As the cross-sectional sample increases,  $\hat{\Delta}^n$  converges in distribution to a  $(T/2)$ -vector of independent, normally distributed random variables with means of zero and equal variances. We

<sup>4</sup> An alternative approach to testing for the number of factors relies on the fact that an approximate factor structure implies that well-diversified portfolios will have betas relative to the pseudofactor that approach zero as  $n$  approaches infinity. This approach is used in Korajczyk and Viallet (1989) and is extended by Heston (1992).

<sup>5</sup> To see this note that:

$$\sqrt{n} (\mu_t^n - \mu_t^{n*}) = \frac{f_t^{*2} \beta^{n'} \beta^n}{\sqrt{n}}$$

which is equal to 0 in the limit.

<sup>6</sup> We assume, without loss of generality, that  $T$  is even.

first need the following definition. Let

$$\psi^{a,n} = \text{var} \left( n^{-1/2} \cdot \sum_{i=a+1}^{a+n} \varepsilon_{it}^2 \right).$$

We assume that there exists  $\psi$ ,  $0 < \psi < \infty$  such that  $\psi^{a,n} \rightarrow \psi$  as  $n \rightarrow \infty$  uniformly in  $a$ .

**THEOREM 3:** *Given the assumptions of Theorem 1 and the convergence of  $\psi^{a,n}$ , then,*

$$\text{dlim}_{n \rightarrow \infty} \sqrt{n} \hat{\Delta}^n \sim N(0, 2\psi I)$$

where *dlim* denotes convergence in distribution and  $N(\cdot, \cdot)$  denotes the multivariate normal distribution.

*Proof:* Note that, under the null hypothesis,  $E[\hat{\Delta}_s^n | f_{2s-1}^{*2}] = f_{2s-1}^{*2} \beta^{n'} \beta^n / n \rightarrow 0$ , as  $n \rightarrow \infty$  because  $\beta^{n'} \beta^n / n \rightarrow 0$  while  $f_{2s-1}^{*2}$  is fixed. The mixing assumption and moment restrictions of Theorem 1 plus the convergence of  $\psi^{a,n}$  satisfy the conditions of Theorem 5.19 of White (1984). It follows immediately from that theorem that  $\text{dlim}_{n \rightarrow \infty} (n^{-1/2} \varepsilon'_t \varepsilon_t - \sigma_\varepsilon) \sim N(0, \psi)$ . The  $(T/2)$ -vector  $\hat{\Delta}^n$  consists of differences between these independent, identically distributed Normals and so has zero means and twice the variances.  $\square$

Standard cross-sectional variance estimates (which assume that the  $\varepsilon$ 's are uncorrelated across cross-sections) will not yield valid estimates of  $\psi$  because of the cross-sectional dependence allowed in an approximate factor structure.<sup>7</sup> Instead, we use the time series of the estimates  $\hat{\Delta}_s^n$  to estimate the mean and variance. This is the same as the approach taken by Fama and MacBeth (1973) to estimate the mean and variance of the equity risk premium in the capital asset pricing model. Define the time-series sample mean and variance of  $\hat{\Delta}^n$ , respectively, by:

$$\bar{\Delta}^n = \frac{2}{T} \sum_{s=1}^{T/2} \hat{\Delta}_s^n$$

and

$$2\hat{\psi}^n = \frac{\sum_{s=1}^{T/2} (\hat{\Delta}_s^n - \bar{\Delta}^n)^2}{\frac{T}{2} - 1}.$$

It follows immediately from Theorem 3 that, as  $n \rightarrow \infty$ , the time series statistic  $\bar{\Delta}^n (2\hat{\psi}^n)^{-1/2}$  is asymptotically  $t$ -distributed with  $T/2$  degrees of freedom.

<sup>7</sup> We assume that  $\{\varepsilon_{it}\}_{i=-\infty}^{\infty}$  is mixing but do not assume that we know the appropriate ordering. If we knew the appropriate ordering, consistent estimates of  $\psi$  could be obtained using the methods of White and Domowitz (1984). An extension of that approach is Mei and Newey (1990).



*B. Test Statistics with Estimated Idiosyncratic Returns*

Theorem 3 cannot be directly applied to our asset return model since, in practice,  $\varepsilon^n$  and  $\varepsilon^{*n}$  must be estimated rather than observed directly. We will now modify the procedure to allow for estimated  $\varepsilon^n$  and  $\varepsilon^{*n}$ .

We assume that we have data on a set of prespecified factors or a set of factor estimates. The prespecification of factors is the approach taken, for example, by Chen, Roll, and Ross (1986). Factors can be estimated in a number of ways, the most common of which are factor analysis (e.g., Roll and Ross (1980) and Lehmann and Modest (1988)) or principal components techniques (e.g., Luedecke (1984) or Connor and Korajczyk (1988)). Our test statistic requires that the factor estimates are  $n$ -consistent. For simplicity of exposition, in the main text we treat the case in which the factors are observed exactly, and extend the result to the case of  $n$ -consistent estimated factors in the appendix.

Given the factor structure in (1) and the observed  $k$ -vector of factors,  $f_t$ , a natural way to obtain estimates of the idiosyncratic returns is through time series regressions of returns on the factors and a constant. For each asset,  $i$ , we estimate the regression:

$$\tilde{r}_{it} = c_i + B_i \tilde{f}_t + \tilde{\varepsilon}_{it} \quad t = 1, 2, 3, \dots, T \tag{5}$$

which can be expressed as a system of regressions:

$$\tilde{r} = [c \ B]F + \tilde{\varepsilon}$$

where  $r$  is an  $n \times T$  matrix of asset excess returns,  $c$  is an  $n$ -vector of constants,  $B$  is an  $n \times k$  matrix of factor sensitivities,  $F$  is a  $k + 1 \times T$  matrix whose first row has entries equal to one and whose rows 2 through  $k + 1$  are time series for factors 1 through  $k$ , and  $\varepsilon$  is an  $n \times T$  matrix of idiosyncratic returns.

The regressions give ordinary least squares regression residuals:

$$\hat{\varepsilon}_{it} = \varepsilon_i (I_T - F'(FF')^{-1}F)_{.t} \tag{6}$$

where  $F'_i = (1 \ \tilde{f}'_t)$ . The same regression can be repeated with the additional pseudofactor  $f_t^*$  added to the matrix of regressors:

$$\hat{\varepsilon}_{it}^* = \varepsilon_i^* (I_T - F^{*'}(F^*F^{*'})^{-1}F^*)_{.t}$$

Ordinary least squares induces a bias in the squared residuals, but we can eliminate this with a simple adjustment (see Theil (1971), Theorem 5.2). For each asset, define the  $T$ -vector of adjusted squared residuals:

$$\hat{\sigma}_{it} = \hat{\varepsilon}_{it}^2 / (I_T - F'(FF')^{-1}F)_{tt} \tag{7}$$

Define  $\hat{\sigma}_{it}^*$  in the same way using  $F^*$ . This is a generalization of the usual degrees of freedom adjustment. For example, consider the case where we just fit a mean to each cross-section. In this case  $F = \iota^{T'}$  (where  $\iota^T$  denotes a  $T$ -vector of ones) and  $1/(I_T - F'(FF')^{-1}F)_{tt} = T/(T - 1)$  which is the

standard degrees of freedom correction for variance estimators after fitting a mean. More generally, if  $F$  has  $k + 1$  rows then, by the properties of the trace of an idempotent matrix, the average of  $1/(I_T - F'(FF')^{-1}F)_{tt}$  across  $t$  is  $T/(T - k - 1)$ , which is the standard degrees of freedom correction for variance estimators from ordinary least squares.

Let  $\hat{\sigma}_{iM}$  denote the  $T$ -vector whose odd components are the odd components of  $\hat{\sigma}_i$ , and whose even components are the even components of  $\hat{\sigma}_i^*$ . For any  $T$ -vector  $\mathbf{x}$  such that  $\mathbf{x}'\mathbf{x} = 1$  define  $\hat{\sigma}_{i\mathbf{x}} = \mathbf{x}'\hat{\sigma}_{iM}$ . For any integer  $a$ , define

$$\psi_x^{a,n} = \text{var} \left( \frac{1}{\sqrt{n}} \sum_{i=a}^{a+n} \hat{\sigma}_{i\mathbf{x}} \right).$$

We make two assumptions about  $\sigma_{\mathbf{x}}$ . First, for every  $\mathbf{x}$  there exists a  $\psi_{\mathbf{x}}$  such that:

$$\lim_{n \rightarrow \infty} \psi_x^{a,n} = \psi_x \tag{8}$$

uniformly in  $a$ . Second, for every  $\mathbf{x}$  there exists  $\gamma_{\mathbf{x}} < \infty$  and  $\delta_{\mathbf{x}} > 2/(\lambda - 2)$  such that:

$$E \left[ |\hat{\sigma}_{i\mathbf{x}}|^{2\delta_{\mathbf{x}}} \right] < \gamma_{\mathbf{x}} \tag{9}$$

for all  $i$ , where  $\lambda$  denotes the size of the mixing process  $\{\tilde{\varepsilon}_i\}$ .

**THEOREM 4:** *Given (8) and (9) together with the conditions for Theorem 3, there exists a  $(T/2) \times (T/2)$  matrix  $\Gamma$  such that:*

$$\text{dlim}_{n \rightarrow \infty} \sqrt{n} \cdot \hat{\Delta}^n \sim N(0, \Gamma).$$

*Proof:* See the Appendix.

Suppose now that we replace the true factors with  $n$ -consistent estimated factors  $\hat{F}^n$ . Replacing the true factors by these estimates does not affect the asymptotic distribution of the statistic, as we show in Theorem 5 in the Appendix.

To adjust for the fact that we are using estimated factors rather than true ones, we use a simple degrees of freedom correction:<sup>8</sup>

$$\hat{\sigma}_{it} = \hat{\varepsilon}_{it}^2 / (1 - (k + 1)/T - k/n). \tag{10}$$

<sup>8</sup> We can write the  $n \times T$  matrix of returns as  $R = [c B]F + \varepsilon$ , where  $c$  is an  $n$ -vector of constants and  $F$  is a  $(k + 1) \times T$  matrix consisting of a  $T$ -vector of ones and the  $k \times T$  matrix of factor realizations. Given that  $c$ ,  $B$ , and  $F$  (except the constant vector in  $F$ ) are estimated, this system of equations has  $nT$  observations and  $n(k + 1) + kT$  estimated parameters. Taking an analogy from least squares regression, we use the degrees of freedom correction (number of observations)/(number of observations - number of estimated parameters) which is  $nT/(nT - n(k + 1) - kT) = 1/(1 - (k + 1)/T - k/n)$ . Note that as  $n \rightarrow \infty$  we get the standard time series regression degrees of freedom correction  $T/(T - k - 1)$ . We found via simulation analysis that this simple correction worked better than the more complicated correction given by (7), which does not adjust for the estimation of  $F$ .

Although we have only derived the asymptotic distribution, in Section III.B below we provide simulation evidence on the finite sample distribution.

We implement our procedure as follows:

1. Given a time series of returns on asset  $i$  and a time series of  $k$  factors and a  $k + 1$ st pseudofactor, estimate  $\hat{\varepsilon}_{it}$  and  $\hat{\varepsilon}_{it}^*$  by ordinary least squares.
2. Calculate adjusted squared residuals  $\hat{\sigma}_{it} = \hat{\varepsilon}_{it}^2/[1 - (k + 1)/T - k/n]$  and  $\hat{\sigma}_{it}^* = \hat{\varepsilon}_{it}^{*2}/[1 - (k + 2)/T - (k + 1)/n]$ .
3. Calculate the  $(T/2)$ -vector  $\hat{\Delta}^n$  by subtracting the cross-sectional means of  $\hat{\sigma}_{it}$  in odd months from the cross-sectional means of  $\hat{\sigma}_{it}^*$  in even months. Under the null hypothesis of  $k$  factors,  $\hat{\Delta}^n \Gamma^{-1/2}$  is asymptotically standard normal as  $n \rightarrow \infty$ .
4. Use the time series  $\hat{\Delta}^n$  to calculate its time series mean  $\bar{\Delta}^n$  and covariance matrix  $\hat{\Gamma}$  and test for a positive mean.

We apply this algorithm in Section III.

### *C. Comparison to Previous Empirical Studies*

Note that, with an approximate factor structure, the standard likelihood ratio tests (which assume  $V^n$  is diagonal) will tend to extract too many factors. Also, as  $n$ , the number of assets used to estimate  $\Sigma^n$ , increases the probability of including multiple assets with correlated idiosyncratic returns will increase. This may be one reason why Dhrymes, Friend, and Gultekin (1984) find that the number of statistically significant factors, identified by the standard likelihood ratio test, grows with the number of assets studied.

Let  $\Sigma^n$  denote the covariance matrix of the return vector,  $r_t^n$ . A primary result from the analyses of Chamberlain and Rothschild (1983) and Ingersoll (1984) is that an approximate  $k$ -factor structure implies that the largest  $k$  eigenvalues of  $\Sigma^n$  grow without bound (with  $n$ ) while the remaining eigenvalues are bounded. This suggests that a natural test would be to investigate the behavior of the eigenvalues of the sample covariance matrix,  $\hat{\Sigma}^n$ , as  $n$  increases. This type of analysis is done in Luedecke (1984) and Trzcinka (1986). As noted by both of these authors, testing that the eigenvalues of  $\hat{\Sigma}^n$  are bounded, as  $n$  approaches infinity, is not a well-posed problem with a finite amount of data. In addition, interpreting the behavior of sample eigenvalues as if they are population eigenvalues can be misleading, as we argue below.

Let  $\Lambda_k^n$  denote the  $k$ th largest eigenvalue of the true covariance matrix  $\Sigma^n$  and  $\hat{\Lambda}_k^n$  denote the  $k$ th largest eigenvalue of the sample covariance matrix  $\hat{\Sigma}^n$ . The general empirical findings in Luedecke (1984) and Trzcinka (1986), regarding the behavior of eigenvalues, are that  $\hat{\Lambda}_1$  dominates the remaining eigenvalues but that all of the eigenvalues increase as  $n$  increases. Thus, by one metric (dominance of the largest eigenvalue) it is possible to argue that there is only one pervasive factor while by another metric (growing eigenvalues) it is possible to argue that there are many factors.

Brown (1989) provides an example of a strict factor economy in which there are  $k$  equally important factors and in which he is able to describe the population eigenvalues. He shows that the *population* eigenvalues,  $\Lambda_1, \dots, \Lambda_k$ , increase with  $n$  with  $\Lambda_1^n$  dominating the remaining ones while the eigenvalues beyond the  $k$ th do not grow with  $n$ . One implication of this analysis is that the dominance of the largest eigenvalue should not be taken as evidence that there is only one factor.

The empirical work in Luedecke (1984) and Trzcinka (1986) shows all of the sample eigenvalues growing with  $n$ . Brown (1989) shows through simulations of his  $k$ -factor model that *all* of the sample eigenvalues grow with  $n$  (as found by Luedecke (1984) and Trzcinka (1986)) *even though only the first  $k$  population eigenvalues grow with  $n$ .*

Our analysis, which allows for large cross-sectional samples with fixed time periods, predicts the results from Brown's simulations. In particular, we show that for fixed time series sample size,  $T$ , and growing  $n$ , all of the *sample* eigenvalues grow with  $n$  even though only  $k$  of the *population* eigenvalues are growing.

Let  $\underline{R}^n$  be the  $n \times T$  matrix of demeaned asset returns,<sup>9</sup>  $\underline{F}$  be the  $k \times T$  matrix of demeaned factor realizations,  $B^n$  be the  $n \times k$  matrix of factor loadings, and  $\underline{\varepsilon}^n$  be the  $n \times T$  matrix of demeaned idiosyncratic returns. By definition,  $\underline{R}^n = B^n \underline{F} + \underline{\varepsilon}^n$  and  $\hat{\Sigma}^n = \underline{R}^n \underline{R}^{n'} / T$ . We use the fact that the nonzero eigenvalues of  $\underline{R}^n \underline{R}^{n'}$  are equal to the nonzero eigenvalues of  $\underline{R}^{n'} \underline{R}^n$ . Now, as  $n$  approaches infinity:

$$\underline{R}^{n'} \underline{R}^n / T = (n/T) \cdot [\underline{F}' (B^{n'} B^n / n) \underline{F} + \underline{\varepsilon}^{n'} \underline{\varepsilon}^n / n + \underline{\varepsilon}^{n'} B^n \underline{F} / n + \underline{F}' B^{n'} \underline{\varepsilon}^n / n] \tag{11}$$

$$\rightarrow (n/T) \cdot (\underline{F}' \Xi \underline{F} + \sigma_\varepsilon I_T) \tag{12}$$

where  $\Xi$  is as defined in (3), and  $\sigma_\varepsilon$  is as defined in Theorem 2. The limiting result in (12) follows from (3), Theorem 2, and the fact that  $(\underline{\varepsilon}^{n'} B^n \underline{F}) / n \rightarrow 0$ . Note that  $\underline{F}' \Xi \underline{F}$  is positive semidefinite and of rank  $k$  and, hence, has no effect on the  $k + 1$ st through the  $T$ th eigenvalues. The  $k + 1$ st to the  $T$ th eigenvalues of  $\hat{\Sigma}^n$  approach  $(n/T)\sigma_\varepsilon$  as  $n$  becomes large (the  $T + 1$ st through  $n$ th eigenvalues are zero for  $n > T$ ). Thus, even though we have a  $k$ -factor economy, *sample* eigenvalues beyond the  $k$ th largest will increase as  $n$  increases, since for fixed  $T$ , and growing  $n$ ,  $(n/T)\sigma_\varepsilon$  will increase. Thus, the empirical observation that all of the sample eigenvalues increase with  $n$  should not be interpreted as evidence in favor of a very large number of factors. In our analysis  $\hat{\Lambda}_{k+1}$  through  $\hat{\Lambda}_T$  increase with  $n$  even though  $\Lambda_{k+1}$  through  $\Lambda_T$  do not increase with  $n$ .

Some intuition for why sample eigenvalues, for fixed  $T$ , behave differently from population eigenvalues can be obtained from (11) and (12). The esti-

<sup>9</sup> That is  $\underline{R}_{it}^n = R_{it}^n - \bar{R}_i$ , where  $\bar{R}_i$  is the time series mean return for asset  $i$ . In matrix notation,  $\underline{R}^n = R^n [I - (1/T)\iota\iota']$  where  $I$  is the  $T \times T$  identity matrix and  $\iota$  is a  $T$ -vector of ones.

mated covariance matrix  $\hat{\Sigma}^n$  approaches  $\Sigma^n$  as  $T$  approaches infinity. However, the error in estimating the covariance matrix does not go away as  $n$  increases, for fixed  $T$ . In (12), the largest  $k$  eigenvalues of  $\underline{F}'\Xi\underline{F}$  increase with  $T$  while the largest  $k$  eigenvalues of  $\sigma_\varepsilon I_T$  do not. Thus, letting  $T$  approach infinity before considering  $n$  (which is equivalent to assuming we have the true covariance matrix,  $\Sigma$ ) gives us a matrix whose largest  $k$  eigenvalues increase with  $n$  while the remaining eigenvalues do not. Evidence presented by Trzcinka (1986) indicates that the values of  $\hat{\Lambda}_j^n$ , increase (as  $n \rightarrow \infty$ ) at a faster rate for  $j \leq 6$  than for  $j > 6$  and that the eigenvalues increase at approximately the same rate for all  $j > 6$ . Using the analysis of this section, his results are consistent with an economy with six factors.

### III. Empirical Results

#### A. The Number of Factors for New York Stock Exchange and American Stock Exchange Stocks

We apply our test to a sample of monthly returns on common stocks traded on the New York Stock Exchange (NYSE) and American Stock Exchange (AMEX) for the period January 1967 to December 1991. There are 300 months in the sample, which we divide into five blocks of 60 months each. Within each block, we use all the securities which have full monthly return records on the Center for Research in Security Prices (CRSP) monthly returns file, giving 1647, 1796, 1869, 1651, and 1626 securities in the periods 1967–1971, 1972–1976, 1977–1981, 1982–1986, and 1987–1991, respectively. We use the monthly Treasury bill returns from CRSP to calculate excess returns.

The first step in the analysis is to determine the set of factors to be used. There are a variety of different approaches which can be taken here. One could derive factor returns from a set of macroeconomic variables (as in Chen, Roll, and Ross (1986)) or use statistically generated factors. We use statistically generated factors based on the asymptotic principal components technique from Connor and Korajczyk (1986). The  $k$ -factor-estimates equal the first  $k$  eigenvectors of the  $T \times T$  cross-product matrix of security excess returns,  $R^{n'}R^n/n$ .

The factor estimates from the asymptotic principal components procedure are used as  $f_t$  in the time series regression (5). For each five-year block, we estimate the idiosyncratic returns for a  $k$ -factor model via a set of  $n$  time series ordinary least squares regressions, one for each security in the sample. We regress each security's excess return on a constant and the  $k$  factors. The residuals from this regression are the estimated idiosyncratic returns of the stock. We then find the mean squared idiosyncratic return for each month as the cross-sectional average of these residuals squared:

$$\hat{\mu}_{kt} = \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{\sigma}_{it}^2$$

where  $\hat{\sigma}_{it}$  is the adjusted squared residual return of security  $i$  in month  $t$  (see equation (10)). The estimated time series  $\hat{\mu}_{kt}$ ,  $t = 1, \dots, 60$  from the five sixty-month blocks are linked together to form  $\hat{\mu}_{kt}$   $t = 1$  to 300.

In Table I we present the time series averages of  $\hat{\mu}_{kt}$ ,  $\bar{\mu}_k$ , for  $k = 0, \dots, 15$ . We also show the change in  $\bar{\mu}_k$  as we increase the number of factors and the change in  $\bar{\mu}_k$  as a proportion of average return variance. These statistics are presented for all months, January separately, and all non-January months.

There is a large decrease in extra explanatory power in moving from one to two factors, and then a slow decline in explanatory power from adding each additional factor. As so often in empirical analysis of returns data, January stands out. It has twice the average return variance of other months, and the statistical factors have substantially higher explanatory power, both in absolute terms and as a percentage of total variance. The greater marginal explanatory power of multiple factors in January compared to other months conforms to findings elsewhere in the APT literature. Using daily data, Cho and Taylor (1987) find that the covariance matrix of returns differs in January from other months. Gultekin and Gultekin (1987) and Connor and Korajczyk (1988) find that the factors beyond the first factor are most important to APT pricing tests in January. Given that these factors are more significant for pricing in January, it is perhaps not surprising that they are more important in explaining cross-sectional variance as well. It is clear from Table I that we must adjust for the difference between January and non-January months. While the statistics presented in Table I do not constitute formal tests, we do feel that Table I provides useful descriptive statistics about factor structure in the context of an approximate factor model.

There is a definite seasonal pattern in idiosyncratic variances. We test for this by regressing  $\hat{\mu}_{kt}$  on a constant and eleven monthly dummy variables that are equal to one in their respective months and zero otherwise. The seasonality in idiosyncratic variance is confined to January. In all cases, no other month carries a significant coefficient; the January dummy is usually significant. The detailed results from these regressions are not reported here but are available from the authors.

We treat January and non-January months separately to calculate  $\hat{\Delta}_s$ . For January months, we define  $\hat{\Delta}_s$  by taking the year-to-year differences in the January values of  $\hat{\mu}_{kt}$ . For non-January months we take month-to-month differences, but delete the December-January and January-February differences. We show two configurations (A and B) of the sample, corresponding to the two ways to construct  $\hat{\Delta}$  (A: odd months for  $k$  and even months for  $k + 1$  and B: even months for  $k$ , odd months for  $k + 1$ ). The analysis is equally valid using configuration A or B; for completeness we perform the analysis both ways, although one should bear in mind that the results are not independent (they are negatively correlated) across configurations.

Table II shows the test results. For both the January and non-January tests, we try both the standard ordinary least squares  $t$ -test and a test with the robust covariance estimator of Newey and West (1987), which is consistent with time series correlation and heteroskedasticity. We do not adjust for

**Table I**  
**Mean Square Idiosyncratic Risk for 0 to 15 Factors  $\bar{\mu}_k$ .**  
**(Time Series Averages for All Months, January, and Non-January Months)**

This table shows the pooled time series cross-sectional average mean-squared idiosyncratic return after  $k$  factors have been extracted from the returns ( $k = 0, \dots, 15$ ).  $\bar{\mu}_k$  is the time series mean of  $\mu_k^i = \hat{\varepsilon}'_t \hat{\varepsilon}_t / n_t$  is the cross-sectional average of adjusted estimated residual returns from a  $k$ -factor model. For  $k$  equal to zero the time series sample mean of each asset is subtracted from its time series returns and the number shown is the cross-sectional average of the assets' time series variances. The idiosyncratic returns are adjusted for degrees of freedom using the correction given by equation (10). The monthly numbers have been converted to annualized percentages by multiplying by 1200. The time series sample differs across the three sets of columns: all months, Januaries only, and all months except Januaries. The data are monthly returns on NYSE and AMEX stocks from January 1967 to December 1991.

$k$	All Months			January		Non-January	
	$\bar{\mu}_k \times 1200$	$\frac{\bar{\mu}_k - 1 - \bar{\mu}_k}{\times 1200} (\bar{\mu}_k - 1 - \bar{\mu}_k) / \bar{\mu}_0 \times 100$	$\frac{\bar{\mu}_k}{\times 1200}$	$\frac{\bar{\mu}_k - 1 - \bar{\mu}_k}{\times 1200}$	$\frac{(\bar{\mu}_k - 1 - \bar{\mu}_k) / \bar{\mu}_0}{\times 100}$	$\frac{\bar{\mu}_k - 1 - \bar{\mu}_k}{\times 1200}$	$\frac{(\bar{\mu}_k - 1 - \bar{\mu}_k) / \bar{\mu}_0}{\times 100}$
0	18.08		32.67			16.61	
1	12.86	5.16	16.84	14.62	44.76	12.39	23.12
2	12.23	0.62	15.49	0.41	1.24	11.84	0.77
3	11.75	0.46	13.18	2.64	8.09	11.64	-0.05
4	11.43	0.31	12.52	1.02	3.11	11.35	0.74
5	11.17	0.23	11.36	1.22	3.72	11.17	0.42
6	10.91	0.24	10.96	0.63	1.92	10.93	0.17
7	10.67	0.22	10.21	1.01	3.09	10.74	0.18
8	10.45	0.18	9.39	1.00	3.05	10.58	1.30
9	10.25	0.17	8.69	0.86	2.64	10.42	1.54
10	10.07	0.15	8.27	0.58	1.78	10.25	1.81
11	9.91	0.14	8.16	0.29	0.88	10.09	1.80
12	9.76	0.14	7.68	0.63	1.92	9.97	1.89
13	9.62	0.14	7.61	0.22	0.67	9.82	2.08
14	9.49	0.13	7.43	0.33	1.00	9.70	2.09
15	9.36	0.13	7.19	0.36	1.12	9.58	2.16

**Table II**  
**Tests for  $k - 1$  vs.  $k$  Pervasive Factors**  
**(January and Non-January Months Considered Separately)**

The January and non-January samples are considered separately because of the higher variances evident in January.  $\bar{\Delta}_{k-1,k}$  is the difference between the cross-sectional average mean-squared idiosyncratic return with  $k$  and  $k - 1$  factors estimated on separate subsamples (odd and even months). The third column ( $t$ -value) divides column two by its time series estimated standard error (scaled by the square root of the time series sample size) to form a  $t$ -statistic. The fourth column (robust- $z$ ) recalculates the  $t$ -statistic from column three using a Newey-West (1987) robust estimate of the time series standard error. The fifth column ( $p$ -value) gives the one-sided tail area of the  $t$ -statistic found in the third column. The next four columns repeat columns two through five on the January sample. The last column aggregates the  $p$ -values of the two  $t$ -statistics. The first half of the table (Configuration A) estimates the idiosyncratic mean-squared return for  $k$  factors using the even months of the sample and for  $k - 1$  factors using odd months. The second half of the table (Configuration B) repeats the analysis using odd months to estimate the idiosyncratic mean square return for  $k$  factors and even months for  $k - 1$  factors. The data are monthly returns on NYSE and AMEX stocks from January 1967 to December 1991.

$k$	Non-January				January				Aggregate $p$ -Value
	$\bar{\Delta}_{k-1,k} \times 1200$	$t$ -Value	Robust- $z$	$p$ -Value	$\bar{\Delta}_{k-1,k} \times 1200$	$t$ -Value	Robust- $z$	$p$ -Value	
1	3.48	5.05	4.94	0.00	19.59	1.50	1.57	0.08	0.00
2	0.30	0.82	0.96	0.21	-0.71	-0.35	-0.36	0.63	0.37
3	0.03	0.09	0.10	0.47	1.18	1.02	1.06	0.17	0.22
4	0.20	0.74	0.90	0.23	0.16	0.10	0.10	0.46	0.28
5	0.11	0.45	0.53	0.33	-0.18	-0.10	-0.10	0.54	0.40
6	0.18	0.62	0.78	0.27	-1.10	-0.56	-0.58	0.71	0.48
7	0.26	0.77	0.70	0.22	0.29	0.23	0.24	0.41	0.24
8	0.33	0.98	0.87	0.17	0.67	0.54	0.56	0.30	0.14
9	0.32	1.04	0.94	0.15	0.54	0.42	0.44	0.34	0.15
10	0.36	1.18	1.12	0.12	0.81	0.62	0.65	0.27	0.10
11	0.44	1.56	1.67	0.06	0.81	0.62	0.65	0.27	0.06
12	0.52	1.83	1.96	0.04	1.44	1.24	1.29	0.12	0.02
13	0.60	2.15	2.19	0.02	1.21	1.02	1.06	0.17	0.01
14	0.59	2.10	2.21	0.02	1.35	1.21	1.26	0.13	0.01
15	0.47	1.63	1.73	0.05	1.16	1.02	1.07	0.16	0.03

Configuration A



Table II—Continued

$k$	Non-January				January				Aggregate $p$ -Value
	$\bar{\Delta}_{k-1,k} \times 1200$	$t$ -Value	Robust- $z$	$p$ -Value	$\bar{\Delta}_{k-1,k} \times 1200$	$t$ -Value	Robust- $z$	$p$ -Value	
	Configuration B								
1	5.43	5.08	5.07	0.00	9.65	1.51	1.57	0.08	0.00
2	0.72	1.74	1.79	0.04	1.52	0.50	0.52	0.31	0.06
3	0.16	0.52	0.54	0.30	4.11	1.90	1.98	0.04	0.04
4	0.15	0.48	0.48	0.32	1.87	1.47	1.54	0.09	0.08
5	-0.05	-0.14	-0.16	0.56	2.61	1.80	1.88	0.05	0.12
6	0.01	0.04	0.04	0.49	2.35	1.44	1.50	0.09	0.15
7	-0.14	-0.45	-0.50	0.67	1.73	0.94	0.98	0.18	0.37
8	-0.29	-0.91	-0.83	0.82	1.32	0.81	0.85	0.22	0.53
9	-0.31	-0.97	-0.87	0.83	1.19	0.75	0.78	0.24	0.56
10	-0.28	-0.97	-0.92	0.88	0.35	0.22	0.23	0.42	0.70
11	-0.33	-1.21	-1.19	0.89	-0.24	-0.14	-0.14	0.55	0.83
12	-0.38	-1.48	-1.53	0.93	-0.18	-0.10	-0.10	0.54	0.87
13	-0.37	-1.53	-1.57	0.94	-0.77	-0.49	-0.51	0.68	0.92
14	-0.33	-1.36	-1.41	0.91	-0.70	-0.43	-0.45	0.66	0.90
15	-0.20	-0.81	-0.81	0.79	-0.44	-0.27	-0.28	0.61	0.78

the autocorrelation in the January  $\hat{\Delta}$ 's. We use a first-order lag for the non-January tests.

We find one or two significant factors in the non-January months. In the January months, we find up to six significant factors (at 10 percent confidence) for one configuration, and one significant factor in the other. The aggregated test statistic, equally weighting the January and non-January tests,<sup>10</sup> gives one significant factor in one configuration and four factors in the other.

For non-January months, a one-factor or two-factor model seems adequate to describe stock returns. Including January, up to six factors are necessary to provide an adequate description. Since January mean returns and variances are unusually large, and many interesting asset-pricing phenomena are concentrated in this month, we argue for a three- to six-factor model.

### *B. Simulations*

In order to examine the finite sample distribution of our model we simulated the empirical analysis with random draws from an exact four-factor model calibrated to U.S. equity returns data. In our actual data, the five 60-month blocks of returns contained 1646, 1796, 1869, 1651, and 1626 securities; we simulated using 1718 securities (the average number from the five actual blocks). For simplicity we let each of the four factors have the same variance, and let the average cross-sectional beta of each factor equal one. We assume that the cross-sectional distribution of each of the betas is independent normal, each with variance  $\sigma_b$ . We assume that the time series distributions of the factors and idiosyncratic returns are also independent normal. We assume that the exact version of the APT holds (that is, expected excess returns are linear in betas) and that the four factors have the same risk premium, denoted  $\pi$ . Let  $MVN(\cdot, \cdot)$  denote the multivariate normal distribution. Our model of excess returns for the generated data is:

$$\begin{aligned} r_{it} &= B_i f_t + \varepsilon_{it}, i = 1, \dots, n; t = 1, \dots, T \\ B_i &\sim MVN(\iota^4, \sigma_b I_4) \\ f_t &\sim MVN(\pi \iota^4, \sigma_f I_4) \\ \varepsilon_{.t} &\sim MVN(0^n, \sigma_\varepsilon I_n) \end{aligned}$$

where  $\varepsilon_{.t}$  is the  $n \times 1$  vector of idiosyncratic returns for period  $t$  and  $0^n$  is an  $n \times 1$  vector of zeros.

<sup>10</sup> Let  $z_{nj}$  and  $z_j$  denote the  $t$ -statistics for non-January and January months. We compute  $z = (z_{nj} + z_j)/\sqrt{2}$  and derive the  $p$ -value of this statistic from a one-sided standard normal distribution.

The simulation model has four parameters,  $\sigma_b$ ,  $\pi$ ,  $\sigma_f$ , and  $\sigma_\varepsilon$ . If there are four true factors, then Table I, row 5 ( $k = 4$ ), column 2 ( $\bar{\mu}_4$ ) gives an (annualized percentage) estimate of  $\sigma_\varepsilon$ ; dividing this figure by 1200 to undo the annualized percentages gives  $\sigma_\varepsilon = 0.009525$ . We calibrated the other three parameters to match three empirical observations over our 1967 to 1991 sample period: the average excess return to the equally weighted portfolio, denoted  $E[r_{EW}]$  (0.004167 per month, using the CRSP Equally Weighted Index), the variance of the equally weighted portfolio, denoted  $\sigma_{EW}$  (0.0039162 per month, again using the CRSP Index), and the variance of the average asset, denoted  $\sigma_r$  [0.01507 per month; see Table I, row 1 ( $k = 0$ ), column 2 ( $\bar{\mu}_0$ )]. Note that the equally weighted portfolio has a factor beta of 1 for each of the four factors (since, by assumption, this is the average cross-sectional beta for each factor). Therefore excess return to the equally weighted portfolio is  $r_{EW} = (f_1 + \pi) + (f_2 + \pi) + (f_3 + \pi) + (f_4 + \pi) + \varepsilon_{EW}$ . It is easy to see from this equation that  $\pi = (1/4)E[r_{EW}]$ . We treat the equally weighted portfolio as approximately well diversified, so that its idiosyncratic variance is approximately zero. Using  $\text{var}(\varepsilon_{EW}) \approx 0$  gives  $\sigma_{EW} \approx 4\sigma_f$ . The return variance of a given asset  $i$  is  $(B_{i1}^2 + B_{i2}^2 + B_{i3}^2 + B_{i4}^2)\sigma_f + \text{var}(\varepsilon_i)$ ; therefore the cross-sectional average variance is  $\sigma_r = 4E[B_{ij}^2]\sigma_f + \sigma_\varepsilon$ . Since the  $B_{ij}$ 's are assumed to be independently and identically distributed cross-sectionally with means of one,  $E[B_{ij}^2] = (1 + \sigma_b)$ . Rearranging the equations gives  $\sigma_b = (\sigma_r - \sigma_\varepsilon)/4\sigma_f - 1$ . Using the values for  $\sigma_r$ ,  $\sigma_\varepsilon$ , and  $\sigma_f$  gives  $\sigma_b = 0.4151$ . To summarize the results of this calibration exercise: we used the parameter values  $\sigma_b = 0.4151$ ,  $\sigma_f = 0.0009569$ ,  $\pi = 0.0010417$ , and  $\sigma_\varepsilon = 0.009525$ .

We generated 1000 random draws of 60 months of returns data on 1718 assets. Different betas were drawn randomly for each of the 1000 blocks. Recall that our empirical work used 5 blocks of 60 months each. To match the simulation exercise with the actual empirical work, we partitioned the generated data into 200 sets of 5 blocks each and ran our tests using the same estimation method used in the previous section. Table III shows the averages over 200 realizations of the statistics reported in Table I.

Not surprisingly, the first factor carries much higher explanatory power than factors two through four. This does not violate the asymptotic model: in the calibrated model the first four factors have equal explanatory power, but there exist nonsingular rotations of the factors which assign higher power to the first factor over sixty observations (but always with nonzero explanatory power to all four factors).

Table III shows some positive explanatory power for factors five through fifteen. This implies a small  $n$ -sample bias in our model: for  $k \geq 5$ , the values of  $\bar{\mu}_{k-1} - \bar{\mu}_k$  should be close to zero, and approximately half of them should be negative and half positive; the actual values are small, but consistently positive. In our simulated model, the distinction between "Januaries" (every twelfth month) and other months is arbitrary, except for the smaller number of observations for "Januaries."

**Table III**  
**Average Simulated Values of Statistics Reported in Table I**  
**(Mean Square Idiosyncratic Risk for 0 to 15 Factors,  $\bar{\mu}_k$ )**

This table shows the averages from 200 simulations of a strict factor model with four equally important factors. The table reports the pooled time series cross-sectional average mean-squared idiosyncratic return after  $k$  factors have been extracted from the returns ( $k = 0, \dots, 15$ ).  $\bar{\mu}_k$  is the time series zero of  $\mu_t^n = \hat{\varepsilon}'_t \hat{\varepsilon}_t / n_t$ , where  $\hat{\varepsilon}'_t \hat{\varepsilon}_t / n_t$  is the cross-sectional average of adjusted estimated residual returns from a  $k$ -factor model. For  $k$  equals zero the time series sample mean of each asset is subtracted from its time series returns and the number shown is the cross-sectional average of the assets' time series variances. The idiosyncratic returns are adjusted for degrees of freedom using the correction given by equation (10). The monthly numbers have been converted to annualized percentages by multiplying by 1200. The time series sample differs across the three sets of columns: all months, "Januaries" only, and all months except "Januaries." There are 1718 assets in the cross-section and 300 time series observations (corresponding to 25 years of monthly data). Each number in the table is the simple average from 200 simulations of the statistics.

$k$	All Months				"January"		"Non-January"		
	$\bar{\mu}_k \times 1200$	$\bar{\mu}_{k-1} - \bar{\mu}_k \times 1200$	$(\bar{\mu}_{k-1} - \bar{\mu}_k) / \bar{\mu}_0 \times 100$	$\bar{\mu}_k \times 1200$	$\bar{\mu}_{k-1} - \bar{\mu}_k \times 1200$	$(\bar{\mu}_{k-1} - \bar{\mu}_k) / \bar{\mu}_0 \times 100$	$\bar{\mu}_k \times 1200$	$\bar{\mu}_{k-1} - \bar{\mu}_k \times 1200$	$(\bar{\mu}_{k-1} - \bar{\mu}_k) / \bar{\mu}_0 \times 100$
0	17.81			18.11			17.96		
1	12.85	5.12	29.02	12.81	5.29	28.68	12.85	5.11	28.41
2	12.25	0.60	3.39	12.24	0.57	3.16	12.25	0.60	3.35
3	11.78	0.47	2.66	11.78	0.47	2.60	11.78	0.47	2.61
4	11.42	0.36	2.04	11.42	0.36	1.99	11.42	0.36	2.01
5	11.35	0.07	0.40	11.35	0.07	0.40	11.35	0.07	0.40
6	11.29	0.07	0.37	11.29	0.06	0.35	11.29	0.07	0.37
7	11.22	0.06	0.36	11.22	0.07	0.38	11.22	0.06	0.35
8	11.16	0.06	0.35	11.16	0.06	0.32	11.16	0.06	0.34
9	11.10	0.06	0.34	11.11	0.06	0.33	11.10	0.06	0.33
10	11.05	0.06	0.33	11.05	0.06	0.30	11.04	0.06	0.32
11	10.99	0.06	0.32	11.01	0.04	0.25	10.99	0.06	0.31
12	10.93	0.06	0.32	10.97	0.05	0.27	10.93	0.05	0.29
13	10.88	0.05	0.31	10.91	0.06	0.32	10.88	0.05	0.28
14	10.83	0.05	0.30	10.86	0.06	0.32	10.82	0.05	0.27
15	10.77	0.05	0.30	10.80	0.06	0.35	10.77	0.05	0.27

Table IV shows the average values of the simulated statistics corresponding to those reported in Table II. The average  $t$ -statistics for one through four factors are quite large indicating that the statistic has reasonable power. As one would expect, there is a sharp drop in the average  $t$ -statistic in moving from four to five factors. The average  $t$ -statistics for factors five through fifteen are all slightly positive. The average  $p$ -values for factors five through fifteen (the last column) should be 0.5 but are below that value, because of the same small sample bias as discussed above.

Table V shows the Type I and Type II error frequencies for our simulation. The model always finds four factors, but often finds more than four. The test has high power, but the size of the test is somewhat too large, because of the small sample bias.

Table VI describes the average and minimum coefficients of determination,  $R^2$ , of regressions of the  $k$ th estimated factors,  $k = 1, \dots, 4$ , on the four true factors. The table indicates that asymptotic principal components estimation procedure provides accurate estimates of the factors for the cross-sectional sample sizes typical of U.S. equity returns data. The same basic simulation finding appears in Connor and Korajczyk (1988) and McCulloch and Rossi (1989) using different approaches.

The simulation results are reasonably satisfactory, at least for the case of a strict factor model with equally important factors. The power of the test is good, but the small sample size is somewhat too large for finite  $n$ . The test might tend to show slightly more factors than are actually in the data. The magnitude of the small sample bias is not large in our simulations (see Table III, column 3), but the low standard error of  $\hat{\Delta}_{k-1, k}$  and the long time series sample makes the bias have some effect on the test size.

#### IV. Conclusions

In this paper we develop a new approach to testing for the appropriate number of factors in an approximate factor model of asset returns. We show that the assumptions of an approximate factor model can be related to strong mixing conditions on the cross-sectional dependence of idiosyncratic returns. We use the theory of strong mixing to develop a new statistic which is asymptotically valid as the number of cross-sections grows large.

Our test does not require a strict factor structure, and is valid in large cross-sections. Test statistics which assume a strict factor model, such as the standard likelihood ratio tests from the factor analysis literature, will tend to identify too many pervasive factors (even in large time series samples) if asset returns follow an approximate factor model. Alternative procedures which investigate the behavior of eigenvalues for large cross-sectional sample sizes may give misleading inferences for finite time series samples.

We apply our test to the cross-section of monthly stock returns on the NYSE and AMEX over the period 1967 to 1991. We find evidence for one to six pervasive factors generating returns on NYSE and AMEX stocks. The influence of the factors beyond the first is particularly strong in January. In

**Table IV**  
**Average Simulated Values of Statistics Reported in Table II**  
**(Tests for  $k - 1$  vs.  $k$  Pervasive Factors  $\bar{\Delta}_{k-1, k}$ )**

This table shows the averages from 200 simulations of a strict factor model with four equally important factors. This table shows the results from applying our test for the number of factors.  $\bar{\Delta}_{k-1, k}$  is the difference between the cross-sectional average mean-squared idiosyncratic return with  $k$  and  $k - 1$  factors estimated on separate subsamples (odd and even months). The third column ( $t$ -value) divides column two by its time series estimated standard error (scaled by the square root of the time series sample size) to form a  $t$ -statistic. The fourth column (robust- $z$ ) recalculates the  $t$ -statistic from column three using a Newey-West (1987) robust estimate of the time series standard error. The fifth column ( $p$ -value) gives the one-sided tail area of the  $t$ -statistic found in the third column. The next four columns repeat columns two through five on the "January" sample. The last column aggregates the  $p$ -values of the two  $t$ -statistics. The first half of the table (Configuration A) estimates the idiosyncratic mean-squared return for  $k$  factors using even months of the sample and for  $k - 1$  factors using odd months. The second half of the table (Configuration B) repeats the analysis using odd months to estimate the idiosyncratic mean square return for  $k$  factors and even months for  $k - 1$  factors. There are 1718 assets in the cross-section and 300 time series observations (corresponding to 25 years of monthly data). Each number in the table is the simple average from 200 simulations of the statistics.

$k$	"Non-January"				"January"				Aggregate $p$ -Value
	$\bar{\Delta}_{k-1, k} \times 1200$	$t$ -Value	Robust- $z$	$p$ -Value	$\bar{\Delta}_{k-1, k} \times 1200$	$t$ -Value	Robust- $z$	$p$ -Value	
	Configuration A								
1	5.08	7.76	7.84	0.00	5.50	2.76	3.02	0.02	0.00
2	0.59	4.11	4.27	0.00	0.55	1.24	1.72	0.21	0.00
3	0.46	4.02	4.16	0.00	0.44	1.26	1.73	0.21	0.00
4	0.37	3.88	4.02	0.00	0.33	1.13	1.64	0.22	0.00
5	0.08	0.91	1.31	0.25	0.06	0.19	1.15	0.45	0.28
6	0.08	0.81	1.25	0.28	0.06	0.18	1.18	0.46	0.31
7	0.08	0.75	1.25	0.30	0.06	0.17	1.11	0.46	0.32
8	0.08	0.72	1.26	0.30	0.04	0.11	1.26	0.47	0.34
9	0.08	0.68	1.23	0.32	0.05	0.13	1.25	0.46	0.35
10	0.07	0.61	1.19	0.34	0.07	0.18	1.28	0.46	0.36
11	0.07	0.54	1.13	0.35	0.06	0.17	1.25	0.47	0.37
12	0.07	0.52	1.12	0.36	0.06	0.17	1.25	0.47	0.38
13	0.06	0.46	1.12	0.38	0.07	0.16	1.23	0.47	0.39
14	0.06	0.41	1.11	0.39	0.07	0.17	1.25	0.46	0.40
15	0.06	0.40	1.09	0.40	0.07	0.16	1.23	0.46	0.40

Table IV—Continued

k	"Non-January"				"January"				Aggregate p-Value
	$\bar{\Delta}_{k-1,k} \times 1200$	t-Value	Robust-z	p-Value	$\bar{\Delta}_{k-1,k} \times 1200$	t-Value	Robust-z	p-Value	
	Configuration B								
1	5.14	7.78	7.89	0.00	5.09	2.76	3.01	0.02	0.00
2	0.61	4.23	4.41	0.00	0.59	1.34	1.83	0.19	0.00
3	0.47	4.10	4.27	0.00	0.50	1.44	1.85	0.16	0.00
4	0.35	3.74	3.88	0.00	0.39	1.33	1.81	0.20	0.00
5	0.06	0.67	1.14	0.31	0.09	0.33	1.21	0.42	0.31
6	0.06	0.57	1.07	0.34	0.07	0.24	1.10	0.44	0.34
7	0.05	0.47	1.06	0.37	0.08	0.28	1.24	0.44	0.36
8	0.04	0.40	1.04	0.39	0.07	0.26	1.20	0.44	0.37
9	0.04	0.37	1.05	0.40	0.07	0.22	1.22	0.44	0.39
10	0.05	0.37	1.04	0.39	0.05	0.15	1.22	0.46	0.40
11	0.05	0.38	1.01	0.39	0.03	0.07	1.23	0.48	0.41
12	0.04	0.33	1.03	0.40	0.04	0.11	1.29	0.47	0.41
13	0.04	0.32	1.05	0.41	0.05	0.11	1.23	0.47	0.41
14	0.05	0.32	1.08	0.41	0.04	0.10	1.23	0.47	0.42
15	0.04	0.28	1.06	0.42	0.06	0.14	1.29	0.47	0.42

**Table V**  
**Percentages of Type I and Type II Errors in 200 Simulations**  
**(5 Percent and 10 Percent Confidence Levels)**

This table shows the frequency of Type I and Type II errors in tests of  $k$  factor vs.  $k - 1$  factor models found in 200 simulations. The data are generated by a strict factor model with four equally important factors. Type I error occurs when the simulated  $t$ -statistic is significantly greater than zero for  $k$  greater than four. A Type II error occurs when this  $t$ -statistic is not significantly greater than zero for  $k$  less than or equal to four. The table shows the frequency of these two errors in 200 simulations. The first half of the table (Configuration A) estimates the  $t$ -statistics using idiosyncratic mean-squared return for  $k$  factors from even months of the sample and for  $k - 1$  factors from odd months. The second half of the table (Configuration B) repeats the analysis using odd months to estimate the idiosyncratic mean square return for  $k$  factors and even months for  $k - 1$  factors. There are 1718 assets in the cross-section and 300 time series observations (corresponding to 25 years of monthly data).

$k$	"Non-January"			"January"		
	5% Confidence	10% Confidence	10% Confidence	5% Confidence	5% Confidence	10% Confidence
Configuration A						
$k$	Empirical type II error frequencies (1-empirical power)					
1	0.00	0.00	0.00	0.12	0.02	0.02
2	0.01	0.01	0.01	0.70	0.53	0.53
3	0.00	0.00	0.00	0.71	0.54	0.54
4	0.00	0.00	0.00	0.75	0.61	0.61
$k$	Empirical type I error frequencies (empirical size)					
5	0.22	0.36	0.36	0.06	0.13	0.13
6	0.17	0.27	0.27	0.07	0.13	0.13
7	0.17	0.29	0.29	0.06	0.13	0.13
8	0.17	0.29	0.29	0.06	0.12	0.12
9	0.18	0.26	0.26	0.06	0.13	0.13
10	0.14	0.27	0.27	0.10	0.16	0.16
11	0.13	0.25	0.25	0.09	0.18	0.18
12	0.12	0.24	0.24	0.09	0.16	0.16
13	0.12	0.22	0.22	0.08	0.17	0.17
14	0.13	0.21	0.21	0.08	0.15	0.15
15	0.12	0.19	0.19	0.06	0.13	0.13



Table V — Continued

$k$	"Non-January"		"January"	
	5% Confidence	10% Confidence	5% Confidence	10% Confidence
	Configuration B			
	Empirical type II error frequencies (1-empirical power)			
1	0.00	0.00	0.11	0.03
2	0.01	0.00	0.67	0.51
3	0.01	0.00	0.68	0.50
4	0.01	0.00	0.69	0.53
	Empirical type I error frequencies (empirical size)			
5	0.15	0.27	0.09	0.15
6	0.12	0.24	0.06	0.15
7	0.10	0.20	0.09	0.16
8	0.09	0.16	0.10	0.16
9	0.10	0.18	0.09	0.15
10	0.11	0.18	0.07	0.14
11	0.08	0.17	0.06	0.12
12	0.08	0.15	0.07	0.13
13	0.08	0.14	0.06	0.12
14	0.08	0.19	0.06	0.11
15	0.09	0.18	0.08	0.12

**Table VI**  
**Regressions of the  $k$ th Estimated Factor on the**  
**Four True Factors**  
**(Average and Minimum  $R^2$  from 1000 Simulations)**

This table examines the fit of the asymptotic principal components estimated factors using 1000 simulations of an exact factor model with four equally important factors. Sixty months of returns data are simulated for 1718 securities. Four factors are estimated on each simulated data set. For each of the 1000 simulations, each of the four estimated factors is regressed against the four true factors which generated the returns data. The table shows the averages of the 1000  $R^2$ 's from these regressions and the minimum of the 1000  $R^2$ 's.

Estimated Factor: $k$	Average $R^2$ (%)	Minimum $R^2$ (%)
1	99.870	99.727
2	98.620	97.368
3	98.058	95.236
4	97.117	92.855

non-January months, we find only one or two significant factors. Given the importance of January returns (both for expected returns and per annum volatility), we argue for a three- to six-factor model.

**Appendix**

*Proof of Theorem 4:* First we show that

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{i\mathbf{x}} = \mathbf{x}' \boldsymbol{\iota}^T \sigma_\varepsilon$$

for all  $\mathbf{x}$  with  $\mathbf{x}'\mathbf{x} = 1$ . Using (6) note that:

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{it}^2 = \frac{1}{n} (\hat{\varepsilon}'\hat{\varepsilon})_{tt} = \left[ (I - F'(FF')^{-1}F) \left( \frac{1}{n} \varepsilon' \varepsilon \right) (I - F'(FF')^{-1}F) \right]_{tt}$$

Since  $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \varepsilon' \varepsilon = \sigma_\varepsilon I_T$  and  $(I - F'(FF')^{-1}F)$  is idempotent this gives:

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} (\hat{\varepsilon}'\hat{\varepsilon})_{tt} = \sigma_\varepsilon (I - F'(FF')^{-1}F)_{tt}$$

Dividing by  $(I - F'(FF')^{-1}F)_{tt}$  gives

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{it} = \sigma_\varepsilon$$

The steps are identical to show that

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{it}^* = \sigma_\varepsilon$$

Since  $\hat{\sigma}_{i\mathbf{x}}$  is a linear combination of  $T$  different series each with plim equal to  $\sigma_\varepsilon$ , this implies

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{i\mathbf{x}} = x' \iota^T \sigma_\varepsilon.$$

Next we show that  $\hat{\sigma}_{i\mathbf{x}}$  is alpha-mixing of the same size as  $\tilde{\varepsilon}_i$ . Consider the scalar series of random variables  $\{\varepsilon_j\}_{j=1}^{nT} = \{\varepsilon_{i1}, \dots, \varepsilon_{iT}\}_{i=1}^n$ . Using the definition of alpha-mixing coefficients (see White (1984, Definition 3.42)) it is easy to see that  $\{\varepsilon_j\}$  is alpha-mixing of the same size as  $\{\varepsilon_i\}$ . Note from (6) and (7) that  $\{\sigma_{i\mathbf{x}}\}$  is a function of  $\{\varepsilon_j\}$  using fixed finite lags. By White (1984, Theorem 3.49)  $\{\sigma_{i\mathbf{x}}\}$  is an alpha-mixing series of the same size as  $\{\varepsilon_i\}$ .

Given that  $\sigma_{i\mathbf{x}}$  is alpha-mixing of the same size as  $\tilde{\varepsilon}_i$ , and given (8) and (9), it follows from White (1984, Theorem 5.19) that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\sigma_{i\mathbf{x}} - x' \iota^T \sigma_\varepsilon)$$

converges in distribution to a zero mean univariate Normal.

We have shown that any linear combination  $\mathbf{x}$  of  $n^{-1/2}(\hat{\sigma}_{iM} - \sigma_\varepsilon \iota^T)$  with  $\mathbf{x}'\mathbf{x} = 1$  converges in distribution to a zero mean univariate Normal. By the Cramer-Wold device (see, e.g., White (1984, p. 108)) this implies that  $n^{-1/2}(\hat{\sigma}_{iM} - \sigma_\varepsilon \iota^T)$  approaches a multivariate Normal with zero means. Since  $n^{-1/2}\hat{\Delta}^n$  consists of the differences between the odd and even components of  $n^{-1/2}(\hat{\sigma}_{iM} - \sigma_\varepsilon \iota^T)$ , it approaches a  $T/2$ -element multivariate Normal with zero means.  $\square$

### Test Statistics with Estimated Factors

The matrix of estimated factors has dimension  $k \times T$ , and so is “small,” relative to the sample size  $nT$ , as  $n$  goes to infinity. This allows us to prove that the estimation error in  $F$  does not affect the  $n$ -asymptotic distribution of our statistic, as long as the estimate of  $F$  is  $n$ -consistent.

Define  $M = (I - F'(FF')^{-1}F)$  and  $\hat{M}^n = (I - \hat{F}'(\hat{F}\hat{F}')^{-1}\hat{F})$ . Let  $\hat{\varepsilon}^n$  denote the residuals estimated using the true factors,  $\hat{\varepsilon}^n = \varepsilon^n M$ , and  $\tilde{\varepsilon}^n$  the residuals using the estimated factors,  $\tilde{\varepsilon}^n = \varepsilon^n \hat{M}^n + B^n F \hat{M}$ . Note that the heteroskedasticity-corrected mean-squared residuals are

$$\text{diag}(\hat{\varepsilon}^{n'} \hat{\varepsilon}^n / n) (\text{diag}(M))^{-1} \text{ and } \text{diag}(\tilde{\varepsilon}^{n'} \tilde{\varepsilon}^n / n) (\text{diag}(\hat{M}^n))^{-1}.$$

**THEOREM 5:** *Given that  $\text{plim}_{n \rightarrow \infty} (\hat{F}^n - F) = 0$  then*

$$\text{dlim}_{n \rightarrow \infty} n^{1/2} \left[ \text{diag}(\hat{\varepsilon}^{n'} \hat{\varepsilon}^n / n) (\text{diag}(M))^{-1} - \text{diag}(\tilde{\varepsilon}^{n'} \tilde{\varepsilon}^n / n) (\text{diag}(\hat{M}^n))^{-1} \right] = 0$$

*Proof of Theorem 5:* We show that  $\text{dlim}_{n \rightarrow \infty} n^{1/2}(n^{-1} \tilde{\varepsilon}^{n'} \tilde{\varepsilon}^n - \sigma_\varepsilon \hat{M}^n) = \text{dlim}_{n \rightarrow \infty} n^{1/2}(n^{-1} \hat{\varepsilon}^{n'} \hat{\varepsilon}^n - \sigma_\varepsilon M)$  and then the result follows by considering the

diagonal elements of these two matrices. First consider the cross-product matrix of the residuals estimated with the true factors:

$$\begin{aligned} \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \hat{\varepsilon}^{n'} \hat{\varepsilon}^n - \sigma_\varepsilon M) &= \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} M \varepsilon^{n'} \varepsilon^n M - \sigma_\varepsilon M) \\ &= M [\text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \varepsilon^{n'} \varepsilon^n - \sigma_\varepsilon I_T)] M. \end{aligned}$$

The second equality uses the fact that  $M$  is symmetric and idempotent so that  $MM = M$ . Next consider the cross-product matrix of residuals estimated using the  $n$ -consistent estimated factors:

$$\begin{aligned} \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \tilde{\varepsilon}^{n'} \tilde{\varepsilon}^n - \sigma_\varepsilon \hat{M}) \\ = \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \hat{M}^n \varepsilon^{n'} \varepsilon^n \hat{M}^n + \hat{M}^n F n^{-1} B^{n'} B^n F \hat{M}^n - \sigma_\varepsilon \hat{M}). \end{aligned} \tag{A1}$$

Next we show that the middle term of (A1) disappears from the distribution limit, that is:

$$\text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \hat{M}^n F B^{n'} B^n F \hat{M}^n) = 0. \tag{A2}$$

If  $g(\cdot)$  is a function differentiable at  $\text{plim}_{n \rightarrow \infty} x^n$ , then  $\text{plim}_{n \rightarrow \infty} g(x^n) = g(\text{plim}_{n \rightarrow \infty} x^n)$ . Therefore,  $\text{plim}_{n \rightarrow \infty} \hat{F}^n = F$  implies  $\text{plim}_{n \rightarrow \infty} \hat{M}^n = M$ . Recall that, if  $L^n$  and  $K^n$  are sequences of fixed dimension random matrices, then  $\text{dlim}_{n \rightarrow \infty} L^n K^n = (\text{plim}_{n \rightarrow \infty} L^n)(\text{dlim}_{n \rightarrow \infty} K^n)$  as long as  $\text{plim}_{n \rightarrow \infty} L^n$  and  $\text{dlim}_{n \rightarrow \infty} K^n$  exist. Applying this to (A2):

$$\begin{aligned} \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \hat{M}^n F B^{n'} B^n F \hat{M}^n) \\ = \text{plim}_{n \rightarrow \infty} \hat{M}^n F [\text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} B^{n'} B^n)] F \text{plim}_{n \rightarrow \infty} \hat{M}^n \\ = MF [\text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} B^{n'} B^n)] FM = 0 \end{aligned}$$

where the last equality follows from  $MF = 0$ . Substituting (A2) into (A1):

$$\begin{aligned} \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \tilde{\varepsilon}^{n'} \tilde{\varepsilon}^n - \sigma_\varepsilon \hat{M}) &= \text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \hat{M}^n \varepsilon^{n'} \varepsilon^n \hat{M}^n - \sigma_\varepsilon \hat{M}) \\ &= \text{dlim}_{n \rightarrow \infty} \hat{M}^n n^{1/2} (n^{-1} \varepsilon^{n'} \varepsilon^n - \sigma_\varepsilon I_T) \hat{M}^n, \end{aligned} \tag{A3}$$

where the last equality uses that  $\hat{M}^n$  is symmetric and idempotent. Applying  $\text{dlim}_{n \rightarrow \infty} L^n K^n = (\text{plim}_{n \rightarrow \infty} L^n)(\text{dlim}_{n \rightarrow \infty} K^n)$  to (A3) implies that (A3) is equal to:

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{M}^n [\text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \varepsilon^{n'} \varepsilon^n - \sigma_\varepsilon I_T)] \text{plim}_{n \rightarrow \infty} \hat{M}^n \\ = M [\text{dlim}_{n \rightarrow \infty} n^{1/2} (n^{-1} \varepsilon^{n'} \varepsilon^n - \sigma_\varepsilon I_T)] M. \quad \square \end{aligned}$$

REFERENCES

Anderson, T. W., 1984, *An Introduction to Multivariate Statistical Analysis*, 2d. ed. (Wiley, New York).

- Brown, Stephen J., 1989, The number of factors in security returns, *Journal of Finance* 44, 1247-1262.
- Chamberlain, Gary, and Michael Rothschild, 1983, Arbitrage and mean variance analysis on large asset markets, *Econometrica* 51, 1281-1304.
- Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383-403.
- Cho, D. Chinhung, and William M. Taylor, 1987, The seasonal stability of the factor structure of stock returns, *Journal of Finance* 42, 1195-1211.
- Connor, Gregory, and Robert A. Korajczyk, 1986, Performance measurement with the arbitrage pricing theory: A new framework for analysis, *Journal of Financial Economics* 15, 373-394.
- , 1988, Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics* 21, 255-289.
- Conway, Delores A., and Marc R. Reinganum, 1988, Stable factors in security returns: Identification through cross validation, *Journal of Business & Economic Statistics* 6, 1-15.
- Dhrymes, Phoebus J., Irwin Friend, and N. Bulent Gultekin, 1984, A critical reexamination of the empirical evidence on the arbitrage pricing theory, *Journal of Finance* 39, 323-346.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607-636.
- Gultekin, Mustafa, and N. Bulent Gultekin, 1987, Stock return anomalies and tests of the APT, *Journal of Finance* 42, 1213-1224.
- Heston, Steve, 1992, Testing approximate linear asset pricing models, Working paper, Yale School of Management.
- Ingersoll, Jonathan E., Jr., 1984, Some results in the theory of arbitrage pricing, *Journal of Finance* 39, 1021-1039.
- Korajczyk, Robert A., and Claude J. Viallet, 1989, An empirical investigation of international asset pricing, *Review of Financial Studies* 2, 553-585.
- Lehmann, Bruce N., and David M. Modest, 1988, The empirical foundations of the arbitrage pricing theory, *Journal of Financial Economics* 21, 213-254.
- Luedecke, Bernd P., 1984, An empirical investigation into arbitrage and approximate  $K$ -factor structure on large asset markets, Doctoral dissertation, Department of Economics, University of Wisconsin.
- McCulloch, Robert, and Peter E. Rossi, 1989, Posterior, predictive, and utility-based approaches to testing the arbitrage pricing theory, Working Paper No. 269, Center for Research in Security Prices, University of Chicago.
- Mei, Jianping, and Whitney Newey, 1990, Application of the Newey-West matrix for correction of heteroscedasticity and cross-sectional correlation, Working paper, New York University.
- Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.
- Roll, Richard, and Stephen A. Ross, 1980, An empirical investigation of the Arbitrage Pricing Theory, *Journal of Finance* 35, 1073-1103.
- , 1984, A critical reexamination of the empirical evidence on the arbitrage pricing theory: A reply, *Journal of Finance* 39, 347-350.
- Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.
- Schwarz, H. R., H. Rutishauser, and E. Stiefel, 1973, *Numerical Analysis of Symmetric Matrices* (Prentice-Hall, Englewood Cliffs, N.J.).
- Stewart, G. W., 1973, *Introduction to Matrix Computations* (Academic Press, Orlando, Fla.).
- Theil, Henri, 1971, *Principles of Econometrics* (Wiley, New York).
- Trzcinka, Charles, 1986, On the number of factors in the arbitrage pricing model, *Journal of Finance* 41, 347-368.
- White, Halbert, 1984, *Asymptotic Theory for Econometricians* (Academic Press, Orlando, Fla.).
- and Ian Domowitz, 1984, Nonlinear regression with dependent observations, *Econometrica* 52, 143-161.